ATHENS UNIVERSITY OF ECONOMICS AND BUSINESS

DEPARTMENT OF STATISTICS

POSTGRADUATE PROGRAM

Extensions of Latent Variables Models with applications on econometric and educational models

By

Michalis Linardakis

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 PhD in Statistics

Athens, Greece
March 2004

ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ

Διεύρυνση Μοντέλων Λανθανουσών Μεταβλητών με Εφαρμογές στην Οικονομετρία και την Εκπαιδευτική Στατιστική

Μιχαήλ Γ. Λιναρδάκης

ΔΙΑΤΡΙΒΗ

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

Αθήνα
Μάρτιος 2004

ACKNOWLEDGEMENTS

If any of us honestly reflects on who we are, how we got here, what we think we might do well, we discover a debt to all those that have directly shaped our lives and our work. Hardship is often one of our teachers; that I acknowledge. But brilliance and inspiration are always the best teachers; for these I thank Petros Dellaportas. You urged me on by way of your untiring support. I'm grateful.

I owe special thanks to quite a few others who have helped with my research. I would like to thank Prof. Ioannis Panaretos who aided me in obtaining the SAT data at short notice. I would also like to thank Alexandros Deloukas for helpful discussions on the transportation policy and Attiko Metro, A.E., for providing the data for the illustration of the multiranked probit model. I would additionally like to acknowledge financial support by the European Union's research network "Statistical and Computational Methods for the Analysis of Spatial Data, ERB-FMRX-CT96-0096".

Furthermore, I would like to thank Dimitris Karlis and Ioannis Ntzoufras who provided me with valuable help in programming codes at the outset of this work, as well as helpful discussions throughout, and their exemplary desire for work.

Special thanks should also be given to Irini Moustaki for her remarks and support during the final stage of this work.

I'm also indebted to a good friend, John Eleftheriou, who spent time editing my use of English in this thesis.

There are many who, from behind the scenes, have encouraged and supported my efforts, and made things meaningful which I wish to thank: my family for all their support which I hope I can recompense some day; my sister (and best friend) Vana, for encouraging me and for being supportive and caring throughout my life; my brother Babis (I had my most prolific years when we were housemates, it's true).

Also, I had always the warmest support by my family of friends in Athens: Mina Mavrou was always there, whenever any difficulties arose during this work and whenever things had to be in high spirits; Thelxi Vogiatzi was looking after my work and also my mood; Stathis Photis was calming in situations full of anxiety and was consecutively supporting me, especially during the last two years of my work; Giorgos Kouvatseas was a close friend and fellow-student from the undergraduate studies, he decided that I would follow the postgraduate studies before I did. To all these people, I am honestly grateful.

VITA

Born in Crete, in 1973, I lived there till 1990. In 1990 I moved to Athens for my undergraduate studies, at the department of Statistics, Athens University of Economics and Business.

I graduated in 1994 and, from 1995 to 1997, I studied on the postgraduate programme for the degree of Master of Science in Statistics, at the same department. Prof. Petros Dellaportas supervised my thesis entitled “A contribution to the Bayesian Analysis of discrete Choice Models”.

At the end of 1998, I started my Ph.D. studies at department of Statistics, under the supervision of Prof. Petros Dellaportas.

ABSTRACT

Michalis Linardakis

Extensions of Latent Variables Models with applications on econometric and educational models

March 2004

The present thesis deals with some extensions of latent variable models. We present the family of discrete choice models, we analyze complicated but realistic ranking data generation structures that have either not been considered before or have not been adequately handled, and we propose novel methodological approaches and MCMC technicalities. We enhance the multinomial probit model by including ranking responses and we refer to this model as the multiranked probit model. We also adopt the notion of utility threshold parameter, which deals realistically with ranking responses and ties, and we enrich the model with random effects on the utility thresholds. To illustrate the proposed model, a real data set is analysed.

Under the general frame of the latent variable models, another family, the IRT models, is presented. We propose a multidimensional IRT model with thresholds; we analyse multiple-choice responses in multiple-choice tests when there are penalties for each wrong answer such as a subtraction of points (a widely used technique that attempts to prevent students from guessing). The literature is still sparse in analysing data sets of multiple-choice answers with omissions. We extend the use of item response models to capture this situation by including guessing and threshold latent parameters. We also separate the ability of each student into several parts, which express different cognitive tasks by a multidimensional scaling approach. A Pseudo-Bayes factor model choice approach (based on cross-validation predictive densities) is used to select the number of dimensions that fit the data better. The proposed model is illustrated with two real data sets.

ΠΕΡΙΛΗΨΗ

Μιχάλης Λιναρδάκης

Διεύρυνση Μοντέλων Λανθανουσών Μεταβλητών με Εφαρμογές στην Οικονομετρία και την Εκπαιδευτική Στατιστική

Μάρτιος 2004

Η διατριβή ασχολείται με υποδείγματα λανθανουσών μεταβλητών. Παρουσιάζεται η οικογένεια μοντέλων διακριτής επιλογής αλλά και αναλύονται δεδομένα, η δομή των οποίων δεν έχει αντιμετωπισθεί επαρκώς από τα υπάρχοντα υποδείγματα. Για το σκοπό αυτό, προτείνεται η μεθοδολογία που βασίζεται στις Monte Carlo Μαρκοβιανές Αλυσίδες της Στατιστικής κατά Bayes. Στην προτεινόμενη μεθοδολογία, διευρύνουμε το υπόδειγμα multinomial probit έτσι ώστε να αντιμετωπίζει αποκρίσεις διάταξης και ονομάζουμε αυτό το μοντέλο multiranked probit. Επίσης, χρησιμοποιούμε την έννοια του «ορίου χρησιμότητας» (utility thresholds), όταν στα δεδομένα διάταξης παρατηρούνται ισοπαλίες. Το μοντέλο που προτείνεται, χρησιμοποιείται για να αναλυθεί ένα σετ πραγματικών δεδομένων.

Επίσης παρουσιάζεται μια άλλη οικογένεια μοντέλων λανθανουσών μεταβλητών, τα IRT μοντέλα. Προτείνουμε ένα πολυδιάστατο IRT μοντέλο με όρια (thresholds). Με το μοντέλο αυτό, αναλύουμε δεδομένα από τεστ πολλαπλών επιλογών σε μαθητές, στις περιπτώσεις που υπάρχει μείωση στη βαθμολογία (penalty) για κάθε λανθασμένη απάντηση (μια τεχνική που χρησιμοποιείται για να αποτρέψει τους μαθητές από την απάντηση «κατά τύχη»). Στη βιβλιογραφία δεν έχει ακόμη εμφανιστεί μοντέλο που να επιτρέπει τη λεπτομερή ανάλυση τέτοιου είδους δεδομένων. Για το λόγο αυτό προτείνουμε το μοντέλο με χρήση παραμέτρων «ορίων» (thresholds). Επίσης, χωρίζουμε την εκτιμώμενη γενική ικανότητα κάθε μαθητή σε επιμέρους γνωστικές ικανότητες - παράγοντες. Η διαδικασία επιλογής μοντέλου –

αριθμού παραγόντων γίνεται μέσω της μεθόδου Pseudo-Bayes factor. Τέλος, το προτεινόμενο μοντέλο χρησιμοποιείται για να αναλυθούν δύο πραγματικά σετ δεδομένων.

Contents

List of Tables

List of Figures

Chapter 1

Introduction: basic ideas and utilization of latent variables

1.1 Defining a latent variable

The term latent variable covers the idea of something underlying what is observed, something that is hidden behind the measurements we record. In this context, our ability to answer a question that measures intelligence is a latent variable; we observe a variable that denotes whether our answer was correct or not, and this categorical variable is what we record. Nevertheless, this indication is only what we can observe from the underlying continuous variable named "intelligence".

Our decision on how to behave in a given situation also has to do with something latent; we finally decide to follow one way or another, meanwhile we have thought out the utilities of any possible ways (probably unconsciously) and we follow the one that maximizes the utility function. In other words, the utility is a summarizing concept that comes prior to its indicators which we measure.

Is it so common? Are there latent variables in many situations, whether this situation can be measured or not? In fact, even in the simplest examples of a statistical problem, we may be faced with latent variables. A usual practise, for example, is to record the age or the income in a questionairre as categorical (ordinal) variables. In fact, these are continuous variables that we probably cannot or we are not willing to record. We simplify the process by writing them down as ordinal but an underlying latent variable is hidden behind them. However, in the following chapters, we will not pay attention to this kind of latent variables (although the way to work with them will be the same) but to those which cannot be directly observed in any way, simplifying a process or not.

1.2 The use of latent variables

A next plausible question is why it is useful to include latent variables in a model since they are not directly revealed, instead of using the indicator variables we have already observed. As Bartholomew and Knott (1999) explain:

> *One reason, common to many techniques of multivariate analysis, is to reduce dimensionality. If, in some sense, the information contained in the interrelationships of many variables can be conveyed, to a good approximation, in a much smaller set, our ability to 'see' the structure in the data will be much improved. This is the idea which lies behind much factor analysis and the newer applications of linear structural models.*

The authors give an example that illustrates the use of a large questionnaire that attempts to investigate the basic political position, which is, in a sense, the underlying latent variable, via a large set of questions:

> *One's view about the desirability of private health care and of tax levels for high earners might both be regarded as a reflection of a basic political position. Indeed, many enquires are designed to probe such basic attitudes from a variety of angles. The question is then one of how to condense the many variables with which we start into a much smaller number of indices with as little loss of information as possible. Latent variable models provide on way of doing this.*

A second reason to use latent variables in a statistical model is inherent to the nature of the variables themselves; though these are important for the research, they cannot be directly observed, as it was pointed out from the beginning of this chapter. Bartholomew and Knott (1999) state:

> *Latent quantities figure prominently in many fields to which statistical methods are applied. This is especially true of the social sciences. A cursory inspection of the literature of social research or of public discussion in newspapers or on television will show that much of it centres on entities which are handled as if they were measurable quantities but for which no measuring instrument exists. Business confidence, for example, is spoken of as though it were a real variable, changes in which affect share prices or the value of the currency. Yet business confidence is an ill-defined concept which may be regarded as a convenient shorthand for a whole complex of beliefs and attitudes. The same is true of quality of life, conservatism, and general intelligence. It is virtually impossible to theorize about social phenomena without invoking such hypothetical variables. If such reasoning is to be expressed in the language of mathematics and thus made rigorous, some way must be found of representing such 'quantities' by numbers. The statistician's problem is to establish a theoretical framework within which this can be done. 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.*

The latent variables may either be continuous (metrical) or categorical. The same is valid for the manifest variables which are observed. Depending on the different types of the variables which are involved in a statistical problem, a different analysis of the same "family" arises. Hence, when the manifest variables are continuous and the resulting latent variables are continuous, we are faced with the well-known factor analysis. In the case where the manifest variables are continuous and the latent variables are categorical, the analysis is called latent profile analysis. On the other hand, when the manifest variables are categorical, we use the latent class analysis, in the case of categorical latent variables, or the latent trait analysis, in the case of continuous latent variables. This later case may involve many different alternative cases which arise from different forms of the categorical manifest variables. Therefore, when the categorical manifest variables are in an ordinal form, the conjoint analysis results. A special case of this analysis is formed when the categorical manifest response variable is binary, which gives rise to the so called discrete choice model. If the ranking responses-manifest variables include ties, we end up with a multiranked model, which is proposed in the present thesis.

There are many other models within the wide family of the latent variables models. Depending on the scientific field, econometrics, marketing, psychology, education etc, we meet the appropriate latent variable models, which have been designed to deal with the specific problems. One of the analyses of this wide family is the item response theory, that is used for psychometric and educational purposes. This model, in its multidimensional version, is also under consideration in the present thesis.

1.3 Outline of the thesis

Each chapter of the following in this thesis, begins with an introductory section, where an outline of the chapter is provided.

The proposed methodology of the thesis is Bayesian and the implementation tool adopted is Markov Chain Monte Carlo (MCMC). To introduce the reader with some basic ideas of the Bayesian methodology, Chapter 2 presents two widely used methods for sampling from conditional distributions: the Gibbs Sampler and the Metropolis-Hastings algorithm. Chapter 3 reviews the discrete choice models and the modifications that improve these models, which have been proposed in the literature (a detailed review and an application of these models is also given in Linardakis, 1997; see also Linardakis, 2001).

The multiranked probit model is proposed in the thesis, that is used to analyze data sets where the responses are of ranking form with ties. That is, the respondents rank a number of alternatives, based on the utility maximization. In the case where the difference of the utilities of two or more alternatives does not exceed a threshold quantity, the alternatives are considered to be indistinguishable with respect to their utility and, hence, a tie occurs. Thus, we analyze complicated but realistic ranking data generation structures that either have not been considered before or have not been adequately handled, and we propose novel methodological approaches and MCMC technicalities. The proposed model is presented in Chapter 4 and an application of this model is provided in Chapter 5. Selected contents of these chapters have also been presented in the articles by Linardakis and Dellaportas (1998, 1999 and 2003).

In the multiranked probit model:

- we ensure identifiable parameters for the covariance matrix of the underlying utility vectors

- we adopt the notion of utility threshold parameter which deals realistically with ranking responses and ties

- we enrich the model with random effects on the utility thresholds (we include a hierarchical step that models the unit-specific utility thresholds as exchangeably distributed)

- we permit the use of heavy tail distributions for the stochastic error term. This is proved to be fruitful when small data sets (that is, sets with a small number of responses) are analysed. In this case, the use of normal errors leads to inexplicable results

- to illustrate the proposed model, we analyze a complicated ranking data set from a stated preference experiment for Attiko Metro

- we present practical graphical answers on pivotal questions of the transportation policy with respect to value of some specific attributes (for example, the price of the ticket) that results into maximization of the profits.

Under the general family of the latent variable models, the Item Response Theory (IRT) is also considered. Chapter 6 reviews the most widely used IRT models and presents the proposed model that analyses data sets with omissions in the responses. Chapters 7 and 8 illustrate the proposed model with two real data set applications (selected contents of chapter 7 have also been presented in Linardakis and Dellaportas, 2000).

In the proposed IRT model:

- we analyze multiple choice responses in multiple choice tests when there are penalties for each wrong answer and we extend the use of item response models to capture this situation by including latent threshold parameters and appropriate guessing parameters,

- we separate the ability of each student into several parts, which express different cognitive tasks by a factor analysis approach (this is a factor analysis that is applied on latent variables, the abilities of the students, and some constraints have to be placed on the parameters for identification and rotation purposes),

- we use a Pseudo-Bayes factor model choice approach (based on cross-validation predictive densities) in order to select the number of dimensions that fit the data better,

- we illustrate the proposed model with a real data set from a multiple-choice test that we constructed from us, and was used as an exam paper in the Department of Statistics, of Athens University of Economics and Business. The analysis of this data also illustrates the interpretation of the factors into the IRT models, with quite reasonable results,

- we compare the model with δ_i parameters with the three parameter normal ogive model, in terms of predicted students ranking. That is, we construct the credible intervals of the predicted ranking of the students based on the two models outputs, and we check the agreement of the intervals with the grades and the grades' ranking the students received,
- another large data set that is analysed (the SAT data) illustrates the convergence and the interpretation of the parameters of the proposed model.

Chapter 2

MCMC Methods and Inference

2.1 Introduction

The approach of this thesis will be Bayesian and the Markov Chain Monte Carlo (MCMC) method will be used throughout. For this reason, we will briefly present the Gibbs Sampler and the Metropolis-Hastings algorithm in this chapter; especially, the Gibbs Sampler is a necessary background for the multinomial probit (MNP) Sampler, which is essential in the presentation of the multiranked probit model that we propose.

2.2 The problem

Suppose that a posterior density $f(\mathbf{y})$ with $\mathbf{y} = (y_1, ..., y_p)$, $\mathbf{y} \in R^p$, is what we would like to obtain for a particular problem. However, the calculation of such a density may require extremely difficult integrations and, moreover, these may, by no means, be done analytically but only numerically. Furthermore, the calculation of a marginal density that is obtained from the joint density $f(\mathbf{y})$ is of particular interest. However, for instance, think of the joint density mentioned above. The marginal density $f(y_1)$ is given by:

$$f(y_1) = \int ... \int f(y_1, ..., y_p) dy_2 ... dy_p.$$

Then, for p large, it is obvious that direct calculation would be either time-consuming or simply impossible! Nevertheless, instead of calculating the posterior density (or the marginal posterior densities) directly, one can draw from conditional densities for all the parameters, given the remaining, in order to obtain a draw from the desired posterior density. This is what the Gibbs Sampler obtains. The method is discussed below.

2.3 The Gibbs Sampler

Suppose we need to calculate the joint density $f(\mathbf{y}) = f(y_1, ..., y_p)$ (or the marginal posterior distributions from a desired density) but only the full conditional distributions $f(y_i|\mathbf{y}_{-i}), \mathbf{y}_{-i}$ denotes all the components except y_i, of the desired quantities are available to sample from. These conditionals can be easily calculated from the product: *likelihood* $\times$ *priors* for each parameter, if one ignores the terms which do not include the parameter of interest . Then, given some initial values for all of the quantities of interest, we can take a random value from the first conditional distribution as if the other quantities were equal to the respective initial values. We then replace the initial value of the first quantity with the first sampled value and the same work is repeated for the next conditional distributions until a value has been sampled for all the desired quantities (given, in each step, the updated values-or the most recent value of the quantity that has been generated -for all the remaining).

So, an iteration has been completed and if the algorithm runs for a sufficient number of iterations, say R, then a Markov chain for each quantity will have been generated. It can be proved that these draws converge to the posterior distribution of interest.

The algorithm can be presented as follows:

$$1.\textit{pick initial values } \mathbf{y}^0 = (y_1^0, ..., y_p^0)$$

$$2.\textit{for } i = 1 \textit{ to a sufficiently large } \mathrm{R} \textit{ do}$$

$$\textit{make the random drawings}$$

$$\left\{ \begin{array}{c} y_1^i \text{ from } f(y_1|\mathbf{y}_{-1}^{i-1}) \\ y_2^i \text{ from } f(y_2|y_1^i, y_3^{i-1}, ..., y_p^{i-1}) \\ y_3^i \text{ from } f(y_3|y_1^i, y_2^i, y_4^{i-1}, ..., y_p^{i-1}) \\ . \\ . \\ . \\ y_p^i \text{ from } f(y_p|\mathbf{y}_{-p}^i) \end{array} \right\}$$

As R approaches to infinity, the joint distribution of $\mathbf{y}^i$ can be shown to approach the joint distribution of $\mathbf{y}$ so that, these drawings can be used to compute posterior moments and density estimates.

It must be mentioned that the total number R of iterations which are needed in order to obtain convergence depends on the nature of the problem, the correlation between the quantities, the autocorrelation of each quantity etc. With respect to the initial values that are used, the Sampler will finally converge to the true posterior distribution even with starting values which are far from the reality but, in this case, the total number of the required iterations will be much larger due to the higher autocorrelation within the chains which will be formed.

For details about the Gibbs Sampler, see Geman and Geman (1984), Gelfand et al.

(1990), Casella and George (1992), Smith and Roberts (1993) among others.

2.4 The Metropolis-Hastings Algorithm

In the Metropolis-Hastings algorithm, the aim is the same as in the Gibbs Sampler, but this algorithm is used when the conditional distributions, which we would like to sample from, do not have the form of a known distribution. Then, a transition probability function $q(y,y')$ is used in order to accept a generated value y' and replace y with y'.

So, if $y^i = y$, y' drawn from $q(y,y')$ is considered as a proposed possible value for y^{i+1} and is accepted with probability

$$\alpha(y,y') = \left\{ \begin{array}{ll} \min\left\{f(y')q(y',y)/f(y)q(y,y'),1\right\} & \text{if } f(y)q(y,y') > 0 \\ 1 & \text{if } f(y)q(y,y') < 0 \end{array} \right\}$$

A sufficient condition for $f(y)$ to be the equilibrium distribution of the constructed chain is that if $q(y,y')$ to be chosen to be irreducible and aperiodic on a suitable state space. Different functions for $q(y,y')$ can be used and can lead to different specific algorithms. For instance, if $q(y,y') = q(y',y)$, then $\alpha(y,y') = \min\left\{\frac{f(y')}{f(y)},1\right\}$ which is the Metropolis algorithm.

For details about the Metropolis-Hastings algorithm, see Chib and Greenberg (1995).

Chapter 3

Discrete Choice Modelling

3.1 Introduction

The discrete choice models are widely used in econometrics, as a special case of the general latent variables models. It is a case where there is a number of explanatory variables, in the form of vectors which are supposed to influence a latent binary vector (the dependent variable-vector). Thus, it is a form of the latent multivariate regression model. This chapter presents the general framework about the discrete choice models and a review on the most widely used models; the multinomial logit model (MNL), the nested logit model and the multinomial probit model. For presentations of discrete choice models, see Maddala (1983), McFadden (1980).

This background and the review that is presented, is necessary for the complete presentation of the discrete choice models and it will lead us to the multiranked probit model that is proposed in the present thesis (which is the subject of the next chapter). The first two models that are presented in sections 3.3 and 3.4 were mostly used in the past, due to the complexity of the multinomial probit model (chapter 3.5). Nevertheless, since this model can also be easily applied, due to the multinomial probit Sampler (MNP), its advantages make it more promising.

3.2 Discrete Choice Models

Let N be the number of individuals of the sample. Suppose that the ith ($i = 1, ..., N$) individual chooses among a set of p mutually exclusive alternatives (for example, brands of a product) ($j = 1, ..., p$), using some given values of K attributes (for example, price). Alternative j is chosen if the utility of this alternative is larger than the utilities of the remaining $(p-1)$ alternatives, given some particular values for the K attributes for each alternative and some characteristics of the ith individual which may also influence the choice (for example, income). For convenience of notation, the total number of attributes and the individuals' characteristics will be denoted as K, since both the alternatives' attributes and individuals' characteristics are the regressor values for the utilities which are formed.

For each choice occasion, a $(p \times 1)$ latent normal vector $\mathbf{z}_i$ is present which represents the unobserved utilities, and choice of alternative j is observed if the jth component of $\mathbf{z}_i$ is larger than all other components. Note that, the investigator finally observes N multinomial vectors, one per individual, where each multinomial vector $\mathbf{y}_i$, ($i = 1, ..., N$ individuals) consists of p rows; each row corresponds to an alternative and its element is equal to 1 if the alternative is the most preferable by the respective individual, and 0 otherwise. However, he does not observe the individual utilities $\mathbf{z}_i$. The model can be written as:

$$\mathbf{z}_i = \mathbf{R}_i \boldsymbol{\beta} + \mathbf{u}_i \text{ , i=1,...,N,} \tag{3.1}$$

and the element of the (p$\times$1) vector $\mathbf{y}_i$ which corresponds to the jth alternative is given by (having in mind that the probability of ties in the normal vector is practically zero):

$$y_{ij}= \begin{cases} 1 \text{ if } z_{ij} = \max \text{ of all the elements in } \mathbf{z}_i, \\ 0 \text{ otherwise} \end{cases}$$

where $\mathbf{R}_i$ is a $p{\times}K$ matrix which consists of the ith individual characteristics and of the regressor values which the ith individual faces for the alternatives (alternative-specific attributes) and $\mathbf{u}_i$ is an error term; a stochastic element the distribution of which we shall discuss later. In fact, different assumptions for the distribution of this term lead to the different models which will be discussed.

So, we need to estimate the vector $\boldsymbol{\beta}$, the individual utilities and, consequently, the probability P_{ij} that individual i will choose alternative j. Nevertheless, one can easily notice that an identification problem arises here; any shift with location m and scale $c{>}0$ from $\mathbf{z}_i$ to $m\iota + c\mathbf{z}_i$, $\iota'_{(1\times p)} = (1, ..., 1)$, will leave the observed choices unchanged and, hence, the problem is unidentifiable. The standard solution for this problem of the latent variables is to work with the differenced system which is achieved by letting $\mathbf{w}'_i = (w_{i1}, ..., w_{i,p-1})$ where $\mathbf{w}_i$ is a (p-1)×1 latent normal vector and $w_{ij} = z_{ij} - z_{ip}$, or, in words, $\mathbf{w}_i$ is the latent normal vector which comes from $\mathbf{z}_i$ if we subtract each element of $\mathbf{z}_i$ with the last element of the vector (i.e. $\mathbf{w}_i$ is the vector of the differences). The last element of $\mathbf{w}_i$ is, obviously, always 0 and, hence, it is omitted.

Hence, equation (3.1) can be rewritten as:

$$\mathbf{w}_i = \mathbf{X}_i\boldsymbol{\beta} + \boldsymbol{\varepsilon}_i \ , \tag{3.2}$$

$$\mathrm{d}_i = \begin{cases} 0 \text{ if } \mathbf{w}_i < 0 \\ \\ \text{index of } \ \max(w_{ij}, \ j = 1, ..., p-1) \text{ otherwise} \end{cases}$$

where $\mathbf{X}_i$ is of dimension (p-1)×K which is obtained from $\mathbf{R}_i$ of dimension p×K by subtracting the pth row from the first (p-1) rows, $\boldsymbol{\varepsilon}_i$ is the stochastic part which corresponds to $\mathbf{u}_i$ in equation (3.1) and d_i is a number which simply includes the information of the multinomial vector $\mathbf{y}_i$; it is equal to 0 if the last alternative (whose utility has been subtracted from the remaining) has been chosen by individual i, or, in the case where one of the first $(p-1)$ alternatives has been chosen, it is equal to the index which corresponds to the row that represents the chosen alternative.

Note that $\mathbf{w}_i$ can be thought of as the set of utilities for individual i, given that the utility of the last alternative is zero, or, in other words, the utilities of the $(p-1)$ first alternatives (the value that is formed for these utilities) with respect to the last utility if it is fixed to the constant value 0.

But, even in this differenced system, one can notice that, scaling $\mathbf{w}_i$ by a positive constant c, the value of d_i (i.e. the choice) will not change. Then, for example, if we suppose for the error term that $\boldsymbol{\varepsilon}_i \sim N(\mathbf{0}, \Sigma)$, the parameters $(\boldsymbol{\beta}, \Sigma)$ have to be estimated for the model but they are observationally equivalent to $(c\boldsymbol{\beta}, c^2\Sigma)$ for all c>0. The proposed multiranked probit model in chapter 4, deals with this problem and works with fully identified parameters.

With respect to the probability P_{ij}, let $V_i = X_i\boldsymbol{\beta}$ be the non-stochastic part in (3.2) that reflects the "representative" tastes of the population, whereas $\boldsymbol{\varepsilon}_i$ is stochastic and reflects the idiosyncrasy of individual i in tastes for the alternative with some particular attributes.

Then, the probability of the jth alternative to be chosen by individual i is:

$$\begin{aligned} P_{ij} &= P[V_{ij} + \varepsilon_{ij} > V_{ik} + \varepsilon_{ik},\ for\ all\ k \neq j] \\ &= P[\varepsilon_{ik} - \varepsilon_{ij} < V_{ij} - V_{ik},\ for\ all\ k \neq j]. \end{aligned}$$

In addition, if $F(\varepsilon_1, ..., \varepsilon_p)$ denotes the joint cumulative distribution function over the values ε_j for $j = 1, ...,$p and F_j the partial derivative of F with respect to the jth argument, the probability P_{ij} can be written:

$$P_{ij} = \int_{\varepsilon=-\infty}^{+\infty} F_j(\varepsilon + V_j - V_1, ..., \varepsilon + V_j - V_p) d\varepsilon \tag{3.3}$$

Different discrete choice models can yield if one assumes F to be a different cumulative distribution; the multivariate normal distribution and the extreme value distribution will be mentioned.

3.3 The Multinomial Logit Model (MNL)

The assumption that each ε_{ij} is distributed independently and identically in accordance with the extreme value distribution yields the multinomial logit model. This is given in the following lemma (see McFadden, 1973).

Lemma 1 *Suppose that each member of a population of utility-maximizing consumers has a utility function* $\mathbf{z}_i = \mathbf{V}_i + \boldsymbol{\varepsilon}_i$, *i=1,...,N which represents individuals, where* $\boldsymbol{\varepsilon}_i$ *is a function that varies randomly in the population with the property that in each possible alternative set* $B = \{x_1, ..., x_p\}$, *the values* $\varepsilon_{ij}, j = 1, ..., p$ *are independently identically distributed with the extreme value distribution with density* $f(\varepsilon_{ij}) = \exp\{-\varepsilon_{ij}\} \exp\{-exp\{-\varepsilon_{ij}\}\}$ *and cumulative function* $F(\varepsilon_{ij}) = exp\{-exp\{-\varepsilon_{ij}\}\}$. *Then, the selection probabilities given by (3.3) satisfy the equation* $P_{ij} = \frac{e^{V_{ij}}}{\sum_{k \in B} e^{V_{ik}}}$.

Proof. The probability of alternative j to be chosen from individual i is

$$P_{ij} = P[V_{ij} + \varepsilon_{ij} > V_{ik} + \varepsilon_{ik},\ for\ all\ k \neq j],\ (k, j = 1, ..., p\ and\ i = 1, ..., N)$$

$$= P[\varepsilon_{ik} - \varepsilon_{ij} < V_{ij} - V_{ik}\ for\ all\ k \neq j].$$

Suppose that ε_{ij} takes, for the moment, the value s. Then, the probability P_{ij}, i.e. the probability that alternative j is chosen, is the probability that each ε_{ik} is less than

$s + V_{ij} - V_{ik}$ for all k, k≠j, and is equal to the probability:

$$P[\varepsilon_{ik} < s + V_{ij} - V_{ik},\ for\ all\ k \neq j].$$

The probability that $\varepsilon_{ij} = s$ and, simultaneously, that $\varepsilon_{ik} < s + V_{ij} - V_{ik}$ is *the density of* ε_{ij} *evaluated at* s *multiplied by the cumulative distribution for each* ε_{ik} *except* ε_{ij} *evaluated at* $s + V_{ij} - V_{ik}$:

$$e^{-s}e^{-e^{-s}}\Pi_{k\neq j}\exp(-e^{-(s+V_{ij}-V_{ik})})$$

which is equal to:

$$e^{-s}\Pi_{k\in B}\exp(-e^{-(s+V_{ij}-V_{ik})})$$

because, when $k = j$, then

$$\exp(-e^{-(s+V_{ij}-V_{ik})}) = e^{-e^{-s}}.$$

Since the random variable ε_{ij} is not necessarily equal to s, but it can take any value within its range, the probability P_{ij} is then:

$$P_{ij} = \int_{s=-\infty}^{+\infty} \exp^{-s} \Pi_{k\in B}\exp(-e^{-(s+V_{ij}-V_{ik})})ds$$

$$= \int_{s=-\infty}^{+\infty} \exp^{-s} \exp(-\sum_{k\in B} e^{-(s+V_{ij}-V_{ik})})ds$$

$$= \int_{s=-\infty}^{+\infty} \exp^{-s} \exp(-e^{-s}\sum_{k\in B} e^{-(V_{ij}-V_{ik})})ds.$$

Let $e^{-s} = t$. Then,

$$-e^{-s} ds = dt \Longrightarrow ds = -\frac{dt}{e^{-s}} = -\frac{dt}{t}.$$

When s$\to \infty$, then t$\to$0 and, when s$\to$-∞, t$\to \infty$. So, the probability P_{ij} is:

$$P_{ij} = \int_{-\infty}^{0} t \exp(-t \sum_{k \in B} e^{-(V_{ij}-V_{ik})})(-\frac{dt}{t})$$

$$= \int_{0}^{\infty} \exp(-t \sum_{k \in B} e^{-(V_{ij}-V_{ik})}) dt$$

$$= \left[\frac{\exp(-t \sum_{k \in B} e^{-(V_{ij}-V_{ik})})}{-\sum_{k \in B} e^{-(V_{ij}-V_{ik})}} \right]_{0}^{\infty}$$

$$= \frac{1}{\sum_{k \in B} e^{-(V_{ij}-V_{ik})}}$$

$$= \frac{\exp^{V_{ij}}}{\sum_{k \in B} \exp^{V_{ik}}}. \blacksquare$$

The probabilities for each alternative to be chosen which were found from the MNL model, exhibit the property of Independence of Irrelevant Alternatives (IIA) which is given in the following axiom. This axiom was introduced by Luce (1959) and it is necessary for the specification of the model (see also McFadden, 1973, Train, 1986, Hensher and Johnson, 1979).

Axiom 1 *Independence of Irrelevant Alternatives: For all possible alternative sets B,* *measured attributes s, and members x and y of B,*

$$P(x|s,\{x,y\}) \cdot P(y|s,B) = P(y|s,\{x,y\}) \cdot P(x|s,B).$$

That is, when $P(x|s,B)$ is positive (there is a little loss of generality in assuming that the selection probabilities are positive since, empirically, a zero probability is indistinguishable from one that is very small), then

$$\frac{P(y|s,\{x,y\})}{P(x|s,\{x,y\})} = \frac{P(y|s,B)}{P(x|s,B)}.$$

Also, for two alternatives x and y in B, the ratio

$$\frac{P_{ix}}{P_{iy}} = \frac{\exp^{V_{ix}} / \sum_{k\in B} \exp^{V_{ik}}}{\exp^{V_{iy}} / \sum_{k\in B} \exp^{V_{ik}}}$$

$$= \frac{\exp^{V_{ix}}}{\exp^{V_{iy}}} = \exp^{V_{ix}-V_{iy}}$$

depends only on alternatives x and y, or, in other words, the ratios of probabilities are necessarily the same no matter what other alternatives are in B or what the characteristics of other alternatives are.

Clearly, a model with such a property is very simple to be applied but is obviously inappropriate for some situations and can lead to wrong results. For instance, consider the example given by Train (1986) about a two-choice problem where a traveller has to choose, given some values of the attributes, between auto (a) and blue bus (bb). Suppose that the utilities for these two alternatives are equal and, hence, $P_a = P_{bb} = \frac{1}{2}$, $\frac{P_a}{P_{bb}} = 1$.

Suppose now, that a red bus is introduced and that the traveller considers the red bus (rb) to be exactly like the blue bus. Then, the ratio $\frac{P_{bb}}{P_{rb}} = 1$ but also the ratio $\frac{P_a}{P_{bb}}$ is equal to one due to the IIA property (the presence of a new alternative does not affect the ratio of the probabilities of the two alternatives). Hence, $P_a = P_{bb} = P_{rb} = \frac{1}{3}$. Nevertheless, in this case, we would expect P_a to be unchanged and P_{bb} (before the third alternative is introduced) to be split: $P_{bb} = P_{rb} = \frac{1}{4}$. So, the multinomial logit model overestimates P_{bb} and P_{rb} and underestimates P_a.

So, the MNL model is inappropriate for situations as in the one described above. A solution is to apply the MNL in subsets of the alternative set B (nested logit) taking into account that the coefficients $\boldsymbol{\beta}$ of the attributes should be the same for all the alternatives in B (see section 3.4). Another solution is the multinomial probit model (see section 3.5).

3.3.1 Estimation in MNL

Recall equation (3.1) where $y_{ij} = 1$ if individual i chooses alternative j and 0 otherwise. The response vector is a $(p \times 1)$ multinomial vector. The likelihood function is then:

$$L = \prod_i \prod_k P_{ik}^{y_{ik}},$$

where $i = 1, ..., N$ represents individuals and $j, k = 1, ..., p$ represent alternatives.

The log-likelihood is given by:

$$\ln L = \sum_i \sum_k y_{ik} \ln P_{ik} = \sum_i \sum_k y_{ik} \ln \frac{\exp^{r_{ik}\boldsymbol{\beta}}}{\sum_{j \in B} \exp^{r_{ij}\boldsymbol{\beta}}}$$

$$= \sum_i \sum_k y_{ik}(r_{ik}\boldsymbol{\beta} - \ln \sum_{j \in B} \exp^{r_{ij}\boldsymbol{\beta}})$$

$$= \sum_i [\boldsymbol{\beta}(\sum_k y_{ik} r_{ik}) - \ln \sum_{j \in B} \exp^{r_{ij}\boldsymbol{\beta}}]$$

So,

$$\frac{\partial \ln L}{\partial \boldsymbol{\beta}} = \sum_i [(\sum_k y_{ik} r_{ik}) - \frac{\sum_{j \in B} r_{ij} \exp^{r_{ij}\boldsymbol{\beta}}}{\sum_{j \in B} \exp^{r_{ij}\boldsymbol{\beta}}}]$$

$$=\sum_i [\sum_k (y_{ik} r_{ik}) - \sum_j r_{ij} P_{ij})]$$

and, since k and j both indicate the alternatives, this is equal to:

$$\sum_i \sum_k (y_{ik-} P_{ik}) r_{ik}.$$

Furthermore, the second derivative is given by:

$$\frac{\partial^2 \ln L}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} = - \sum_i \sum_k (r_{ik-} \bar{r_i})' P_{ik} (r_{ik-} \bar{r_i})$$

where $\bar{r_i} = \sum_k r_{ik} P_{ik}$, r_{ik} is the vector of K individual's characteristics and attributes which individual i faced for alternative k.

This last equation is the negative of a weighted moment matrix of independent variables, it is negative semidefinite and lnL is concave in $\boldsymbol{\beta}$, meaning that is maximized at any critical point where $\frac{\partial \ln L}{\partial \boldsymbol{\beta}} = \mathbf{0}$. If further, the matrix $\frac{\partial^2 \ln L}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'}$ is nonsingular, lnL has a unique maximum in $\boldsymbol{\beta}$ (see McFadden, 1973).

This model had been widely used in the past due to its simplicity compared with the multinomial probit model. For some applications, see, for example, Louviere and Woodworth (1983), Eagle (1984), Louviere (1984).

The analysis of these data sets is usually performed by the use of discrete choice models which include such diverse fields in econometrics as shopping behavior, housing choices and travel mode choices (see, for example McFadden, 1978, Louviere, 1984). A widely used model, mainly due to its simplicity, is the multinomial logit model despite its property of independence of irrelevant alternatives that is often violated (see Hensher and Johnson, 1979, Maddala, 1983, Train, 1986).

3.4 The Nested Logit Model (GEV)

As it has been shown, the logit model is inappropriate in situations, where the Independence of Irrelevant Alternatives (IIA) property is assumed but it is not true. However, if the alternatives were partitioned into subsets such that the ratio of probabilities for any two alternatives that are in the same subset is independent of the existence or characteristics of other alternatives, then another type of logit model, the nested logit model (GEV) could be derived. In that case, the IIA property holds within each subset but not between alternatives from different subsets.

In the GEV model, the errors ε_{ij} are assumed to follow the generalized extreme value distribution (that's why, the name GEV is used); the marginal distribution of each ε_{ij} is the extreme value distribution but the $\varepsilon_{ij}s$ in the same subset are correlated and the $\varepsilon_{ij}s$ between different subsets are uncorrelated. We assume that the set of alternatives can be partitioned into K subsets named B_k, k=1,...,K.

The joint cumulative distribution of the random variables ε_{ij} for all j in the set of alternatives is then:

$$\exp\left\{-\sum_{k=1}^{K} a_k(\sum_{b\in B_k} e^{-\frac{b}{\lambda_k}})\right\},$$

where K is the number of subsets, λ_k is a measure which drops when the correlation of ε_{ij}'s within subset k arises ((1-λ_k) is a measure of the correlation). Then, the probability for alternative j in subset B_k to be chosen is

$$P_{ij} = \frac{\exp^{V_{ij}/\lambda_k}(\sum_{j\in B_k}\exp^{V_{ik}/\lambda_k})^{\lambda_k-1}}{\sum_l(\sum_{j\in B_l}\exp^{V_{jk}/\lambda_l})^{\lambda_l-1}}.$$

For a proof, see McFadden (1978).

Moreover, if the observed component of the utility is written as a term common for all alternatives within a subset plus a part that differs for alternatives within subsets, plus ε_{ij}, i.e.

$$U_{ij} = W_{ij}^{(k)} + \lambda_k Y_{ij}^{(k)} + \varepsilon_{ij} \text{ for } \; j \text{ in } \; B_k,$$

then the probability that alternative j in subset B_k is chosen is the product of the probability that subset B_k will be chosen times the probability that alternative j within subset k will be chosen.

That is,

$$P_{ij} = P_{ij|B_k} \cdot P_{B_k} = \frac{\exp(Y_{ij}^{(k)})}{\sum_{j \in B_k} \exp(Y_{ij}^{(k)})} \frac{\exp(W_{ij}^{(k)} + \lambda_k I_k)}{\sum_{l=1}^{K} \exp(W_{ij}^{l} + \lambda_l I_l)}$$

where $I_k = \ln \sum_{j \in B_k} e^{Y_{ij}^k}$, $P_{ij|B_k}$ is the probability that alternative j within subset B_k will be chosen by individual i and P_{B_k} is the probability for B_k to be chosen.

It is clear now, that the GEV model uses the logit model within subsets (note the conditional probability $P_{ij|B_k}$) and the same model to handle the different subsets as alternatives of a general set (note the marginal probability P_{B_k}, where I_k can be thought of as the average utility that an individual expects from subset k; it is usually called "the inclusive value"). So, the IIA property is assumed again for each one of the logit models which arise. Hence, the problem has been reduced but has not been eliminated.

3.5 The Multinomial Probit Model (MNP)

It has already been mentioned that the Independence of Irrelevant Alternatives (IIA) property implies that the unobserved components in ε_{ij} in (3.2), for all j, and for a particular individual i, are assumed to have the same distribution, with the same mean and variance and to be uncorrelated making the MNL model unrealistic for most of the cases.

On the contrary, the multinomial probit model (MNP) arises if one relaxes these assumptions letting ε_{ij} to be jointly normal distributed with zero mean vector and variance-covariance matrix Σ.

The model was first proposed by Thurstone (1927) but has not been widely used due to its computational difficulties.

In this model, the density function of the vector $\boldsymbol{\varepsilon}_i$ is:

$$f(\boldsymbol{\varepsilon}_i) = (2\pi)^{-\frac{1}{2}p} \, |\Sigma|^{-\frac{1}{2}} \exp[-\frac{1}{2} \, \boldsymbol{\varepsilon}_i{}' \, \Sigma^{-1} \, \boldsymbol{\varepsilon}_i]$$

where p is the number of alternatives.

3.5.1 Choice probabilities and likelihood function

Recall the probability

$$P_{ij} = P[V_{ij} + \varepsilon_{ij} > V_{ik} + \varepsilon_{ik} \text{ for all } k \neq j]$$

$$= P[\varepsilon_{ik} < \varepsilon_{ij} + V_{ij} - V_{ik} \text{ for all } k \neq j]$$

and, finally, suppose for simplicity that ε_{ij} is known. That is, the probability is the cumulative distribution function of ε_{ik} evaluated at $\varepsilon_{ij} + V_{ij} - V_{ik}$ for all k, i.e. the probability P_{ij}, given a particular value of ε_{ij}, is given by the multiple integral over all $\boldsymbol{\varepsilon}'s$ except ε_{ij} :

$$\int\limits_{\varepsilon_{i1}=-\infty}^{\varepsilon_{ij}+V_{ij}-V_{i1}} \int\limits_{\varepsilon_{i2}=-\infty}^{\varepsilon_{ij}+V_{ij}-V_{i2}} \cdots \int\limits_{\varepsilon_{ip}=-\infty}^{\varepsilon_{ij}+V_{ij}-V_{ip}} f(\boldsymbol{\varepsilon}_i) d\varepsilon_{ip} ... d\varepsilon_{i1}.$$

Since the value ε_{ij} is not known, the probability P_{ij} is then :

$$\int_{\varepsilon_{ij}=-\infty}^{\infty} \int\limits_{\varepsilon_{i1}=-\infty}^{\varepsilon_{ij}+V_{ij}-V_{i1}} \int\limits_{\varepsilon_{i2}=-\infty}^{\varepsilon_{ij}+V_{ij}-V_{i2}} \cdots \int\limits_{\varepsilon_{ip}=-\infty}^{\varepsilon_{ij}+V_{ij}-V_{ip}} f(\boldsymbol{\varepsilon}_i) d\varepsilon_{ip} ... d\varepsilon_{i1} d\varepsilon_{ij}.$$

Substituting this formula into the log-likelihood function $\sum_i \sum_k y_{ik} \ln P_{ik}$, we take the function which has to be maximized in order to obtain estimators for the parameters:

$$\sum_i \sum_k y_{ik} \ln(\int_{\varepsilon_{ij}=-\infty}^{\infty} \int_{\varepsilon_{i1=-\infty}}^{\varepsilon_{ij}+V_{ij}-V_{i1}} \int_{\varepsilon_{i2=-\infty}}^{\varepsilon_{ij}+V_{ij}-V_{i2}} \ldots \int_{\varepsilon_{ip=-\infty}}^{\varepsilon_{ij}+V_{ij}-V_{ip}} (2\pi)^{-\frac{1}{2}p} \left|\Sigma\right|^{-\frac{1}{2}} \times$$

$$\exp[-\frac{1}{2} \boldsymbol{\varepsilon}_i{}' \, \Sigma^{-1} \, \boldsymbol{\varepsilon}_i] d\varepsilon_{ip} \ldots d\varepsilon_{i1} d\varepsilon_{ij})$$

The expense of the estimation in such a model using classical methods is obvious.

Classical simulation approaches have been proposed to overcome the computational burden of the estimation with respect to the multinomial probit model (see Hajivassiliou and Ruud, 1994, for an overview of these methods; McFadden, 1989; McFadden and Ruud, 1994, the method of simulated scores (Hajivassiliou and McFadden, 1990). Alternatively, Albert and Chib (1993) developed a Bayesian data augmentation method combined with the Gibbs sampler (Smith and Roberts, 1993) to obtain the latent data multinomial probit model. This model reconstructs the latent utilities $\mathbf{w}$ via the available information provided by $\mathbf{y}$ whereas it is based on the edifying note that, conditional on $\mathbf{w}$, the model is simplified to a standard Bayesian multivariate regression analysis. Geweke, Keane and Runkle (1994) have shown that, although these methods are not less computationally demanding than a Bayesian one, their asymptotic approximations can be inaccurate. We will focus on this Bayesian approach in the next section, since it provides slightly better results with respect to the accuracy.

3.6 Bayesian Discrete Choice Models

Interest has recently focused on the more powerful and realistic multinomial probit model so as to ease its computational burden; McCullogh and Rossi (1994); Chib and Greenberg (1998); Nobile (1998); Allenby and Rossi (1999); Chen and Dey (2000).

3.6.1 General Framework and Theoretical background

The MNP Sampler (McCullogh and Rossi, 1994) is the Bayesian approach for estimating the parameters in the multinomial probit model.

Recall that:

$$\mathbf{w}_i = \mathbf{X}_i\boldsymbol{\beta} + \boldsymbol{\varepsilon}_i,$$

$$\mathrm{d}_i = \begin{cases} 0 & \text{if } \mathbf{w}_i < 0 \\ \\ index\ of\ \max(w_{ij},\ j = 1, ..., p-1) & \text{otherwise} \end{cases}$$

where $\boldsymbol{\varepsilon}_i \sim N(\mathbf{0}, \Sigma)$.

The first problem which arises with respect to this equation is that the utilities $\mathbf{w}_i$ (i.e. the utilities of the $(p-1)$ alternatives, given that the pth utility is 0, for individual i) have not been observed. The solution for this provides the data augmentation step; a way of changing the multinomial vector into a normal vector.

So, after this change, the equation $\mathbf{w}_i = \mathbf{X}_i\boldsymbol{\beta} + \boldsymbol{\varepsilon}_i$ is a regression equation with multivariate terms, where the parameters $\boldsymbol{\beta}$ and Σ and the latent variables $\mathbf{w}_i$ have to be estimated.

In the Bayesian framework, this requires draws from the conditional distributions of each one of the parameters $\boldsymbol{\beta}$ and Σ and the latent variables $\mathbf{w}_i$. Hence, $\boldsymbol{\beta}$ can be drawn from a multivariate normal distribution, Σ can be drawn from an inverted Wishart distribution or, for convenience, $\Sigma^{-1} = G$ can be drawn from a Wishart distribution and, finally, the latent vector $\mathbf{w}_i$ can be drawn from a multivariate normal distribution, with covariance matrix Σ or, for convenience, each w_{ij}, j=1,...,p-1, can be drawn from a univariate normal distribution, conditional on the $(p-2)$ remaining latent variables (see theorem below).

The definition of the Wishart distribution and the theorem that was mentioned, with respect to the multivariate normal distribution, are given in the following:

Definition 2 *Wishart distribution $W(k-p-1, R)$: A $p \times p$ positive definite matrix X has a Wishart distribution if*

$$f(X|k,R) \propto |R|^{k/2}\,|X|\,\frac{k-p-1}{2}\exp[-\frac{1}{2}tr(RX)]$$

with k>p degrees of freedom and R a $p \times p$ symmetric non-singular matrix (scale matrix). E(X)=$\left(\frac{R^{-1}}{k}\right)$ and X^{-1} is a covariance matrix $\left(E(X^{-1}) = \frac{R}{k-p-1}\right)$.

The Wishart distribution is a multivariate extension of a chi-square distribution. For details, see De Groot (1970).

Theorem 3 *Let X=$\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \sim N_p(\boldsymbol{\mu}, \Sigma)$ a partition of X into two parts, where*

$$\boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix}, \Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}, |\Sigma_{22}| > 0$$

Then, the conditional distribution of X_1, given $X_2 = \mathbf{x}_2$ is multivariate normal with

$$mean = \boldsymbol{\mu}_1 + \Sigma_{12}(\Sigma_{22})^{-1}(\mathbf{x}_2 - \boldsymbol{\mu}_2)$$

and

$$variance = \Sigma_{11} - \Sigma_{12}(\Sigma_{22})^{-1}\Sigma_{21}.$$

For a proof and for other details about the theorem and the definition above, see Johnson and Wichern (1988), pp. 131-132 and 143-144.

The MNP Sampler can now be described in detail in the next section.

3.6.2 Prior Distributions for the MNP Sampler

The Bayesian approach that deals with the MNP model, requires the specification of the priors over the parameters $(\boldsymbol{\beta}, \Sigma)$, computation of the posterior density and the MNP likelihood. Setting a normal prior on $\boldsymbol{\beta}$ and an independent Wishart on Σ^{-1}:

$$\boldsymbol{\beta} \sim N(\bar{\boldsymbol{\beta}}, A^{-1}) \text{ and } G = \Sigma^{-1} \sim W(v, V),$$

the formulae of these prior distributions for the parameters $\boldsymbol{\beta}$ and Σ^{-1} are analogous to:

$$p(\boldsymbol{\beta}|\bar{\boldsymbol{\beta}}, A) \propto |A|^{0.5} \exp\{-\frac{1}{2}(\boldsymbol{\beta} - \bar{\boldsymbol{\beta}})' A(\boldsymbol{\beta} - \bar{\boldsymbol{\beta}})\}$$

and

$$p(G|v, V) \propto |G|^{(v-p-1)/2} etr\{-\frac{1}{2}GV\}.$$

Then, the posterior density, which has to be calculated, is given by:

$$p(\boldsymbol{\beta}, \Sigma|\mathbf{y}_1, ..., \mathbf{y}_N, \mathbf{X}) \propto p(\boldsymbol{\beta}, \Sigma) \cdot \log -likelihood.$$

The MNP Sampler avoids the direct evaluation of multiple integrals that mentioned before.

3.6.3 The Data Augmentation Step

Recall the problem that has been mentioned; the investigator has observed $\mathbf{y}_i$ and $\mathbf{X}_i$ but $\mathbf{w}_i$ are unknown and, hence, the dependent variables' vector is, up to now, discrete. Albert and Chib (1993) suggested the data augmentation step as a solution for this problem; instead of the use of the multinomial vector $\mathbf{Y}_i$ they introduce an independent unobserved latent vector $\mathbf{Z}_i, i = 1, ..., N$, where $\mathbf{Z}_i = (z_{i1}, ..., z_{ip})$ are specified in such a way that, alternative j is chosen if $\mathrm{z}_{ij} > \mathrm{z}_{ik}$ for $k \neq j$, and $k, j = 1, ..., p$.

In the differenced system that is described here, this idea yields the latent vector $\mathbf{w}_i$,where $\mathbf{w}_i|\boldsymbol{\beta}, G_i,d_i$ is a (p-1)variate normal distribution truncated over the appropriate cone in R^{p-1} .The truncation is used so that, if $y_{ij} = 1$ (or, equivalently, $d_i = j$), then $w_{ij} > \max(w_{ik}, 0)$, $\forall\ k \neq j$, and, if $y_{ip} = 1$ (or, equivalently, $d_i = 0$), then $w_{ij} < 0$, $\forall\ j$ *and* $k,\ j = 1, ...p-1$.

For simplicity, one can draw from all the conditional distributions of each component of vector $\mathbf{w}_i$, given all the others and which will yield draws from the truncated (p-1)variate normal distribution (see theorem 1). In such a way, the discrete multinomial vector $\mathbf{Y}_i$ has been substituted by $\mathbf{w}_i$ which yields (after the convergence) the differences of the unobserved utilities among the (p-1) alternatives and the last pth alternative for the ith individual. Note that, the equation $\mathbf{w}_i = \mathbf{X}_i\boldsymbol{\beta} + \boldsymbol{\varepsilon}_i$ is now a multivariate regression equation with a continuous dependent (latent) vector.

3.6.4 The Steps of the Algorithm

Now since the differences among the unobserved utilities are not unknown any more, the MNP Sampler (McCullogh and Rossi, 1994) can be presented.

The Sampler consists of three main steps from which the successive draws from the conditional distributions will finally converge to the posterior distribution.

Step 1

Draw from N×(p-1) conditionals on $w_{ij}|\mathbf{w}_{i,(-j)}, \boldsymbol{\beta}, G_{,}d_i$, for $i = 1, ..., N$, $j = 1, ..., p-1$, where $\mathbf{w}_{i,(-j)}$ is the vector $\mathbf{w}_i$, of dimension (p-1)×1 except the jth component:

Using theorem 1 and setting

$$X_1 = w_{ij}\ and\ X_2 = \mathbf{w}_{i,(-j)},$$

the distribution of $\mathbf{w}_i$ can be written

$$\begin{bmatrix} w_{ij} \\ \mathbf{w}_{i(-j)} \end{bmatrix} \sim N_{p-1}\left(\begin{bmatrix} \mathbf{x}'_{ij}\boldsymbol{\beta} \\ X_{i(-j)}\boldsymbol{\beta} \end{bmatrix}, \begin{bmatrix} \sigma_{jj} & \sigma_{j(-j)} \\ \sigma_{(-j)j} & \Sigma_{(-j)(-j)} \end{bmatrix}\right)$$

and w_{ij} has a truncated normal distribution over the appropriate cone in R^{p-1}:

$$w_{ij}|\mathbf{w}_{i,-j}, \boldsymbol{\beta}, G, d_i \sim N(m_{ij}, \tau^2_{ij}),$$

where

$$m_{ij} = \mathbf{x}'_{ij}\boldsymbol{\beta} + (\sigma_{j(-j)}\Sigma^{-1}_{(-j)(-j)})(\mathbf{w}_{i(-j)} - X_{i(-j)}\boldsymbol{\beta}),$$

$$\tau^2_{ij} = \sigma_{jj} - \sigma_{j(-j)}\Sigma^{-1}_{(-j)(-j)}\sigma_{(-j)j}$$

and $\mathbf{x}_{ij}$ is the jth row of $\mathbf{X}_i$, $\mathbf{X}_{i(-j)}$ is the submatrix created by deleting the jth row from $\mathbf{X}_i$ and the submatrices in the covariance matrix of the distribution of $\mathbf{w}_i$ come from the partitioning of G, with respect to the jth argument:

$$G = \begin{bmatrix} \Sigma_{(-j)(-j)} & \sigma_{(-j)j} \\ \sigma_{j(-j)} & \sigma_{jj} \end{bmatrix}^{-1}.$$

Then, perform the following acceptance-rejection procedure which corresponds to the data augmentation step: if $\mathrm{d}_i \neq 0$, accept the sampled values if the highest value is positive and corresponds to the chosen alternative, i.e. the alternative which corresponds to the d_ith row. If $\mathrm{d}_i = 0$ (i.e. the last alternative, which has been subtracted and corresponds to utility equal to zero, has been chosen, so that, all the remaining utilities have to be lower than the chosen one, i.e. lower than 0), accept the sampled values if they are all negative. Otherwise, reject them and sample again from step 1.

Step 2

Draw from $\boldsymbol{\beta}|\mathbf{w}, G$:

Transform the regression equation

$$\mathbf{w}_i = \mathbf{X}_i\boldsymbol{\beta} + \boldsymbol{\varepsilon}_i,\ \ \boldsymbol{\varepsilon}_i \sim N(\mathbf{0}, G^{-1})$$

to an equation with i.i.d. errors using the Cholesky decomposition of G=CC$'$. To obtain this, premultiply the equation by C$'$:

$$C'\mathbf{w}_i = C'X_i\boldsymbol{\beta} + C'\boldsymbol{\varepsilon}_i, C'\boldsymbol{\varepsilon}_i \sim N(\mathbf{0}, I_{p-1}),$$

or, in stacked form:

$$C'\mathbf{w} = C'X\boldsymbol{\beta} + C'\boldsymbol{\varepsilon}, C'\boldsymbol{\varepsilon} \sim N(\mathbf{0}, I_{N(p-1)}).$$

Then, draw from the distribution:

$$\boldsymbol{\beta}|\mathbf{w}, G \sim N(\hat{\boldsymbol{\beta}}, \Sigma_{\beta}),$$

where

$$\Sigma_{\boldsymbol{\beta}} = [(C'X)'(C'X) + A]^{-1}$$

and

$$\hat{\boldsymbol{\beta}} = \Sigma_{\beta}[(C'X)'C'\mathbf{w} + A\bar{\boldsymbol{\beta}}].$$

Step 3

Draw from $G|\mathbf{w}, \boldsymbol{\beta}$:

Given $\boldsymbol{\beta}, \mathbf{w}$ (from steps 1 and 2) and X ,the $\boldsymbol{\varepsilon}_i' s$ can be calculated: $\boldsymbol{\varepsilon}_i = \mathbf{w}_i - X_i\boldsymbol{\beta}$.

The conditional Wishart distribution is then:

$$G|\mathbf{w},\boldsymbol{\beta} \sim W(v+N, V+\sum_{i=1}^{N}\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i').$$

In order to obtain a random matrix G drawn from this distribution we can use again the Cholesky decomposition of $(V+\sum_{i=1}^{N}\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i')^{-1} = S^{-1}$:

$$S^{-1} = LL'.$$

Moreover, if T is a lower triangular matrix with diagonal elements $t_{ii} \sim \sqrt{X^2_{v+N-i+1}}$ and draws from N(0,1) for the elements below the diagonal, we obtain the matrix

$$\tilde{V} = TT',$$

which is a random matrix from the standardized Wishart distribution with v+N degrees of freedom (i.e. Wishart with scale matrix the identity one):

$$\tilde{V} \sim W(v+N, I).$$

Finally, the random matrix G is given by:

$$G = L\tilde{V}L'.$$

For details of the procedure in step 3, see Johnson (1987), pp. 203-204.

Note that d_i has not been used in steps 2 and 3 because this information is included in $\mathbf{w}_i$.

3.6.5 Identification Problem for the Parameters

However, another problem arises here. Note that the full posterior over $\boldsymbol{\beta}$ and Σ has been computed and we have to report the marginal posterior of the identified (properly normalized) parameters $\frac{\boldsymbol{\beta}}{\sigma_{p-1}}$ and variance-covariance coefficients:

$$\frac{Cov(utility\ l,\ utility\ m)}{\sigma_{p-1}^2},$$

where l, m=1,...,p-1 and σ_{p-1} is the standard deviation of the utility for alternative $p-1$. That is, in order to obtain convergence, we have to divide each parameter with, say, the standard deviation (for $\boldsymbol{\beta}$) or the variance (for the covariances and the remaining variances) of the utility of one alternative. In other words, we have to keep this variance constant (equal to one). Otherwise, if we examine each of the parameters without division, the sequence of draws will follow a random walk without convergence due to the identification problem (recall that all of the sets $(c\boldsymbol{\beta}, c^2\Sigma)$ are observationally equivalent to $(\boldsymbol{\beta}, \Sigma)$). Moreover, proper priors have to be used to keep the sequences for the parameters away from going to infinity (c can go to infinity in the set $(c\boldsymbol{\beta}, c^2\Sigma)$ if one uses improper priors).

The necessity of the division of each sequence with the sequence of a variance in order to obtain convergence, demonstrates that the algorithm is a hybrid Gibbs Sampler and requires specific handling. Also, the use of proper priors may not be suitable if no prior information is available.

We believe that these are serious drawbacks of the method of McCullogh and Rossi (1994). In the proposed multiranked model in chapter 4, we provide an algorithm which improves the McCullogh and Rossi methodology.

3.7 Modifications for the MNP Sampler

Some modifications of the MNP Sampler, which attempt to solve the identification problem for the parameters of the model, have been proposed over the years for this identification problem; see Nobile (1998), McCullogh, Polson and Rossi (2000), Chen and Dey (1998, 2000), Chib and Greenberg (1998).

3.7.1 Gibbs Sampler with a Metropolis Step

It has been mentioned that, multiplying the set of parameters $(\boldsymbol{\beta}, \Sigma)$ by a positive constant c, a new set of parameters is yielded which is also valid. The restriction for c (to avoid the case where $c \to \infty$) comes, up to now, from the use of proper priors. Nobile (1998) suggests another restriction for the constant c by rescaling the set of parameters. That is, after a Gibbs cycle has yielded $\psi_0 = (\boldsymbol{\beta}, \Sigma)$, and before the next cycle, one may change the scale of ψ_0, taking care to respect the equilibrium distribution of the sampling chain. The algorithm is described below.

At first, perform a MNP Sampler step, as it has already been described, and take

$$\psi_0 = (\vartheta, w_{ij}),\ \vartheta = (\boldsymbol{\beta}, \Sigma).$$

Then, before the next MNP Sampler, perform a Metropolis step. Draw $\psi^* = c\psi_0$ from some distribution $h(\psi|\psi_0)$ and accept it with probability:

$$\min\left[\frac{f(\psi^*)}{f(\psi_0)}\frac{h(\psi_0|\psi^*)}{h(\psi^*|\psi_0)}, 1\right],$$

where $f(\psi) \propto k(data|\vartheta)\pi(\vartheta)$ is the equilibrium distribution of the Markov chain and $\frac{f(\psi^*)}{f(\psi_0)}$ depends only on the ratio between the prior distribution at ϑ_0 and $\vartheta^* = c\vartheta_0$ ($\frac{f(\psi^*)}{f(\psi_0)} = \frac{\pi(\vartheta^*)}{\pi(\vartheta_0)}$). Also, the candidate ψ^* is best obtained by sampling c. For instance, if c is drawn from an Exp(1) distribution, the acceptance probability of the Metropolis step is

$$\min\left[\frac{\pi(\vartheta^*)}{\pi(\vartheta_0)}\frac{1}{c}\exp(\frac{c^2-1}{c}),1\right].$$

This algorithm, with some slightly more expensive programming requirements, provides a faster convergence because c is restricted to vary in a smaller set of values.

Nevertheless, the problems which were mentioned at the end of the previous chapter are still unsolved.

3.7.2 A Modification by McCullogh and Rossi

An alternative algorithm has been suggested by McCullogh and Rossi (1994) in order to obtain fully identified parameters. In this algorithm, the first and the second step remain the same as in the MNP Sampler, that has been described, but the third step (draw for Σ) has been modified in order to restrict $\sigma_{11}(=\sigma_1^2)$ equal to 1.

Let $U=\varepsilon_1$ and $Z=(\varepsilon_2,...\varepsilon_{p-1})'$ so that $\boldsymbol{\varepsilon}=(U,Z)$. Σ indexes the joint distribution of (U,Z) which is $N(0,\Sigma)$. Also, let $\gamma=E(UZ)$ so as $Z|U\sim N(\frac{\gamma}{\sigma_{11}}U,\Sigma_Z-\gamma\gamma/\sigma_{11})$. Finally, let $\Phi=\Sigma_z-\gamma\gamma/\sigma_{11}$. There is one to one correspondence between Σ and $(\sigma_{11},\gamma,\Phi)$, hence, setting $\sigma_{11}=1$ and putting priors on γ and Φ, we obtain a prior on $(\Sigma|\sigma_{11}=1)$. Σ is then equal to $\begin{bmatrix}1 & \gamma' \\ \gamma & \Phi+\gamma\gamma'\end{bmatrix}$. Setting the priors

$$\gamma\sim N(\bar{\gamma},B^{-1})$$

and

$$\Phi^{-1}\sim W(k,C),$$

we obtain the conditional distributions which replace the third step of the MNP algorithm by using:

$$\Phi^{-1} \sim W(k+n, C + (Z - U\gamma')'(Z - U\gamma'))$$

$$\gamma \sim N(A_\gamma(vec(\Phi^{-1}Z'U) + B\gamma), A_\gamma)$$

where $A_\gamma = (U'U\Phi^{-1} + B)^{-1}$.

Although the use of this algorithm remedies the problems which have been mentioned, there is a difficulty in setting priors for γ and Φ^{-1}; these parameters are not very natural to think about. To see how one can get in trouble with the priors for γ and Φ^{-1}, consider the case where $p = 3$, Σ is 2×2 and the correlation $\rho = \frac{\gamma}{\sqrt{\Phi+\gamma^2}}$. If the prior for γ is much more diffuse than that of Φ, the prior distribution for ρ will concentrate near ±1.

Chapter 4

The proposed multiranked probit model

4.1 Introduction

This chapter presents the proposed multiranked probit model. Almost all the existing analysis techniques are concerned with the modeling of the utilities using the information of just the discrete choice (the response that indicates the one most preferable alternative) of each multinomial response data point, whereas the literature in modeling ranking or rating responses is still sparse. Of course, in each case of different alternatives attributes, only the first choice is the one that will actually be used by a particular decision maker in practice. However, compared to the discrete choice, a ranking response contains useful additional information about the structure of the relation between the utilities and the attributes of the competing alternatives.

The model we propose deals with data which are of ranking form, thus, it is a generalization of the discrete choice models described in the previous chapter. Hence, this model can also be easily used in the data of the discrete choice form. The resulting model formulations give rise to the so-called multiranked probit model which emerges from a series of ranking responses in a set of hypothetical scenarios. In summary, in this model, we enhance the multinomial probit model with the embodiment of a utility threshold parameter which deals realistically with ranking responses, intransitivity of indifference among alternatives, or ties. Moreover, we ensure identifiable parameters for the covariance matrix of the underlying utility vectors, we include a hierarchical step that models the unit-specific utility thresholds as exchangeably distributed and, finally, we permit the use of heavy tail distributions for the stochastic error term. Our proposed methodology is Bayesian and the implementation tool adopted is MCMC.

4.2 Methodology

4.2.1 Aggregate Analyses

An essential element of the first step in stated preference data collection is the design of an appropriate experiment that includes some prespecified values of the continuous or discrete covariates-attributes of the alternatives. These attributes normally cover as many aspects transport experts believe that affect the probabilities of the alternatives chosen by a decision maker as possible. Depending on the experimental design that is used, some different combinations of these attributes values form different hypothetical scenarios and each decision maker may face all or a subset of the hypothetical scenarios originated by the design.

At the data collection step, each respondent is faced with a set of mutually exclusive and collectively exhaustive alternatives and either selects the most preferable alternative or orders all the competitive alternatives or rates them, according to the corresponding utilities. Hence, the observed responses may be one of three types; discrete choices or ranking vectors or rating vectors respectively. We denote the observed responses as $\mathbf{y}_{im}$ of dimension $(J \times 1)$, for the mth decision maker in the ith response of the sample, where J, $J > 2$, is the number of the alternatives used in the experiment. In addition, if an additive "main effects" compositional rule forms the utilities vector $\mathbf{u}_{im}$ of the same dimension as $\mathbf{y}_{im}$, then

$$\mathbf{u}_{im} = (u_{im1}, ..., u_{imJ})' = R_i \boldsymbol{\beta}_{\mathbf{m}}, \tag{4.1}$$

where R_i is a $(J \times K)$ matrix of the values of the K attributes in the hypothetical scenario of situation i, $\boldsymbol{\beta}_m$ is a vector of dimension $(K \times 1)$ that depicts the mth respondent attributes weights, $i = 1, ..., N$, $m = 1, ..., M$. Note that, in equation (4.1), not only $\boldsymbol{\beta}_m$ is to be estimated but also $\mathbf{u}_{im}$, the continuous underlying (latent) utility vector which is hidden behind the observed response vector $\mathbf{y}_{im}$. The way the underlying (latent) vector is estimated is in conjunction with the information derived from $\mathbf{y}_{im}$. For instance, adopting the concept of random utility (Thurstone, 1927), if alternative j appears to be the first choice in $\mathbf{y}_{im}$, then u_{imj} should be equal to $\max_z(u_{imz})$. The model described in (4.1), which is based upon an "individual-by-individual" basis, is usually estimated using classical conjoint analysis but it may be absolutely useless or difficult to be handled when, for instance, a manager is interested in making decisions for the whole market or segments of it (see Moore, 1980, Louviere, 1984).

Almost all the existing aggregate analysis techniques are concerned with the modeling of the utilities using the information of just the discrete choice (the response that indicates the one most preferable alternative) of each multinomial response data point, whereas the literature in modeling ranking or rating responses is still sparse. Of course, in each case of different alternatives attributes, only the first choice is the one that will actually be used by a particular decision maker in practice. However, beyond a shadow of doubt, compared to the discrete choice, a ranking response contains useful additional information about the structure of the relation between the utilities and the attributes of the competing alternatives. Unlike this limitation of the discrete choice models, their applications (besides the travel mode choice) include such diverse fields as housing choices and shopping behavior (see, for example, McFadden, 1978, Louviere, 1984). This wide use causes the requirement for the evaluation of the statistics to meet its obligation.

In this chapter, we are motivated by a challenging real ranking response data set to demonstrate that complicated but realistic data generation structures have either not been considered before or have not been adequately handled.

In the aggregate analyses, we are not interested any more in estimating the vector $\boldsymbol{\beta}_{\mathbf{m}}$ per decision maker but a vector $\boldsymbol{\beta}$. Hence, in discrete choice models, where the responses of different decision makers as well as the responses on different hypothetical scenarios by a decision maker are assumed independent, equation (4.1) can be written, by dropping the subscript m, as

$$\mathbf{u}_i = R_i \boldsymbol{\beta} + \mathbf{e_i}, \tag{4.2}$$

where $\mathbf{e}_i$ is a stochastic term vector of dimension $(J \times 1)$.

4.2.2 The multiranked probit model

We denote the observed responses as $\mathbf{y}_i$ of dimension $(J \times 1)$, for the $i - th$ response of the sample, where J, $J > 2$, is the number of the alternatives used in the experiment. In addition, if an additive "main effects" compositional rule forms the utilities vector $\mathbf{u}_i$ (of the same dimension as $\mathbf{y}_i$) and the responses of different decision makers as well as the responses on different hypothetical scenarios by a decision maker are assumed independent, then

$$\mathbf{u}_i = (u_{i1}, ..., u_{iJ})' = R_i \boldsymbol{\beta} + \mathbf{e}_i, \tag{4.3}$$

where R_i is a $(J \times K)$ matrix of the values of the K attributes in the hypothetical scenario of situation i, $\boldsymbol{\beta}$ is a vector of dimension $(K \times 1)$ that captures the "general tastes" of the attributes weights throughout the population of interest or particular market segments, $i = 1, ..., N$, and $\mathbf{e}_i$ is a stochastic term vector of dimension $(J \times 1)$. As in equation (4.1), note that, in equation (4.3), not only $\boldsymbol{\beta}$ is to be estimated but also $\mathbf{u}_i$, the continuous latent utility vector which is related to the observed response vector $\mathbf{y}_i$. Adopting the concept of random utility (Thurstone, 1927), this relationship states that if alternative j appears to be the first choice in $\mathbf{y}_i$, then u_{ij} should be equal to $\max_z(u_{iz})$.

However, if $\mathbf{u}_i = (u_{i1}, ..., u_{iJ})^T$, $J > 2$, any shift with location γ and scale $\tau > 0$ from $\mathbf{u}_i$ to $\gamma I_J + \tau \mathbf{u}_i$, where I_J is the unit vector of dimension J, will leave the observed choices $\mathbf{y}_i$ unchanged. Hence, as in equations (3.1) and (3.2), for these identification purposes (see McCullogh and Rossi, 1994), we usually work with the differenced system which is achieved by letting $\mathbf{w}_i = (u_{i1} - u_{iJ}, ..., u_{i(J-1)} - u_{iJ})^T$ and re-expressing (4.3) as

$$\mathbf{w}_i = X_i \boldsymbol{\beta} + \boldsymbol{\epsilon}_i, \tag{4.4}$$

where $\mathbf{w}_i$ is a latent vector of dimension $(J-1) \times 1$, X_i is the matrix of dimension $(J-1) \times K$ that derives from R_i if we subtract the first $(J-1)$ rows by the last one and $\boldsymbol{\epsilon}_i$ is the error term of dimension $(J-1)$ which characterizes the link function density; recall that, in the multinomial probit model $\boldsymbol{\epsilon}_i$ is assumed to follow $N_{J-1}(\mathbf{0}, \Sigma)$, i.e. the multivariate normal distribution of dimension $(J-1)$ (Aitchison and Bennett, 1970). The corresponding multinomial probability p_{ij} of alternative j to be chosen in situation i is given by

$$p_{ij} = P[\mathbf{x}_{ij}\boldsymbol{\beta} + \varepsilon_{ij} > \mathbf{x}_{iq}\boldsymbol{\beta} + \varepsilon_{iq}], \text{for all } q \neq j, \tag{4.5}$$

where $\mathbf{x}_{ij}$ is the $j-th$ row of X_i, $j, q = 1, ..., J$ and ε_{ij} is the $j-th$ element of $\boldsymbol{\epsilon}_i$.

The likelihood function of this model is formed by the product of N independent multinomial distributions

$$L(\boldsymbol{\beta}, \Sigma | y_1, ..., y_N, X_i) = \prod_{i=1}^{N} \prod_{j=1}^{J} p_{ij}^{y_{ij}}. \tag{4.6}$$

It is clear, however, that the computation of the multinomial probabilities entails the calculation of multiple integrals of the multivariate normal density making a maximum likelihood estimation burdensome.

In section 3.5, it was mentioned that, within the above Bayesian data augmentation framework, McCullogh and Rossi (1994) developed a Gibbs sampler procedure to obtain draws from the posterior distributions of $\boldsymbol{\beta}$, Σ and $\mathbf{w}_i$ for the multinomial probit model. Since the latent vectors $\mathbf{w}_i$ in the model are not directly observed, each set of the parameters $(\boldsymbol{\beta}, \Sigma)$ is observationally equivalent to $(c\boldsymbol{\beta}, c^2\Sigma)$, for any $c > 0$. We chose to obtain the re-parametrization of Σ, with almost no additional computational cost, where $\sigma^2_{(J-1)\times(J-1)} = 1$ throughout the algorithm, and ensure identifiable parameters by using a theorem of standard multivariate analysis, proved by Dawid (1988). In the context of MCMC sampling this was first applied in Dellaportas (1998). A similar, but more complicated procedure was used by McCullogh *et al.* (2000) who define a prior density on the elements of the conditional Wishart. As pointed out by Nobile (2000) both algorithms produce similar results. Although we deal with ranking responses, discrete choices can be treated similarly since they just form a special case of our approach.

There is no suitable statistical method to deal with ranking responses with ties in the data structure, even though ties usually appear in real data sets. Georgesku-Roegen (1958) introduced the threshold parameters into the theory of consumer choice so that a choice between any two alternatives will be considered only when their utilities exceed some necessary minimum (see also Krishnan, 1977, Lioukas, 1984). We model ties in the data adopting this idea. The threshold δ is defined such that, for two alternatives a and b, with utilities w_{ia} and w_{ib} respectively,a is preferred to b when

$$w_{ia} > w_{ib} + \delta, \tag{4.7}$$

and a and b are equally preferred when

$$|w_{ia} - w_{ib}| \leq \delta. \tag{4.8}$$

In addition, a probability of intransitivity of indifference among some alternatives may be reported when, for instance, among three alternatives a, b and c, with respective utilities w_a, w_b and w_c,

$$\begin{aligned} w_a - w_b &\leq \delta, \\ w_b - w_c &\leq \delta \\ w_a - w_c &> \delta. \end{aligned} \tag{4.9}$$

Then, the decision maker will be indifferent between a and b and between b and c, though a will be clearly preferred to c. Such deviations from the behavior imposed by the discrete choice models may be part of the decision making mechanism and, hence, it seems reasonable to systematize the treatment of thresholds and threshold-associated intransitivities.

We adopt this threshold parameter δ into the multinomial probit model. We also propose a hierarchical model which treats the threshold parameters of all decision makers as exchangeable. Thus, there are thresholds δ_m, $m = 1, ...M$, where M is the number of the decision makers ($M \leq N$; equality holds in case where each decision maker responses to just one hypothetical scenario). We denote as N_m, the number of responses of decision maker m, i.e. the decision maker m provides the responses from 1_m (first response) to N_m, $\sum_m N_m = N$).

Finally, rather than assuming the multiranked probit model with normal errors, we will be concerned with a general error structure model that is obtained by assuming that the N vectors of the utilities follow a multivariate Student-t distribution with ν degrees of freedom, i.e.

$$\mathbf{w}_i \sim N_{J-1}(X_i \boldsymbol{\beta}, \lambda_i^{-1} \Sigma), \tag{4.10}$$

where λ_i follows a gamma distribution:

$$\lambda_i \sim G(\frac{\nu}{2}, \frac{\nu}{2}), \tag{4.11}$$

$i = 1, ..., N$; see, for example, Chen and Dey (1998). The degrees of freedom are used in our approach as an additional continuous parameter to be estimated.

4.2.3 The proposed model

In summary, equation (4.4), enriched by the approaches described above, yields the model:

$$\begin{aligned} \mathbf{w}_i &= X_i\boldsymbol{\beta} + \boldsymbol{\epsilon}_i, \\ \boldsymbol{\epsilon}_i &\sim N(0, \frac{1}{\lambda_i}\Sigma | \sigma^2_{(J-1)\times(J-1)} = 1), \\ \lambda_i &\sim G(\frac{\nu}{2}, \frac{\nu}{2}), \end{aligned} \tag{4.12}$$

$w_{ij} > w_{iq} + \delta_m$, if, in y_i, alternative j is preffered to q by decision maker m; $|w_{ij} - w_{iq}| \leq \delta_m$, if, in y_i, alternatives j and q are equally preferred by decision maker m; $w_{iJ} = 0$ for all i; $\delta_m \sim N(\mu_\delta, \sigma^2_\delta)$, where $j, q = 1, ..., J$ alternatives, $i = 1, ..., N$ responses, $m = 1, ..., M$ decision makers.

4.3 Implementation

In this section we provide MCMC implementation details. To facilitate the derivation of the MCMC steps, we first break down the model probabilistic structure into its conditional independence parts.

4.3.1 The link function

Rather than assuming the multivariate probit model, we will be concerned with a general error structure model that is obtained by assuming that the N vectors of the utilities follow a scale mixture of multivariate normal link functions, i.e. $\mathbf{w}_i \sim N_{J-1}(X_i\boldsymbol{\beta}, \kappa(\lambda_i)\Sigma)$ and $\lambda_i \sim \pi(\lambda_i)$, where $\kappa(\lambda_i)$ is a positive function of one-dimensional positive-valued scale mixing variable λ_i and $\pi(\lambda_i)$ is a mixing distribution, $i = 1, ..., N$. The special case where

$\kappa(\lambda_i) = \frac{1}{\lambda_i}$ and $\lambda_i \sim G(\frac{\nu}{2}, \frac{\nu}{2})$ yields the multivariate Student-t distribution with ν degrees of freedom. Using different degrees of freedom, a variety of link functions arises; when $\nu = 1$ we obtain the multivariate Cauchy distribution, when $\nu \to \infty$ the multivariate normal arises etc. Hence, the multivariate probit model is a special case when $\lambda_i = 1$.

4.3.2 The conditional independence structure

The proposed model specification models the latent utilities, where the coefficient vector $\boldsymbol{\beta}$ of the covariates, the covariance matrix Σ of the error term, the mixing parameters λ_i and the degrees of freedom ν are parameters to be estimated. Recall that, given the vectors of the utilities $\mathbf{w}_i$, $i = 1, ..., N$, the above parameters are of the standard Bayesian multivariate regression form.

The vectors $\mathbf{w}_i$, which are hidden behind the stated preferences, can be thought of as auxiliary quantities for the procedure but, at heart, they do reveal additional important information as it will be evident in section 2.4. We recover the latent vectors $\mathbf{w}_i$ using the observed ranking responses $\mathbf{y}_i$ and the latent threshold parameters δ_m, normally distributed around the mean μ_δ with variance σ_δ^2.

The conditional independence structure of the model gives rise to three MCMC sampling steps. First, the regression parameters ($\boldsymbol{\beta}$, Σ^{-1}, $\lambda_1, ..., \lambda_N$, ν), conditional on $\mathbf{w}_1, ...,$ $\mathbf{w}_N$ and the covariates $X_1, ..., X_N$ are easily obtained from standard Bayesian multivariate regression results. Next, note that the threshold-related parameters $\delta_1, ..., \delta_M$, μ_δ and σ_δ^2, are conditionally independent of all other parameters given $\mathbf{w}_1, ..., \mathbf{w}_N$. Finally, we need to sample $\mathbf{w}_1, ..., \mathbf{w}_N$, given the information obtained from the remaining parameters of the model and $\mathbf{y}_1, ..., \mathbf{y}_N$.

The conditional independence structure described above yields the following full conditional distributions,

$$[\boldsymbol{\beta}|\cdot] = [\boldsymbol{\beta}|\Sigma^{-1}, \lambda_1, ..., \lambda_N, \mathbf{w}_1, ..., \mathbf{w}_N, X_1, ..., X_N],$$

$$[\Sigma^{-1}|\cdot] = [\Sigma^{-1}|\boldsymbol{\beta}, \lambda_1, ..., \lambda_N, \mathbf{w}_1, ..., \mathbf{w}_N, X_1, ..., X_N],$$

$$[\lambda_i|\cdot] = [\lambda_i|\boldsymbol{\beta}, \Sigma^{-1}, \mathbf{w}_i, \nu, X_i],\ i = 1, ..., N,$$

$$[\nu|\cdot] = [\nu|\lambda_1, ..., \lambda_N],$$

for the sampling step of the regression parameters,

$$[\delta_m|\cdot] = [\delta_m|\mu_\delta, \sigma_\delta^2, \mathbf{w}_{1_m}, ..., \mathbf{w}_{N_m}, \mathbf{y}_{1_m}, ..., \mathbf{y}_{N_m}],\ m = 1, ..., M,$$

$$[\mu_\delta|\cdot] = [\mu_\delta|\delta_1, ..., \delta_M, \sigma_\delta^2],$$

$$[\sigma_\delta^2|\cdot] = [\sigma_\delta^2|\delta_1, ..., \delta_M, \mu_\delta],$$

for the threshold-related parameters and

$$[\mathbf{w}_i|\cdot] = [\mathbf{w}_i|\boldsymbol{\beta}, \Sigma^{-1}, \lambda_i, \delta_m, \mathbf{y}_i, X_i],\ i = 1, ..., N,$$

for the latent utilities calculation. To complete our model formulation, we need to specify prior densities for the parameters $\boldsymbol{\beta}$, Σ^{-1},λ_1, ...,,λ_N, ν, μ_δ and σ_δ^2. In the application of the next chapter, we use either non informative improper or locally diffuse priors.

4.3.3 Sampling the Parameters of the multivariate regression

Setting a prior $[\boldsymbol{\beta}] = N_K(\boldsymbol{\beta}^\star, A^{-1})$ on $\boldsymbol{\beta}$, its required full conditional distribution is

$$\boldsymbol{\beta}|\Sigma^{-1}, \lambda_1, ..., \lambda_N, \mathbf{w}_1, ..., \mathbf{w}_N, X_1, ..., X_N \sim N_K(\hat{\boldsymbol{\beta}}, \Sigma_\beta^{-1}),$$

where $\Sigma_{\boldsymbol{\beta}} = \sum_i [\lambda_i {X_i}^T \Sigma^{-1} X_i] + A$ and $\hat{\boldsymbol{\beta}} = \Sigma_\beta \{\sum_i [\lambda_i {X_i}^T \Sigma^{-1} \mathbf{w}_i] + A\boldsymbol{\beta}^\star\}$.

A Wishart prior $W(v, V)$ with v degrees of freedom and scale matrix V on Σ^{-1} yields a Wishart conditional distribution

$$\Sigma^{-1}|\boldsymbol{\beta}, \lambda_1, ..., \lambda_N, \mathbf{w}_1, ..., \mathbf{w}_N, X_1, ..., X_N \sim W(v + N, V + \sum_{i=1}^{N} \lambda_i \boldsymbol{\epsilon}_i \boldsymbol{\epsilon}_i^T)), \tag{4.13}$$

where $\boldsymbol{\epsilon}_i = \mathbf{w}_i - X_i\boldsymbol{\beta}$.

To ensure identifiable parameters in the model, we sample the matrix Σ such that the $(J-1)-th$ element of the diagonal of Σ is fixed to 1. That is, the variance of the utilities of the $(J-1)-th$ alternative is equal to 1 throughout the algorithm. To obtain it, we use lemma (2) of Dawid (1988) (see also Dempster, 1969, for a slightly different version). According to it, instead of sampling Σ^{-1} from the conditional Wishart distribution given in (4.13), we use the steps outlined in the following to sample indirectly from $\Sigma^{-1}|\sigma^2_{J-1} = 1$:

-Split the scale matrix $V + \sum_{i=1}^{N} \lambda_i \boldsymbol{\epsilon}_i \boldsymbol{\epsilon}_i^T$ in (4.13) into the matrices

$$\begin{bmatrix} G^{11}_{(J-2)\times(J-2)} & G^{12}_{(J-2)\times 1} \\ G^{21}_{1\times(J-2)} & G^{22}_{1\times 1} \end{bmatrix}.$$

-Draw Γ from $W^{-1}(v + N + 1, G^{11} - G^{12}(G^{22})^{-1}G^{21})$.

-Draw B^T from $N_{J-2}((G^{22})^{-1}G^{21}, (G^{22})^{-1} \otimes \Gamma)$.

-Perform the inverse transformation

$$\Sigma \leftarrow \begin{bmatrix} \Gamma + B^T B & B^T \\ B & 1 \end{bmatrix}.$$

The next step in the algorithm is to draw the mixing variables λ_i. Setting $[\lambda_i] = G(\frac{\nu}{2}, \frac{\nu}{2})$, the conditional distribution of λ_i is

$$\lambda_i|\boldsymbol{\beta}, \Sigma^{-1}, \mathbf{w}_i, \nu, X_i \sim G\left(\frac{\nu + J - 1}{2}, \frac{1}{2}\left[\nu + (\mathbf{w}_i - X_i\boldsymbol{\beta})^T \Sigma^{-1}(\mathbf{w}_i - X_i\boldsymbol{\beta})\right]\right), i = 1, ..., N.$$

Examining thoroughly the proposed model, a natural next step is to find out the link function of the error term that fits the data as good as possible. In our approach, we avoid the usual computational burden of fitting several models with different degrees of freedom on the data by using instead an additional Metropolis-Hastings step in the procedure that samples from the continuous positive distribution of the degrees of freedom ν. Given λ_i, and setting $[\nu] = G(a, b)$, the conditional distribution of ν is

$$[\nu|\lambda_1, ..., \lambda_N] \propto \nu^{a-1}\Gamma(\frac{\nu}{2})^{-N}(\frac{\nu}{2})^{\frac{\nu N}{2}}(\prod_i \lambda_i)^{\frac{\nu}{2}-1}\exp[-\frac{\nu}{2}(2b + \sum_i \lambda_i)]. \qquad (4.14)$$

The unknown form of this conditional density requires specific handling that can be obtained by an additional Metropolis-Hastings step. We propose the use of a random walk step with a proposal taken as a normal density (truncated to positive values) with variance appropriately tuned to achieve accepted probability between 0.25-0.50.

4.3.4 Threshold parameters

Recall that the N_m responses of the $m-th$ decision maker is a subset of the N responses of the sample. Then, δ_m, $m = 1, ...M$, conditional on the latent responses $\mathbf{w}_i$ that correspond to the $m - th$ decision maker, are assumed to follow a normal distribution with common mean μ_δ and variance σ_δ^2, appropriately truncated in an interval (l_m, u_m) defined for each decision maker. These bounds are defined such that

$$l_m = \max_{jq} |w_{ij} - w_{iq}| \qquad (4.15)$$

for all pairs j and q of distinct alternatives at which a tie has been observed (over all the response vectors of the $m - th$ decision maker) and

$$u_m = \min_{jq} |w_{ij} - w_{iq}| \qquad (4.16)$$

for all pairs j and q of distinct alternatives which have been clearly ranked without ties by the $m - th$ decision maker. Note that, if the ranking response vectors of a particular decision maker do not include a tie, $l_m = 0$. That is, discrimination has been obtained, no matter how small δ_m is.

With respect to the random effects in the model, if we set $[\mu_\delta] = 1$, the conditional distribution of $\mu_\delta | \delta_1, ..., \delta_M, \sigma_\delta^2$, is

$$N\left(\frac{\sum_m \delta_m}{M}, \frac{\sigma_\delta^2}{M}\right).$$

In addition, setting $[\sigma_\delta^2] = IG(a^\star, b^\star)$, the inverse gamma conditional of $\sigma_\delta^2 | \mu_\delta$, δ_m is

$$IG\left(a^\star + \frac{1}{2}M, b^\star + \frac{1}{2}\sum_m (\delta_m - \mu_\delta)^2\right).$$

4.3.5 Latent utilities

The conditional distribution of $\mathbf{w}_i$, is multivariate normal of dimension $J-1$, truncated over the appropriate cone in $\mathcal{R}^{J-1}$, or, alternatively, each element w_{ij} of $\mathbf{w}_i$ is drawn from a univariate normal distribution truncated over an appropriate interval. Hence, we draw from $N \times (J-1)$ conditionals

$$w_{ij} | \mathbf{w}_{i(-j)}, \boldsymbol{\beta}, \Sigma^{-1}, \lambda_i, \delta_m, \mathbf{y}_i, X_i \sim N(m_{ij}, \tau_{ij}^2), \tag{4.17}$$

for $i = 1, ..., N$, $j = 1, ..., J-1$, where

$$m_{ij} = \mathbf{x}_{ij}^T \boldsymbol{\beta} + (\sigma_{j(-j)} \Sigma_{(-j)(-j)}^{-1})(\mathbf{w}_{i(-j)} - X_{i(-j)}\boldsymbol{\beta}),$$

$$\tau_{ij}^2 = \frac{1}{\lambda_i}(\sigma_{jj} - \sigma_{j(-j)} \Sigma_{(-j)(-j)}^{-1} \sigma_{(-j)j}),$$

$\mathbf{w}_{i(-j)}$ is the vector $\mathbf{w}_i$ of dimension $(J-1) \times 1$ except the $j-th$ element, $\mathbf{x}_{ij}$ is the $j-th$ row of X_i, $X_{i(-j)}$ is the submatrix created by deleting the $j-th$ row from X_i and the submatrices $\Sigma_{(-j)(-j)}$, $\sigma_{j(-j)}$, $\sigma_{(-j)j}$ and σ_{jj} come from the partitioning of Σ with respect to the $j-th$ argument, such that

$$\Sigma = \begin{bmatrix} \Sigma_{(-j)(-j)} & \sigma_{(-j)j} \\ \sigma_{j(-j)} & \sigma_{jj} \end{bmatrix}.$$

These conditional distributions are appropriately truncated so that the drawn utilities are consistent with the ranking of the alternatives obtained by the corresponding decision maker. For given δ_m, the intervals on the real line which truncate the draws of w_{ij} are defined as follows. Let $w_{ij}^{(k)}$ denote the utility of alternative j in the $i-th$ response such that $y_{ij} = k$ ($k = 1$ for the most preferable alternative etc.). Then, the conditional distribution of $w_{ij}^{(k)}$, given by (4.17), is truncated in the interval (l_w, u_w), where

$$l_w = (w_{ij}^{(k+1)} + \delta_m) \tag{4.18}$$

and

$$u_w = (w_{ij}^{(k-1)} - \delta_m). \tag{4.19}$$

We define $w_{ij}^{(0)} = \infty$ and, for $p > max_j(y_{ij})$, for $j = 1, ..., J$, $w_{ij}^{(p)} = -\infty$. Then, using as a base line the condition $w_{iJ} = 0$, one can generate the truncation intervals of each value w_{ij} given the remaining elements of $\mathbf{w}_i$. For example, in a set with three alternatives, if $\mathbf{y}_i = [1, 3, 2]^T$, the truncation we use is $w_{i1} > \delta_m$ and $w_{i2} < -\delta_m$. If $\mathbf{y}_i = [2, 3, 1]^T$, the truncation is $w_{i2} + \delta_m < w_{i1} < -\delta_m$ and $w_{i2} < w_{i1} - \delta_m$. This procedure guarantees that unequally preferred alternatives correspond to utilities with difference larger than δ and that a higher utility and, consequently, a higher probability of alternative to be chosen corresponds to a more preferred alternative.

However, the formulae just outlined capture the situation where no ties have been observed in $\mathbf{y}_i$. Note that $max_j(y_{ij})$, for $j = 1, ..., J$ is not necessarily equal to J because a tie may take place for all j where $y_{ij} = k$. Some modifications are required on the procedure that calculates the constraints of the utilities distribution of (4.17) in the case of ties or apparent intransitivities.

To demonstrate the utilities constraints in cases of intransitivity and ties in response vectors, we denote as

$$\mathbf{t}_{iz}^{(k)} = (t_{i1}^{(k)}, ..., t_{iZ}^{(k)}) \tag{4.20}$$

the vector of the Z elements of $\mathbf{w}_i$ which correspond to $y_{ij} = k$. Then, one can generate the truncation interval of $t_{i1}^{(k)}$ using the formulae (4.18) and (4.19) (unless $y_{iJ} = k$, where one can set $t_{i1}^{(k)} = w_{iJ} = 0$) and the truncation intervals (l_t, u_t) of the remaining elements of $\mathbf{t}_i^{(k)}$ using the sampled value $t_{i1}^{(k)}$ and the bounds

$$l_t = \max[(t_{i1}^{(k)} - \delta_m), (w_{ij}^{(k+1)} + \delta_m)]$$

and

$$u_t = \min[(t_{i1}^{(k)} + \delta_m), (w_{ij}^{(k-1)} - \delta_m)].$$

Nevertheless, in this case, if one attempts to form the truncation intervals for the remaining non-tied responses using (4.18) and (4.19), more than one utilities correspond to $y_{ij} = k + 1$ or to $y_{ij} = k - 1$. Then, if $k > 1$ or $k < \max_j(y_{ij})$, one should use

$$w_{ij}^{(k+1)} = \max_z(t_{iz}^{(k+1)}) \tag{4.21}$$

and

$$w_{ij}^{(k-1)} = \min_z(t_{iz}^{(k-1)}) \tag{4.22}$$

in (4.18) and (4.19), respectively. For example, in the set with three alternatives, if $\mathbf{y}_i = [1, 1, 2]^T$, then the truncation one could use is $w_{i1} > \delta_m$ and $max(w_{i1} - \delta_m, \delta_m) < w_{i2} < w_{i1} + \delta_m$, whereas, if $\mathbf{y}_i = [1, 2, 2]^T$, the truncation one could use is $w_{i1} > \max(w_{i2} + \delta_m, \delta_m)$ and $-\delta_m < w_{i2} < \min(\delta_m, w_{i1} - \delta_m)$.

Another possible response is the apparent intransitivity. In this case, if $y_{ij} = k$ is an intransitive choice (in the sense that $w_{ij}^{(k-1)} - w_{ij}^{(k)} \leq \delta_m$, $w_{ij}^{(k)} - w_{ij}^{(k+1)} \leq \delta_m$ though $w_{ij}^{(k-1)} - w_{ij}^{(k+1)} > \delta_m$), the truncation intervals for $w_{ij}^{(k-1)}$, $w_{ij}^{(k)}$ and $w_{ij}^{(k+1)}$ are, respectively,

$$(w_{ij}^{(k+1)} + \delta_m, w_{ij}^{(k)} + \delta_m),$$

$$(w_{ij}^{(k-1)} - \delta_m, w_{ij}^{(k+1)} + \delta_m) \tag{4.23}$$

and

$$(w_{ij}^{(k)} - \delta_m, w_{ij}^{(k-1)} - \delta_m).$$

Hence, in a three alternatives response, if the first and the second alternatives are tied, the second and the third are tied, but the first alternative is ranked ahead of the third one, then the truncation intervals for w_{i1} and w_{i2} are, respectively, $w_{i2} + \delta_m < w_{i1} < w_{i2} + 2\delta_m$ and $w_{i1} - \delta_m < w_{i2} < \delta_m$. In addition, if the first and the second alternatives are tied with the third one but the second is ranked ahead of the first, the truncation one could use is $-\delta_m < w_{i1} < w_{i2} - \delta_m$ and $w_{i1} + \delta_m < w_{i2} < \delta_m$. Table 4.1 shows some examples of truncation intervals in a four-alternatives example, that includes ranking responses, ties or apparent intransitivities. The element "2.5" of the last choice vector indicates that this alternative (the second one) is equally preffered with the first and the third alternative (an apparent intransitivity occurs), the first is more preferable than the third though.

Finally, the sampled values w_{ij} are obtained by an "one-for-one" sampling method by Devroye (1986) along with the exponential rejection sampling suggested by Robert (1995).

Table 4.1: Example of truncation intervals of the utilities in a four-alternatives choice.

Choice vector y_i	**Utility** w_{ij}	**Lower bound**	**Upper bound**
1	w_{i1}	$w_{i2} + \delta_m$	$+\infty$
2	w_{i2}	$w_{i3} + \delta_m$	$w_{i1} - \delta_m$
3	w_{i3}	δ_m	$w_{i2} - \delta_m$
4	w_{i4}	0	0
2	w_{i1}	δ_m	$w_{i3} - \delta_m$
4	w_{i2}	$-\infty$	$-\delta_m$
1	w_{i3}	$w_{i1} + \delta_m$	$+\infty$
3	w_{i4}	0	0
3	w_{i1}	$w_{i3} + \delta_m$	$-\delta_m$
1	w_{i2}	δ_m	$+\infty$
4	w_{i3}	$-\infty$	$w_{i1} - \delta_m$
2	w_{i4}	0	0
3	w_{i1}	$w_{i2} + \delta_m$	$w_{i3} - \delta_m$
4	w_{i2}	$-\infty$	$w_{i1} - \delta_m$
2	w_{i3}	$w_{i1} + \delta_m$	$-\delta_m$
1	w_{i4}	0	0
1	w_{i1}	$max(w_{i2}, w_{i3}) + \delta_m$	$+\infty$
2	w_{i2}	δ_m	$w_{i1} - \delta_m$
2	w_{i3}	$max(w_{i2} - \delta_m, \delta_m)$	$min(w_{i2} + \delta_m, w_{i1} - \delta_m)$
3	w_{i4}	0	0
2	w_{i1}	$max(0, w_{i3}) + \delta_m$	$w_{i2} - \delta_m$
1	w_{i2}	$w_{i1} + \delta_m$	$+\infty$
3	w_{i3}	$-\delta_m$	$min(\delta_m, w_{i1} - \delta_m)$
3	w_{i4}	0	0
2	w_{i1}	$w_{i3} + \delta_m$	$min(w_{i2} + \delta_m, -\delta_m)$
2.5	w_{i2}	$w_{i1} - \delta_m$	$w_{i3} + \delta_m$
3	w_{i3}	$w_{i2} - \delta_m$	$w_{i1} - \delta_m$
1	w_{i4}	0	0

Chapter 5

Latent Utilities for Transportation Services; Analysis of Attiko Metro Data

5.1 Introduction

In this chapter, we analyze a real data set from a stated preference experiment which was designed to explain and predict passengers behavior towards three main transportation modes in the city of Athens. This data set will illustrate the proposed model and the practical information that can be derived from the output. A key feature of the analysis of transportation systems is the prediction of the passengers' behavior when changes are brought about by new services, investments in infrastructure or changes in operating and pricing policies. The focus of these analyses is usually on the estimation of the "value of time" and the trade-off between travel cost and travel characteristics such as in-vehicle time, walking time, waiting time, parking search time, etc., of some competitive alternatives (e.g. metro, car, bus). Such estimations are based upon the continuous utilities of the alternatives which are hidden behind a stated preferences designed experiment aimed to the transportation services users.

5.2 A stated preference experiment for transportation modes

In the development study of Athens' metro, a designed stated choice experiment aimed to assess the value and relative importance of certain travel characteristics (attributes) with respect to transportation modes "car", "metro" and "bus". The set of characteristics included the travel cost in drachmas, the total walk time to/from transportation modes in minutes, the total in-vehicle time in minutes excluding parking search time of private means of transport, the total waiting time in minutes for public transportation, the parking search time in minutes (applicable to cars only) and the inconvenience associated with a mode transfer in the course of a journey. To investigate individual behavior in terms of trip purpose, income and car availability, the sample was partitioned into five segments based on the trip purpose (home-based work, home-based social / recreation, home-based education, all non home-based purposes and home-based other, like business, shopping etc.), based on monthly household income (3 groups: up to 200 thousand drachmas, between 201-400 thousand drachmas and above 400 thousand drachmas) and, finally, based on car availability (daily-occasionally or never). The survey was designed to collect about 100 effective interviews per segment. This required a minimum total of 500 effective interviews. The set of characteristics with their corresponding levels are given in Table 5.1 (inconvenience of a transfer = 1 denotes that more than one transportation modes were needed).

The stated choice experiment was elaborated using hypothetical scenarios about the values of the attributes which the respondents faced in order to form their choices. The "main effects" orthogonal fractional factorial design consisted of 32 different combinations of the attribute levels of Table 5.1. The 32 runs of the experiment (i.e. hypothetical

Table 5.1: Attribute levels for travel characteristics.

	Walking time (min)	In-vehicle time (min)	Parking search time (min)	cost (drachmas)	waiting time (min)	inconvenience of a transfer
Car	2 or 7	10 or 30	5 or 15	200 or 400	-	-
Metro	10 or 20	15 or 25	-	100 or 300	2 or 7	0 or 1
Bus	5 or 10	25 or 40	-	75 or 150	10 or 25	0 or 1

scenarios or, otherwise stated, triplets of mode alternatives metro, bus and car) were clustered in 8 blocks of 4 runs each. Each respondent was faced with one of the eight blocks. The resulting data structure for each run has the form of Table 5.2.

Table 5.2: A part of the data. In this particular hypothetical scenario, the decision maker's first choice was "Metro" and the second one was "Car".

	Choice	Walking time (min)	In-vehicle time (min)	Parking search time (min)	Cost	Waiting time (min)	Inconvenience of a transfer
Car	2	2	30	5	400	0	0
Metro	1	10	15	0	300	7	1
Bus	3	5	25	0	75	25	1

One may wonder why the data contain a ranked rather than a discrete choice preference. In fact, the actual experiment did not use the structure appeared in Table 5.2 and the ranking responses were caused by necessity rather than desire. As indicated by a pilot study, it is very difficult for a decision maker to take into account the complexity of the 14 attribute values (of Table 5.1) per scenario, so it was decided that all three possible combinations of pairs of the triplet (car-metro, metro-bus and car-bus) should be shown instead of one triplet. In each paired choice, the "none" alternative was added to avoid forcing the respondent into selecting a non-preferred response. Thus, each respondent faced 12 cards (4 runs by 3 cards per run). The three pairwise comparisons were converted to a single ranked triplet, possibly with ties, inconsistencies and lexicographic responses. Inconsistencies and lexicographic responses were eliminated; see Wardman (1988). The former inconsistent responses yielded when the rational choice axioms were violated. For instance, in a hypothetical scenario, the decision maker may had preferred car compared to metro, metro compared to bus, but bus compared to car whereas, in the logical ranking, car should had been more preferable than bus. Furthermore, the lexicographic responses were caused when, for instance, the ranking responses of an individual were based only on the values of a particular attribute, such as the lowest in-vehicle time, whereas the trade-offs among attributes were ignored. The number of the triplets of valid responses and the number of the decision makers per trip purpose and based on car availability are shown in Table 5.3. For a detailed description of this stated choice experiment, see Spanos, Deloukas and Anastassaki (1997).

Table 5.3: Number of triplets of valid responses and decision makers.

Trip Purpose	Valid responses	Decision Makers
Work	664	176
Recreation	282	74
Other	323	87
Education	310	83
Non home-based	270	73
Car availability		
Daily	648	169
Occasionally	348	94
Never	853	230

When analysing such data, a primary goal is to estimate the relative importance of travel characteristics for the five different trip purposes. Rather than dealing with the whole data set using appropriate explanatory dummy variables for the trip purpose, we chose to analyze these five data sets separately. This is a usual strategy in the transportation literature. For example, Wardman (1988) points out that the calibration of separate models for each category of interest allows a more detailed examination of the responses. Moreover, another potential problem caused when the analysis is based on all data is the "majority fallacy". As Moore (1980) writes (see also Huber and Moore, 1980):

The majority fallacy is caused by heterogeneity of preferences; for example, *if half of the people like large cars and the other half like small cars, the 'average' person may like medium sized cars best, even though no real person wants one.*

Hence, if we include all the different groups in just one analysis, we may conclude to an aggregation of the different tastes that the groups may have into an average "taste" that will represent none of the respondents.

5.2.1 Prior specifications and MCMC output

The five data sets were analyzed using the proposed model with multivariate student-t distribution for the error term and with random effects for the sampling step of δ_m. For a series of diffuse prior specifications (gamma densities) for the degrees of freedom ν, we noticed that the MCMC algorithm does not converge due to the uncontrollably large values of ν. The analysis showed the parameter for the degree of freedom to take large values as if it was trying to approach the normal distribution against the gamma prior information that was given. Using several values for the mean and the variance for the gamma prior, the posterior density was taking the prior shape providing evidence for the posterior density's smothering within the prior frames. In addition, when the variance of the proposal distribution was very large, this parameter was taking uncontrollably large values and it was varying without convergence. In fact, all of these values of the degrees of freedom were providing good approximations of the multivariate normal distribution. Hence,we based our algorithm on $\lambda_i = 1$. Hence, we ended up assuming normal errors for all the data subsets except one that will be discussed later.

For the rest of the parameters, we used a normal prior for $\boldsymbol{\beta}$ with mean $\mathbf{0}$ and covariance matrix diag(100), a conjugate Wishart prior for Σ^{-1} with $v = 3$ and scale matrix diag(3) which results to a mean equal to diag(1), and an inverse gamma prior for σ_δ^2 with $a^\star = 0.5$ and $b^\star = 1$. When a Student-t distribution was used, the prior for λ_i was $G(\frac{\nu}{2}, \frac{\nu}{2})$. These prior specifications provided fairly vague priors.

We used 4000 sampled values from the posterior densities for each data set. After a burn-in of 5000 iterations, a lag 20 was used to save computer space. The Raftery and Lewis, the Geweke and the Heilderberger-Welch diagnostics (see, for example, Brooks and Roberts, 1998) were used and they provided evidence for the convergence of all the parameters.

5.2.2 Calibration of separate models

Our first goal is to estimate the relative importance of travel characteristics for the five different trip purposes. Rather than dealing with the whole data set using appropriate explanatory dummy variables for the specification of the trip purpose, we chose to analyze these five data sets separately. Wardman (1988) pointed out that the calibration of separate models for each category of interest allows a more detailed examination of the responses; this is a usual strategy of the transport experts and that was also used in the analysis with the multinomial logit model in our data set in the case where only the discrete (first) choice was reported (Spanos *et al.*, 1997). Hence, the use of the same strategy in our analysis is considered as a facilitation for the transport experts since, at least, rough comparisons of the results may be performed.

5.2.3 Results

The posterior output consists of practical information such as travel characteristics (e.g. walking time, waiting time etc.), expressed either in drachmas per hour or in minutes of in-vehicle time, and 95% credible intervals of the probability of choosing a particular transportation mode: key factors in determining whether a policy has positive or negative net benefits.

Table 5.4 shows the estimated posterior parameters along with their 95% credible intervals (the 2.5th and 97.5th percentiles of the sampled values of the parameters). INT denotes the inconvenience associated with the mode transfer in the course of a journey, PST denotes the parking search time, INVT denotes the in-vehicle time, AASC denotes the alternative specific constant for car and MASC denotes the alternative specific constant for metro.

The coefficients of all the travel characteristics are negative for all the five purposes as they were plausibly expected to be. In addition, almost all of these 95% credible intervals do not include 0. Exception is the credible intervals of parking search time for the trip purposes "Recreation" and "Other". Also, the coefficients of the alternative specific constants for car and metro are all close to 0 which shows that the travel characteristics which were included in the stated choice experiment can sufficiently explain the choices of the decision makers and the utilities which were formed.

Furthermore, note that the credible intervals for the covariance σ_{12} of all the five trip purposes include only positive values which shows that a simple model which assumes uncorrelated utilities of the alternatives within each response (such as the multinomial logit or a multivariate probit with zero off-diagonal elements of Σ) would not had been a suitable choice for these particular data sets.

The parameter μ_δ is larger (equal to 0.223) for decision makers whose trip purpose is non home-based whereas the smallest value for μ_δ appears when the trip purpose is "work". That is, non home-based commuters seem more indifferent in their choice of transportation mode whereas commuters to work seem more decisive in their choices. This result provides evidence for the way the utilities are quantified; 15% of the responses by non-home based commuters include ties or intransitivities whereas the respective percentages for the remaining data sets are 6-9%.

We will concentrate on some of the results and we will illustrate some examples of the way we provided answers to some questions posed by the transportation engineers. Assume, for example, that the travel characteristic "Inconvenience of transfer" (IoT) is of interest. Figure 5-1a shows the posterior error bars of the elements of β that correspond to IoT for the five trip purposes. These represent the 95% credible intervals of this parameter. The existence of IoT reduces the utility of transportation modes; note that commuters to recreation are more discouraged when using more than one transportation

Table 5.4: Posterior means of the parameters for the five trip purposes. In parentheses, the 2.5th and 97.5th percentiles are presented.

	Work	**Recreation**	**Other**	**Education**	**Non home-based**
walk	-0.053	-0.041	-0.068	-0.083	-0.063
	(-0.069,-0.038)	(-0.067,-0.016)	(-0.093,-0.044)	(-0.108,-0.058)	(-0.090,-0.037)
wait	-0.031	-0.023	-0.020	-0.037	-0.046
	(-0.043,-0.020)	(-0.045,-0.002)	(-0.038,-0.002)	(-0.055,-0.019)	(-0.065,-0.027)
INT	-0.248	-0.556	-0.230	-0.264	-0.370
	(-0.379,-0.118)	(-0.777,-0.353)	(-0.429,-0.030)	(-0.466,-0.063)	(-0.602,-0.153)
PST	-0.024	-0.026	-0.022	-0.061	-0.053
	(-0.046,-0.001)	(-0.073,0.020)	(-0.063,0.016)	(-0.104,-0.020)	(-0.102,-0.007)
INVT	-0.031	-0.034	-0.042	-0.047	-0.028
	(-0.039,-0.023)	(-0.048,-0.020)	(-0.055,-0.028)	(-0.062,-0.034)	(-0.043,-0.014)
cost	-0.004	-0.003	-0.003	-0.006	-0.004
	(-0.005,-0.003)	(-0.005,-0.002)	(-0.004,-0.002)	(-0.007,-0.005)	(-0.005,-0.003
AASC	0.171	0.647	0.024	0.805	0.087
	(-0.197,0.544)	(-0.040,1.372)	(-0.593,0.658)	(0.197,1.464)	(-0.590,0.776)
MASC	0.587	0.711	0.574	0.584	0.440
	(0.350,0.824)	(0.316,1.102)	(0.214,0.930)	(0.211,0.979)	(0.022,0.835)
σ_{11}^2	1.942	3.498	2.578	2.719	3.033
	(1.536,2.415)	(2.400,4.993)	(1.760,3.614)	(1.915,3.813)	(2.040,4.405)
σ_{12}	0.610	0.761	0.468	0.729	0.498
	(0.475,0.750)	(0.482,1.018)	(0.237,0.705)	(0.493,0.979)	(0.214,0.774)
μ_δ	0.123	0.178	0.164	0.155	0.223
	(0.095,0.151)	(0.116,0.240)	(0.111,0.220)	(0.102,0.209)	(0.152,0.299)

modes.

The transport experts' inference is usually based on ratios of the estimated parameters of the travel characteristics. This emerges from the identification problem and the restriction on the covariance matrix of $\mathbf{w}_i$ that permit the inference to be based on ratios of the estimated parameters and not on the parameters themselves. The usual, and sensible ratios used in such applications are created by dividing each travel characteristic by "cost" or by "in-vehicle time". The corresponding error bars similar to IoT are given in Figures 5-1b and 5-1c and the resulting ratios in Figures 5-2a and 5-2b. These ratios were constructed by manipulating the MCMC output in the usual way; see,

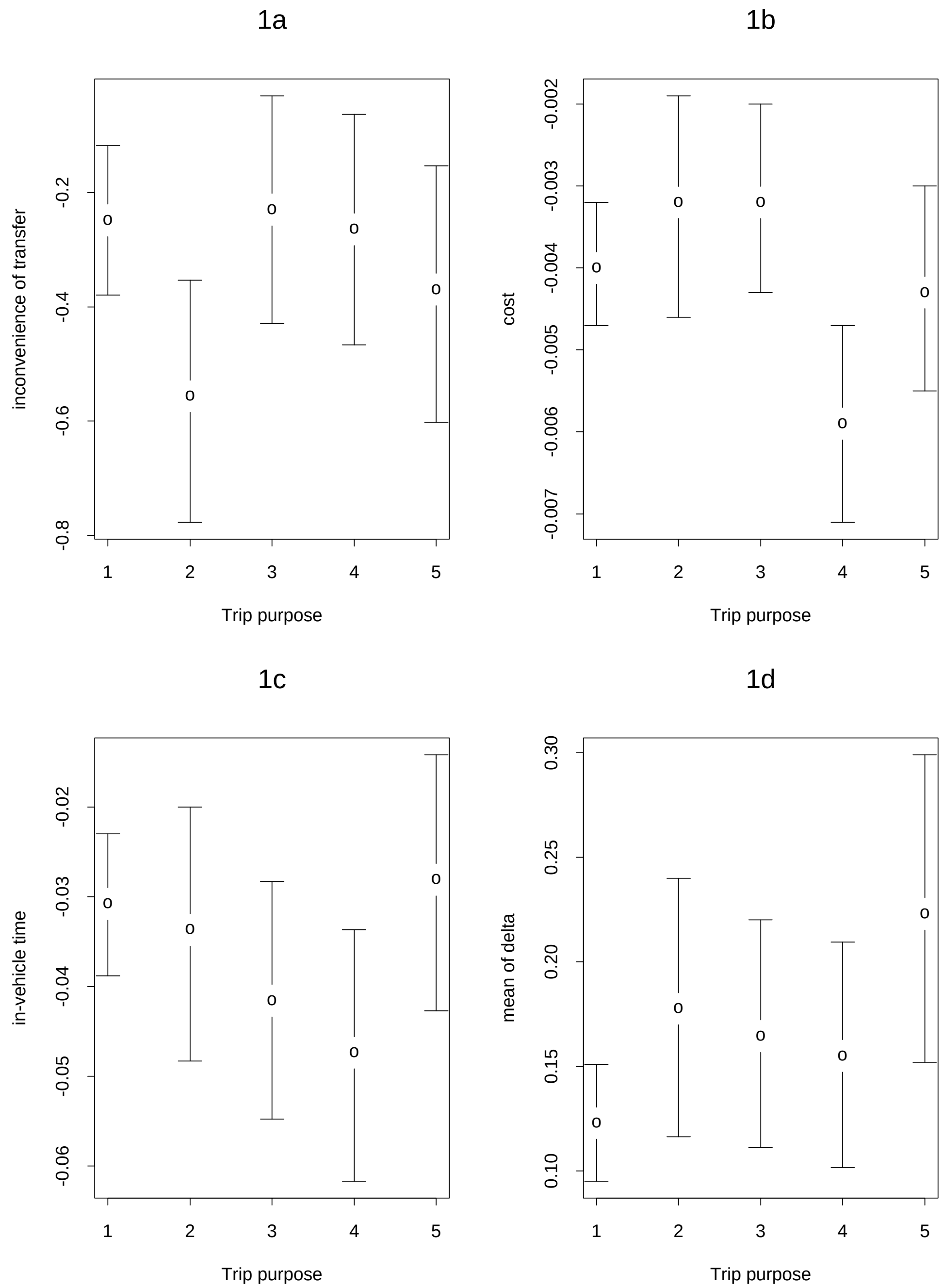


Figure 5-1: Posterior error bars. 1a: Inconvenience of a transfer. 1b: Cost. 1c: In-vehicle time. 1d: μ_δ. Trip purpose 1:Work, 2:Recreation, 3:Other trip purposes, 4:Education, 5:Non home-based.

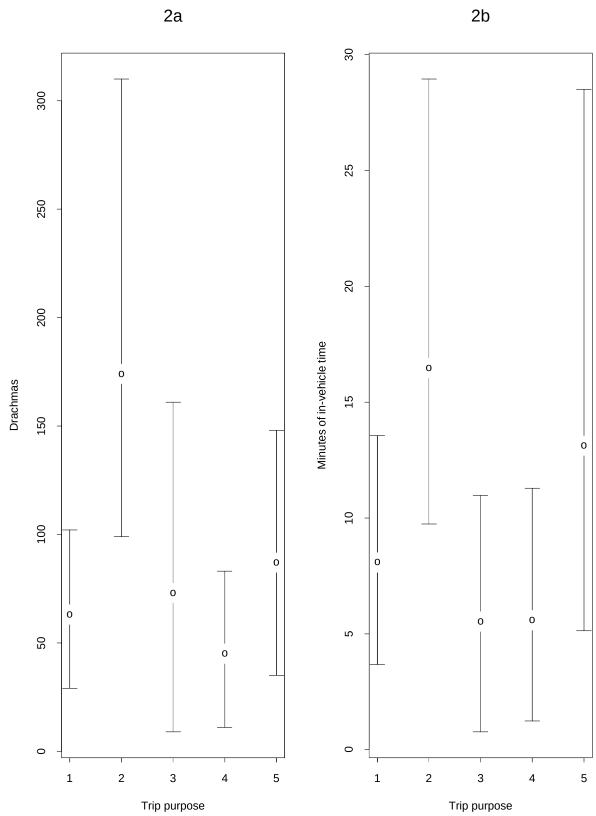


Figure 5-2: Medians and 95% credible intervals of ratios. 2a: Inconvenience of a transfer expressed in drachmas. 2b: Inconvenience of a transfer expressed in in minutes of in-vehicle time. Trip purpose 1:Work, 2:Recreation, 3:Other trip purposes, 4:Education, 5:Non home-based.

for example, Smith and Roberts (1993). Figure 5-1b provides evidence that commuters for educational purposes are more discouraged to use a transportation mode when the cost is high. Moreover, note that commuters to recreation would be willing to pay about 174 drachmas to avoid IoT (Figure 5-2a) and that IoT is, on average, 16 times more important for them than a minute of in-vehicle time (Figure 5-2b).

Table 5.5 provides the relation between the in-vehicle time and the remaining travel characteristics via the median and 2.5th and 97.5th percentiles of the ratio of walk, wait, IoT, and parking search time (PST) with the parameter "in vehicle time" (INVT). We report the median rather than the mean of each ratio because the normal densities of the parameters of the travel characteristics generate a Cauchy distributed ratio where the mean does not exist. This yields an expression of these travel characteristics equivalent to minutes of in-vehicle time.

Table 5.5: Travel characteristics in minutes of in-vehicle time.

	Work	*Recreation*	*Other*	*Education*	*Non home-based*
walk	1.72	1.23	1.64	1.74	2.24
	(1.16,2.48)	(0.47,2.44)	(1.06,2.51)	(1.16,2.65)	(1.11,4.90)
wait	1.01	0.69	0.48	0.78	1.64
	(0.63,1.55)	(0.07,1.58)	(0.05,1.01)	(0.40,1.26)	(0.86,3.37)
IoT	8.09	16.47	5.53	5.58	13.12
	(3.73,13.54)	(9.74,28.95)	(0.77,10.97)	(1.24,11.28)	(5.13,28.51)
PST	0.77	0.76	0.52	1.28	1.90
	(0.03,1.62)	(-0.63,2.36)	(-0.36,1.66)	(0.42,2.33)	(0.23,4.58)

It is usually considered by transport experts that the travel characteristics "walk", "wait" and "PST" should be up to 2.5 times higher than "INVT" and that the penalty of a transfer in the course of a journey should be 3-4 times higher than "INVT". However, only the parameter of "walk" seems to be consistent to such a consideration for all the trip purposes. The values of "wait" and "PST" are, in general, lower than 1 (which is the value that corresponds to "INVT" for the purposes of the comparison). That is, a minute of in-vehicle time may be more important than a minute of waiting time or a minute of parking search time for that particular population. However, the corresponding 95% credible intervals include the unity so there is evidence that the values of "wait" and "PST" are, at least, equally important to the value of "INVT". It is also noteworthy that the value of "IoT" in minutes of in-vehicle time is higher than it was expected for all the trip purposes.

The value of time (the money in drachmas that the decision makers would pay in order to avoid an hour of the travel characteristic "walk", "wait", "PST" or "INVT") is also a ratio of interest which we calculated as the median of the ratio 60·(travel characteristic)/cost. In addition, the money in drachmas the decision makers would pay in order to avoid the penalty associated with a transfer in the course of a journey was calculated as INT/cost. The medians and the 2.5th and 97.5th percentiles of these ratios are given in Table 5.6.

Table 5.6: Travel characteristics in drachmas per hour.

	Work	*Recreation*	*Other*	*Education*	*Non home-based*
walk	800	776	1301	847	884
	(564,1094)	(282,1579)	(775,2233)	(607,1128)	(483,1431)
wait	471	440	377	377	645
	(297,676)	(46,967)	(46,808)	(191,587)	(370,1009)
INT	63	174	73	45	87
	(29,102)	(99,310)	(9,161)	(11,83)	(35,148)
PST	355	484	412	617	747
	(14,717)	(-398,1495)	(-302,1293)	(203,1075)	(90,1502)
INVT	465	633	788	484	396
	(335,619)	(355,1064)	(497,1286)	(346,655)	(196,658)

Moreover, note that the in-vehicle time is almost equally important for commuters to work and recreation. However, the ratio (in-vehicle time/cost) showed that commuters to recreation are willing to pay more in order to avoid an hour of in-vehicle time. Nevertheless, these two densities were highly overlapped, indicating that this difference is, at least, not significant.

5.2.4 Segmentation of the sample based on car availability

Another segmentation of the sample (besides the trip purpose) is based on the car availability. We ran the programs for the respondents who had a car available daily or occasionally and for those who never have a car available. We also provide the results of the model for the whole data set (without any segmentation) for purposes of comparison.

Table 5.7 shows the posterior means (and the credible intervals) of the two levels of "car availability" and the respective results for the whole sample. Note the narrower credible intervals of the parameters for the whole sample (as they should plausibly be, due to the large sample size).

In addition, Tables 5.8 and 5.9 show the travel characteristics expressed in minutes

Table 5.7: Posterior means of the parameters for the three levels of "Car Availability" and for the whole data set.

	Car Availability		**Whole sample**
	Daily/Occasionally	*Never*	
walk	-0.0551	-0.0596	-0.0550
	(-0.0676,-0.0421)	(-0.0736,-0.0455)	(-0.0644,-0.0457)
wait	-0.0310	-0.0279	-0.0287
	(-0.0409,-0.0213)	(-0.0385,-0.0174)	(-0.0358,-0.0213)
INT	-0.3297	-0.2402	-0.2853
	(-0.4395,-0.2224)	(-0.3550,-0.1297)	(-0.33628,-0.2079)
PST	-0.0221	-0.0480	-0.0317
	(-0.0419,-0.0025)	(-0.0710,-0.0239)	(-0.0468,-0.0180)
INVT	-0.0319	-0.0361	-0.0325
	(-0.0384,-0.0253)	(-0.0441,-0.0286)	(-0.0373,-0.0277)
cost	-0.0033	-0.0045	-0.0037
	(-0.0039,-0.0027)	(-0.0052,-0.0038)	(-0.0042,-0.0033)
AASC	-0.2292	0.3915	0.2824
	(-0.0722,0.5353)	(0.0218,0.7528)	(0.0519,0.5141)
MASC	0.6192	0.4831	0.5433
	(0.4172,0.8183)	(0.2524,0.6939)	(0.3980,0.6895)
σ_{11}^2	2.1578	2.4544	2.1443
	(1.7846,2.6047)	(1.9576,3.0320)	(1.8659,2.4656)
σ_{12}	0.5836	0.5949	0.6029
	(0.4610,0.7067)	(0.4534,0.7330)	(0.5095,0.6923)
μ_δ	0.1326	0.1395	0.1205
	(0.1107,0.1560)	(0.1161,0.1658)	(0.1068,0.1348)

of in-vehicle time and in drachmas per hour respectively. It is noteworthy that those who use a car daily or occasionally, believe that a minute of walking time and a waiting time and an inconvenience of a transfer are more important for them (than for those who do not have a car available) and they would be willing to pay more in order to avoid them. The only characteristic that seems not to annoy them is the parking search time (compared to those who do not have a car!).

Table 5.8: Travel characteristics in minutes of in-vehicle time.

	Car Availability		**Whole sample**
	Daily/Occasionally	*Never*	
walk	1.73	1.65	1.69
	(1.26,2.35)	(1.22,2.22)	(1.35,2.09)
wait	0.97	0.77	0.88
	(0.66,1.38)	(0.48,1.15)	(0.65,1.16)
INT	10.38	6.67	8.82
	(6.91,14.66)	(3.40,10.47)	(6.23,11.70)
PST	0.71	1.34	0.97
	(0.08,1.35)	(0.66,2.08)	(0.55,1.49)

Table 5.9: Travel characteristics in drachmas per hour.

	Car Availability		**Whole sample**
	Daily/Occasionally	*Never*	
walk	995	785	884
	(731,1325)	(596,1004)	(719,1063)
wait	562	364	459
	(375,773)	(228,525)	(338,590)
INT	100	53	77
	(65,140)	(28,80)	(54,99)
PST	403	640	509
	(45,767)	(317,950)	(286,756)
INVT	577	476	521
	(438,741)	(367,604)	(434,621)

5.2.5 Model validation

We used a simple method of discrepancy measurements to detect possible outliers (see Chen and Dey, 2000). In particular, we calculated the difference between the expected and the observed ranking responses. Thus, if R is the number of the drawn values from the posterior distribution of each parameter of interest, then R vectors $\hat{\mathbf{w}}_i$ are reproduced for each one of the N responses of the sample. Combining the information of δ_m and $\hat{\mathbf{w}}_i$, we can obtain the expected responses $\hat{\mathbf{y}}_{ir}$, $r = 1, ..., R$, for the $i - th$ response. The discrepancy D between observed and expected responses can then be measured as

$$D = \frac{1}{N \cdot R} \sum_i \sum_r (\hat{\mathbf{y}}_{ir}^T - \mathbf{y}_i^T)(\hat{\mathbf{y}}_{ir} - \mathbf{y}_i).$$

The standardized observational level discrepancy measures d_i were calculated for each alternative separately. When $|d_i| > 3$ (Chen and Dey, 2000), the observation is considered as aberrant.

This detection of possible outliers was applied to the commuters to work dataset of 664 responses. The number of aberrant observations with respect to the alternatives car, metro and bus were 9, 9 and 3 respectively. The examination of all these responses showed that those decision makers either did not choose the respective alternative although its attributes had the lowest possible values specified by the experiment or chose the alternative although its attributes had the highest possible values, whereas the characteristics of the remaining alternatives were much more attractive.

5.2.6 Models with Student-t Errors for Small Data Sets

As it was previously mentioned, the surprising result for these data sets (that was also appeared in the analysis with the multinomial logit model where only the discrete choices where analyzed, see Spanos *et al.*, 1997) is that the value of in-vehicle time (in drachmas per hour) is higher for the trip purpose "recreation" than for "work". Nevertheless, these two densities are highly overlapped (Figure 5-3).

However, the suspicion of such a peculiarity for that particular population could have dramatically influenced the transportation politics. So, we tried to go deeper into that problem by reanalyzing these two data sets (trips for work and recreation) using the additional information of the household monthly income categories. This resulted in three subsets for each of the two trip purposes.

The analysis of these six data sets first required the investigation of the appropriate distribution for the degrees of freedom of the error term (see equation 4.14). This pro-

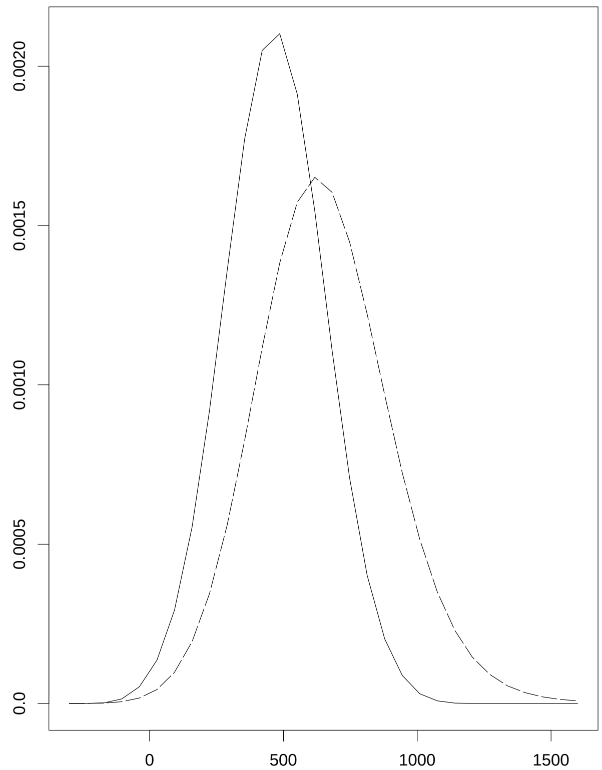


Figure 5-3: Densities of the value of time in drachmas per hour. The solid line represents the trip purpose "Work" and the dashed line the trip purpose "Recreation".

cedure showed only the model for the smallest data set, that corresponds to recreation trip purposes for individuals with household monthly income above 400 thousand drachmas (Table 5.10 provides the sample sizes), to prefer a distribution for the error term with heavier tails than those of the normal distribution. This data set contains 41% of responses with intransitivities and ties whereas such responses in the five remaining data sets where much fewer (1-11%). In particular, we tried gamma priors with several values of a and b which resulted to similar posterior densities. We ran the programs using different initial values to avoid absorption points in the Metropolis-Hastings step of the algorithm, see Gelman, Carlin, Stern and Rubin (1995, pp. 362). Finally, a gamma prior for the degrees of freedom with $a = 1.1$ and $b = 0.011$ was used and resulted to a posterior mean equal to 2.07 and posterior variance 0.016. In the Metropolis within Gibbs step for the parameter of the degrees of freedom, a normal proposal distribution was used with variance equal to 0.005, which yielded a probability of acceptance about 0.49 in the Metropolis-Hastings algorithm. This proposal density was truncated to values larger than 2 so as not to include long-tailed distributions while they have infinite variance and are not realistic in the far tails (see Gelman *et al.*, 1995, pp. 350). To obtain 4000 posterior values, a lag equal to 100 was used after a burn-in of 10000 iterations.

Table 5.10: Sample sizes by Household Monthly Income (in thousand drachmas) for trip purposes "Work" and "Recreation"

	Work		*Recreation*	
Income	Responses	Respondents	Responses	Respondents
up to 200	214	58	89	24
201-400	308	80	112	29
above 400	142	38	81	21

Having satisfied ourselves that the data sets themselves chose the distribution for the error term that fits them better, we proceed on the estimation of the parameters and we focus on the calculation of the value of in-vehicle time (in drachmas per hour). As it is shown (Table 5.11), the 95% credible intervals of the two estimations are highly overlapped for all of the three categories of the monthly household income and, hence, none of these groups behaves in a particular way that requires special treatment.

It is important to mention, however, that the model with normal errors for the commuters for a recreation purpose with a monthly income of above 400 thousand drachmas provided inexplicable coefficients (for example, almost all the posterior values of the parameter of "walk" had positive signs) whereas, when used the model with varying degrees of freedom, no such problems were reported.

Table 5.11: Value of in-vehicle time in drachmas per hour for each monthly household income category.

Income	*Work*	*Recreation*	
up to 200	368	695	
	(174,614)	(289,1677)	
201-400	501	417	
	(308,767)	(-308,2620)	
above 400	531	405[a]	782[b]
	(297,889)	(267,620)	(411,1915)

[a] Normal errors.
[b] Student-t errors with varying ν.

5.3 On the decisions of transportation policy

The last step in our attempt is the investigation of the transportation policy that is proposed in the light of the information we derived from this analysis. A manager may be interested in investigating the behavior of the decision makers against changes of the value of an alternative's attribute (or the values of some alternative attributes) when all the remaining attributes values are known.

For illustrative purposes, we focus on the trip purpose "Work" and on the relation between travel cost and the probabilities of the three alternatives to be chosen. A straightforward question is related to the way the cost may influence these probabilities and, finally, to the optimal ticket price of the recently constructed metro in Athens. For this reason, in a presumptive situation, we created a possible scenario to be investigated. We fixed each of the remaining travel characteristics to the mean value of the values used at the experimental design, whereas the travel cost was varying from 100 to 300 drachmas (which were the two values specified in the experiment for the travel cost of metro) with an increment of 4 drachmas. Using the posterior quantities output at hand, we generated 200 trials per each of the 4000 posterior sampled draws. In each trial, we reported the alternative with the highest utility; if this utility was greater than the second greater utility plus a threshold δ (which is also available by the output at hand via μ_δ and σ_δ^2), then success of this alternative was obtained, otherwise, a tie occurred. Recall that the utility of the last alternative "Bus" was fixed to 0 throughout the whole procedure as it was mentioned in the presentation of the model. Then, from the 200 trials of each posterior draw, the probabilities of the three alternatives to be chosen were estimated. Finally, approximate credible intervals of these probabilities were calculated as the 2.5th and the 97.5th percentiles of the 4000 estimated values of the three probabilities. The results are reflected in Figure 5-4.

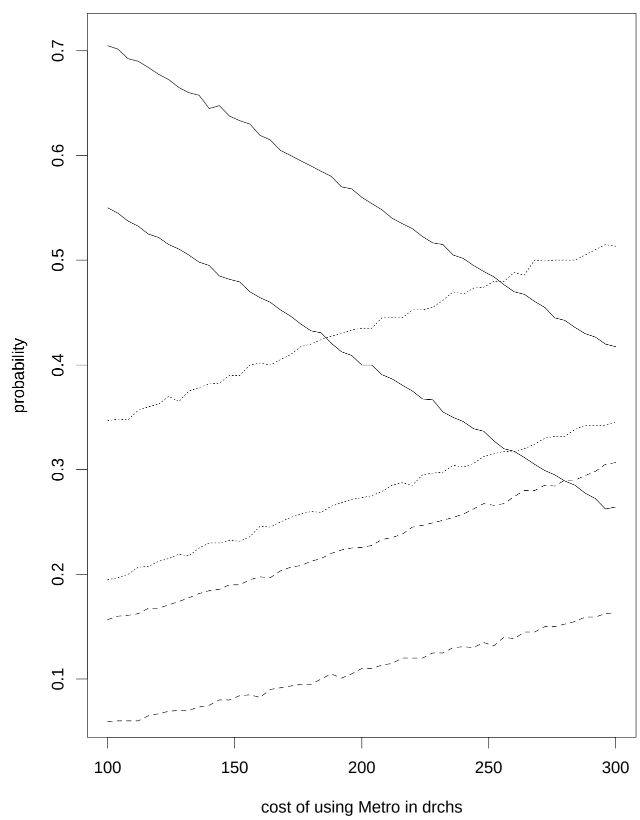


Figure 5-4: 95% credible intervals of the probability of choosing a transportation mode with varying cost of using Metro. Solid lines correspond to metro, dotted lines to car and dashed lines to bus.

Note that, in our hypothetical scenario, the probability of bus to be the first choice is lower than the respective probability of metro even if the ticket price of metro takes the highest value specified by the experiment. In addition, the alternative "car" is able to compete with the metro when the travel cost of the latter is above of about 250 drachmas. Incidentally, the Athens Metro came into operation on the 31th of January 2000 with a ticket price of 250 drachmas for all routes.

One more question that can be investigated using the output, is related to the optimal waiting time for Metro. The same procedure as before was repeated where the waiting time for the Metro varied from 2 to 7 minutes with an increment of 0.1 minutes. We fixed each of the remaining travel characteristics to the mean value of the values used at the experimental design, except the cost of using Metro which was set equal to 250 drachmas. Moreover, since the behaviors of commuters to work and recreation are different, we investigated both of these groups. For commuters to work, the optimal waiting time for Metro is, on average, 5 minutes (Figure 5-5a). On the other hand, for commuters for recreational purposes, "Metro" is not able to compete "Car" even if the waiting time for the former is just 2 minutes (Figure 5-5b).

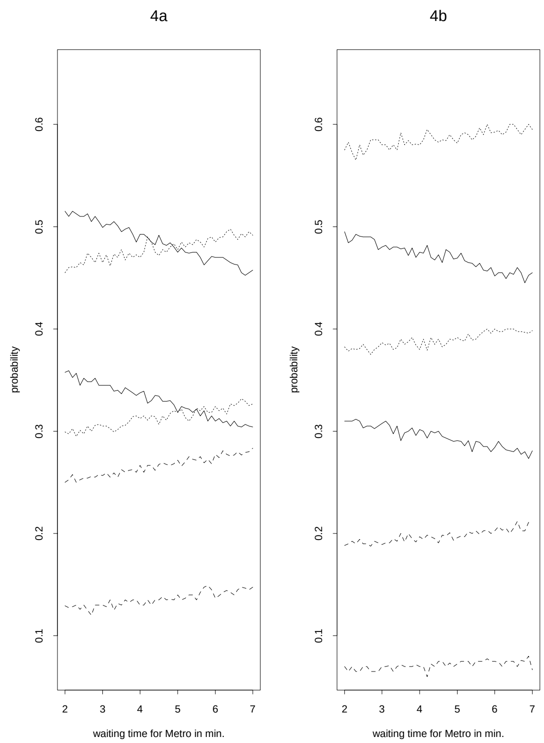


Figure 5-5: 95% credible intervals of the probability of choosing a transportation mode with varying waiting time for Metro. 4a: Trip purpose "Work". 4b: Trip purpose "Recreation". Solid lines correspond to metro, dotted lines to car and dashed lines to bus.

Chapter 6

Item response Theory

6.1 Introduction

Item Response Theory (IRT) models are widely used in psychological and educational statistical analyses that aim to quantify parameters of a person, named abilities, and parameters of items, named discrimination and difficulty. In the educational framework, the multiple choice tests are subject of IRT models, where the data consist of series of correct or incorrect responses of each student per item included in the test. Thus, IRT is the study of test and item scores based on assumptions concerning the mathematical relationship between abilities (or other hypothesized traits) and item responses.

A simple model that is usually used is the Rasch model (Rasch, 1961) that quantifies the general ability of each person and the difficulty parameter of each item. More general models include the discrimination parameters of each item and the item-guessing parameters (see, Albert and Ghosh, 2000). However, in most of the cases, the performance of a student and, hence, the respective ability can not be assumed to be unidimensional; multiple skills of a student are involved in producing the manifest responses.

Moreover, the literature is still sparse in the analysis of data that are not dichotomous. Nevertheless, in multiple choice tests, a common practice is that a penalty is being imposed (e.g. subtraction of points for each incorrect answer) so as to prevent students from guessing behavior. This yields a number of omissions in the responses. In usual item response models, these omissions are regarded as incorrect answers. However, the omissions include some kind of different information than the incorrect responses; in the case where a penalty is used for each wrong answer, omissions may yield due to uncertainty of the students whereas incorrect responses may yield due to misinformed or confused students.

In this chapter, we present a widely used IRT model, the three parameters logistic model. Special cases of this model are the Rasch model and the two parameters logistic model. We propose a Bayesian approach for a multidimensional item response model, that can be seen as a generalization of the three parameters model.

In section 6.3, a general multidimensional item response model is proposed for use in data with "non-answer" responses. In chapter 7, a real data set is analyzed using the proposed model. Another real data set illustrates the multidimensional item response model in chapter 8.

6.2 Review of IRT Modes

Item response theory models deal with the item level responses rather than the total scores of the students on a test. The first IRT models were analysing data from tests in which the responses were of a dichotomous correct-incorrect format. The distribution of this "correct-incorrect" variable or, otherwise stated, the probability of a correct response is decomposed on the general ability of the examinee's parameters $\theta \in (-\infty, +\infty)$, the difficulty parameters $b_j \in (-\infty, +\infty)$ and the discrimination parameters $a_j \in (0, +\infty)$. The ability is a structural parameter; the probability of success on item j is usually presented as $P_i(\theta)$, whereas the item parameters (difficulty and discrimination) are considered as nuisance parameters. The function of the probability of success is known as item response function.

6.2.1 The Normal-Ogive Model

The first IRT model was the normal-ogive model. The idea of the model is first presented by Thurstone's discriminal dispersions theory of stimulus perception (Thurstone, 1927), also, the same basic model has been studied among others by Richardson (1936), Ferguson (1942) and Lawley(1943).

The response function for the jth item, based on this model, is given by a normal cumulative distribution function:

$$P_j(\theta) = \int_{-\infty}^{a_j(\theta - b_j)} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz. \tag{6.1}$$

6.2.2 The Rasch Model

Rasch introduced what he called "a structural model for items in a test", (see Rasch, 1960). As van der Linden and Hambleton (1997) state:

> *Rasch's main motivation for his models was his desire to eliminate references to populations of examinees in analyses of tests (Rasch, 1960, Preface; Chap. 1). Test analysis would only be worthwhile if it were individual-centered, with separate parameters for the items and the examinees. To make his point, Rasch often referred to the work of Skinner, who was also known for his dislike of the use of population based statistics and always experimented with individual cases.*

The probability of a correct answer ($U_{ij} = 1$) is formulated by:

$$P(U_{ij} = 1|\theta_i) = \frac{1}{1 + e^{-(\theta_i - b_j)}}, \tag{6.2}$$

and, hence, the joint likelihood function of this model is given by:

$$L(\theta, b; u) = \frac{\prod_j b_j^{u_{.j}} \prod_i \theta_i^{u_{i.}}}{\prod_i \prod_j (1 + \theta_i / b_j)}, \tag{6.3}$$

where $i = 1, ..., n$ is the index of the examinees, $j = 1, ..., k$ is the index of the items and $u_{i.}$ and $u_{.j}$ are the marginal sums of the data matrix, i.e. the total scores of student i and item j respectively.

Several methods have been proposed for the estimation of the parameters in this model; the EM algorithm (Thissen, 1982) calculates the marginal maximum-likelihood to eliminate the impact of the ability parameters when estimating item parameters, an iterative least-squares method (Verhelst and Molenaar, 1988), semiparametric methods (De Leeuw and Verhelst, 1986) and Bayesian methods that use a hierarchical modeling (see Swaminathan and Gifford, 1982).

For the Rasch model, item totals and individual totals are sufficient statistics for the model parameters.

6.2.3 The Two and Three Parameters Logistic Models

Birnbaum proposed a logistic model instead of the one parameter normal ogive model (see Birnbaum, 1968):

$$P_j(\theta) = \frac{1}{1 + e^{a_j(\theta - b_j)}}, \tag{6.4}$$

that also includes the discrimination parameters a_j.

One of the most commonly used models among applied psychologists is the Three-Parameter Logistic Model. This model has been used primarily for modeling cognitive ability data, but recently it has been applied to personality data as well (see Embretson and Reise, 2000, for some other applications).

This model is a more general form of the one parameter (Rasch Model) and two parameter logistic model. The formula of the model consists of three parameters: item discrimination (parameter a), item location or difficulty (parameter b), and the height of the lower asymptote of the response function (parameter c). Note that the two parameter logistic model can be obtained from three parameters model be setting $c = 0$; the Rasch model, with one parameter, can be obtained by setting $c = 0$ and $a = 1$. The formula of the three parameter model is:

$$P_j(\theta) = c_j + (1 - c_j)\frac{1}{1 + e^{-Da_j(\theta - b_j)}}, \tag{6.5}$$

where θ represents the value of the latent trait (e.g., conscientiousness or cognitive ability), $P(\theta)$ represents the probability of a correct response, D is a scaling constant equal to 1.702, and a, b, and c are the parameters characterizing an item.

Under this formulation, larger "a" parameters provide better discrimination among examinees. The item difficulty parameter "b" (or threshold parameter) is related to the proportion-correct score, "p," in classical test theory. Obviously, "p" and "b" are inversely related. Large values of "p" indicate relatively "easy" items, whereas large values of "b" indicate "difficult" items. Finally, the guessing parameter c indicates the probability of responding correctly for examinees who have very low θ.

The likelihood function of the two-parameter logistic model can be written as:

$$L(\theta, a, b; u) = \prod_i \prod_j P_j(\theta_i; a_j, b_j)^{u_{ij}} [1 - P_j(\theta_i; a_j, b_j)]^{1-u_{ij}}. \tag{6.6}$$

The parameter estimation of this model requires the maximization of the logarithm of the likelihood function which leads to the following estimation equations:

$$\begin{aligned} \sum_j a_j(u_{ij} - P_j(\theta_i; a_j, b_j)) = 0, i = 1, ..., n, \\ \sum_i a_j(u_{ij} - P_j(\theta_i; a_j, b_j)) = 0, j = 1, ..., k, \\ \sum_i (u_{ij} - P_j(\theta_i; a_j, b_j))(\theta_i - b_j) = 0, j = 1, ..., k., \end{aligned} \tag{6.7}$$

Birnbaum suggested the joint maximum likelihood estimates to jointly solve the equations for the values of the unknown parameters iteratively, by starting with initial values for the ability parameters, solves the equations for the item parameters, fixes the item parameters, and solves the equations for improved estimates of the values of the ability parameters etc. However, the convergence behaviour of this algorithm was not satisfactory; Bock and Lieberman (1970) introduced the marginal maximum likelihood method for the normal-ogive model, which was also applicable for the two and three parameters logistic models. According to this method, for a density function $f(\theta)$ of the ability distribution, the marginal probability of obtaining a response vector $\mathbf{u} \equiv (u_1, ...u_k)$ on the items in a test is equal to:

$$P(u|a, b) = \int_{-\infty}^{\infty} \prod_j [P_j(\theta; a_j, b_j)]^{u_j} [1 - P_j(\theta; a_j, b_j)]^{1-u_j} f(\theta) d\theta. \tag{6.8}$$

The marginal likelihood function associated with the full data matrix $\mathbf{u}$ is given by the multinomial kernel

$$L(a, b; u) = \prod_{u=1}^{2^n} \pi_u^{r_u}, \tag{6.9}$$

where π_u is the probability of the response function in (6.8), which has frequency r_u in the data matrix. However, also this algorithm is slow, especially for large number of items. For this reason, Bock and Aitkin (1981) proposed a version of the method in which an EM algorithm is implemented, in such a way that a realistic larger number of items can be handled. Moreover, this method characterizes the ability distribution empirically and, hence, arbitrary assumptions about its form are avoided.

In addition to the methods described above, a variety of Bayesian approaches that estimate the parameters of the two and three parameters logistic models have been proposed (see, for example, Tsutakawa, 1992, Tsutakawa and Lin, 1986, Swaminathan and Gifford 1985, 1986, Tsutakawa and Johnson, 1990, Tsutakawa and Soltys, 1988).

6.2.4 A Bayesian estimation of the normal ogive model

Rather than assuming the logistic representation, one can adopt the normal one, which leads to the so called normal ogive model. We will focus on the Bayesian approach for estimating the parameters of the model, since this will be essential for the presentation of section 6.3.

The approach to estimating the two-parameter normal ogive model has been presented by Albert (1992). A generalization for the inclusion of the guessing parameter can be found in the work of Sahu (2002) (see also Béguin and Glas, 2001).

Following the notation of the previous sections, the Bayesian estimation of the three parameter normal ogive model is outlined in the following.

Let $Y_{ij} = 1$ denotes a correct response of a person i on an item j. The probability of a correct response is given by:

$$\begin{aligned} P(Y_{ij} = 1; \theta_i, a_j, b_j, c_j) &= c_j + (1 - c_j)\Phi(\eta_{ij}) \\ &= \Phi(\eta_{ij}) + c_j(1 - \Phi(\eta_{ij})), \end{aligned} \tag{6.10}$$

where Φ denotes the standard normal cumulative distribution function, $\eta_{ij} = a_j\theta_i - b_j$

and c_j is the "pseudo-guessing parameter". Under this scheme, the respondent either knows the correct answer with probability $\Phi(\eta_{ij})$ or guesses correctly with probability $(1 - \Phi(\eta_{ij}))$. Hence, a correct response does not necessarily means knowledge of the respondent. Suppose that K_{ij} is a binary vector that is equal to 1 if person i knows the correct answer to item j and 0 otherwise (in the two-parameter logistic or normal ogive models, K_{ij} are identical to the actual response Y_{ij}, this data augmentation with the latent K_{ij} is fruitful in the three parameter models though). Thus, the conditional probability of K_{ij} given Y_{ij} is:

$$\begin{aligned} P(K_{ij} = 1|Y_{ij} = 1, \eta_{ij}, c_j) &\propto \Phi(\eta_{ij}) \\ P(K_{ij} = 0|Y_{ij} = 1, \eta_{ij}, c_j) &\propto c_j(1 - \Phi(\eta_{ij})) \\ P(K_{ij} = 1|Y_{ij} = 0, \eta_{ij}, c_j) &= 0 \\ P(K_{ij} = 0|Y_{ij} = 0, \eta_{ij}, c_j) &= 1. \end{aligned}$$

In addition, the latent variables K_{ij} are further augmented by the underlying (latent) variables $Z_{i,j}$ for all i, j, which are independent and normally distributed with mean η_{ij} and standard deviation equal to 1, such that $Z_{ij} > 0$ if $K_{ij} = 1$ and $Z_{ij} \leq 0$ if $K_{ij} = 0$. This leads to the following conditional distribution of Z_{ij}:

$$p(Z_{ij}|K_{ij}, \eta_{ij}) = \phi(Z_{ij}; \eta_{ij}, 1)(\mathbf{I}(Z_{ij} > 0)\mathbf{I}(K_{ij} = 1) + \mathbf{I}(Z_{ij} \leq 0)\mathbf{I}(K_{ij} = 0)), \quad (6.11)$$

where $\phi(Z_{ij}; \eta_{ij}, 1)$ is the normal density with mean η_{ij} and standard deviation 1.

The joint posterior distribution which we need to simulate from is:

$$p(\mathbf{Z}, \mathbf{K}, \boldsymbol{\theta}, \mathbf{a}, \mathbf{b}|\mathbf{y}) = \mathbf{p}(\mathbf{Z}, \mathbf{K}|\mathbf{y}; \mathbf{a}, \mathbf{b}, \mathbf{c}, \boldsymbol{\theta})\mathbf{p}(\boldsymbol{\theta})\mathbf{p}(\mathbf{a}, \mathbf{b})\mathbf{p}(\mathbf{c}). \quad (6.12)$$

If we set a prior $p(a, b) = \prod \mathbf{I(a_j > 0)}$ for the block parameters a and b, and a conjugate non informative prior $Beta(1, 1)$ on the guessing parameter c_j, the draws can be obtained from the following full conditional distributions:

- Draw $Z_{ij}|\mathbf{K}, \boldsymbol{\theta}, \mathbf{a}, \mathbf{b}, \mathbf{y})$ from the conditional distribution $N(\eta_{ij}$, truncated on the left by 0 if $K_{ij} = 1$ or truncated on the right otherwise.

- Draw from the conditional distribution of $\boldsymbol{\theta}$, given the remaining parameters:

$$N\left(\frac{\sum_j a_j(Z_{ij} + b_j) + \mu/\sigma^2}{\sum_j a_j^2 + 1/\sigma^2}, \frac{1}{\sum_j a_j^2 + 1/\sigma^2}\right) \tag{6.13}$$

  where μ and σ are the parameters of the normal prior on $\boldsymbol{\theta}$.

- Draw from the conditional distribution of the block parameters $\mathbf{a}, \mathbf{b}$, which is $N((\mathbf{X^TX})^{-1}\mathbf{X^TZ_j}, (\mathbf{X^TX})^{-1}\mathbf{I(a_j > 0)})$, where $\mathbf{X} = (\boldsymbol{\theta} - \mathbf{1})$, $\mathbf{1}$ is the n-dimensional column vector with elements 1.

- Draw from the conditional distribution of K, given the remaining parameters:

$$\begin{aligned} P(K_{ij} = 1|Y_{ij} = 1, \eta_{ij}, c_j) &\propto \Phi(\eta_{ij}) \\ P(K_{ij} = 0|Y_{ij} = 1, \eta_{ij}, c_j) &\propto c_j(1 - \Phi(\eta_{ij})) \\ P(K_{ij} = 1|Y_{ij} = 0, \eta_{ij}, c_j) &= 0 \\ P(K_{ij} = 0|Y_{ij} = 0, \eta_{ij}, c_j) &= 1. \end{aligned}$$

- Draw from the conditional distribution of the probability for guessing c, given $\mathbf{K}$ and $\mathbf{y}$: $Beta(s_j + 1, t_j - s_j + 1)$, where t_j denotes the number of the students who do not know the correct answer and guess, and s_j is the number of correct responses of those who do not know the answer and guess for the response.

In addition, one may assume that the ability of the students is not unidimensional; then the ability to answer may require several different cognitive tasks, or, otherwise stated, several factor to contribute to the correct response. Thus, the above model can be extended to the multidimensional case, where a latent factor analysis is performed. A model under this assumption is described in section 6.3.

6.2.5 Data structures other than the dichotomous case

The models that presented in the previous sections cover the dichotomous case of the data, where the responses are of the "correct-incorrect" form. Several modifications of the above models have been proposed in the literature that deal with other variations of the data types. Incomplete responses is one of these cases, where the students do not respond to the same set of items. Also, the rating scale model (see Andersen, 1997) is a modification of the Rasch model that deals with polytomous responses to a set of test items. In this case, the response categories are scored in such a way that the total score of the items constitute a rating of the respondents on a latent scale. Typical examples of such data structures are often found in the psychological tests where, for example, the respondent answers questions like "Feeling nervous making decisions" or "Feeling lonely" where the possible responses could be " Always", "Often", "Moderately", "A little bit" and "Not at all".

Another variation of the Rasch family of models is the partial credit model (see Masters and Wright, 1997). This model deals with responses recorded in two or more ordered categories in such a way that combines results across items to obtain measures on some underlying variable. An example of this data structure could arise when several markers (which are considered as different "items" grade essays of students in a 5-scale evaluation (e.g. A for "excellent", B, C, D, or E for "very bad"). By combining the different evaluations of the markers, the partial credit model can estimate the ability of each student in writing essays.

These are two typical examples of the plethora of the data structures that can arise and the respective models to deal with. However, there is still no appropriate model in the literature that can deal with responses which include omissions. For example, in the American SAT examination, there is a penalty for each wrong response, whereas the omission is acceptable (and does not subtract points which is not the case of a wrong response). Nonetheless, omissions provide useful additional information on the students' ability and have to be treated in a different manner than the wrong responses. Moreover, these are definitely not missing data which are not to be included in the analysis; they do give a different aspect of the general ability. A model that deals with this kind of data structure is proposed in the next section.

6.3 The proposed model

Suppose that there are n students ($i = 1, ..., n$), each facing k multiple choice questions ($j = 1, ...k$). In addition, suppose that d latent groups of different abilities (factors) form the general ability of each student. The latent ability w_{ij} of student i on item j is supposed to follow a normal distribution with unit variance, i.e.

$$N(\sum_{p=1}^{d} a_{jp}\theta_{ip} - b_j, 1). \tag{6.14}$$

The actual responses restrict the underlying (latent) quantities w_{ij} as follows:

- y_{ij}=1 (correct response), then $w_{ij} > \delta_i$
- y_{ij}=0 (incorrect response), then $w_{ij} < 0$
- y_{ij}=omission, then $w_{ij} \leq \delta_i$,

where δ_i is the positive threshold ability that a student has to exceed so as to respond correctly due to uncertainty (the same across the items). Moreover, there is a probability c_j that $w_{ij} \leq \delta_i$ even if $y_{ij} = 1$; we call it guessing probability for item j.

Under this scheme, the parameters to be estimated are:

- difficulty parameter b_j of each item: the higher the value of this parameter, the more difficult the item is.

- guessing parameter c_j for each item. Estimation of guessing parameters is important since students' chance success due to guessing contributes to error variance and diminishes the reliability of multiple choice tests (see Zimmerman and Williams, 2003) .

- threshold parameter δ_i for each student: it represents the difference between the ability that corresponds to the correct answers and the ability that corresponds to omissions. In other words, δ_i shows how risky the student is.

- Discrimination parameters a_{jp} for each item. Since the multiple choice test consists of several items which require different cognitive abilities to be answered, we estimate different discrimination parameters for different factors (factor loadings) per item. We assume that most of the items require a combination of cognitive abilities and the form of this combination is of major interest. Therefore, we use a d-dimensional factor item response model (Béguin and Glas, 2001).

- The abilities θ_{ip} for each student: thus, we can find the students with good performance but also students with a high general performance in one cognitive task but low performance in another. Also, sometimes we need to discriminate between students who perform well due to their discernment and students who perform well just because are good to remember things by heart.

The conjunction of the latent abilities and the probabilities that student i will answer item j correctly, is straightforward (for a nice presentation, see Albert and Ghosh, 2000). Given these constraints, the likelihood of the model is proportional to

$$\prod_{ij} \exp(-\frac{1}{2}(w_{ij} - \sum_{p=1}^{d} a_{jp}\theta_{ip} + b_j)^2). \tag{6.15}$$

Note that the likelihood is free from the data y_{ij}; they influence the likelihood only indirectly through the constraints that form the underlying variables w_{ij}.

6.3.1 Prior and Conditional Distributions

We place a normal prior distribution for $\Theta_i = (\theta_{i1}, \theta_{i2}, ..., \theta_{id})^T$, say $N_d(\mathbf{0}, T^{-1})$. Then, the conditional distribution of Θ_i given $B = (b_1, b_2, \ldots, b_k)$ and $W_{i.} = (w_{i1}, w_{i2}, \ldots, w_{ik})$ is

$$\Theta_i^T | A, B, W_{i.} \sim N_d((AA' + T)^{-1} \sum_j A_j(w_{ij} + b_j), (AA' + T)^{-1}),$$

where A is a $d \times k$ matrix of the discrimination parameters a_{ij} and A_j is the $j-th$ column of A. To ensure identifiability, we set T to be equal to the identity matrix.

Using a normal prior distribution $N_{(d+1)}(\mathbf{0}, V)$ for the block parameter vector $(A_j^T \; b_j)^T$, the conditional distribution of the block parameters is

$$(A_j^T \; b_j)^T | \mathbf{\Theta}, W_{.j} \sim N((X'X + V^{-1})^{-1} X' W_{.j}, (X'X + V^{-1})^{-1}),$$

where $X = (\mathbf{\Theta} \; -\mathbf{1})$ is an $n \times (d+1)$ matrix, $\mathbf{\Theta} = [\Theta_1^T \; \Theta_2^T \; \ldots \; \Theta_n^T]^T$, $(\mathbf{1})$ is a column vector with elements (1), and $W_{.j} = (w_{1j}, w_{2j}, \ldots, w_{nj})$. To make the model identifiable, we set each element of A A_{ij}=0, if $i > j$ (hence, we fix one element equal to 0 for a two-factor model, three elements for a three-factor model etc., see Fraser, 1988).

In the model, we treat the parameters δ_i as random effects which are drawn from a normal distribution. Suppose we set an inverse gamma prior on the variance of the random effects, with parameters 0.5 and 1. Then, the variance of the conditional distribution for the random effects is:

$$\sigma_\delta^2 \sim 1/gamma(\frac{n+1}{2}, 1 + \frac{\sum_i (\delta_i - \mu_\delta)^2}{2}).$$

The mean of the random effects is given by:

$$\mu_\delta \sim N(\sum_i \delta_i / n, \sigma_\delta^2 / n).$$

The distribution of δ_i is normal with mean μ_δ and variance σ_δ^2, appropriately truncated between the values lo_i and up_i, where

$$lo_i = max(0, max_j[w_{ij} I(y_{ij} = omission)])$$

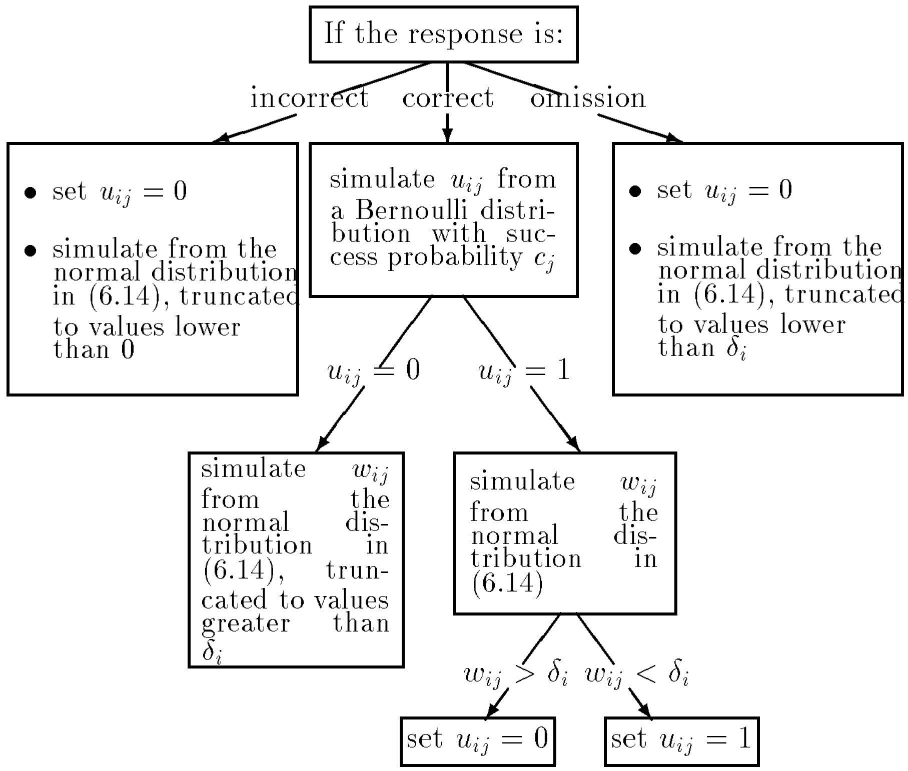
and

$$up_i = min_j(w_{ij} I(w_{ij} > \delta_i)),$$

where $I(.)$ is an indicator function with values 1 if (.) is true, otherwise w_{ij} is omitted from the calculations. If $I(w_{ij} > \delta_i)$ is false for all j, δ_i is set to be equal to lo_i (see Linardakis and Dellaportas, 2003). In addition, δ_i are considered to be random effects with mean and variance to be estimated.

To simulate the guessing parameters c_j, we introduce Bernoulli random variables u_{ij} with success probability c_j. Depending on the actual response, we simulate the values of u_{ij} and w_{ij}, as it is shown in the diagram below.

Since c_j is the probability of success, we assume a conjugate beta prior distribution with parameters κ and λ. Hence, the posterior conditional distribution is Beta($\kappa + \sum_i u_{ij}$ +(number of incorrect responses in item j), λ+(number of omissions in j)+(number of correct responses in j)-$\sum_i u_{ij}$).



Finally, the latent variables w_{ij} are drawn from (6.14), appropriately truncated using the information given by the original data y_{ij} and the quantities δ_i.

6.4 Code for the δ IRT Model

We will concentrate our presentation a bit more on the parameters w_{ij} and c_j, by looking at the codes that can be used, since these parameters simulated in a different way in the δ IRT model, compared to the other widely used models. For simplicity of the presentation, we will focus on the unidimensional model, the generalization to the multivariate "factor model" case is straightforward.

The latent variables w_{ij} are truncated such that:

if response=wrong then

begin

$u_{i,j} := 0;$

$w_{i,j} := normal(\theta_i * a_j - b_j, 1);$ *truncated to negative values*

end

if response=omission then

begin

$u_{i,j} := 0;$

$w_{i,j} := normal(\theta_i * a_j - b_j, 1);$ *truncated to the interval* $(-\infty, \delta_i)$

end

if response=correct then

begin

$u_{i,j} := bernoulli(c_j);$

if $u_{i,j} = 0$ *then*

begin

$w_{i,j} := normal(\theta_i * a_j - b_j, 1);$ *truncated to the interval* (δ_i, ∞)

end

else if $u_{i,j} = 1$ *then*

begin

$w_{i,j} := normal(\theta_i * a_j - b_j, 1);$

if $w_{i,j} < \delta_i$ *then* $u_{i,j} := 1$ *else* $u_{i,j} := 0;$

end;

end;

Finally, the guessing parameters c_j are drawn such that:

- $n_{.j}$ =number of correct responses and omissions of item j
- $u_{.j} = \sum_i u_{ij}$
- $c_j := beta(\kappa + u_{.j} + n - n_{.j}, \lambda + n_{.j} - u_{.j});$

Chapter 7

An Application of the Proposed IRT Model

7.1 Introduction

In this chapter, we analyze multiple choice response data in multiple choice tests when there are penalties for each wrong answer such as a subtraction of points (a widely used technique that attempts to prevent students from guessing); the literature is still sparse and the usual item response theory models are inappropriately used. We extend the use of item response models to capture this situation by including guessing and threshold latent parameters. We also separate the ability of each student into several parts, which express different cognitive tasks by a multidimensional scaling approach. A Pseudo-Bayes factor model choice approach (based on cross-validation predictive densities) is used to select the number of dimensions that fit the data better. The model with δ_i parameters is also compared with the three parameter normal ogive model, in terms of predicted students ranking.

7.2 Data and Models Used

We analyze a data set obtained from the exam paper of the course "Data Analysis" of the third and fourth year (102 students) in the department of Statistics, Athens University of Economics and Business. Some selected items of this test are given in Appendix A.

We constructed the test of 33 items using the terms:

- 3 points for each correct response
- 0 points for omission
- -1 point for each incorrect response.

The total score for each student was the sum of the points in each of the 33 items.

The MCMC of the conditional distributions mentioned in section 6.3, as well as in section 6.2.4 in its multidimensional form, was implemented using the Gibbs sampler. We applied four different models on the data set, with two and three factors and with and without threshold parameter δ (sections 6.3 and 6.2.4 respectively) .

7.2.1 Prior specifications

To ensure identifiability of the model, we set the prior variance of the parameters a_{jp} equal to 1 (the first d elements of the diagonal of the prior variance matrix V). A diffuse prior was used for b_j; so the $(d+1)th$ element of the diagonal was set to 10^4. Finally, the parameters of the Beta prior distribution of c_j were set equal to 2 and 6 for κ and λ respectively. These parameters were chosen so that $E(c_j)$ is a prespecified value, for example 0.25, see Sahu (2002) and references therein.

7.2.2 Convergence of the parameters

Taking sampled values with lag 5, we produced MCMC chains with autocorrelation close to 0. A burn-in equal to 100 was eliminated from the sampled values and convergence was verified both graphically and with the Raftery-Lewis, the Heidelberger and Welch, and the Geweke convergence criteria.

Raftery and Lewis convergence diagnostic uses a small number of iterations and the diagnostic responses the total number of iterations that is needed for the specific problem in order to obtain convergence, the number of iterations which have to be discarded (burn-in), a dependence factor (values close to 1 indicate independence within each chain) and a thinning value k that may be used in order to reduce the dependence within each chain. The value of k is set to a value high enough that successive draws of the parameter are approximately independent (for details, see Raftery and Lewis, 1992).

> "*This strategy, known as thinning, can be useful when the set of simulated values is so large that reducing the number of simulations by a factor of k gives important savings in storage and computation time. Except for storage and the cost of handling the simulations, however, there is no advantage in discarding intermediate simulation draws, even if highly correlated*" (Gelman, 1995).

3000 sampled values from each posterior distribution were available for inference. These values seem to be a stationary chain with low autocorrelation. We used the Bayesian Output Analysis Program (BOA)[1] of Brian Smith that provides the convergence criteria for the MCMC chains. Table 7.1 shows the results of the Raftery and Lewis criterion, of an accuracy = ± 0.01 and probability = 0.9. To save space, we present the results of some randomly selected parameters of the three-factor model. The results of all the remaining models which are used in the following sections, behaved in a similar way, with respect to the convergence.

We also used the Geweke convergence diagnostic (which responses z-scores between the interval (-2,2) in the case where convergence is obtained) and the Heidelberger and

[1]The program may be found and is freely available at www.public-health.uiowa.edu. It runs under S-plus and R

Table 7.1: Raftery and Lewis Convergence diagnostic for the parameters for the IRT model with delta

Thin	*Burn-in*	*Total*	*Dependence Factor*
	ability parameters		
2	2	1456	2.206061
1	1	710	1.075758
1	2	777	1.177273
1	1	705	1.068182
1	2	741	1.122727
1	1	693	1.050000
	discrimination parameters		
1	2	741	1.122727
1	1	800	1.212121
1	2	805	1.219697
1	1	666	1.009091
1	2	772	1.169697
1	2	741	1.122727
	difficulty parameters		
1	2	816	1.2363636
1	1	693	1.0500000
1	1	655	0.9924242
1	1	628	0.9515152
1	1	682	1.0333333

Welch stationarity and interval halfwidth test (which also suggests the number of iterations to be kept and to be discarded). For details about these diagnostics, see Cowles and Carlin (1994).

Table 7.2 shows the Geweke diagnostic criterion results. All of the parameters' sampled values that are shown, resulted to a z-score between the interval (-2,2), indicating that convergence has been achieved. The same results were obtained by the Heidelberger and Welch diagnostic; all of the parameters passed the test, where there was no need in discarding additional sampled values.

Finally, Table 7.3 shows the autocorrelations of the same selected parameters. All of them are close to 0, even with lag equal to 5. Hence, since the diagnostic tests also suggested not to use any additional lag, we used the 3000 sampled values for inference.

Table 7.2: Geweke Convergence diagnostic for the parameters for the IRT model with delta

Z-Score	*p-value*	*Z-Score*	*p-value*	*Z-Score*	*p-value*
ability parameters		*discrimination parameters*		*difficulty parameters*	
1.5574	0.1193	1.5485	0.1214	-0.8569	0.3914
0.1594	0.8732	0.0198	0.9841	-1.4124	0.1578
-0.6182	0.5363	-0.0137	0.9890	0.9409	0.3467
-0.7757	0.4378	-0.6014	0.5475	-0.9356	0.3494
-0.8172	0.4137	-1.1989	0.2305	-1.4046	0.1601
-0.2037	0.8385	-1.1595	0.2462		

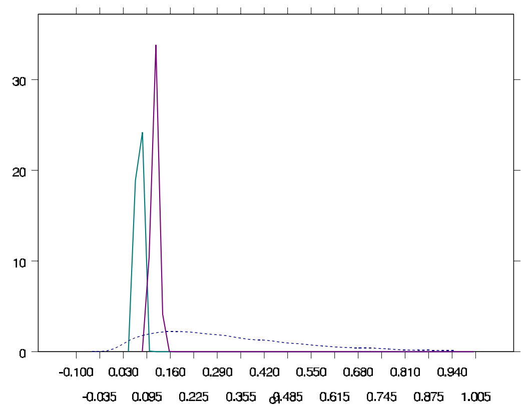


Figure 7-1: Prior and posterior densities of guessing parameters.

7.2.3 Sensitivity Analysis

A necessary feature of the model is that the prior specification does not force the posterior densities. We examine the guessing parameters behaviour graphically in Figure 7-1. The solid lines show densities of two randomly selected guessing parameters and the dashed line shows the prior density that was used. Obviously, the prior variance is much greater than the posterior ones, hence evidence is provided that the densities have not been forced by the prior one.

In addition, recall that the prior variance of the difficulty parameters was equal to 1000. The posterior variance of these parameters was close to 0.005. Even for the

Table 7.3: Autocorrelations for the parameters for the IRT model with delta

Lag 1	*Lag 5*	*Lag 10*	*Lag 50*
	ability parameters		
0.1485	0.0273	-0.0285	0.0017
0.1026	-0.0160	0.0226	-0.0059
0.1287	-0.0174	0.0351	0.0088
0.1624	0.0211	0.0096	-0.0155
0.0568	0.0151	-0.0017	-0.0165
0.1290	0.0143	0.0532	0.0212
	discrimination parameters		
0.3330	0.0149	0.0430	0.0253
0.0824	-0.0139	-0.0162	-0.0109
0.1981	0.0251	-0.0155	-0.0226
0.1776	0.0090	-0.0103	0.0116
0.3136	0.0052	-0.0097	0.0094
0.3018	-0.0027	0.0039	-0.0031
	difficulty parameters		
0.1766	0.0201	0.0424	-0.0135
0.1001	0.0154	-0.0042	-0.0346
0.2881	0.0432	0.0130	-0.0052
0.0758	0.0031	0.0336	0.0212
0.4469	0.1038	0.0060	0.0009

discriminatory parameters, which had a prior variance equal to 1 for purposes of model identification, the posterior variance of them was close to 0.007.

7.2.4 The Pseudo-Bayes Factor

To compare the four models, we used the pseudo-Bayes factor (see Sahu, 2002) that is defined as the ratio of the cross-validation predictive densities under the two competitive models. The cross-validation predictive density is defined as:

$$\pi(y_r|\mathbf{y_{(r),obs}}) = \int \boldsymbol{\pi}(\mathbf{y_r}|\boldsymbol{\zeta}, \mathbf{y_{(r),obs}})\boldsymbol{\pi}(\boldsymbol{\zeta}|\mathbf{y_{(r),obs}})\mathbf{d}\boldsymbol{\zeta}, \tag{7.1}$$

where $\mathbf{y_{(r),obs}}$ denotes the set of observations $\mathbf{y_{obs}}$ with rth component deleted, and ζ is the set of the parameters in the model.

The Pseudo-Bayes factor is then defined as the ratio of (7.1) under the two models. That is:

$$PsBF = \prod_{r=1}^{N} \frac{\pi(y_{r,obs}|\mathbf{y_{(r),obs}}, \mathbf{M_1})}{\pi(y_{r,obs}|\mathbf{y_{(r),obs}}, \mathbf{M_2})}, \tag{7.2}$$

where N is the total number of observations (i.e. the product of the number of items and the number of students). This quantity is a surrogate for the Bayes factor and has similar interpretation.

The computation of (7.2) involves the evaluation of (7.1) for each binary response. Having the MCMC sampled values for the parameters (say B sampled values of each parameter of the model), a Monte Carlo estimate of (7.1) is given by:

$$\hat{\pi}(y_r|\mathbf{y}_{(r),obs}) = \frac{1}{B}\sum_{t=1}^{B} \frac{1}{p_{ij}^{y_{ij}}(1-p_{ij})^{1-y_{ij}}}, \tag{7.3}$$

where p_{ij} is the probability of correct response under the assumed model and it is evaluated at the simulated parameter values ζ. This estimation is quite accurate when B is large (see Gelfand and Dey, 1994).

The logarithm of the quantities which are needed to calculate the Pseudo-Bayes factors in our models are given in Table 7.4.

Table 7.4: Log(cross-validation predictive densities).

model	*without* δ	*with* δ
2 factors	-836.81	-828.03
3 factors	-817.11	-810.29

Note that the Pseudo-Bayes factor for the three-factor model with the inclusion of δ indicated supremacy over the three other models, producing pseudo-Bayes factors greater than exp(6.82) (for example, exp(-810.29+817.11)=exp(6.82) That is, the three-factor model with delta parameters is exp(6.82) times more probable than the three-factor model without delta parameters). In the following, we use the three-factor model with δ

for inference.

7.3 Predicted ranking of the students

Recall that the Pseudo Bayes factor showed that the model with the inclusion of the parameters δ_i is preferable compared to the model without these parameters. We can also check the behaviour of the models in terms of the ranking of the students (the student with the highest ability takes the ranking number "1", etc.). Since the grade the students received (i.e. 3 points per correct answer, 0 points per omission and -1 point per wrong response) does not take into account any weight of the items, we consider it as a "unidimensional" calculation. Thus, to make it more comparable with the results of the models, we run the unidimensional models (that is, without the inclusion of the factors), with and without the parameters δ_i. The former is the unidimensional case of the model presented in section 6.3 and the latter is the unidimensional three parameter normal ogive model of section 6.2.4. It must be noted that in the case of the latter model, the omissions in the data set were treated as wrong responses, as it is the case in this model, without the help of the δ parameters.

After convergence had achieved, we ran the two models in order to obtain 1000 sampled values from the posterior distributions of the parameters. To save space, a lag equal to 10 was used. Using the values of the abilities of the students θ_i for each iteration, we sorted them in descending order (such that the highest performance or, otherwise stated, the highest θ_i received ranking number 1, etc.), From these 1000 ranking vectors per model, we constructed the 95% credible intervals of the rankings (i.e. 102 credible intervals, one per student).

We also constructed the ranking of the scores received on the test. These scores had many ties, especially in the middle of the rankings. That is, many students with an "average performance" received the same grade. In these cases, the mean ranking of the scores with ties was given to these students. Then, we counted the number of the

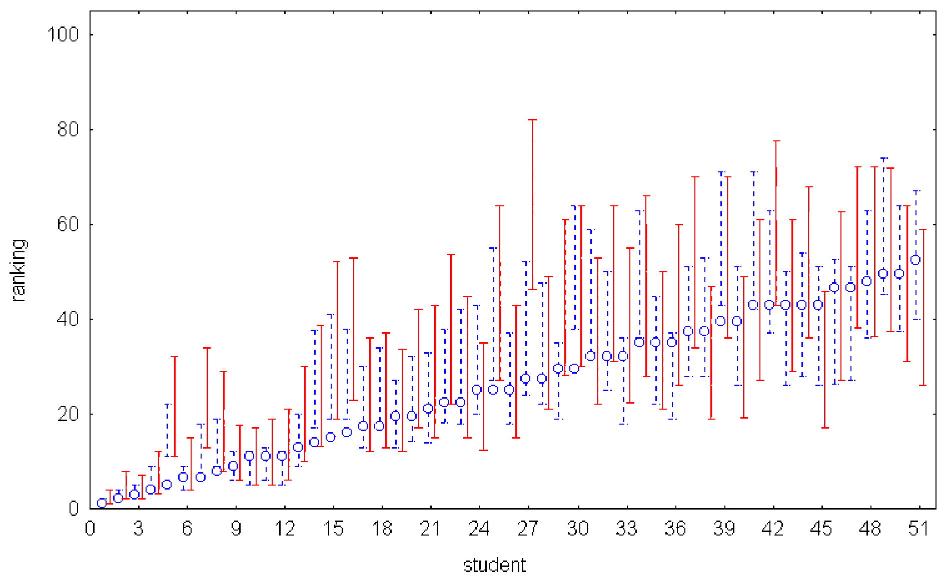


Figure 7-2: Credible intervals of the students rankings. Students ranked 1-51.

intervals per model which include the "based on the grade received" ranking. Both of the two models agreed with this observed ranking; the intervals of the model without δ_i parameters all included the observed ranking, whereas 101 out of 102 intervals of the model with δ_i included the observed ranking.

To make things more difficult, we also constructed the 50% credible intervals per student ranking for each of the two models. The results are shown in Figures 7-2 and 7-3.

In the graphs, the horizontal axis represents the students in ascending order of the ranks. The dotted lines represent the credible intervals which were constructed from the model without δ parameters, the solid lines, on the right of the dotted ones, represent the respective intervals from the model with δ parameters and, finally, the cycles represent the "based on the grade received" ranking.

Based on these results, 80 out of 102 intervals of the model with δ_i included the observed ranking. On the other hand, 69 out of 102 intervals of the model without

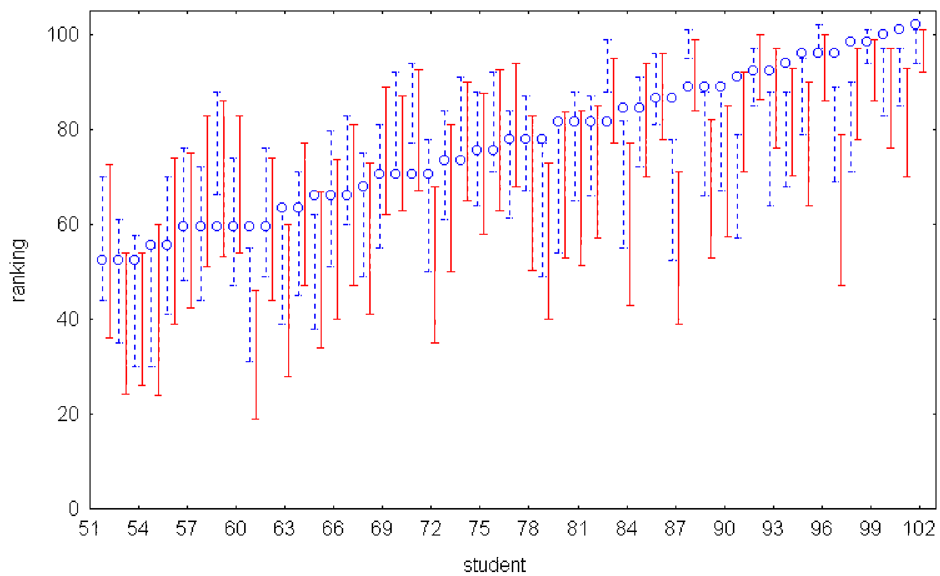


Figure 7-3: Credible intervals of the students rankings. Students ranked 52-102.

δ_i included the observed ranking. These results indicate a difference on the estimated parameters of the two models. Another important outcome is that the credible intervals of the model with δ_i parameters were, in general, wider than the respective intervals of the model without δ_i. Hence, the model that has taken into account the omissions (with the inclusion of δ_i) expresses a higher uncertainty (i.e. higher variability) on the estimated rankings, which may be interpreted as a larger number of ties on the ranks, or, otherwise stated, as a lower variability on the ability parameters. Finally note that the credible intervals of the low or high rankings (high and low performance respectively) are quite tighter compared to the intervals of the "mean performance" students (in the middle of the graphs). This is quite plausible since (as it was mentioned before) a lot of ties were observed at the "mean performance" students, which means that the estimated abilities were close to each other, producing wider credible intervals of the rankings.

7.4 Output of the three factor IRT model with δ_i parameters

In section 7.2.4, it was shown that the three factor model with delta is preferable compared to the remaining models that were examined. Hence, our interpretation of the results will be based on the output of this model.

The three factors seem to have valuable physical interpretation. Factor 1 seems to include items that require the recovery of specific numbers from output tables. Items of this factor require students to remember and use formulae (see Table 7.5 for some selected items, their loadings to the factors and the difficulty and guessing parameters)[2]. Factor 2 seems to include items that require both memory and discernment. Finally, factor 3 includes methodological items (appropriate analysis etc.).

We calculated the total estimated ability of each student as the sum of the abilities of the three factors (the mean of the 3000 sampled values per person of the estimands θ_{i1}, θ_{i2} and θ_{i3}). A scatter plot of these abilities against the actual total score is given in Figure 7-4. Note that the estimated abilities are close to the actual score but not identical. It seems that, using a single sum to construct the observed total is unfair for some students.

In Table 7.6, we provide the characteristics and estimated parameters of students who appear to have interesting deviations from the general behavior. Note that students no. 66 and 72 are much better in items of factor 1 compared to their general performance. Moreover, student no. 26 has much better performance in items that are methodological or require discernment than items which require one to remember things by heart. Student 102 is better in methodological items whereas student no. 92 is worse in them compared to his general score. Student no. 18 would be unfairly dealt with a test with only methodological or discernment items. On the other hand, students no. 26 and 61 would be unfairly dealt with a test with only items that require recovery of numbers from

[2]More details on items that load on each particular factor as well as characteristics of the items can be found at the end of Appendix A

Table 7.5: Indicative items and parameters

j	Item	a_{j1}	a_{j2}	a_{j3}	b_j	c_{j1}
1	The cross-tabulation provides the relation between gender and smoking (yes-no). The % of female who smoke is...	**0.81**	0	0	-0.50	0.15
14	The ANOVA table of a linear model is given. R^2 was calculated as:...	**1.06**	0.44	0.43	0.31	0.15
16	The ANOVA table of a linear model is given. The sum of squared residuals is...	**1.50**	0.50	0.51	-0.80	0.18
7	*Find the linear model that has been used by just looking at the graph of the unstandardized predicted values versus a continuous independent variable. (answer:linear regression with one independent dummy variable; the continuous one was not used)*	0.39	**0.46**	0.25	1.30	0.13
17	*Find the linear model that has been used by just looking at the graph of the unstandardized predicted values versus a continuous independent variable. (answer:linear regression with the continuous variable of the graph as independent)*	0.12	**0.41**	0.29	0.62	0.21
32	A simple linear regression equation was estimated as $Y = 4.15 - 0.96X + \epsilon$ with $R^2 = 0.81$. The correlation between X and Y is equal to...	0.69	**1.00**	0.89	0.91	0.17
10	In a linear model, in order to include an independent categorical variable with 4 levels, where each level denotes a different profession category, we have to...	0.20	0.55	**0.88**	-0.55	0.15
12	A data set includes the results of a medical test with 2 possible outcomes and a variable that denotes whether the person is a smoker or not. We will try to analyze the data using...	0.24	0.35	**0.75**	-0.27	0.15
33	In a data set, the dependent variable is the grade in a test (max 105) and the independent variables are the gender and the marital status (single, married, divorced). A possible analysis could be...	0.38	0.59	**0.62**	0.85	0.17

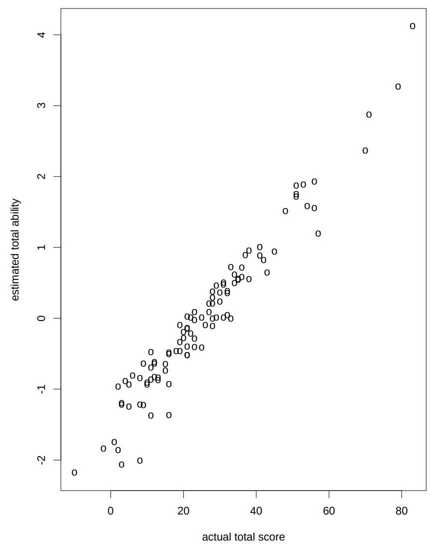


Figure 7-4: Scatter plot between estimated total ability and actual total score.

tables.

Table 7.6: Students' parameters and characteristics.

id #	correct	incorrect	omissions	total score	δ	ability 1	ability 2	ability 3
18	16	16	1	32	0.08	0.70	-0.37	-0.29
26	16	14	3	34	0.08	-1.18	0.74	1.05
57	20	6	7	54	0.11	0.62	0.37	0.59
61	20	12	1	48	0.09	-0.08	0.83	0.76
66	15	13	5	32	0.10	0.96	-0.13	-0.45
72	16	13	4	35	0.09	0.98	0.03	-0.46
92	26	7	0	71	0.09	1.62	0.85	0.40
102	21	10	2	53	0.09	0.29	0.62	0.98

The correlations between the mean of θ_{i1} per person and the corresponding means of θ_{i2} and θ_{i3} are 0.295 and 0.064 respectively. Moreover, the correlation between the means per person of θ_{i2} and θ_{i3} is 0.677; the ability in items of factor 1 (which require memory or recovery of appropriate numbers from tables) is almost uncorrelated to the

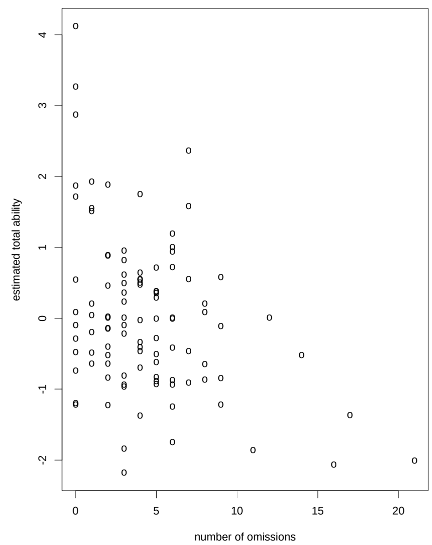


Figure 7-5: Scatter plot between estimated total ability and number of omissions per student.

abilities related to methodology or discernment. On the other hand, methodology and discernment are positively correlated (as it is plausible). This relation also provides some evidence that one more factor in the model (a forth one) may not be useful.

Figure 7-5 shows the relation between the estimated total ability and the number of omissions; a negative relation shows that the greater the ability is, the more confident the student is (which is also plausible). However, note, in Table 7.6, that student no. 57 has a high performance but he is unexpectedly highly uncertain.

In Figure 7-6 it is shown that the parameters δ_i are positively related to the number of omitted items per student (which is also plausible; the greater δ_i means the less risky student and, hence, the greater number of omissions).

Moreover, we calculated the correlations of the ability per factor with the parameters δ_i; the greater the ability is, the lower δ_i (the more risky the student is); δ_i does not seem to be related to the abilities of factor 1 (the correlation is equal to -0.126), whereas, it is negatively related to the abilities of factor 2 and 3 (correlations equal to -0.435 and

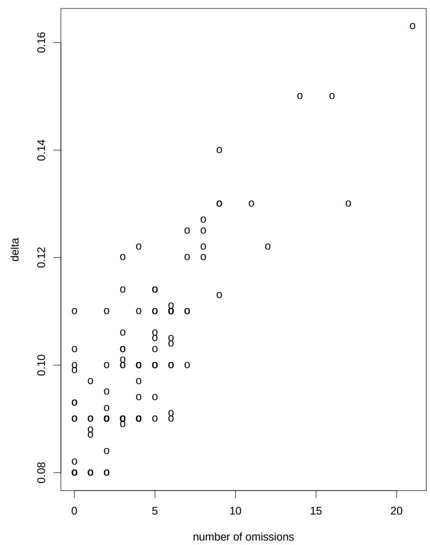


Figure 7-6: Scatter plot between δ_i and number of omissions per student.

-0.357 respectively). Note, in Table 7.6, however, that, for example, student no. 18 is unexpectedly risky if we take into account his low performance in items of factor 2.

Finally, Figure 7-7 presents the difficulty parameters of the items against the number of correct responses per item. A strongly negative relation provides evidence that the data support the model and the estimated parameters.

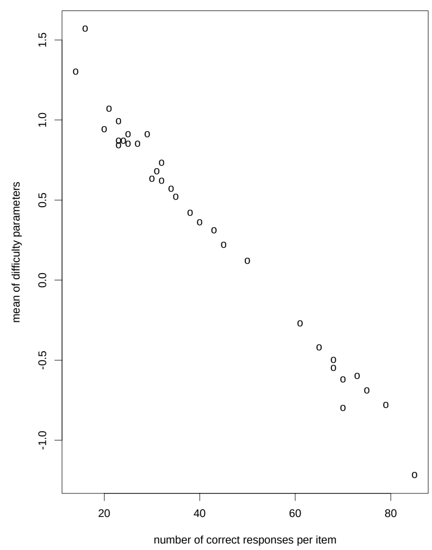


Figure 7-7: Scatter plot between difficulty parameters and number of correct responses per item.

Chapter 8

SAT data analysis via a latent variables factor model with thresholds

8.1 Introduction

In this chapter, the analysis of the SAT data set is presented. A brief presentation of the SAT tests and the data set that is used here are given, as well as of the model that is used and the prior specifications are also provided. The model with δ_i parameters is compared with the three parameter normal ogive model, in terms of students ranking. The features of the model with respect to the convergence (graphically examined) is given. Finally, the outcome of the model and inference on the estimated parameters are given both with Tables and graphically

8.2 The SAT Data

8.2.1 General

The SAT test[1] is offered to American students after or during the high school. A high grade on the test is an admission requirement of the college the students are interested in attending, since it is supposed to measure the student's ability to do college-level work. The basic SAT I tests include the verbal and the math tests, each of them containing different sessions that are supposed to measure different skills and reasoning abilities of the students. The Verbal SAT I test includes different question types-sessions which refer to critical reading, analogies and sentence completions. The Math SAT I test includes arithmetic and algebraic reasoning and geometric reasoning.

The scores are reported on the 200-to-800 scale. The average score of the students in each of the two basic SAT I tests is about 500. The final score is calculated by a two step process. First, the raw score is calculated based on the following rules: Questions answered correctly receive one point. Omitted questions receive no points. For multiple choice questions answered incorrectly, a fraction of a point is subtracted; either 1/4 point is subtracted for five-choice questions or 1/3 point is subtracted for four-choice questions.

To get the raw score, a fraction of the multiple-choice questions answered wrong is subtracted from the number of questions answered correctly. If that resulting score is a fraction, it is rounded to the nearest integer number. At the second step, the raw score is changed into a scaled score; it is weighted to adjust for slight differences in difficulty between test editions and it ensures the comparability of the scores among different tests.

The available time for answering the test is calculated so as the time is enough for most of the students. Studies are done to find out whether most students have enough time to attempt to answer all the questions in each test section. These studies show the time limits are appropriate if all the students taking the test answer 75% of the questions in each section and if 80% reach the last question in the section. Based on these studies,

[1] For more details on the SAT tests, see the official web site of the SAT test www.collegeboard.com

the limits are appropriate for the majority of the students.

8.2.2 The analyzed SAT data set

In chapter 7, we illustrated the proposed IRT model with δ_i parameters, showing the convergence behaviour of the model, the influence of the priors used, the Pseudo Bayes factor model choice, the estimated ranking of the students and the interpretation of the results of the model, including the three ability factors which were resulted. In the data set that is analyzed in the present chapter, the convergence behaviour did not appear to have a different behaviour. Hence, (since the items of the test analyzed are not known, for further interpretation), we will focus on the estimated ranking of the students and on some students and items with specific features (e.g. risky students, items with high discrimination power etc.).

The data set that is used in the analysis refers to the October 1998 test takers. A sample of 1000 test takers is used for the purposes of the illustration. We use the raw responses (correct, incorrect, omit) of the test takers to the SAT I verbal test that consists of 3 sections. Section 1 includes 35 questions, section 2 consists of 31 questions and section 3 contains 12 questions. Thus, there are 78 questions in the whole test and the raw responses to be analyzed are 78000. Table 8.1 shows a summary of the 78 items on the test; it reports the percent of correct and incorrect answers and omissions that were given by the 1000 students to each item. The table also shows the difficulty parameters that were estimated; these will be discussed in section 8.3.5.

Note that there are some questions which are difficult compelling the students to omit them. See, for example, questions number 10 and 35. The percentage of incorrect responses are quite high; 50.9% and 54.7% respectively, so are the percents of omission; 26.8% and 33.1% respectively. So, these questions might be tricky and the students do not realize that they answer incorrectly. On the other hand, there are some other questions that while also difficult, they prevent students from a guessing behaviour. For example, questions 9, 16 and 44 are difficult as the previously mentioned questions are (percents of incorrect responses are 53.6%, 61.9% and 63% respectively). These questions prevent students from answering incorrectly though; the percents of omission are 16.2%, 3.5% and 7.2% respectively.

There are also some questions, for example the questions number 12 and 47, where almost all of the respondents answered them correctly; the percentages of correct responses are 95.8% and 90.1% respectively. Finally, the questions number 23 and 64 seem to be quite difficult to be answered from the majority of the students. The respective percentages of correct responses are 17.5% and 19.7%. It is interesting to see how these particular questions-items, that are interesting with respect to their percentages, will behave in the analysis and the outcome of the model.

8.3 Analysis

8.3.1 The estimated ranking of the students based on the SAT data

In section 7.3 we examined the estimated ranking of the students and we compared the results of the univariate normal ogive model with δ_i parameters, with the results of the univariate normal ogive model without δ_i. The same procedure was followed for the SAT data as well.

We focused on the 100 first students, since a ranking of the whole sample size (1000 test takers) would have included too many ties. We constructed the ranking of the scores received on the test. The scores had some ties, especially in the middle of the rankings, but these were fewer compared with the data in section 7.3. This was plausible since the number of items was larger and the grades were varying in a wider interval. After that, we constructed the 95% credible intervals of the estimated ranking and we counted the number of the intervals per model which include the “based on the grade received” ranking. Both the set of intervals derived from the two models included 99 out of 100 of the observed rankings.

We also constructed the 50% credible intervals per student ranking for each of the two models. The results are shown in Figures 8-1 and 8-2.

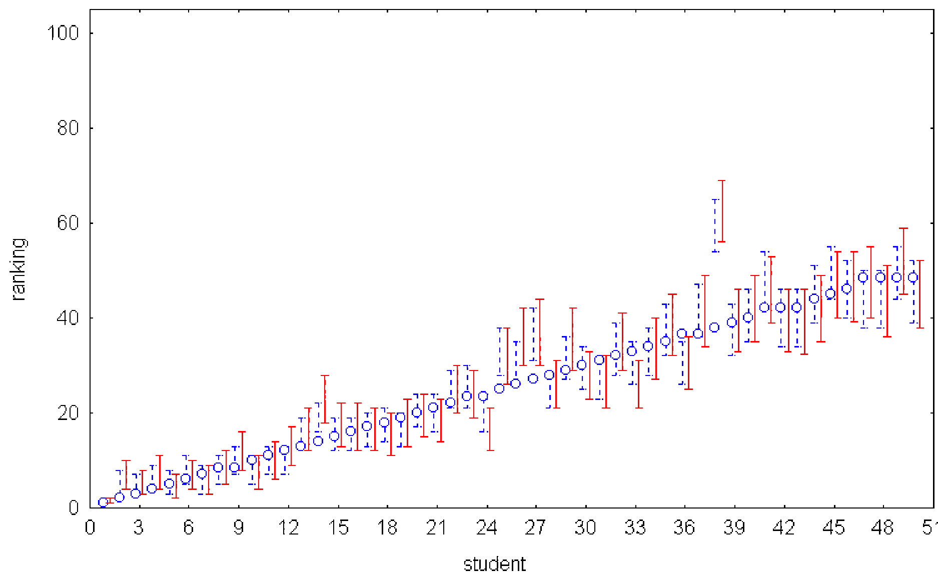


Figure 8-1: Credible intervals of the students rankings. Students ranked 1-50.

In the graphs, the dotted lines represent the credible intervals which were constructed from the model without δ parameters, the solid lines on the right of the dotted ones represent the respective intervals from the model with δ parameters and, finally, the cycles represent the "based on the grade received" ranking.

According to the results, 80 out of 100 intervals of the model with δ_i included the observed ranking, whereas 79 out of 100 intervals of the model without δ_i included the observed ranking. In addition, the credible intervals of the model with δ_i parameters were, in general, wider than the respective intervals of the model without δ_i, as it was also the case in the data set of section 7.3. However, if we compare these intervals with those of section 7.3, we note that the credible intervals of the rankings for the SAT test takers are tighter (probably, due to the larger data set, the larger number of items).

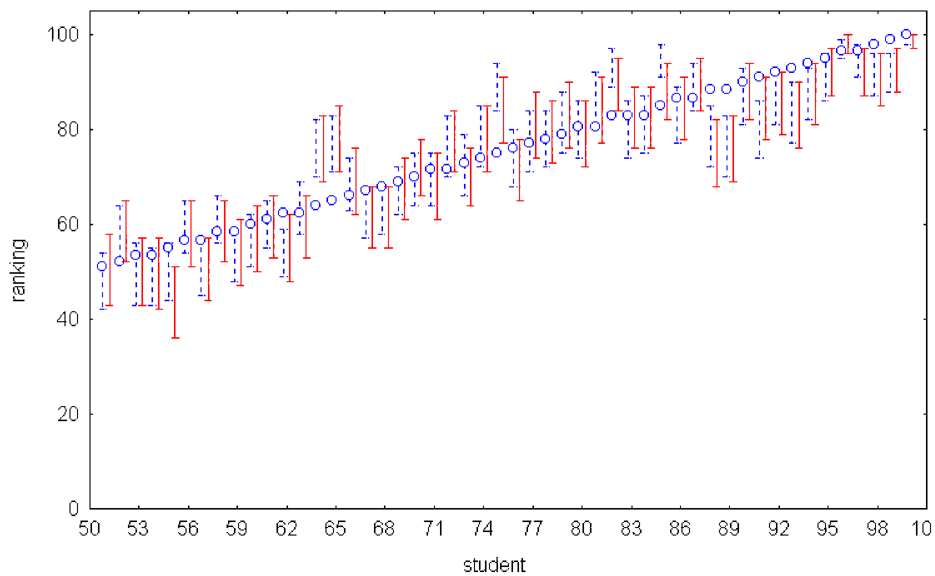


Figure 8-2: Credible intervals of the students rankings. Students ranked 51-100.

8.3.2 A three factor model

A three factor model is also used to analyze the SAT data set. A threshold parameter δ_i (as random effects) is used for each test taker in order to determine how risky a student is. In the same manner, with the same constraints that are applied to the abilities of the different students at the same question (and not to the abilities of each different student at the 78 questions as it is used in the illustration here) and with no additional computational cost, one can estimate a δ_j parameter for each question. This helps the researcher to distinguish among questions which prevent students from a guessing behaviour or not, or, in other words, to find out the tricky questions (low estimated mean of δ) and non-tricky questions (high estimated mean of the parameter δ).

8.3.3 Prior specifications

The discrimination parameters of the items that are examined here were truncated to non negative values from a truncated normal prior with a variance equal to 1 for identification purposes. On the other hand, since the restrictions have been applied on the discrimination parameters, due to identification, the difficulty parameters need not be restricted; a vague normal prior with variance equal to 1000 was used. The guessing parameters conjugate prior was Beta with prior parameters $\kappa = 2$ and $\lambda = 6$. Finally, a vague normal prior (with prior variance equal to 1000) was used for the abilities parameters per factor for each student.

8.3.4 Convergence of the parameters

All the parameters of the model converged rapidly. It seems that if one discards at most the 50 first iterations, convergence has achieved for all of the parameters. The algorithm ran for 20000 iterations. To save space, a lag equal to 5 was used. Convergence of some selected parameters is depicted in the graphs of Appendices B, C and D.

First, the convergence of the parameters that have to do with the items (discrimination and difficulty parameters) are given. Figures 9-1 to 9-5 of Appendix B show the convergence of the discrimination parameters (abilities of the items to discriminate among students) of the first 7 items of the test, for all of the three factors of the model. Hence, for each item, three convergence plots are presented; one per factor. Note that 5 out of 19 graphs that are shown, indicate a rapid convergence where no sampled value need to be discarded in order to achieve convergence. Note, in addition, that the convergence graphs of item 1 for the ability parameters of factors 2 and 3 are not presented; they have been set equal to 0 due to identification problems mentioned in the presentation of the general model.

The graphs in Appendix C show the convergence of the difficulty parameters for the first 20 out of 78 items. All of the parameters shown here need the first iterations to be discarded in order to achieve convergence. However, the number of the initial iterations to discard is small; in all of the cases, it is enough if we discard the first 50 iterations.

With respect to the parameters that refer to the students, the convergence was obtained without discarding the first iterations for all the parameters. The rapid convergence of the parameter for the ability of the first student on the first factor is depicted in the following graph. To save space, the remaining graphs of the students of the sample are not shown here, since they all showed the same immediate convergence.

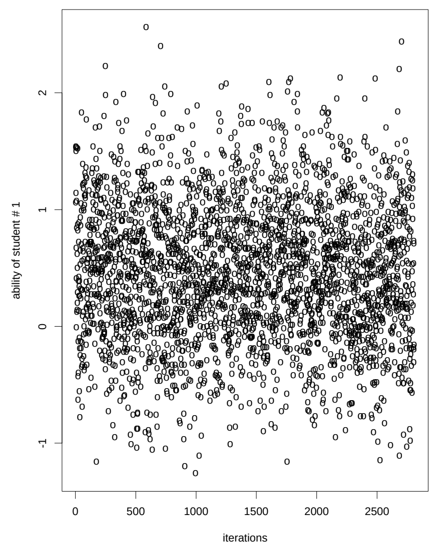


Figure 8-3: Convergence scatter plot of the ability of student #1 on factor 1

Finally, the autocorrelation plots for the convergence of the δ parameters of the first 20 out of 1000 students of the sample are shown in Appendix D. Like in the previous graphs, it is shown that convergence was achieved without high autocorrelation on the sampled values.

8.3.5 Posterior outcome on the difficulty parameters

Since the convergence of the parameters has been achieved, we present the posterior distributions by reporting the mean and the respective credible intervals of selected parameters.

Table 8.1 shows the characteristics of the 78 items of the test (percent of correct and incorrect responses and omissions) for comparison purposes along with the means and the credible intervals of the difficulty parameters.

The easier the item is, the lower the difficulty parameter. For example, questions 1, 2, 12 and 47 have percents of correct responses 92.5, 88.8 and 90.1 respectively. The mean of the estimated difficulty parameters for these items are -2.09, -1.76, -2.04 and -1.83 respectively. On the other hand, items 23 and 64 have low percentages of correct responses; 17.5 and 19.7 respectively. The mean of the estimated difficulty parameters for these items are 1.06 and 1.10 respectively. These results provide evidence with respect to the correctness of the way that the parameters have been estimated. Moreover, note that question number 16 had been mentioned in section 8.2.2 as a question with extremely high percentage of incorrect responses. However, this question does not have a high estimated difficulty parameter (the mean value is 0.43); this is plausible since the number of omissions should be taken into account and the model does it. In this case, this percentage is 3.5%, much lower comparing with the remaining items' percentages. The same is valid for question number 44.

The above results show us that omissions and incorrect responses provide quite different information and should not be treated in the same way during the calculation of the students' scores. On the other hand, the percent of correct responses should be clearly related in a linear way to the difficulty parameter. The relation of the difficulty parameters of the items with the percent of correct responses per item is depicted in Figure 8-4. Note that the points are close to a straight line with a negative slope, as they should be; the difficulty parameters have been estimated correctly, based on the information the data provide via the number of correct responses per item (or, otherwise stated, the number of incorrect responses plus the omissions).

In addition, Figure 8-5 shows the relation of the difficulty parameters of the items with the percent of incorrect responses per item. The scatter of the points is also linear, with a positive slope in this case. This means that the higher the percentage of incorrect responses is, the higher the difficulty parameter is. However, the points of the graph are not so close to a straight line as in the case of Figure 8-4. As it was discussed above,

Table 8.1: Items' characteristics and difficulty parameters

item #	% of correct responses	% of incorrect responses	% of omissions	**difficulty parameter**	credible interval
1	92.5	6.9	0.6	**-2.09**	(-2.43,-1.78)
2	88.8	10.7	0.5	**-1.76**	(-1.98,-1.56)
3	53.5	43.9	2.6	**-0.13**	(-0.23,-0.04)
4	63.9	28.3	7.8	**-0.45**	(-0.55,-0.35)
5	64.4	32.3	3.3	**-0.61**	(-0.74,-0.48)
6	63.3	30.0	6.7	**-0.51**	(-0.62,-0.40)
7	58.6	35.2	6.2	**-0.30**	(-0.40,-0.20)
8	54.5	37.0	8.5	**-0.16**	(-0.25,-0.07)
9	30.2	53.6	16.2	**0.56**	(0.46,0.65)
10	22.3	50.9	26.8	**0.83**	(0.73,0.94)
11	88.4	10.4	1.2	**-1.55**	(-1.72,-1.39)
12	95.8	4.1	0.1	**-2.04**	(-2.33,-1.82)
13	88.1	11.1	0.8	**-1.33**	(-1.45,-1.21)
14	84.7	9.7	5.6	**-1.47**	(-1.64,-1.31)
15	81.0	17.8	1.2	**-1.00**	(-1.10,-0.90)
16	34.6	61.9	3.5	**0.43**	(0.34,0.53)
17	71.1	23.7	5.2	**-0.67**	(-0.77,-0.57)
18	71.7	25.6	2.7	**-0.87**	(-1.02,-0.74)
19	42.3	37.5	20.2	**0.20**	(0.10,0.29)
20	36.8	57.5	5.7	**0.36**	(0.27,0.45)
21	29.9	39.4	30.7	**0.61**	(0.50,0.72)
22	21.3	49.6	29.1	**0.82**	(0.72,0.91)
23	17.5	54.1	28.4	**1.06**	(0.94,1.18)
24	79.2	18.8	2.0	**-1.00**	(-1.12,-0.88)
25	38.7	53.5	7.8	**0.32**	(0.22,0.42)
26	72.9	23.2	3.9	**-0.78**	(-0.89,-0.68)
27	42.6	50.0	7.4	**0.18**	(0.09,0.27)
28	57.4	34.0	8.6	**-0.26**	(-0.35,-0.16)
29	68.4	25.3	6.3	**-0.54**	(-0.62,-0.45)
30	39.7	46.9	13.4	**0.28**	(0.19,0.37)
31	33.0	50.1	16.9	**0.50**	(0.40,0.60)
32	45.4	35.6	19.0	**0.10**	(0.00,0.21)
33	47.8	34.4	17.8	**-0.00**	(-0.14,0.13)
34	26.0	50.3	23.7	**0.82**	(0.70,0.95)
35	12.2	54.7	33.1	**1.29**	(1.17,1.42)
36	82.4	16.7	0.9	**-1.24**	(-1.41,-1.09)

Table 8.1:Items' characteristics and difficulty parameters

item #	% of correct responses	% of incorrect responses	% of omissions	**difficulty parameter**	credible interval
37	81.0	17.0	2.0	**-1.04**	(-1.15,-0.93)
38	52.7	44.7	2.6	**-0.10**	(-0.19,-0.01)
39	55.4	36.7	7.9	**-0.18**	(-0.27,-0.09)
40	87.7	11.6	0.7	**-1.38**	(-1.53,-1.24)
41	56.0	39.0	5.0	**-0.26**	(-0.37,-0.15)
42	48.9	38.1	13.0	**-0.01**	(-0.12,0.09)
43	57.8	36.3	5.9	**-0.29**	(-0.39,-0.19)
44	29.8	63.0	7.2	**0.64**	(0.53,0.75)
45	82.2	14.1	3.7	**-1.11**	(-1.23,-0.99)
46	80.5	19.0	0.5	**-0.96**	(-1.06,-0.85)
47	90.1	6.9	3.0	**-1.83**	(-2.05,-1.62)
48	48.2	47.0	4.8	**0.03**	(-0.06,0.11)
49	28.7	41.4	29.9	**0.61**	(0.51,0.70)
50	23.6	37.6	38.8	**0.81**	(0.70,0.92)
51	81.4	16.9	1.7	**-1.02**	(-1.13,-0.92)
52	59.2	39.2	1.6	**-0.29**	(-0.39,-0.20)
53	76.9	21.5	1.6	**-0.83**	(-0.93,-0.73)
54	67.6	28.1	4.3	**-0.55**	(-0.64,-0.45)
55	64.8	31.0	4.2	**-0.50**	(-0.60,-0.40)
56	59.7	38.8	1.5	**-0.34**	(-0.44,-0.24)
57	58.2	35.8	6.0	**-0.25**	(-0.34,-0.17)
58	36.4	56.4	7.2	**0.38**	(0.29,0.48)
59	68.6	24.5	6.9	**-0.60**	(-0.71,-0.50)
60	86.4	9.1	4.5	**-1.70**	(-2.00,-1.47)
61	60.0	34.3	5.7	**-0.36**	(-0.46,-0.26)
62	52.7	40.0	7.3	**-0.11**	(-0.21,-0.02)
63	58.0	31.7	10.3	**-0.26**	(-0.36,-0.17)
64	19.7	57.8	22.5	**1.10**	(0.97,1.23)
65	58.3	24.7	17.0	**-0.35**	(-0.46,-0.24)
66	56.0	28.5	15.5	**-0.27**	(-0.39,-0.15)
67	72.3	27.2	0.5	**-0.72**	(-0.82,-0.61)
68	74.2	24.6	1.2	**-0.77**	(-0.87,-0.67)
69	30.5	49.0	20.5	**0.56**	(0.47,0.65)
70	60.0	37.1	2.9	**-0.30**	(-0.38,-0.21)
71	79.0	17.9	3.1	**-1.05**	(-1.18,-0.93)

Table 8.1:Items' characteristics and difficulty parameters

item #	% of correct responses	% of incorrect responses	% of omissions	**difficulty parameter**	credible interval
72	37.1	53.0	9.9	**0.34**	(0.25,0.42)
73	49.3	43.7	7.0	**-0.02**	(-0.12,0.08)
74	41.5	46.2	12.3	**0.23**	(0.14,0.33)
75	55.7	36.1	8.2	**-0.18**	(-0.26,-0.09)
76	50.9	34.7	14.4	**-0.07**	(-0.17,0.04)
77	38.8	40.1	21.1	**0.33**	(0.23,0.44)
78	52.5	26.5	21.0	**-0.13**	(-0.24,-0.03)

this happens because the difficulty of a question is shown by the number of incorrect responses but also by the number of omissions; both of these responses show how difficult the question is. The relation of the difficulty parameters of the items with the percent of omissions per item is shown in Figure 8-6. To confirm the above results, if we add the number of incorrect responses and omissions, we obviously conclude with the whole information provided by the number of correct responses, except that the slope of the line in the scatter plot is positive in this case (see Figure 8-7).

8.3.6 Information derived from the discrimination parameters

Table 9.1 of Appendix E shows the discrimination parameters of the 78 items of the test for each of the three factors of the model; the means and the credible intervals (in the parentheses) are reported. Note that the estimated mean of item 1 on factors 2 and 3, as well as the estimated mean of item 36 on factor 3 have been set equal to 0 by the model due to the restrictions that solve the identification problem.

The highest estimated mean among the three (one per factor) means of the item is written in bold; it shows the factor that the particular item matches best. These scores can also be treated exactly as the factor loadings in the classical factor analysis. However, the three factors do not seem to contain the items of each session of test separately. It seems that the factors describe different abilities than those the different session of the test attempt to measure. The questions of the test have not been provided by the test

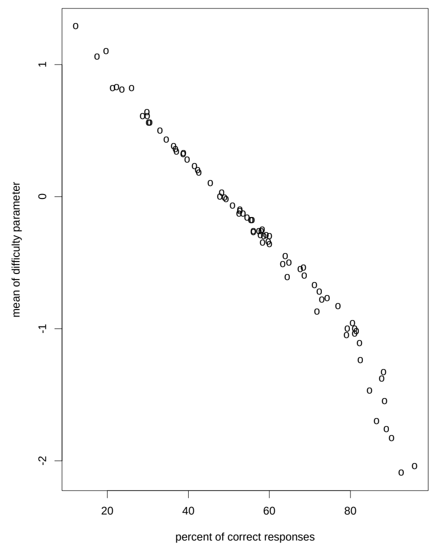


Figure 8-4: Scatter plot of the difficulty parameter and % of correct responses

makers since some questions are repeated in different years; hence, it is not possible to check the exact kind of ability each factor measures (as we did with the data in the illustration of the model described in the previous chapter). However, some parts of the test seem to contain many questions of one factor together. For example, questions 1 to 7 all measure mainly the ability that factor 1 represents (also, questions 11-18 and 52-59). It must be mentioned that the discrimination parameters do not relate with the difficulty parameters in any way; all the factors contain both difficult and easy questions (characterized by the valuation of the difficulty parameters).

From Table 9.1 of Appendix E, it is obvious that there are some questions which do have a significant discrimination power. Question number 33 contributes to all the three factors with relatively high loadings. It was not a rather easy question (percentage of correct responses 47.8%) but also its percentage of incorrect responses was 34.4%, which ranged the question almost in the middle of the difficulty scale (indeed, the difficulty parameter's mean was 0, just in the middle of the latent variable "difficulty"). Moreover,

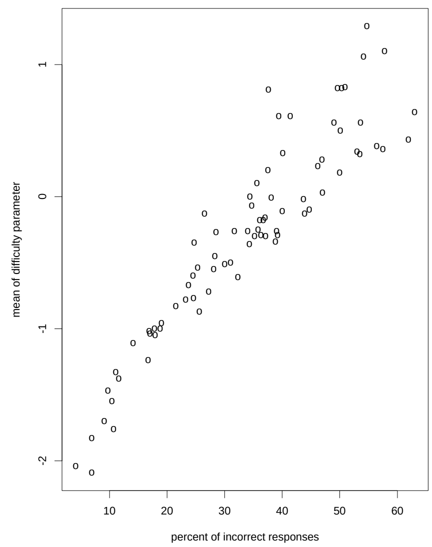


Figure 8-5: Scatter plot of the difficulty parameter and % of incorrect responses

this question seems to necessitate different cognitive abilities in order to be answered, at least all the three different cognitive abilities that the factors of the model represent.

Two of the "difficult" questions, that had been mentioned in the presentation of the data, were the questions number 23 and 64. These two questions do not seem to have an extremely high parameter mean in any of the three factors. The same is valid for the "too easy" question number 12. Hence, the discrimination ability of a question seems to be irrelevant with the difficulty (indeed, in practise there is no use to include questions in a test that can not be answered by anybody or questions which have an obvious answer for anybody). The discrimination parameter of item 1 in factor 1 is an exception since the values of the other two factor parameters were set by default equal to 0.

It must be noted that if we had set different discrimination parameters equal to 1, we would have ended up with a rotation of the matrix of the estimated parameters; however the rotation of the factor loadings in a factor analysis is beyond the scope of this presentation of the IRT model with omitted responses.

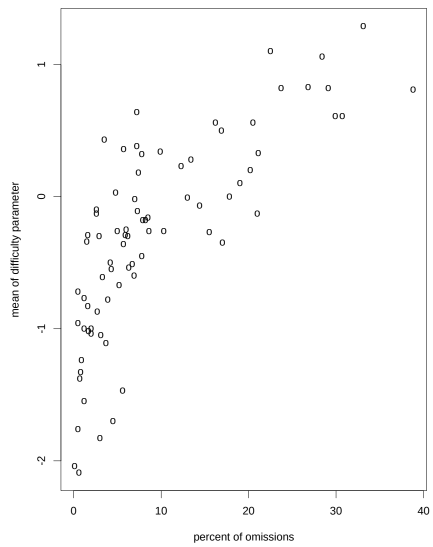


Figure 8-6: Scatter plot of the difficulty parameter and % of omissions

8.3.7 Posterior outcome for the students' performance

After the SAT examination, each student's performance is reported through the general score on the SAT test. However, the model that is proposed here distinguishes the performance of the students on the different items groups that are formed by the latent factor analysis that was used. To save space, the separate performance of the first 100 students of the sample on the three groups are reported in Table 9.2 of Appendix E. The mean abilities along with the respective credible intervals are reported.

Since the abilities of all the students are supposed to follow a normal distribution with mean 0 and variance 1, each positive value of Table 9.2 indicated performance of the student that is greater than the average. From the table, one can note that there are students with high abilities in all the different factors (see, for example, students 1, 39, 54, 59, 83, 87, 93 and 100) and, of course, students with bad performance in all the different factors (see student number 8). However, there are also some students who perform very well in one factor but not in another. Such students receive a medium-level

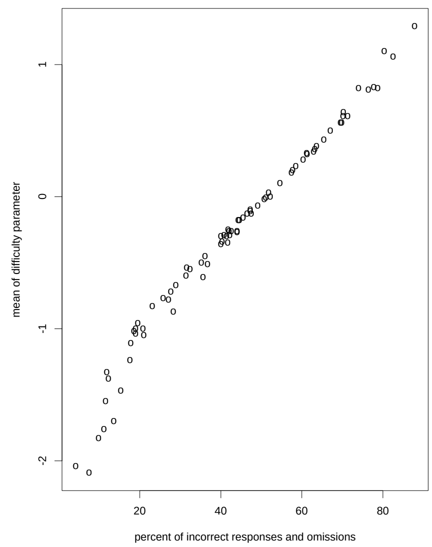


Figure 8-7: Scatter plot of the difficulty parameter and % of incorrect responses and omissions

grade on the test; they can be very good in some particular sciences though, that require high performance on a specific kind of problems. For example, students 14, 19, 74 are very good on items of the third factor but not so good on the remaining items of the test. Also, students 35, 49 and 90 are very good only on the items of the first factor and not on the remaining.

8.3.8 Posterior outcome for the students' δ parameters

The δ parameters for each one of the 100 first students are reported in Table 8.2. These parameters are of great importance for the presentation here, since they are newly imported, theoretically and practically, into the IRT model.

The parameters are all truncated to non-negative values (hence, the lower band of each credible interval that is shown in Table 8.2 is 0; it is presented for the reader to recall this restriction. This lower band does not make any other sense).The lower the mean of the parameter is, the more risky the student is. In other, words, the mean of the parameters show the mean additional ability the students have to have in order to decide to answer a question when they do not know the answer for sure. The higher additional ability that have to have also means that they can recognize which responses are able to answer and which (and how many) of them should be omitted. In addition, an explanation from another point of view can be given in this case: the less risky the student is, the higher ability she/he needs to guess for an answer, and, moreover, the more number of omissions in case that she/he feels that this additional ability that is required, is not available.

Note in Table 8.2 that there are some students with high mean δ, for example students 27, 64, 84, 91, 99 and 100, with respective mean of δ equal to 0.045, 0.045, 0.049, 0.053 and 0.053. These students have a large number of omissions (from 17 to 32 omissions each), thus, they are less risky than the remaining students and they need to feel that they have higher ability than the other in order to try to answer a question that they don't know. On the other hand, there are also some students with high mean value of δ, for example, students 9, 28, 33, 36 and 57 (respective mean value of δ 0.047, 0.044, 0.044, 0.044 and 0.045) who they do have a very small number of omissions (5, 2, 1, 0, and 3 respectively). But as in the previous case, this means that they are not risky at all, since also the number of incorrect responses they have are very small (5, 6, 9, 8 and 7 respectively). Hence, they are students with low risk and good performance on the test. In addition, there are students who are very risky and try to answer a question when they are not sure about the answer. See, for example, students 4, 10, 11, 40, 47 and 74. Their respective mean values of δ are 0.032, 0.030, 0.031, 0.034, 0.033 and 0.031, and their number of omissions are 0, 0, 0, 1, 0 and 0 respectively. This means that they try to answer the questions when they are don't know the answer; this guessing behaviour is also apparent from the large number of incorrect responses these students have given; these are 51, 46, 29, 22 45 and 38 respectively. This indicates that they are risky students with low performance. This is also taken into account by the model in the calculation of their general ability since a portion of their correct responses might had been given by chance, due to their guessing behaviour.

Table 8.2: δ parameters of the students
(continued...)

student number	*number of correct responses*	*number of incorrect responses*	*number of omissions*	*δ parameter*	*credible interval*
1	59	19	0	**0.035**	(0.000,0.105)
2	41	20	17	**0.041**	(0.000,0.120)
3	57	12	9	**0.043**	(0.000,0.120)
4	27	51	0	**0.032**	(0.000,0.100)
5	44	21	13	**0.038**	(0.000,0.120)
6	37	24	17	**0.037**	(0.000,0.120)
7	26	41	11	**0.037**	(0.000,0.100)
8	33	45	0	**0.034**	(0.000,0.100)
9	68	5	5	**0.047**	(0.000,0.130)
10	32	46	0	**0.030**	(0.000,0.095)
11	49	29	0	**0.031**	(0.000,0.100)
12	48	18	12	**0.039**	(0.000,0.110)
13	32	41	5	**0.031**	(0.000,0.100)
14	44	32	2	**0.032**	(0.000,0.090)
15	48	26	4	**0.040**	(0.000,0.120)
16	47	31	0	**0.033**	(0.000,0.090)
17	57	21	0	**0.037**	(0.000,0.110)
18	51	26	1	**0.034**	(0.000,0.110)
19	53	25	0	**0.036**	(0.000,0.100)
20	26	41	11	**0.039**	(0.000,0.120)
21	64	10	4	**0.037**	(0.000,0.110)
22	32	35	11	**0.038**	(0.000,0.110)
23	30	48	0	**0.032**	(0.000,0.100)
24	35	38	5	**0.035**	(0.000,0.100)
25	28	50	0	**0.033**	(0.000,0.100)
26	56	16	6	**0.039**	(0.000,0.110)
27	28	25	25	**0.045**	(0.000,0.125)
28	70	6	2	**0.044**	(0.000,0.130)
29	39	34	5	**0.037**	(0.000,0.110)
30	62	16	0	**0.038**	(0.000,0.110)
31	66	12	0	**0.035**	(0.000,0.105)
32	35	42	1	**0.032**	(0.000,0.105)
33	68	9	1	**0.044**	(0.000,0.120)
34	55	23	0	**0.034**	(0.000,0.110)
35	43	32	3	**0.033**	(0.000,0.100)
36	70	8	0	**0.044**	(0.000,0.120)

Table 8.2: δ parameters of the students
(continued...)

student number	*number of correct responses*	*number of incorrect responses*	*number of omissions*	δ *parameter*	*credible interval*
37	*47*	*21*	*10*	**0.038**	*(0.000,0.110)*
38	*60*	*14*	*4*	**0.041**	*(0.000,0.110)*
39	*67*	*11*	*0*	**0.037**	*(0.000,0.100)*
40	*55*	*22*	*1*	**0.034**	*(0.000,0.095)*
41	*32*	*34*	*12*	**0.037**	*(0.000,0.110)*
42	*55*	*11*	*12*	**0.042**	*(0.000,0.120)*
43	*38*	*21*	*19*	**0.037**	*(0.000,0.110)*
44	*55*	*20*	*3*	**0.040**	*(0.000,0.110)*
45	*56*	*19*	*3*	**0.034**	*(0.000,0.100)*
46	*48*	*23*	*7*	**0.039**	*(0.000,0.110)*
47	*33*	*45*	*0*	**0.033**	*(0.000,0.100)*
48	*40*	*35*	*3*	**0.035**	*(0.000,0.100)*
49	*50*	*24*	*4*	**0.039**	*(0.000,0.120)*
50	*30*	*29*	*19*	**0.040**	*(0.000,0.120)*
51	*37*	*35*	*6*	**0.036**	*(0.000,0.100)*
52	*44*	*29*	*5*	**0.036**	*(0.000,0.100)*
53	*55*	*18*	*5*	**0.041**	*(0.000,0.110)*
54	*43*	*30*	*5*	**0.034**	*(0.000,0.100)*
55	*34*	*36*	*8*	**0.032**	*(0.000,0.100)*
56	*29*	*49*	*0*	**0.033**	*(0.000,0.100)*
57	*68*	*7*	*3*	**0.045**	*(0.000,0.130)*
58	*59*	*19*	*0*	**0.041**	*(0.000,0.120)*
59	*30*	*35*	*13*	**0.038**	*(0.000,0.115)*
60	*67*	*9*	*2*	**0.043**	*(0.000,0.120)*
61	*47*	*22*	*9*	**0.040**	*(0.000,0.115)*
62	*69*	*9*	*0*	**0.043**	*(0.000,0.125)*
63	*64*	*13*	*1*	**0.037**	*(0.000,0.110)*
64	*35*	*16*	*27*	**0.045**	*(0.000,0.130)*
65	*48*	*26*	*4*	**0.033**	*(0.000,0.100)*
66	*63*	*14*	*1*	**0.041**	*(0.000,0.120)*
67	*49*	*29*	*0*	**0.034**	*(0.000,0.100)*
68	*53*	*21*	*4*	**0.037**	*(0.000,0.100)*
69	*62*	*12*	*4*	**0.041**	*(0.000,0.125)*
70	*33*	*40*	*5*	**0.034**	*(0.000,0.100)*
71	*71*	*7*	*0*	**0.045**	*(0.000,0.125)*
72	*20*	*55*	*3*	**0.035**	*(0.000,0.100)*

Table 8.2: δ parameters of the students

student number	number of correct responses	number of incorrect responses	number of omissions	δ parameter	credible interval
73	68	7	3	**0.044**	(0.000,0.125)
74	40	38	0	**0.031**	(0.000,0.100)
75	47	31	0	**0.032**	(0.000,0.100)
76	59	12	7	**0.044**	(0.000,0.120)
77	34	28	16	**0.039**	(0.000,0.110)
78	64	11	3	**0.040**	(0.000,0.115)
79	39	20	19	**0.044**	(0.000,0.130)
80	45	26	7	**0.043**	(0.000,0.120)
81	31	47	0	**0.032**	(0.000,0.100)
82	29	48	1	**0.034**	(0.000,0.105)
83	33	40	5	**0.033**	(0.000,0.100)
84	53	8	17	**0.049**	(0.000,0.135)
85	38	37	3	**0.033**	(0.000,0.110)
86	51	19	8	**0.037**	(0.000,0.110)
87	49	21	8	**0.040**	(0.000,0.120)
88	38	36	4	**0.034**	(0.000,0.100)
89	48	27	3	**0.035**	(0.000,0.110)
90	43	24	11	**0.041**	(0.000,0.110)
91	45	4	29	**0.048**	(0.000,0.140)
92	52	26	0	**0.034**	(0.000,0.105)
93	34	25	19	**0.040**	(0.000,0.110)
94	28	32	18	**0.037**	(0.000,0.110)
95	53	21	4	**0.040**	(0.000,0.110)
96	76	1	1	**0.052**	(0.000,0.140)
97	71	6	1	**0.043**	(0.000,0.120)
98	45	30	3	**0.036**	(0.000,0.120)
99	26	20	32	**0.053**	(0.000,0.135)
100	29	17	32	**0.053**	(0.000,0.135)

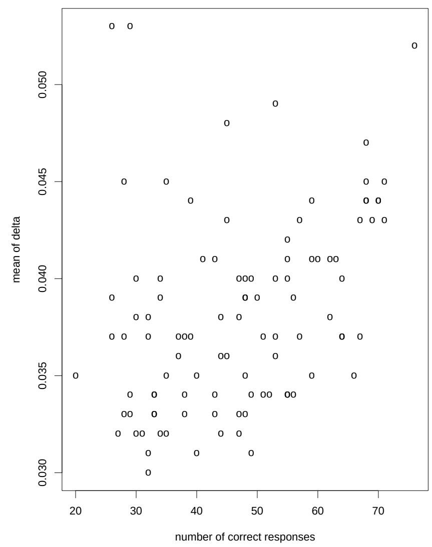


Figure 8-8: Scatter plot of mean of δ with the number of correct responses

It is obvious, from the discussion of Table 8.2, that the parameters δ are not related to a single descriptive measurement that arises from the data but it expresses an informative combination of all the characteristics of the data. Hence, the relation of δ with any single descriptive measurement of the data can not be strong. It seems that a combination of them may result to a stronger relationship. The relations between the parameter δ estimated by the model and the descriptives for each student (number of correct and incorrect responses and omissions) are depicted in the following Figures.

Figure 8-8 shows the weak relation between δ and the number of correct responses. It seems that the information provided by the number of correct responses is not depicted by the parameter δ; this has to do with the ability parameters of each student on each factor that have been presented in the previous discussions. Therefore, both good and bad students may be or may not be risky (this was also obvious from the discussion of Table 8.2).

The same information is provided by the Figures 8-9 and 8-10, although the relation

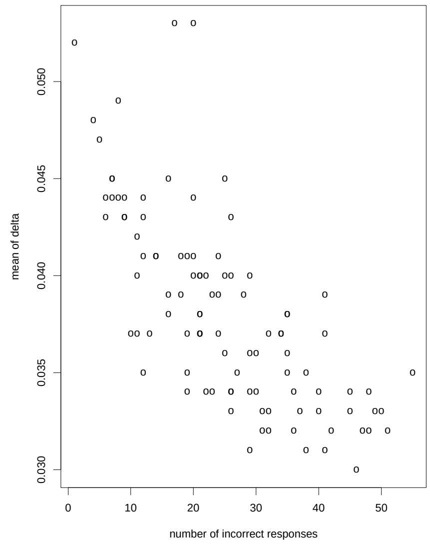


Figure 8-9: Scatter plot of mean of δ with the number of incorrect responses

is not that weak as it was in Figure 8-8. Hence, students with large value of incorrect responses may or may not be risky, depending on whether they also responded with a large value of omissions or not.

Figures 8-11 and 8-12 show the relations of a combination of the descriptive measurements of the sample with the parameter δ. The former graph shows the sum of the incorrect responses and omissions and the latter shows the sum of the correct responses and omissions. Both of these graphs show a weak relation with the parameter δ.

Finally, Figure 8-13 depicts the relation of the parameter δ with the difference between the incorrect responses and omissions. This relation seems to be stronger than all of the previously mentioned graphs. This seems to be evidence for the way the parameter δ has been quantified; the higher the incorrect responses and the lower the omissions are, the more risky the student is.

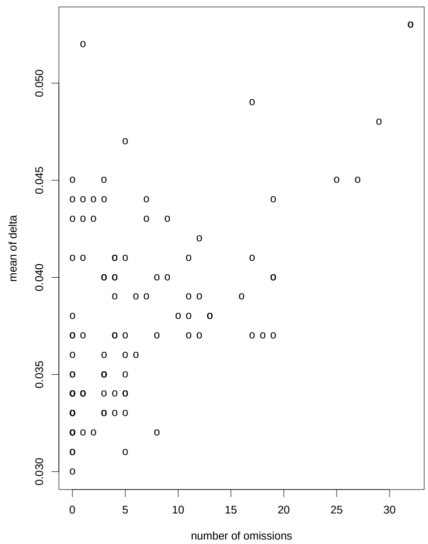


Figure 8-10: Scatter plot of mean of δ with the number of omissions

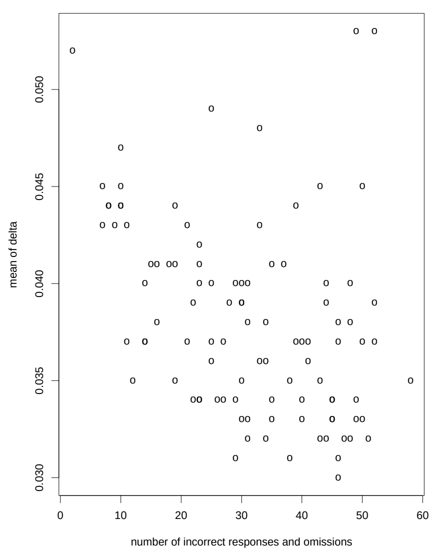


Figure 8-11: Scatter plot of mean of δ with the number of incorrect responses plus omissions

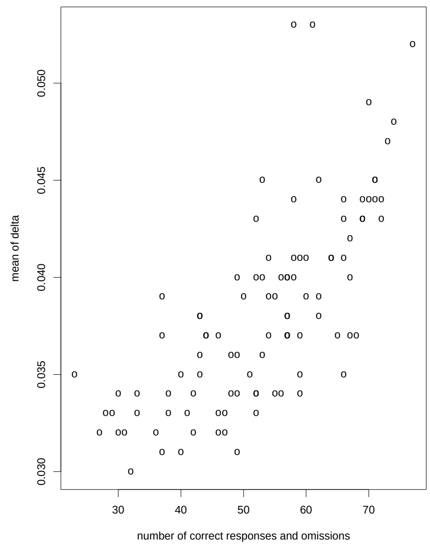


Figure 8-12: Scatter plot of mean of δ with the number of correct responses plus omissions

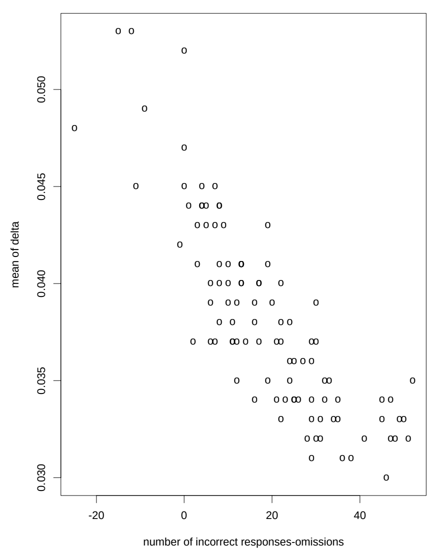


Figure 8-13: Scatter plot of mean of δ with the number of incorrect responses minus omissions

Chapter 9

Further Research

We have presented some latent variable models along with modifications on them which arise from the specific particularities of the data. In the multiranked probit model, we have presented a realistic model generation structure which describes ranking responses collected from stated preference experiments and we have provided MCMC details which implement our proposed models. We have also illustrated that real problems originated by small sample sizes or large percentage of responses with ties and intransitivities can be incorporated by adopting a more general error structure. What led us to carrying out this work was the utilisation of all the ranking responses which included ties and intransitivities. Therefore, the responses that were difficult to handle were not eliminated nor were inappropriate link functions used with the intent of simplifying calculations.

We have also covered many modeling aspects by utilizing a broad family of models (multivariate probit, mixture scale of normal link functions with varying degrees of freedom in the real domain). However, in the model uncertainty and model selection task, one can also use a marginal likelihood approach (see Chib, 1995) to choose or evaluate among different models which can arise. This approach can also be applied in the proposed multidimensional IRT model with thresholds where the number of the dimensions-factors was pre-defined. A marginal likelihood approach can be used to determine the number of factors that fit the data better.

Finally, the results of the proposed IRT model can be compared with the results of a model for nominal data, if we treat the omissions as a different response category, see Bock (1972), Thissen and Steinberg (1984).

Appendix A: Multiple choice test of the "Data Analysis" course

The crosstabulation provides the relation between gender (male, άρρεν- female,θήλυ) with the systematic body exercise (yes,ναι- no,όχι).

			Συστηματική άθληση		
			Ναι	Όχι	Total
Φύλο	Άρρεν	Count	91	10	101
		Expected Count	79,9	21,1	101,0
		% within Φύλο	90,1%	9,9%	100,0%
		% within Συστηματική άθληση	66,9%	27,8%	58,7%
	Θήλυ	Count	45	26	71
		Expected Count	56,1	14,9	71,0
		% within Φύλο	63,4%	36,6%	100,0%
		% within Συστηματική άθληση	33,1%	72,2%	41,3%
Total		Count	136	36	172
		Expected Count	136,0	36,0	172,0
		% within Φύλο	79,1%	20,9%	100,0%
		% within Συστηματική άθληση	100,0%	100,0%	100,0%

The percentage of the female which do not exercise their body systematically is:

a) 14,9%
b) 36,6%
c) 72,2%
d) 33,1%

In a multiple linear regression with independent variables X1 and X2, we construct the ANOVA table of the model.

ANOVA

Model		Sum of Squares	df	F	Sig.
1	Regression	375,097	2	22,469	,000[a]
	Residual	308,834	37		
	Total	683,930	39		

a. *Predictors: (Constant), X1, X2*

The sum of residuals for this model is:

a) 308,834
b) 308,834 / 37
c) 0
d) 37

A data set includes the result of a lung capacity test as "Within accepted (normal) bounds" or "Out of accepted (normal) bounds" as well as whether the person is a smoker or not. We will try to analyze the data set using:

a) Two separate t-tests
b) Regression with the use of dummy variables
c) Two-way ANOVA
d) X^2 test

In a linear regression model, where the dependent variable is the weight of the persons, and independent the continuous variables A, B and C, we obtain the following output with the coefficients

Coefficients[a]

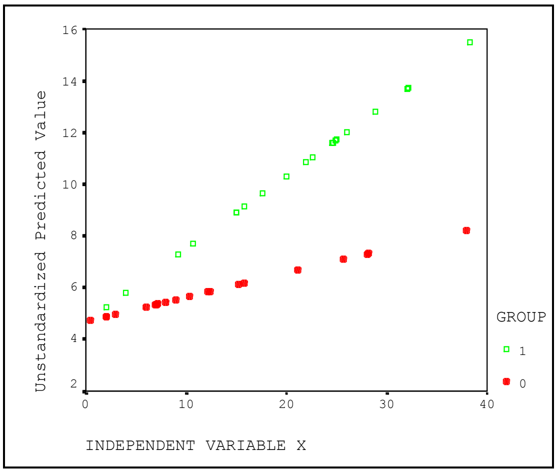
Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	t	Sig.
1	(Constant)	24,674	2,317	10,648	,000
	A	-6,201	,953	-6,506	,000
	B	,119	,414	,287	,776
	C	9,3E-02	,048	1,922	,062

a. Dependent Variable: Βάρος

In order to construct a satisfactory model that explains the weight of the persons, (with hypothesis testing at significance level α= 5%), the next step of our analysis should be:

a) To eliminate the variable A, since the estimated coefficient has the highest standard error (0,953), compared with the remaining independent variables, and to fit the model again.
b) To eliminate the variables B and C since their coefficients are not statistically significant, and to fit the model again.
c) To eliminate the variable B since its coefficient is not statistically significant, and to fit the model again
d) To eliminate the variable C since its coefficient is almost equal to 0 (0,093), and to fit the model again

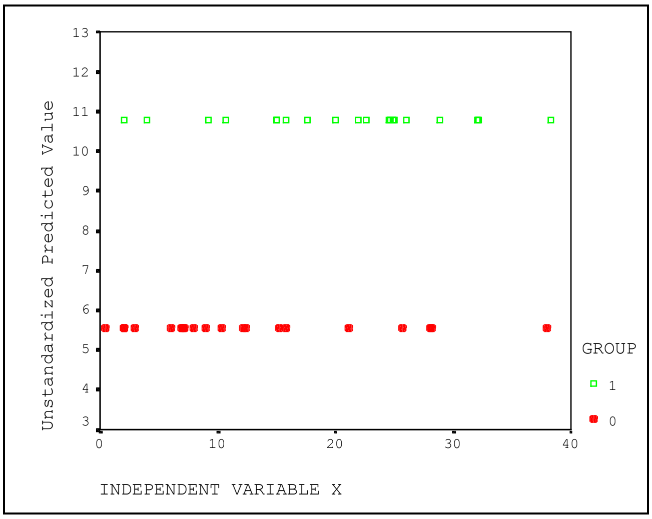
In a data set, the dependent variable Y is continuous. The data points come from 2 different groups (variable «group», with values 0 and 1) where the mean values of Y for the 2 groups differ statistically significantly. There is also an independent continuous variable X that is statistically significant. We use a model (with either «group», or X, or both) and we find the expected values of Y; these expected values along with the values of X are given in the graph below:



The linear model that has been used in that case is:

a) linear regression with independent variables group and X
b) linear regression with independent variables X and the product (X*group)
c) two way ANOVA
d) one way ANOVA (using the variable group as independent)

In a data set, the dependent variable Y is continuous. The data points come from 2 different groups (variable «group», with values 0 and 1) where the mean values of Y for the 2 groups differ statistically significantly. There is also an independent continuous variable X that is statistically significant. We use a model (with either «group», or X, or both) and we find the expected values of Y; these expected values along with the values of X are given in the graph below:



The linear model that has been used in that case is:

a) linear regression with independent variables group and X
b) a separate linear regression for each group with independent variable X
c) two-way ANOVA
d) linear regression with independent variable group

The table below shows the results of a Chi-square test.

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)
Pearson Chi-Square	3,969*	1	,046	
Fisher's Exact Test				,090

*. 1 cells (25,0%) have expected count less than 5. The minimum expected count is 4,35.

If we use a 5% confidence level, we can conclude that:

a) There is independence between rows and columns (according to the Pearson Chi-Square test)
b) There is no independence between rows and columns (according to the Pearson Chi-Square test)
c) There is independence between rows and columns (according to the Fisher's exact test)
d) We can not conclude, since the two tests (Pearson Chi-Square and Fisher's exact) do not agree with respect to the conclusion.

In a linear model, the coefficient of an independent variable is not statistically significant (with α=5%). The 95% CI for this parameter is:

a) (-125,-39)
b) (12,42)
c) (0.04,0.09)
d) (-1,24)

In a linear model (where the dependent variable is continuous) in order to include an independent categorical variable with 4 levels, where each level denotes a different profession category, we have to:

a) multiply the categorical variable with each of the continuous independent variables of the model
b) construct 4 dummy variables
c) construct 3 dummy variables
d) include the variable in the model as it is, considering it as a variable that is ordinal

A linear regression equation was calculated as $Y=4.15-0.96X+\varepsilon$, with coefficient of determination equal to $R^2=0.81$. The correlation coefficient between X and Y is equal to:

a) 0.96
b) -0.96
c) 0.9
d) -0.9

In a data set, the dependent variable is the (continuous) grade on a test, and the independent variables are the gender of the students and the year of studies (3^{rd}, 4^{th}, 5^{th}). We will try to analyze the data set using:

a) Regression analysis with two independent variables
b) Two-way ANOVA
c) ANCOVA of the gender with the year of studies
d) t-test with independent variables the gender and the year of studies

In order to introduce a discrete independent variable into a linear model (where the dependent is continuous), where the discrete variable contains 5 levels, each level denoting the number of children in the family, we should:

a) Multiply the discrete variable with all the continuous independent variables of the model
b) Construct 5 dummy variables
c) Construct one dummy variable instead of the original discrete variable
d) Insert the variable in the model , as it is, given that it is an ordinal one

A data set includes the result of a lung capacity test with values from 40 to 98.4 as well as whether the person is a smoker or not. We will try to analyze the data set using:

a) regression with two dummy variables
b) t-test
c) Two-way ANOVA
d) ANCOVA

In an experiment we record the weight of fishes. A different diet was used for each of the three groups of the experiment. The data set was analyzed with ANOVA and the pairwise comparisons (using the Scheffe test) are given in the table below:

Multiple Comparisons

Dependent Variable: Βάρος

Scheffe

(I) DIAITES	(J) DIAITES	Mean Difference (I-J)	Std. Error	Sig.
Δίαιτα Α	Δίαιτα Β	10,6667 *	1,756	,000
	Δίαιτα Γ	13,0667 *	1,756	,000
Δίαιτα Β	Δίαιτα Α	-10,6667 *	1,756	,000
	Δίαιτα Γ	2,4000	1,756	,401
Δίαιτα Γ	Δίαιτα Α	-13,0667 *	1,756	,000
	Δίαιτα Β	-2,4000	1,756	,401

*. *The mean difference is significant at the .05 level.*

According to the results, if we use α= 5%, we can conclude that, in general, the diet that is suggested in order to obtain higher weight of the fishes, as well as the worst diet are:

a) the differences of the weights are not statistically significant
b) A is a better diet, B is a worse
c) A is a better diet, C is a worse
d) A is a better diet, whereas there is no statistically significant difference between diets B and C

The ANOVA table of an analysis is given below.

ANOVA

Βάρος

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	**x**	2	**z**	31,383	,000
Within Groups	**y**	42	**w**		
Total	2422,578	44			

The value 31,383 of F was calculated as:

a) x/z
b) z/w
c) x/2
d) x/y

A data set includes the result of a lung capacity test with values from 40 to 98.4, whether the person is a smoker or not, and the (continuous) age of the person. We will try to analyze the data set using:

a) ANCOVA
b) Regression analysis with independent variables the age and two dummy variables
c) One-way ANOVA
d) t-test

In a multiple regression model with independent variables X1 and X2, we construct the ANOVA table of the model.

ANOVA

Model		Sum of Squares	df	F	Sig.
1	Regression	375,097	2	22,469	,000[a]
	Residual	308,834	37		
	Total	683,930	39		

a. *Predictors: (Constant), X1, X2*

Using the F test of the table, we conclude that:

a) there is at least one statistically significant independent variable in the model
b) all the coefficients of the independent variables are statistically significant
c) we should not include any other independent variable in the model, since the ones which were used suffice
d) the constant and all the coefficients in the model are statistically significant

In a linear regression with dependent variable Y, and independent the continuous variables A, B and C, we get the following output:

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,754[a]	,568	,537	5,0495

a. *Predictors: (Constant), A, B, C*

b. *Dependent Variable: Y*

The value 0,754 of R can be interpreted as:

a) The correlation coefficient between Y and $\hat{Y}$
b) The correlation coefficient between Y and the independent variables
c) The correlation coefficient between $\hat{Y}$ and the residuals of the model
d) The correlation coefficient between Y and the residuals of the model

In a regression analysis, where Y is the dependent variable, we use the assumption of the normality of y_i. In order to check this assumption, we can:

a) Check the normality of Y graphically (for example, using a histogram)
b) Test the normality of Y using the non parametric test Kolmogorov-Smirnov of one sample
c) Check Y graphically to see whether there are outliers (which are far from normality) and, in case there are, to eliminate them from the sample
d) Test the normality via the residuals using the non parametric test Kolmogorov-Smirnov of one sample

A linear regression analysis, with dependent variable Y and independent var1, var2 and var3, provides the following table.

Model Summary[b]

Model	R	R Square	Adjusted R Square
1	,930[a]	,865	,537

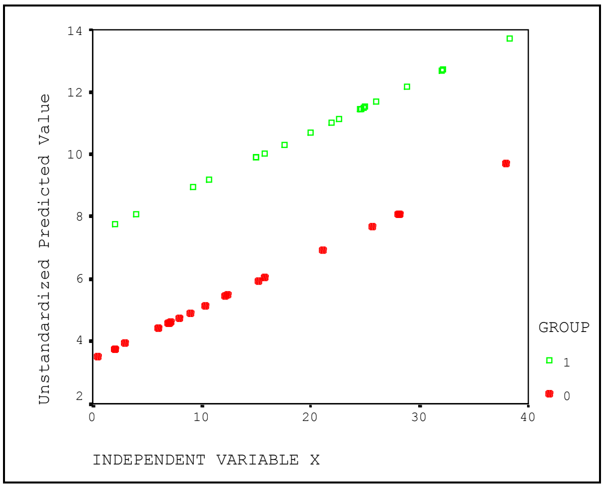
a. Predictors: (Constant), var1, var2, var3

b. Dependent Variable: Y

The large difference between R^2 and adjusted R^2 may be due to:

a) wrong choice of independent variables
b) small sample size
c) the fact that the normality and independence assumptions of y_i are violated
d) either the small sample size or the use of some independent variables which do not explain a significant amount variability of the dependent variable

In a data set, the dependent variable Y is continuous. The data points come from 2 different groups (variable «group», with values 0 and 1) where the mean values of Y for the 2 groups differ statistically significantly. There is also an independent continuous variable X that is statistically significant. We use a model (with either «group», or X, or both) and we find the expected values of Y; these expected values along with the values of X are given in the graph below:



The linear model that has been used in that case is:

a) linear regression with independent variables group and X
b) linear regression with independent variables X and the product (X*group)
c) linear regression with independent variable X
d) one-way ANOVA (with variable group)

Which one of the above is not an assumption for the linear model

a) homoscedasticity of the residuals
b) y_i are independent
c) y_i are normally distributed with mean á+βx_i and variance σ^2_i
d) the errors are uncorrelated

An estimated parameter is tested, in significance level 5% and is found to be statistically significant. What could be the result of the test if we used a significance level equal to 1% and 10%;

a) in α=1%, the parameter would be statistically significant whereas in α=10% it would not.
b) in α=1%, the parameter would not be statistically significant whereas in α=10% it would be.
c) in α=10% and α=1%, the parameter would be statistically significant
d) in α=10%, the parameter would be statistically significant whereas we can not conclude for α=1%

Which of the following is not an assumption of the linear model

a) the errors are normally distributed with mean 0
b) the independent variables are not correlated
c) y_i have mean $\alpha+\beta x_i$
d) the errors are identically distributed

The ANOVA table of a problem is given below:

ANOVA

Βάρος

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	**x**	2	**z**	31,383	,000
Within Groups	**y**	42	**w**		
Total	2422,578	44			

The sum of squares of the residuals is:

a) y
b) x
c) $(x-y)^2$
d) $(x-y)^2/44$

In a multiple regression model, in which of the following cases, we are not faced with a problem:

a) when the independent variables are far from normality
b) when the independent variables are highly correlated or some of them are a linear combination of some other
c) when we perform multiple tests on the parameters of the model
d) when the sample size is at most equal to the number of parameters of the model

The ANOVA table of a linear model is given below.

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	60,975	1	60,975	5,830	,022[a]
	Residual	334,672	32	10,459		
	Total	395,647	33			

a. *Predictors: (Constant), INDEPEND*

b. *Dependent Variable: DEPEND*

The coefficient of determination (R^2) of this model is:

a) 0,154 (=60,975/395,647)
b) 0,171 (=10,459/60,975)
c) 0,846 (=334,672/395,647)
d) it can not be calculated with the numbers provided

In a simple regression analysis, we estimate the growth (παραγωγή) of an agricultural product given the quantity of fertilizer (λίπασμα) (in kilos) that has been used. The table of the coefficients estimated is given below

Coefficients[a]

Model		Unstandardized Coefficients			
		B	Std. Error	t	Sig.
1	(Constant)	,757	,550	1,375	,176
	Λίπασμα σε κιλά	,925	,035	26,726	,000

a. *Dependent Variable: Παραγωγή σε τόνους*

With respect to the coefficient of the variable «fertilizer», we can conclude that

a) the fertilizer does not influence statistically significantly the growth
b) Increase of the fertilizer by 0,925 kilos causes mean increase of growth by 1 ton
c) Increase of the fertilizer by 1 kilos causes mean increase of growth by 0,925 tons
d) Increase of the fertilizer by 1 kilos causes mean increase of growth by 1,682 (=0,757+0,925·1) tons

Items that load on the first factor

Item 1

The crosstabulation provides the relation between gender (male, άρρεν- female,θήλυ) with the systematic body exercise (yes,ναι- no,όχι).

Item 1	
Discrimination 1	0.81
Discrimination 2	0
Discrimination 3	0
Difficulty	-0.50
Guessing pr.	0.15

			Συστηματική άθληση		
			Ναι	Όχι	Total
Φύλο	Άρρεν	Count	91	10	101
		Expected Count	79,9	21,1	101,0
		% within Φύλο	90,1%	9,9%	100,0%
		% within Συστηματική άθληση	66,9%	27,8%	58,7%
	Θήλυ	Count	45	26	71
		Expected Count	56,1	14,9	71,0
		% within Φύλο	63,4%	36,6%	100,0%
		% within Συστηματική άθληση	33,1%	72,2%	41,3%
Total		Count	136	36	172
		Expected Count	136,0	36,0	172,0
		% within Φύλο	79,1%	20,9%	100,0%
		% within Συστηματική άθληση	100,0%	100,0%	100,0%

The percentage of the female which do not exercise their body systematically is:

a) 14,9%
b) 36,6%
c) 72,2%
d) 33,1%

Item 3

In a multiple linear regression with independent variables X1 and X2, we construct the ANOVA table of the model.

Item 3	
Discrimination 1	0.86
Discrimination 2	0.62
Discrimination 3	0.64
Difficulty	0.73
Guessing pr.	0.18

ANOVA

Model		Sum of Squares	df	F	Sig.
1	Regression	375,097	2	22,469	,000[a]
	Residual	308,834	37		
	Total	683,930	39		

a. *Predictors: (Constant), X1, X2*

The sum of residuals for this model is:

a) 308,834
b) 308,834 / 37
c) 0
d) 37

Item 14

The ANOVA table of a linear model is given below

Item 14	
Discrimination 1	1.06
Discrimination 2	0.44
Discrimination 3	0.43
Difficulty	0.31
Guessing pr.	0.15

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regressio	60,975	1	60,975	5,830	,022[a]
	Residual	334,672	32	10,459		
	Total	395,647	33			

a. *Predictors: (Constant), INDEPEND*

b. *Dependent Variable: DEPEND*

The coefficient of determination (R^2) of this model is:

a) 0,154 (=60,975/395,647)
b) 0,171 (=10,459/60,975)
c) 0,846 (=334,672/395,647)
d) it can not be calculated with the numbers provided

Item 16

The ANOVA table of a problem is given below:

Item 16	
Discrimination 1	1.50
Discrimination 2	0.50
Discrimination 3	0.51
Difficulty	-0.80
Guessing pr.	0.18

ANOVA

Βάρος

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	**x**	2	**z**	31,383	,000
Within Groups	**y**	42	**w**		
Total	2422,578	44			

The sum of squares of the residuals is:

a) y
b) x
c) $(x\text{-}y)^2$
d) $(x\text{-}y)^2/44$

Item 19

An estimated parameter is tested, in significance level 5% and is found to be statistically significant. What could be the result of the test if we used a significance level equal to 1% and 10%;

Item 19	
Discrimination 1	0.83
Discrimination 2	0.63
Discrimination 3	0.37
Difficulty	-0.62
Guessing pr.	0.19

a) in α=1%, the parameter would be statistically significant whereas in α=10% it would not.
b) in α=1%, the parameter would not be statistically significant whereas in α=10% it would be.
c) in α=10% and α=1%, the parameter would be statistically significant
d) in α=10%, the parameter would be statistically significant whereas we can not conclude for α=1%

Item 20

The ANOVA table of a regression model is given below.

Item 20	
Discrimination 1	0.78
Discrimination 2	0.29
Discrimination 3	0.29
Difficulty	-1.22
Guessing pr.	0.17

ANOVA

Βάρος

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	**x**	2	**z**	31,383	,000
Within Groups	**y**	42	**w**		
Total	2422,578	44			

The value 31,383 of F was calculated as:

a) x/z
b) z/w
c) x/2
d) x/y

Item 26

In a simple regression analysis, we estimate the growth (παραγωγή) of an agricultural product given the quantity of fertilizer (λίπασμα) (in kilos) that has been used. The table of the coefficients estimated is given below

Item 26	
Discrimination 1	0.66
Discrimination 2	0.63
Discrimination 3	0.56
Difficulty	-0.42
Guessing pr.	0.17

Coefficients[a]

Model		Unstandardized Coefficients B	Std. Error	t	Sig.
1	(Constant)	,757	,550	1,375	,176
	Λίπασμα σε κιλά	,925	,035	26,726	,000

a. *Dependent Variable: Παραγωγή σε τόνους*

With respect to the coefficient of the variable «fertilizer», we can conclude that

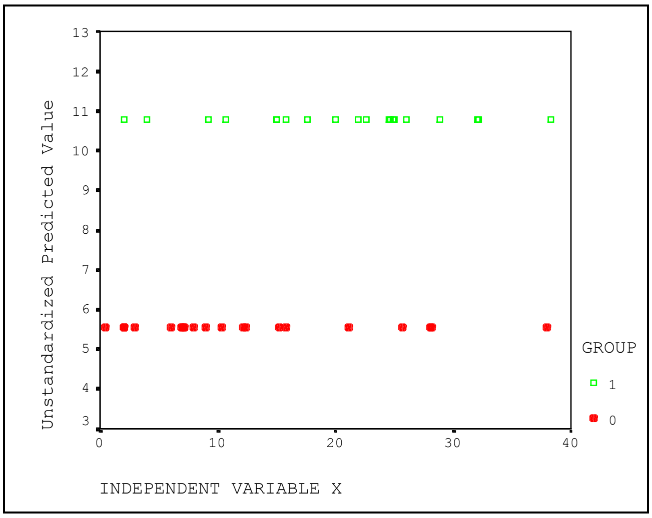
a) the fertilizer does not influence statistically significantly the growth
b) Increase of the fertilizer by 0,925 kilos causes mean increase of growth by 1 ton
c) Increase of the fertilizer by 1 kilos causes mean increase of growth by 0,925 tons
d) Increase of the fertilizer by 1 kilos causes mean increase of growth by 1,682 (=0,757+0,925·1) tons

Items that load on second factor

Item 7

Item 7	
Discrimination 1	0.39
Discrimination 2	0.46
Discrimination 3	0.25
Difficulty	1.30
Guessing pr.	0.13

In a data set, the dependent variable Y is continuous. The data points come from 2 different groups (variable «group», with values 0 and 1) where the mean values of Y for the 2 groups differ statistically significantly. There is also an independent continuous variable X that is statistically significant. We use a model (with either «group», or X, or both) and we find the expected values of Y; these expected values along with the values of X are given in the graph below:



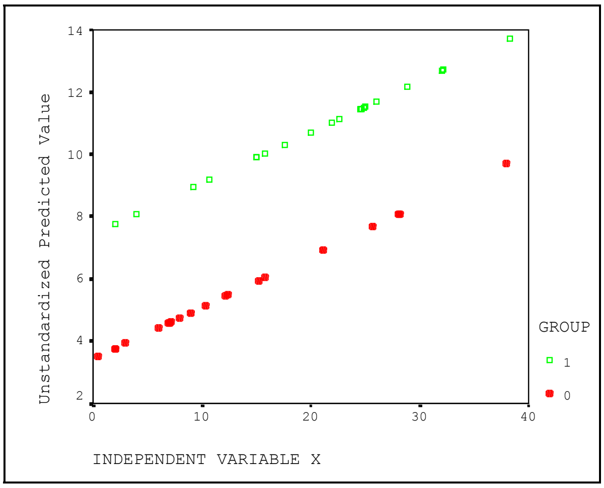
The linear model that has been used in that case is:

a) linear regression with independent variables group and X
b) a separate linear regression for each group with independent variable X
c) two-way ANOVA
d) linear regression with independent variable group

Item 17

Item 17	
Discrimination 1	0.12
Discrimination 2	0.41
Discrimination 3	0.29
Difficulty	0.62
Guessing pr.	0.21

In a data set, the dependent variable Y is continuous. The data points come from 2 different groups (variable «group», with values 0 and 1) where the mean values of Y for the 2 groups differ statistically significantly. There is also an independent continuous variable X that is statistically significant. We use a model (with either «group», or X, or both) and we find the expected values of Y; these expected values along with the values of X are given in the graph below:



The linear model that has been used in that case is:

a) linear regression with independent variables group and X
b) linear regression with independent variables X and the product (X*group)
c) linear regression with independent variable X
d) one-way ANOVA (with variable group)

Item 32

A simple linear regression equation was estimated as Y=4.15-0.96X+ε with coefficient of determination R^2=0.81. The correlation between X and Y is equal to:

a) 0.96
b) 0.96
c) 0.9
d) -0.9

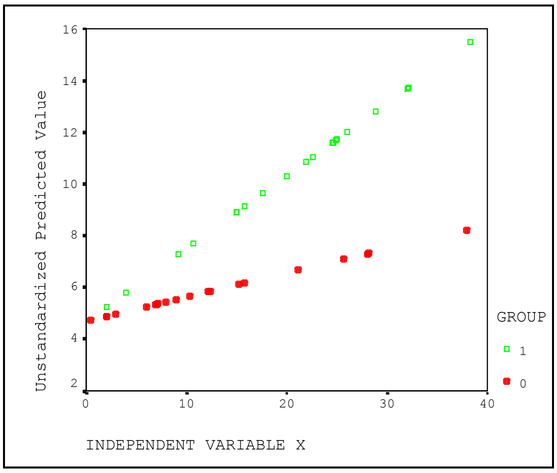
Item 32	
Discrimination 1	0.69
Discrimination 2	1.00
Discrimination 3	0.89
Difficulty	0.91
Guessing pr.	0.17

Items that load on the third factor

Item 6

In a data set, the dependent variable Y is continuous. The data points come from 2 different groups (variable «group», with values 0 and 1) where the mean values of Y for the 2 groups differ statistically significantly. There is also an independent continuous variable X that is statistically significant. We use a model (with either «group», or X, or both) and we find the expected values of Y; these expected values along with the values of X are given in the graph below:

Item 6	
Discrimination 1	0.46
Discrimination 2	0.46
Discrimination 3	0.55
Difficulty	0.36
Guessing pr.	0.13



The linear model that has been used in that case is:

a) linear regression with independent variables group and X
b) linear regression with independent variables X and the product (X*group)
c) two way ANOVA
d) one way ANOVA (using the variable group as independent)

Item 9

In a linear model, the coefficient of an independent variable is not statistically significant (with α=5%). The 95% CI for this parameter is:

a) (-125,-39)
b) (12,42)
c) (0.04,0.09)
d) (-1,24)

Item 9	
Discrimination 1	0.57
Discrimination 2	0.62
Discrimination 3	0.65
Difficulty	0.91
Guessing pr.	0.13

Item 10

In a linear model (where the dependent variable is continuous) in order to include an independent categorical variable with 4 levels, where each level denotes a different profession category, we have to:

a) multiply the categorical variable with each of the continuous independent variables of the model
b) construct 4 dummy variables
c) construct 3 dummy variables
d) include the variable in the model as it is, considering it as a variable that is ordinal

Item 10	
Discrimination 1	0.20
Discrimination 2	0.55
Discrimination 3	0.88
Difficulty	-0.55
Guessing pr.	0.15

Item 12

A data set includes the result of a lung capacity test as "Within accepted (normal) bounds" or "Out of accepted (normal) bounds" as well as whether the person is a smoker or not. We will try to analyze the data set using:

a) Two separate t-tests
b) Regression with the use of dummy variables
c) Two-way ANOVA
d) X^2 test

Item 12	
Discrimination 1	0.24
Discrimination 2	0.35
Discrimination 3	0.75
Difficulty	-0.27
Guessing pr.	0.15

Item 18

Which one of the above is not an assumption for the linear model

a) homoscedasticity of the residuals
b) y_i are independent
c) y_i are normally distributed with mean á+βx_i and variance σ^2_i
d) the errors are uncorrelated

Item 18	
Discrimination 1	0.34
Discrimination 2	0.44
Discrimination 3	0.55
Difficulty	0.68
Guessing pr.	0.17

Item 24

A data set includes the results of a medical test with values from 40 to 98.4 and whether the patient was smoker or not. A possible analysis could be:

a) linear regression with 2 dummy variables
b) t-test
c) two-way anova
d) ancova

Item 24	
Discrimination 1	0.22
Discrimination 2	0.31
Discrimination 3	0.49
Difficulty	0.57
Guessing pr.	0.19

Item 29

In a simple regression analysis, where Y is the dependent variable, we assume the normality of y_i. To test that assumption, we may:

a) check the normality of Y via a graph(e.g. a histogram)
b) check the normality of Y using the non-parametric one sample Kolmogorov-Smirnov test
c) check Y with a graph to see whether there are outlies and, if yes, to remove them from the analysis
d) check the normality of the residuals using the non-parametric one sample Kolmogorov-Smirnov test

Item 29	
Discrimination 1	0.40
Discrimination 2	0.53
Discrimination 3	0.58
Difficulty	0.99
Guessing pr.	0.15

Item 33

In a data set, the dependent variable is the grade in the test (max 105), and the independent variables are the gender and the marital status (with values single, married and divorced). A possible analysis could be:

a) regression analysis with 2 independent variables
b) two way anova
c) ancova of the gender with the marital status
d) t-test with independent variables the gender and the marital status

Item 33	
Discrimination 1	0.38
Discrimination 2	0.59
Discrimination 3	0.62
Difficulty	0.85
Guessing pr.	0.17

Appendix B: Convergence of discrimination and ability parameters

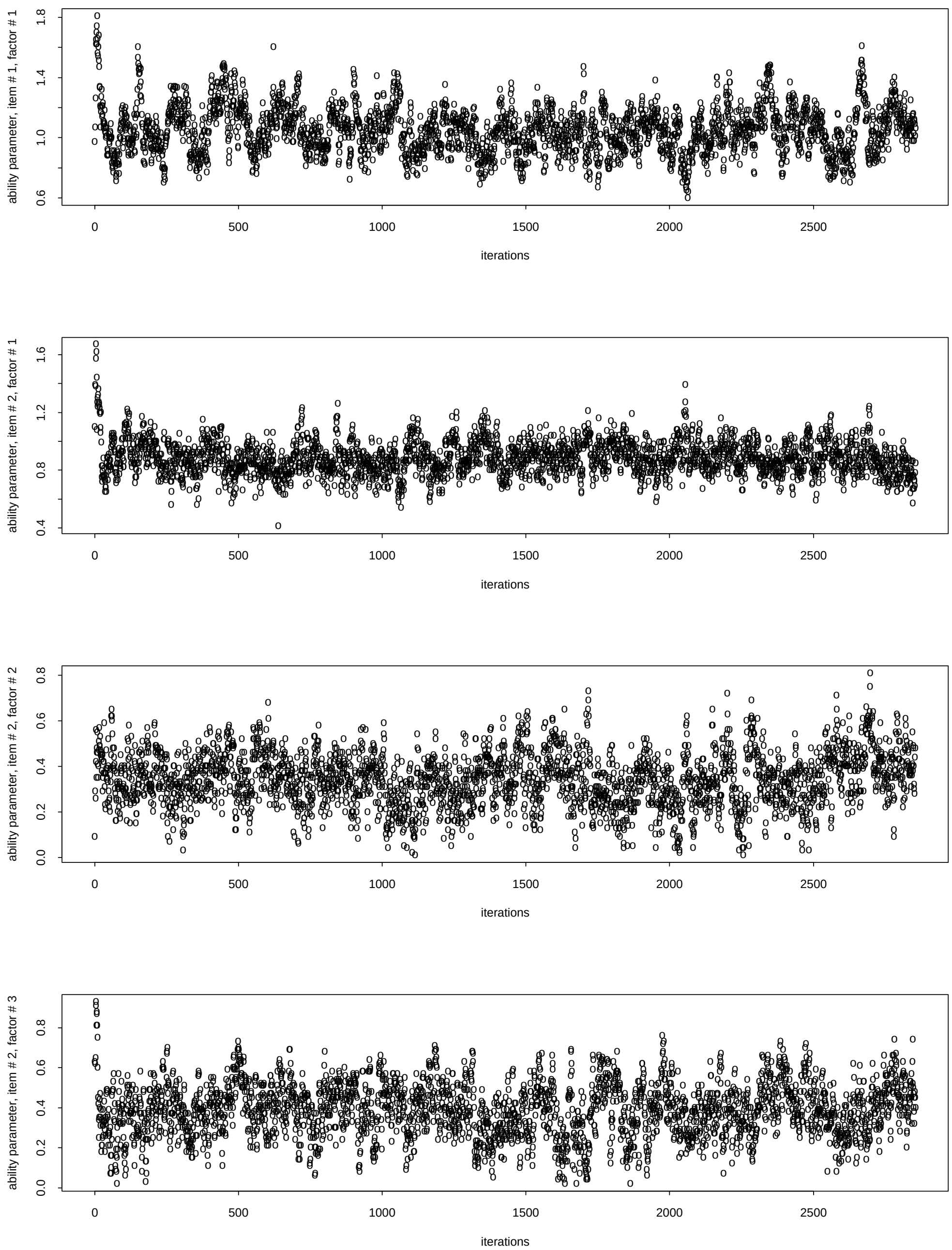


Figure 9-1: Convergence scatter plot of selected ability parameters.

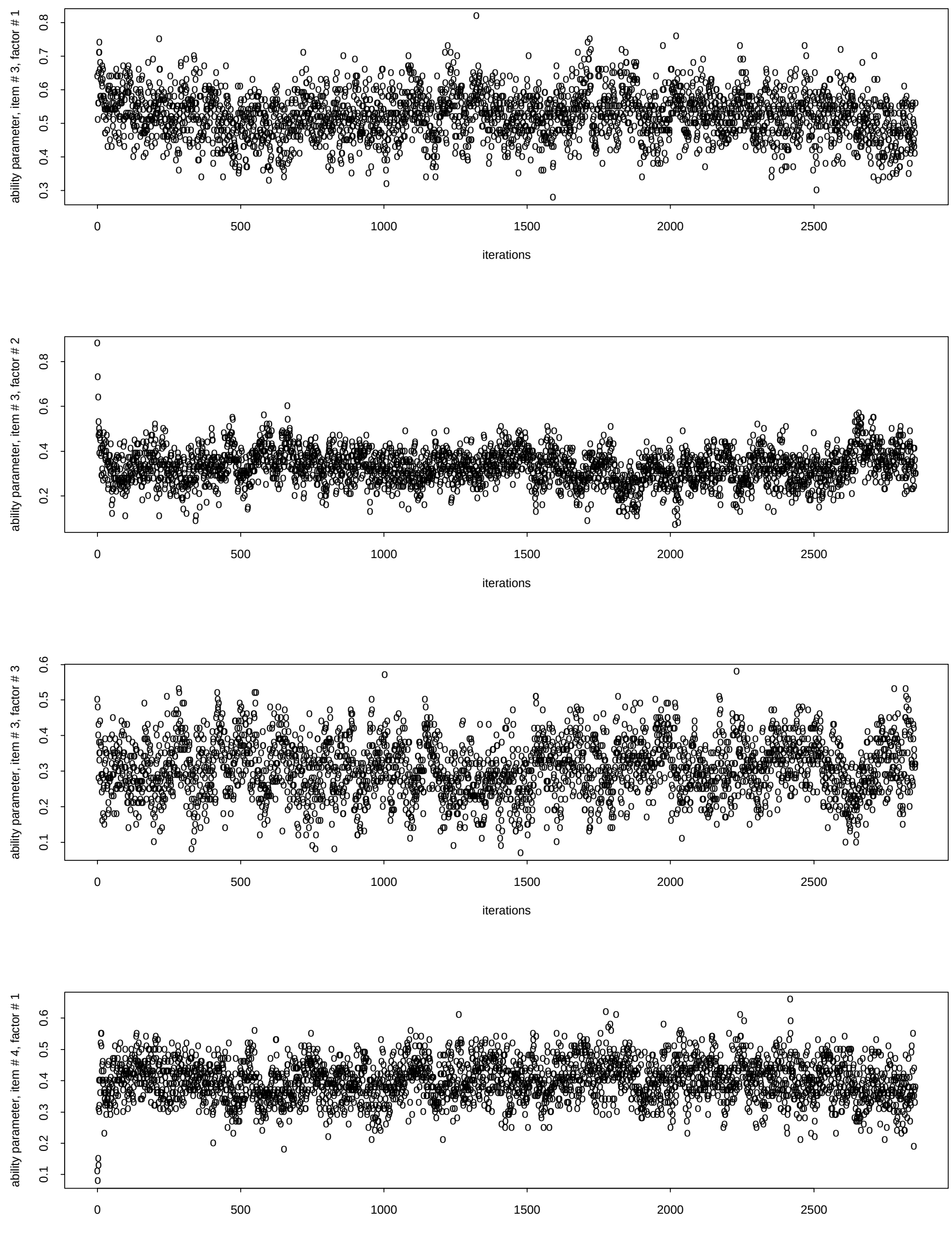


Figure 9-2: Convergence scatter plot of selected ability parameters.

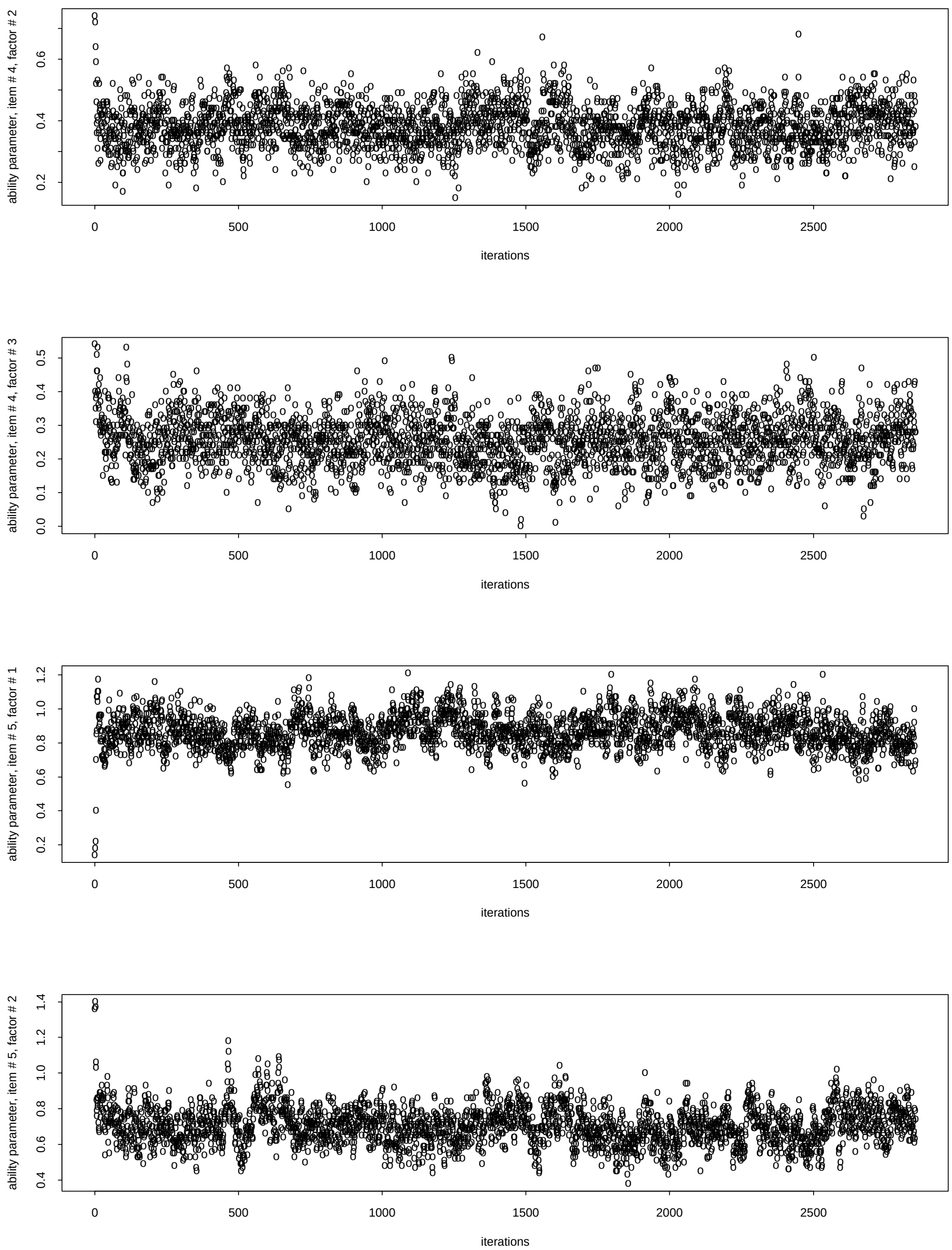


Figure 9-3: Convergence scatter plot of selected ability parameters.

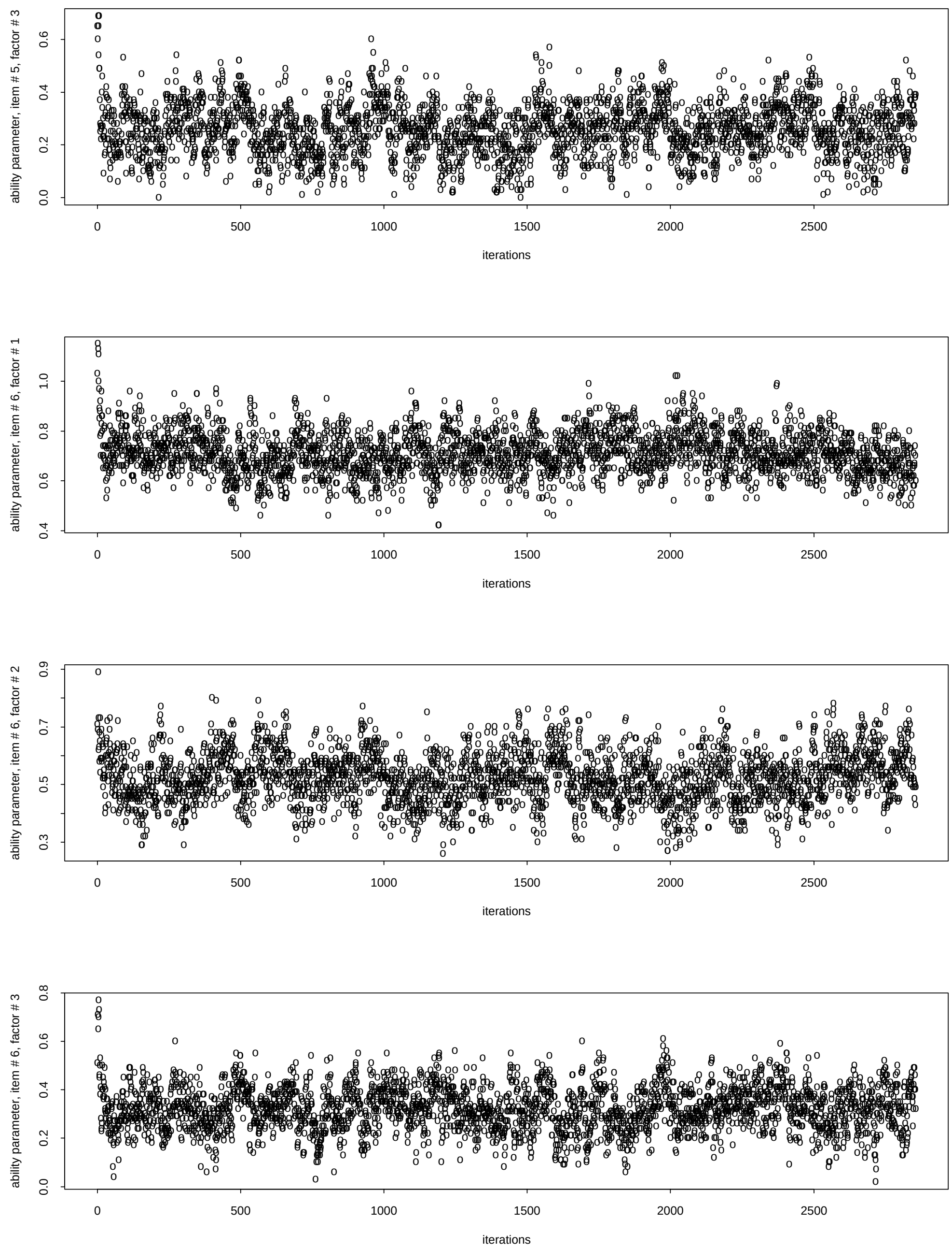


Figure 9-4: Convergence scatter plot of selected ability parameters.

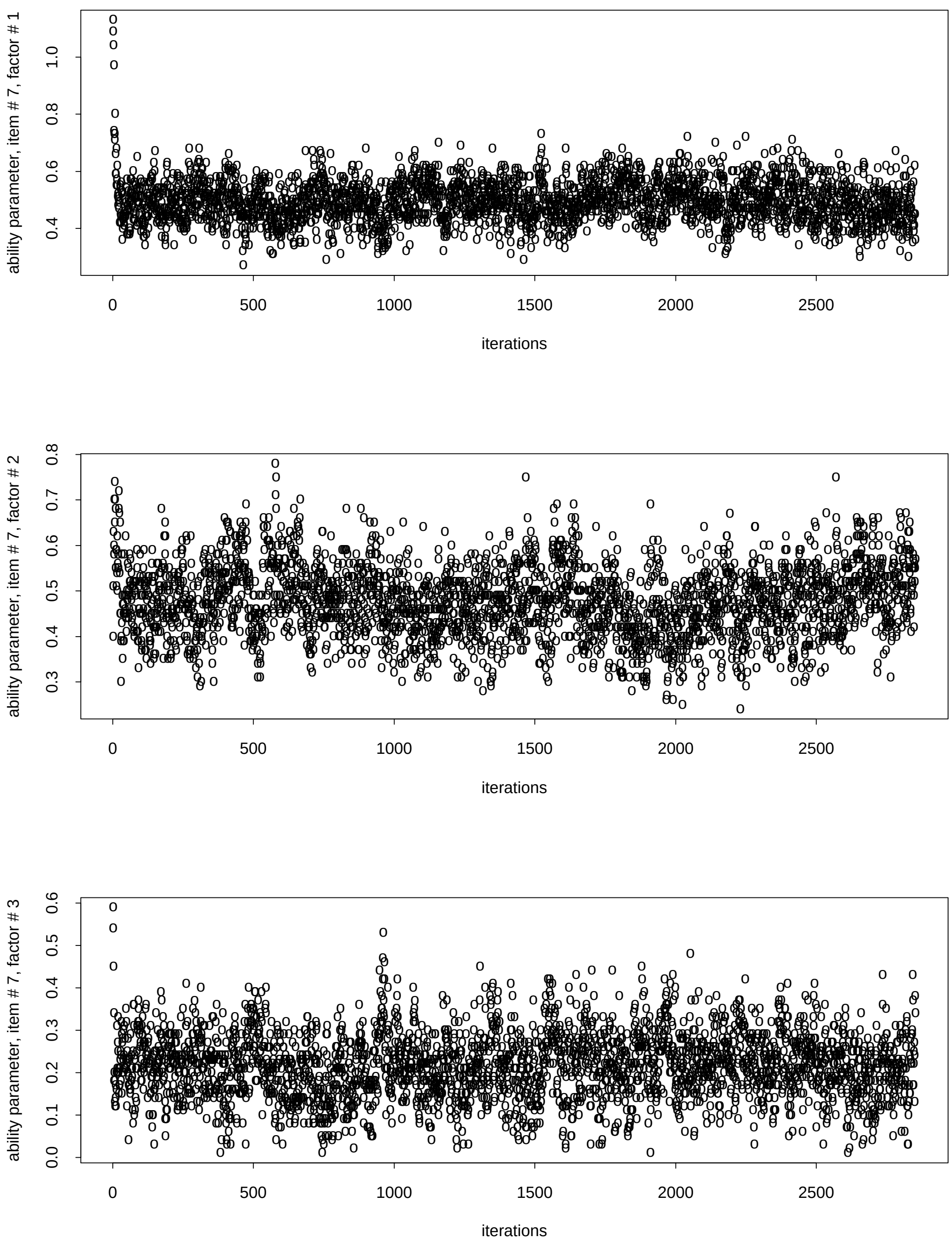


Figure 9-5: Convergence scatter plot of selected ability parameters.

Appendix C: Convergence graphs of the difficulty parameters for the first 20 items

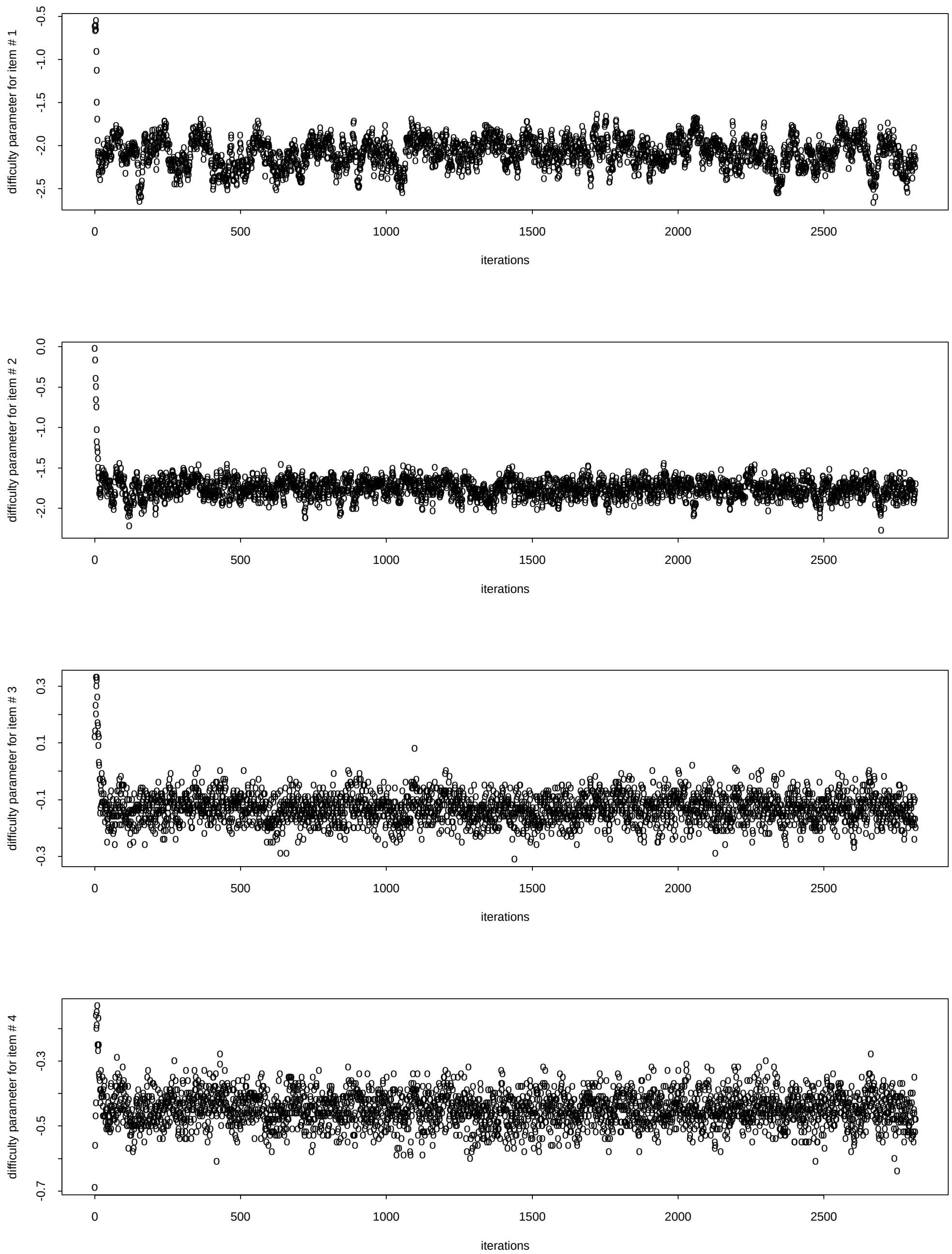


Figure 9-6: Convergence scatter plot of the difficulty parameters.

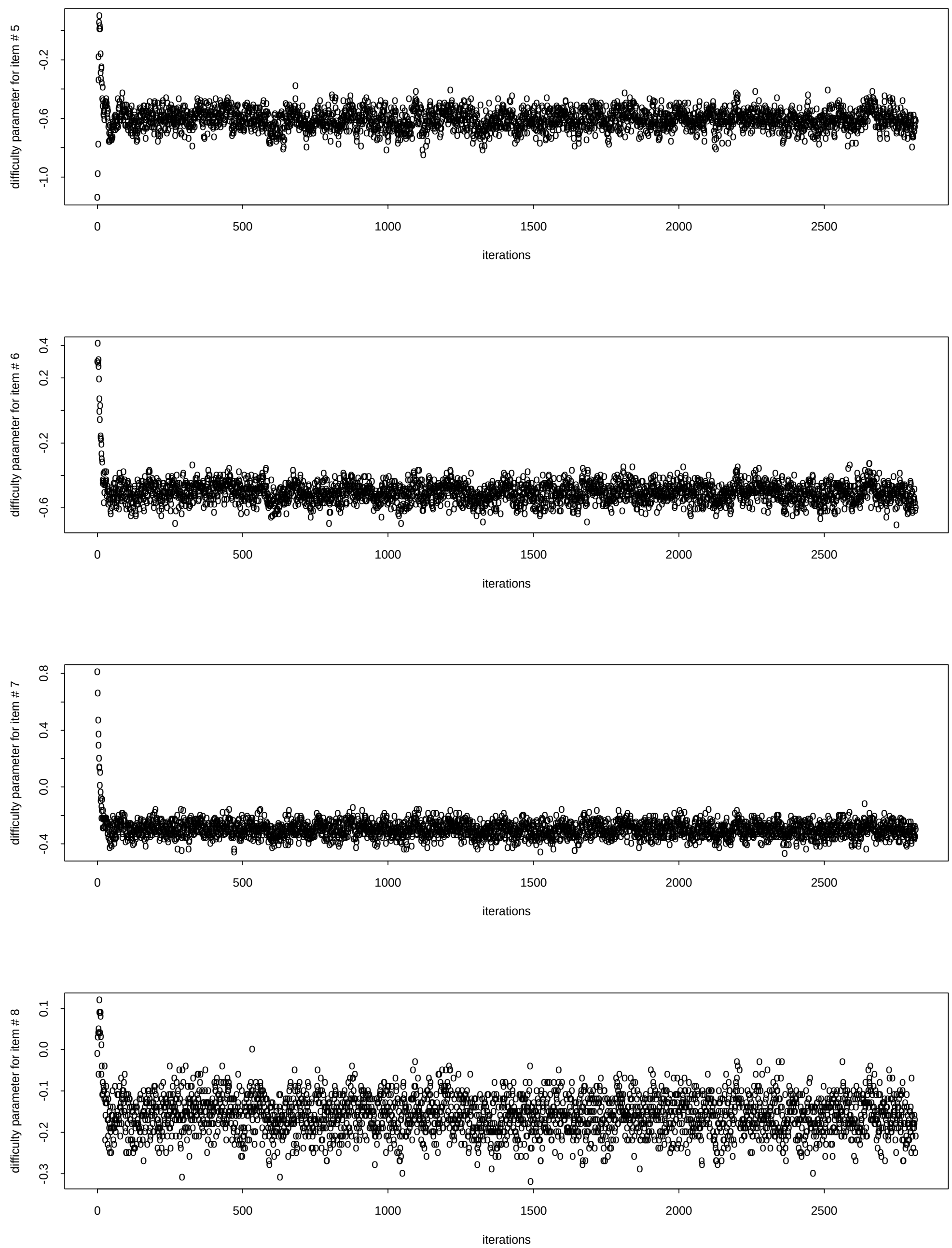


Figure 9-7: Convergence scatter plot of the difficulty parameters.

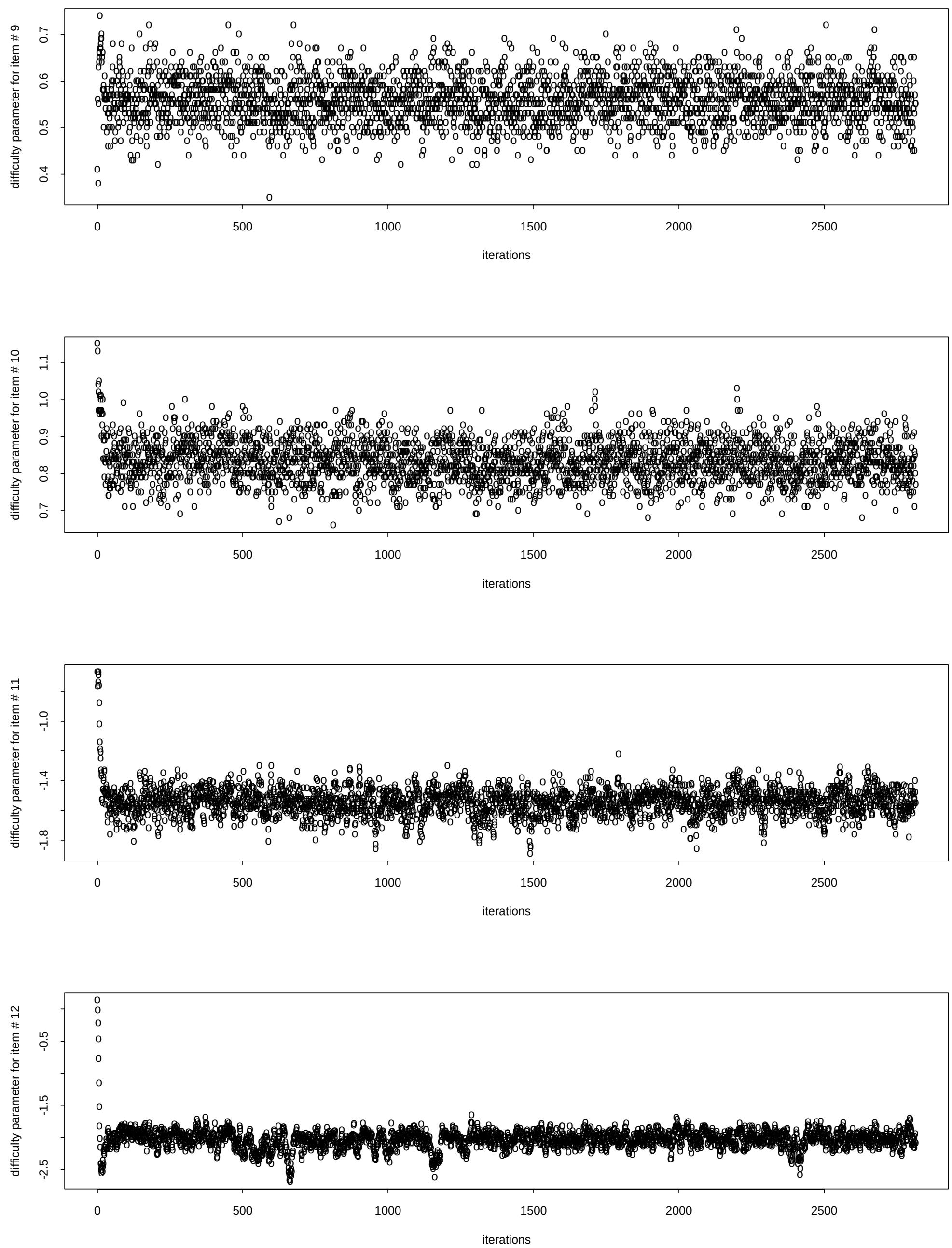


Figure 9-8: Convergence scatter plot of the difficulty parameters.

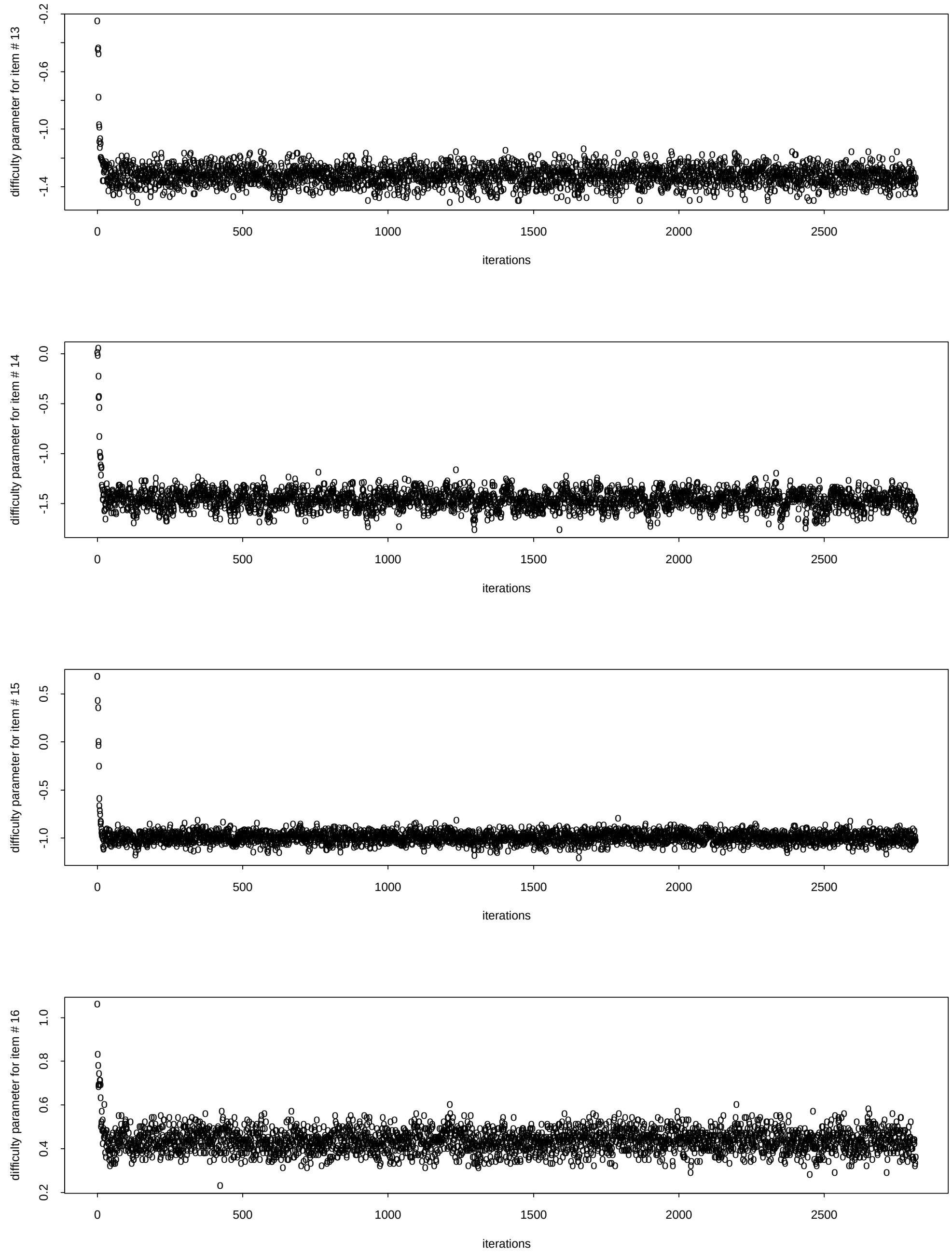


Figure 9-9: Convergence scatter plot of the difficulty parameters.

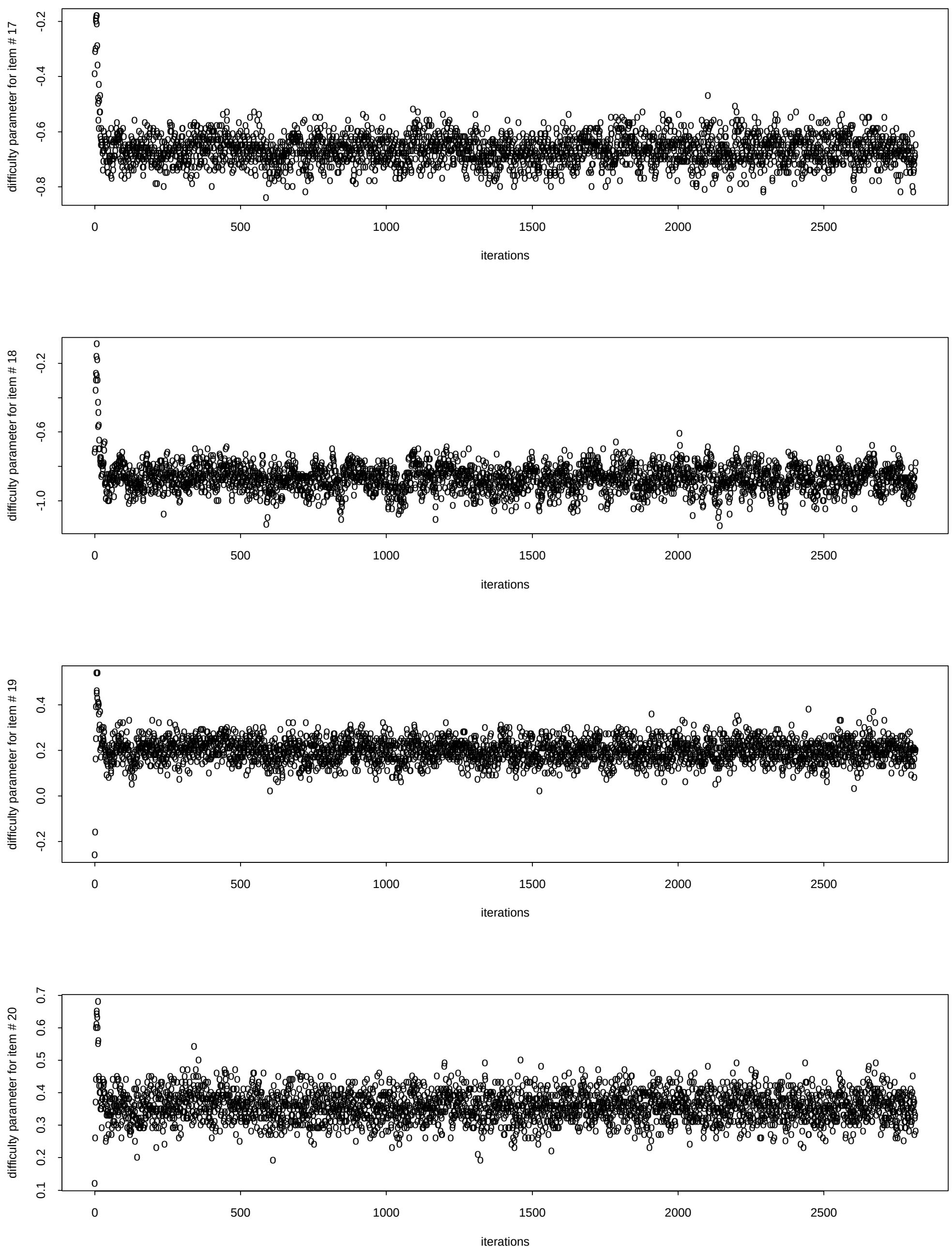


Figure 9-10: Convergence scatter plot of the difficulty parameters.

Appendix D: Autocorrelation plots for the convergence of the δ parameters

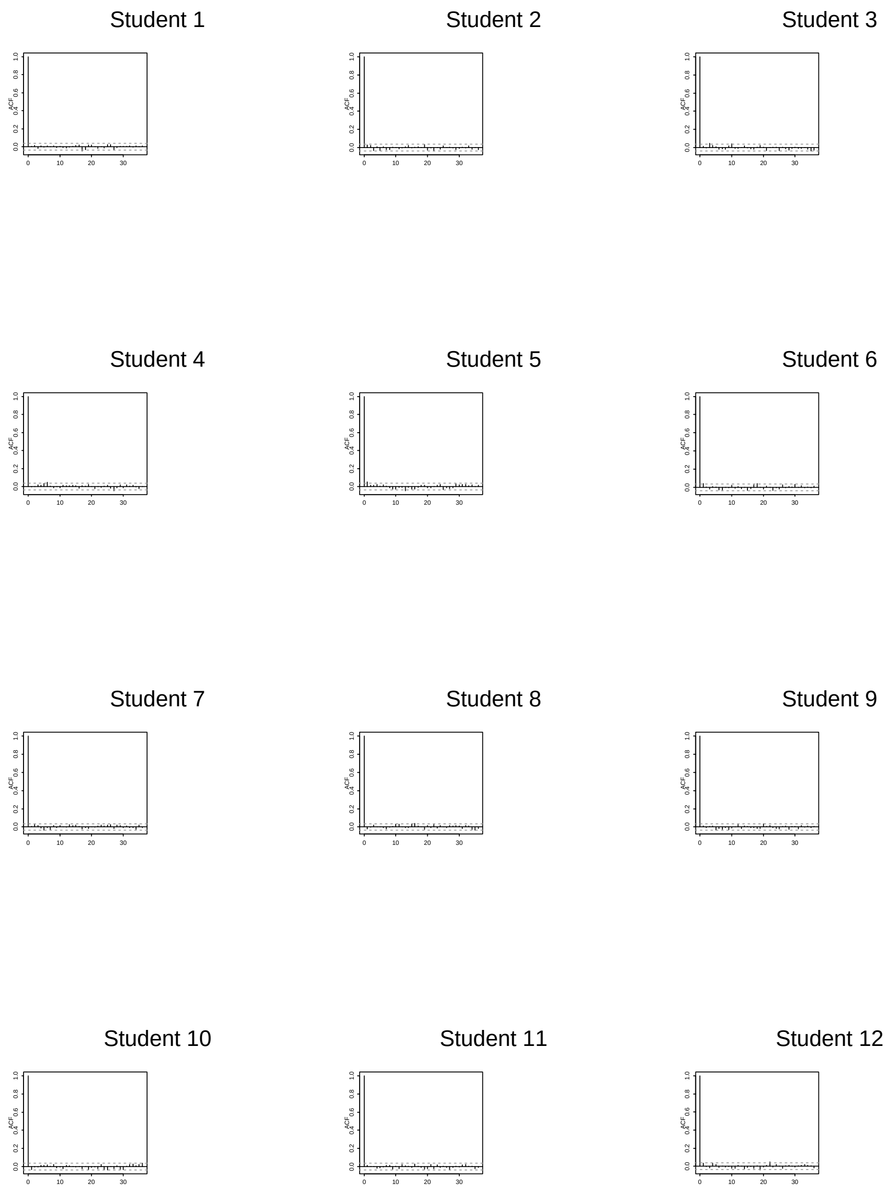


Figure 9-11: Autocorrelation plots of the δ parameters.

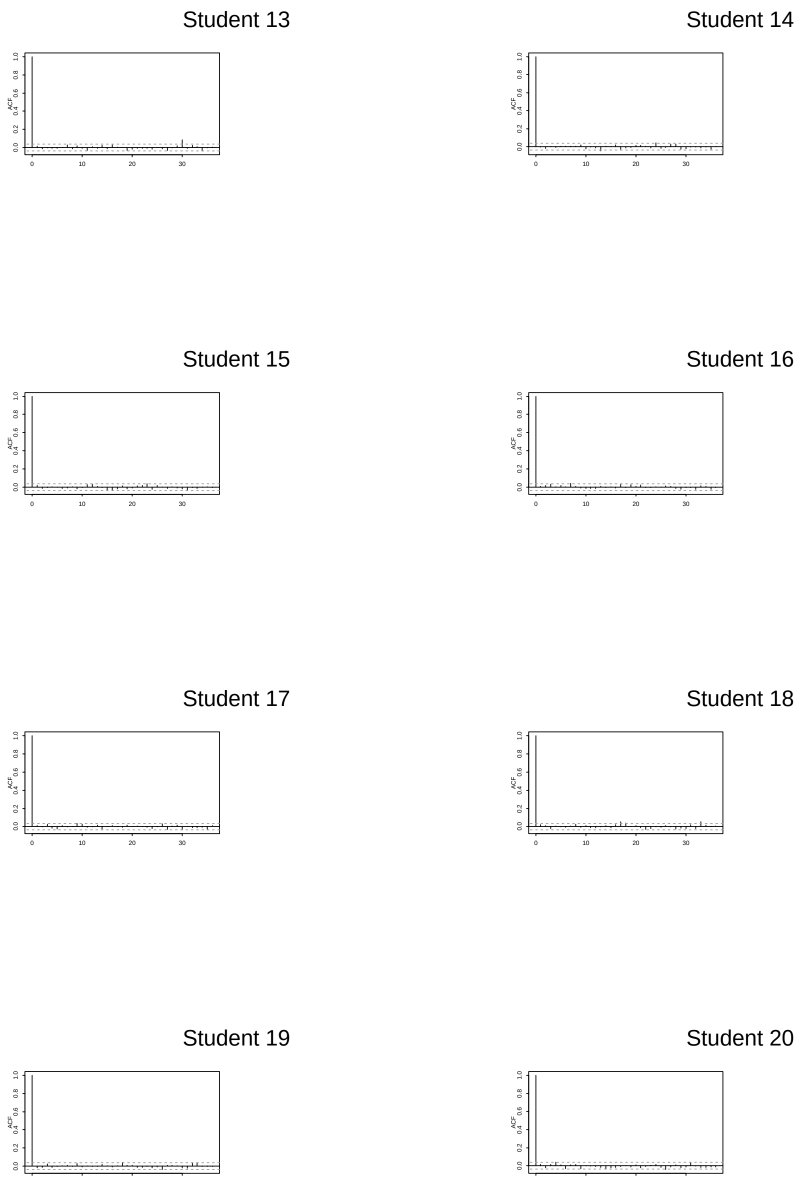


Figure 9-12: Autocorrelation plots of the δ parameters.

Appendix E: Tables of parameter estimates in the SAT data model

Table 9.1: Items' discrimination parameters
(*continued...*)

item #	*discrimination parameter, factor 1*	*discrimination parameter, factor 2*	*discrimination parameter, factor 3*
1	**1.04** (0.77,1.37)	*0.00* (0.00,0.00)	*0.00* (0.00,0.00)
2	**0.88** (0.67,1.10)	*0.34* (0.12,0.58)	*0.39* (0.12,0.65)
3	**0.52** (0.38,0.66)	*0.33* (0.18,0.47)	*0.31* (0.15,0.47)
4	**0.40** (0.28,0.52)	*0.38* (0.25,0.52)	*0.25* (0.12,0.40)
5	**0.87** (0.69,1.07)	*0.70* (0.49,0.90)	*0.25* (0.07,0.45)
6	**0.71** (0.55,0.88)	*0.52* (0.35,0.70)	*0.32* (0.14,0.49)
7	**0.49** (0.35,0.63)	*0.48* (0.32,0.63)	*0.21* (0.06,0.37)
8	*0.35* (0.22,0.48)	**0.42** (0.28,0.58)	*0.15* (0.02,0.30)
9	*0.12* (0.01,0.25)	**0.47** (0.33,0.61)	*0.18* (0.03,0.33)
10	*0.10* (0.01,0.24)	**0.49** (0.35,0.64)	*0.12* (0.01,0.28)

Table 9.1:Items' discrimination parameters
(continued...)

item #	*discrimination parameter, factor 1*	*discrimination parameter, factor 2*	*discrimination parameter, factor 3*
11	**0.72** *(0.53,0.91)*	*0.17* *(0.02,0.36)*	*0.28* *(0.06,0.52)*
12	**0.43** *(0.21,0.64)*	*0.04* *(0.00,0.13)*	*0.40* *(0.09,0.80)*
13	**0.34** *(0.19,0.48)*	*0.26* *(0.09,0.43)*	*0.16* *(0.02,0.34)*
14	**0.85** *(0.65,1.06)*	*0.32* *(0.11,0.55)*	*0.39* *(0.15,0.64)*
15	**0.46** *(0.32,0.60)*	*0.08* *(0.00,0.21)*	*0.15* *(0.01,0.33)*
16	**0.41** *(0.28,0.56)*	*0.39* *(0.25,0.53)*	*0.18* *(0.04,0.35)*
17	**0.44** *(0.30,0.58)*	*0.17* *(0.04,0.31)*	*0.35* *(0.18,0.53)*
18	**0.86** *(0.68,1.06)*	*0.57* *(0.38,0.78)*	*0.25* *(0.06,0.45)*
19	*0.42* *(0.29,0.56)*	**0.50** *(0.35,0.64)*	*0.15* *(0.02,0.30)*
20	**0.40** *(0.27,0.53)*	*0.18* *(0.06,0.30)*	*0.21* *(0.06,0.36)*
21	*0.34* *(0.18,0.51)*	**0.61** *(0.44,0.80)*	*0.09* *(0.00,0.25)*
22	*0.15* *(0.03,0.28)*	**0.29** *(0.16,0.43)*	*0.07* *(0.00,0.19)*
23	*0.09* *(0.00,0.21)*	**0.56** *(0.40,0.72)*	*0.18* *(0.02,0.38)*
24	**0.51** *(0.35,0.68)*	*0.15* *(0.02,0.30)*	*0.40* *(0.19,0.61)*
25	**0.51** *(0.36,0.66)*	*0.18* *(0.04,0.33)*	*0.28* *(0.10,0.46)*
26	*0.38* *(0.23,0.52)*	**0.47** *(0.31,0.63)*	*0.39* *(0.23,0.56)*
27	*0.17* *(0.04,0.30)*	*0.29* *(0.16,0.42)*	**0.36** *(0.21,0.52)*
28	**0.45** *(0.32,0.60)*	*0.31* *(0.17,0.46)*	*0.42* *(0.27,0.57)*
29	*0.21* *(0.08,0.33)*	*0.23* *(0.10,0.37)*	**0.24** *(0.08,0.40)*

Table 9.1:Items' discrimination parameters
(*continued...*)

item #	*discrimination parameter, factor 1*	*discrimination parameter, factor 2*	*discrimination parameter, factor 3*
30	*0.29* (0.14,0.43)	*0.35* (0.20,0.49)	**0.43** (0.24,0.62)
31	*0.23* (0.08,0.38)	**0.51** (0.34,0.66)	*0.42* (0.24,0.62)
32	*0.37* (0.21,0.53)	*0.62* (0.42,0.81)	**0.68** (0.47,0.90)
33	*0.68* (0.47,0.89)	**0.99** (0.72,1.24)	*0.80* (0.56,1.08)
34	*0.29* (0.13,0.46)	**0.69** (0.51,0.88)	*0.52* (0.31,0.74)
35	*0.25* (0.09,0.42)	**0.35** (0.19,0.51)	*0.28* (0.07,0.47)
36	**0.73** (0.55,0.92)	*0.45* (0.26,0.65)	*0.00* (0.00,0.00)
37	**0.47** (0.33,0.61)	*0.23* (0.07,0.39)	*0.27* (0.09,0.45)
38	*0.39* (0.25,0.52)	**0.41** (0.28,0.55)	*0.09* (0.01,0.22)
39	**0.53** (0.40,0.66)	*0.34* (0.20,0.49)	*0.17* (0.03,0.32)
40	**0.55** (0.40,0.72)	*0.18* (0.03,0.35)	*0.16* (0.01,0.36)
41	**0.72** (0.56,0.89)	*0.63* (0.45,0.81)	*0.23* (0.06,0.42)
42	*0.46* (0.30,0.62)	**0.71** (0.55,0.89)	*0.25* (0.08,0.40)
43	*0.50* (0.35,0.65)	**0.54** (0.38,0.72)	*0.15* (0.02,0.33)
44	*0.41* (0.25,0.58)	**0.68** (0.50,0.87)	*0.07* (0.00,0.21)
45	**0.53** (0.39,0.68)	*0.28* (0.13,0.45)	*0.14* (0.01,0.30)
46	**0.39** (0.26,0.53)	*0.13* (0.02,0.27)	*0.09* (0.00,0.22)
47	**0.81** (0.62,1.03)	*0.42* (0.18,0.65)	*0.36* (0.13,0.60)
48	*0.21* (0.10,0.33)	**0.35** (0.23,0.47)	*0.12* (0.01,0.24)

Table 9.1:Items' discrimination parameters (*continued...*)

item #	discrimination parameter, factor 1	discrimination parameter, factor 2	discrimination parameter, factor 3
49	0.30 (0.16,0.44)	**0.40** (0.27,0.53)	0.09 (0.01,0.22)
50	0.15 (0.02,0.29)	**0.56** (0.42,0.73)	0.11 (0.01,0.28)
51	0.27 (0.14,0.40)	0.25 (0.11,0.41)	**0.33** (0.17,0.50)
52	**0.53** (0.40,0.67)	0.28 (0.15,0.42)	0.18 (0.04,0.32)
53	**0.38** (0.23,0.52)	0.05 (0.00,0.15)	0.24 (0.08,0.42)
54	**0.46** (0.32,0.60)	0.14 (0.02,0.27)	0.30 (0.13,0.48)
55	**0.68** (0.54,0.83)	0.16 (0.03,0.31)	0.29 (0.08,0.49)
56	**0.57** (0.43,0.72)	0.38 (0.22,0.53)	0.36 (0.20,0.53)
57	**0.40** (0.27,0.53)	0.14 (0.02,0.27)	0.29 (0.12,0.45)
58	**0.44** (0.32,0.57)	0.30 (0.16,0.45)	0.21 (0.06,0.37)
59	**0.43** (0.30,0.57)	0.27 (0.14,0.43)	0.38 (0.21,0.54)
60	0.81 (0.57,1.06)	0.13 (0.01,0.33)	**0.82** (0.47,1.21)
61	**0.53** (0.38,0.67)	0.48 (0.33,0.64)	0.39 (0.22,0.56)
62	**0.49** (0.34,0.63)	0.34 (0.18,0.50)	0.43 (0.26,0.61)
63	0.27 (0.14,0.40)	0.31 (0.17,0.45)	**0.39** (0.24,0.56)
64	0.48 (0.30,0.66)	**0.60** (0.42,0.78)	0.43 (0.23,0.64)
65	**0.63** (0.46,0.79)	0.55 (0.37,0.72)	0.60 (0.43,0.78)
66	0.42 (0.26,0.58)	0.63 (0.45,0.81)	**0.69** (0.50,0.91)
67	**0.56** (0.42,0.70)	0.22 (0.08,0.36)	0.20 (0.04,0.37)

Table 9.1:Items' discrimination parameters

item #	discrimination parameter, factor 1	discrimination parameter, factor 2	discrimination parameter, factor 3
68	**0.44** (0.31,0.57)	0.24 (0.09,0.40)	0.26 (0.10,0.43)
69	0.32 (0.18,0.46)	**0.36** (0.22,0.50)	0.28 (0.12,0.45)
70	**0.27** (0.16,0.39)	0.27 (0.15,0.40)	0.16 (0.03,0.29)
71	**0.62** (0.46,0.77)	0.33 (0.17,0.48)	0.35 (0.17,0.55)
72	0.17 (0.04,0.29)	0.26 (0.14,0.39)	**0.32** (0.18,0.46)
73	0.46 (0.32,0.60)	**0.55** (0.39,0.70)	0.36 (0.21,0.52)
74	0.41 (0.27,0.55)	**0.41** (0.27,0.55)	0.37 (0.22,0.52)
75	0.27 (0.15,0.39)	**0.27** (0.15,0.39)	0.21 (0.08,0.34)
76	**0.55** (0.39,0.70)	0.49 (0.33,0.64)	0.39 (0.21,0.57)
77	0.29 (0.12,0.44)	**0.64** (0.47,0.83)	0.49 (0.27,0.73)
78	0.36 (0.20,0.51)	**0.64** (0.47,0.83)	0.48 (0.23,0.75)

Table 9.2: Students abilities parameters
(continued...)

student #	*ability parameter on factor 1*	*ability parameter on factor 2*	*ability parameter on factor 3*
1	*0.46* *(-0.70,1.65)*	*0.19* *(-0.83,1.25)*	**1.17** *(-0.01,2.45)*
2	*-1.47* *(-2.40,-0.45)*	**0.70** *(-0.43,1.88)*	*-0.45* *(-1.41,0.58)*
3	*-1.94* *(-2.95,-0.97)*	*-0.15* *(-1.14,0.89)*	**0.83** *(-0.40,2.20)*
4	*-1.01* *(-2.05,0.05)*	*0.30* *(-0.74,1.43)*	**0.62** *(-0.41,1.78)*
5	*-0.91* *(-1.84,0.02)*	*0.12* *(-0.94,1.23)*	**1.29** *(0.10,2.55)*
6	*0.38* *(-0.68,1.52)*	**0.64** *(-0.50,1.84)*	*0.57* *(-0.47,1.75)*
7	*-0.02* *(-1.08,1.10)*	*-0.54* *(-1.51,0.51)*	**0.65** *(-0.51,1.88)*
8	**-0.29** *(-1.33,0.74)*	*-0.67* *(-1.64,0.36)*	*-0.55* *(-1.58,0.50)*
9	*-1.06* *(-2.04,-0.02)*	*-0.72* *(-1.70,0.25)*	**-0.18** *(-1.21,0.93)*
10	*1.11* *(-0.20,2.43)*	*0.36* *(-0.65,1.47)*	**1.25** *(-0.01,2.54)*
11	**0.93** *(-0.32,2.27)*	*-0.23* *(-1.26,0.86)*	*0.35* *(-0.88,1.59)*
12	*0.77* *(-0.34,1.97)*	*-0.03* *(-1.06,1.09)*	**1.24** *(-0.13,2.65)*
13	*0.23* *(-0.85,1.35)*	**1.37** *(0.18,2.63)*	*0.63* *(-0.61,1.93)*
14	*0.71* *(-0.43,1.88)*	*-0.88* *(-1.87,0.08)*	**1.38** *(0.22,2.56)*

Table 9.2:Students abilities parameters
(*continued...*)

student #	*ability parameter on factor 1*	*ability parameter on factor 2*	*ability parameter on factor 3*
15	*-0.21* (-1.19,0.87)	**0.52** (-0.56,1.73)	*-0.27* (-1.30,0.82)
16	*0.16* (-0.92,1.27)	*-0.64* (-1.58,0.37)	**0.26** (-0.80,1.38)
17	**0.53** (-0.61,1.65)	*-0.64* (-1.55,0.30)	*-0.20* (-1.18,0.81)
18	*-0.34* (-1.31,0.70)	**0.34** (-0.76,1.49)	*0.22* (-0.84,1.34)
19	*-0.71* (-1.72,0.29)	*-0.38* (-1.42,0.69)	**1.37** (0.08,2.66)
20	**0.74** (-0.50,2.01)	*-0.29* (-1.32,0.75)	*0.71* (-0.51,2.03)
21	*0.01* (-1.07,1.09)	**1.15** (-0.12,2.53)	*0.56* (-0.57,1.80)
22	*0.34* (-0.72,1.47)	*0.42* (-0.63,1.52)	**1.07** (-0.13,2.34)
23	*0.54* (-0.59,1.71)	*-0.07* (-1.13,1.11)	**0.56** (-0.71,1.85)
24	*-0.81* (-1.76,0.19)	**0.80** (-0.49,2.12)	*-1.43* (-2.39,-0.42)
25	**0.72** (-0.47,2.02)	*-0.15* (-1.14,0.89)	*0.06* (-0.98,1.17)
26	**0.82** (-0.35,2.05)	*0.06* (-1.03,1.19)	*0.38* (-0.77,1.66)
27	**0.53** (-0.60,1.68)	*0.22* (-0.84,1.35)	*-0.39* (-1.39,0.65)
28	*-0.99* (-1.93,0.01)	*-0.66* (-1.60,0.37)	**0.73** (-0.45,1.93)
29	*-0.31* (-1.31,0.70)	*-0.16* (-1.20,0.91)	**0.26** (-0.81,1.34)
30	*-0.83* (-1.77,0.10)	**0.15** (-0.88,1.21)	*0.14* (-0.87,1.22)
31	**0.48** (-0.68,1.74)	*0.28* (-0.78,1.40)	*-0.17* (-1.18,0.90)
32	*-0.72* (-1.71,0.35)	*0.96* (-0.16,2.15)	**1.50** (0.03,2.99)
33	**0.88** (-0.44,2.26)	*-0.10* (-1.13,0.98)	*-0.06* (-1.07,1.02)

Table 9.2:Students abilities parameters
(continued...)

student #	*ability parameter on factor 1*	*ability parameter on factor 2*	*ability parameter on factor 3*
34	*-0.30* (-1.27,0.78)	**-0.18** (-1.23,0.92)	*-1.79* (-2.71,-0.90)
35	**1.32** (0.03,2.64)	*0.31* (-0.87,1.54)	*-0.97* (-1.96,0.06)
36	*-1.05* (-1.98,-0.03)	**0.70** (-0.36,1.78)	*0.53* (-0.55,1.63)
37	*0.73* (-0.38,1.92)	**0.91** (-0.21,2.11)	*-0.09* (-1.11,1.02)
38	**1.00** (-0.17,2.22)	*-0.12* (-1.15,0.98)	*0.35* (-0.66,1.51)
39	*0.29* (-0.80,1.39)	**0.69** (-0.58,2.01)	*0.41* (-0.67,1.53)
40	*-0.55* (-1.55,0.47)	*-0.39* (-1.42,0.68)	**-0.26** (-1.24,0.85)
41	*-0.64* (-1.60,0.39)	**-0.00** (-0.97,1.05)	*-0.42* (-1.41,0.59)
42	*-1.80* (-2.87,-0.78)	**1.58** (0.16,3.03)	*0.17* (-0.87,1.29)
43	*-0.15* (-1.12,0.91)	*-0.99* (-1.95,0.02)	**0.43** (-0.69,1.55)
44	**0.23** (-0.82,1.36)	*0.13* (-1.01,1.38)	*-0.51* (-1.52,0.61)
45	*0.48* (-0.58,1.56)	**1.03** (-0.27,2.34)	*-0.42* (-1.37,0.57)
46	*-1.46* (-2.38,-0.54)	**0.90** (-0.21,2.11)	*-1.06* (-1.99,-0.06)
47	*1.07* (-0.30,2.44)	*0.97* (-0.24,2.23)	**1.45** (0.09,2.84)
48	*-0.59* (-1.55,0.40)	**-0.56** (-1.62,0.58)	*-1.47* (-2.44,-0.48)
49	**0.83** (-0.38,2.03)	*0.14* (-0.93,1.21)	*-1.37* (-2.33,-0.36)
50	*0.81* (-0.36,2.02)	*0.12* (-0.88,1.19)	**1.19** (-0.14,2.55)
51	*-0.08* (-1.10,0.92)	*-0.19* (-1.20,0.89)	**1.21** (0.06,2.44)
52	*-0.52* (-1.52,0.47)	*-1.14* (-2.12,-0.10)	**0.42** (-0.74,1.61)

Table 9.2:Students abilities parameters
(*continued...*)

student #	*ability parameter on factor 1*	*ability parameter on factor 2*	*ability parameter on factor 3*
53	**0.03** *(-0.94,1.15)*	*-0.20* *(-1.22,0.87)*	*-2.51* *(-3.51,-1.49)*
54	**1.38** *(0.07,2.70)*	*0.04* *(-1.03,1.19)*	*0.85* *(-0.41,2.12)*
55	**0.03** *(-1.05,1.16)*	*-0.53* *(-1.56,0.57)*	*-0.02* *(-1.04,1.10)*
56	*0.24* *(-0.80,1.39)*	**0.46** *(-0.65,1.55)*	*-0.84* *(-1.76,0.10)*
57	*-1.56* *(-2.44,-0.64)*	**1.31** *(0.18,2.53)*	*0.89* *(-0.26,2.06)*
58	*0.30* *(-0.76,1.40)*	*-0.41* *(-1.43,0.64)*	**0.45** *(-0.66,1.59)*
59	**0.93** *(-0.26,2.13)*	*0.23* *(-0.97,1.47)*	*0.63* *(-0.54,1.80)*
60	**0.95** *(-0.35,2.27)*	*0.07* *(-0.93,1.12)*	*-0.16* *(-1.20,0.95)*
61	*-0.07* *(-1.12,1.00)*	*0.17* *(-0.88,1.28)*	**0.26** *(-0.85,1.41)*
62	**0.56** *(-0.53,1.71)*	*-0.36* *(-1.42,0.68)*	*-0.29* *(-1.28,0.73)*
63	**0.91** *(-0.39,2.24)*	*-0.08* *(-1.10,0.98)*	*0.08* *(-0.98,1.20)*
64	**0.84** *(-0.35,2.06)*	*0.66* *(-0.46,1.86)*	*-0.31* *(-1.31,0.77)*
65	*-0.60* *(-1.57,0.43)*	**0.39** *(-0.69,1.55)*	*-0.58* *(-1.58,0.45)*
66	*-0.27* *(-1.21,0.73)*	*0.47* *(-0.68,1.65)*	**1.21** *(0.04,2.41)*
67	*0.75* *(-0.59,2.06)*	*-0.66* *(-1.64,0.38)*	**1.03** *(-0.33,2.46)*
68	*0.40* *(-0.72,1.52)*	**0.85** *(-0.40,2.16)*	*-0.72* *(-1.71,0.34)*
69	*-0.44* *(-1.50,0.69)*	*-1.21* *(-2.20,-0.14)*	**-0.01** *(-1.05,1.07)*
70	*0.51* *(-0.55,1.68)*	*-0.43* *(-1.46,0.66)*	**0.82** *(-0.29,1.95)*
71	*-1.37* *(-2.36,-0.41)*	*0.67* *(-0.41,1.80)*	**1.00** *(-0.33,2.35)*

Table 9.2:Students abilities parameters (*continued...*)

student #	ability parameter on factor 1	ability parameter on factor 2	ability parameter on factor 3
72	**-0.41** (-1.39,0.60)	-1.06 (-2.02,-0.08)	-0.67 (-1.61,0.37)
73	**-0.25** (-1.27,0.81)	-0.28 (-1.24,0.82)	-0.59 (-1.60,0.44)
74	0.82 (-0.32,2.01)	-0.72 (-1.70,0.37)	**1.19** (-0.07,2.43)
75	-0.37 (-1.47,0.77)	-0.44 (-1.33,0.50)	**0.97** (-0.22,2.21)
76	-1.20 (-2.19,-0.10)	0.50 (-0.62,1.67)	**0.98** (-0.29,2.30)
77	**0.03** (-0.99,1.13)	-1.98 (-2.95,-0.96)	-0.31 (-1.31,0.72)
78	0.12 (-0.91,1.16)	**0.42** (-0.73,1.60)	-0.87 (-1.90,0.26)
79	-0.03 (-1.05,1.05)	1.06 (-0.10,2.22)	**1.13** (-0.14,2.48)
80	-0.74 (-1.67,0.32)	-0.54 (-1.54,0.51)	**0.09** (-1.00,1.20)
81	-0.27 (-1.32,0.81)	-0.41 (-1.40,0.70)	**0.83** (-0.41,2.17)
82	-0.76 (-1.77,0.30)	-1.21 (-2.15,-0.17)	**0.40** (-0.73,1.62)
83	0.53 (-0.54,1.74)	**0.67** (-0.52,1.87)	0.54 (-0.72,1.84)
84	-0.64 (-1.60,0.34)	**1.53** (0.27,2.82)	0.81 (-0.36,2.02)
85	**0.69** (-0.46,1.84)	0.45 (-0.72,1.68)	-0.08 (-1.06,0.99)
86	**0.43** (-0.66,1.57)	0.09 (-0.95,1.20)	-1.01 (-1.97,-0.01)
87	**0.81** (-0.36,2.09)	0.48 (-0.57,1.59)	0.50 (-0.57,1.67)
88	**0.48** (-0.63,1.66)	-0.47 (-1.47,0.58)	-0.49 (-1.47,0.55)
89	-0.42 (-1.44,0.65)	-0.49 (-1.46,0.52)	**-0.32** (-1.36,0.82)
90	**0.83** (-0.31,2.11)	-0.47 (-1.45,0.60)	-0.02 (-1.10,1.19)

Table 9.2:Students abilities parameters

student #	ability parameter on factor 1	ability parameter on factor 2	ability parameter on factor 3
91	0.14 (-0.94,1.21)	-0.41 (-1.40,0.62)	**1.17** (-0.07,2.43)
92	**0.51** (-0.64,1.72)	0.42 (-0.69,1.57)	-0.98 (-1.95,-0.01)
93	**1.16** (-0.27,2.59)	0.78 (-0.43,2.02)	0.81 (-0.38,2.01)
94	**0.42** (-0.64,1.52)	0.31 (-0.81,1.48)	-0.61 (-1.61,0.44)
95	**0.72** (-0.37,1.89)	-0.53 (-1.54,0.54)	0.10 (-0.92,1.18)
96	**1.07** (-0.07,2.26)	0.93 (-0.25,2.10)	0.32 (-0.73,1.40)
97	**1.11** (-0.02,2.27)	0.21 (-0.87,1.41)	0.72 (-0.47,1.98)
98	-0.30 (-1.28,0.79)	-1.80 (-2.80,-0.73)	**-0.01** (-1.11,1.14)
99	-1.28 (-2.26,-0.30)	**-0.04** (-1.12,1.00)	-0.69 (-1.68,0.34)
100	0.39 (-0.69,1.52)	**0.68** (-0.46,1.85)	0.10 (-1.05,1.27)

References

Aitchison, J. and Bennett, J.A. (1970). Polychotomous Quantal Response by Maximum Indicant, *Biometrika*, 57, 253–262

Albert, J.H. (1992) Bayesian Estimation of Normal Ogive Item Response Functions using Gibbs Sampling, *Journal of Educational Statistics*, 17, 251–269

Albert, J.H. and Chib, S. (1993). Bayesian Analysis of Binary and Polychotomous Response Data, *Journal of the American Statistical Association*, 88, 669–679

Albert, J.H. and Ghosh, M. (2000). Item Response Modeling, *In Generalized Linear Models: A Bayesian Perspective.* (D.K. Dey, S. Ghosh and B. Mallick, eds.), Marcel-Dekker, New York, 173–193

Allenby, G.M. and Rossi, P.E. (1999). Marketing Models of Consumer Heterogeneity, *Journal of Econometrics*, 89, 57–78

Andersen E.B. (1997). The Rating Scale Model. In *Handbook of Modern Item response Theory* (eds var der Linden W. and Hambleton R.K), 67–84, Springer–Verlag, New York

Bartholomew, D.J. and Knott, M. (1999). Latent Variable Models and Factor Analysis, Oxford University Press Inc., New York

Béguin, A.A. and Glas, C.A.W. (2001). MCMC estimation and some model-fit analysis of multidimensional IRT models, *Psychometrika*, 66, 4, 541–561

Birnbaum, A. (1968). Some latent trait models. In *Statistical theories of mental test scores*, eds F.M. Lord and M.R. Novick. Reading, MA: Addison-Wesley

Bock, R.D. (1972). Estimating item parameters and latent ability when the responses are scored in two or more nominal categories, *Psychometrika*, 37, 29–51

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

Bock, R.D. and Lieberman, M. (1970). Fitting a response model for n dichotomously scored items, *Psychometrika*, 35, 179–197

Brooks, S.P. and Roberts, G.O. (1998). Diagnosing convergence of Markov chain Monte Carlo algorithms, *Statistics and computing*, 8, 319–335

Casella, G. and George E. (1992). Explaining the Gibbs Sampler, *American Statistician*, 46, 167–174

Chen, M.H. and Dey, D.K. (1998). Bayesian Modeling of Correlated Binary Responses via Scale Mixture of Multivariate Normal Link Functions, *Sankhyā, Series A*, 60, 322–343

Chen, M.H. and Dey, D.K. (2000). Bayesian Analysis and Computation for Correlated Ordinal Data Models, In *Generalized Linear Models: A Bayesian Perspective* (eds D.K. Dey, S. Ghosh and B. Mallick), 135–162, Marcel-Dekker, New York

Chib, S. (1995). Marginal Likelihood from the Gibbs Output, *Journal of the American Statistical Association*, 90, 1313–1321

Chib, S. and Greenberg, E. (1995). Understanding the Metropolis-Hastings, *American Statistician*, 49, 327–335

Chib, S. and Greenberg, E. (1998). Analysis of Multivariate Probit Models, *Biometrika*, 85, 347–361

Cowles, M.K. and Carlin, B.P. (1994). Markov Chain Monte Carlo Convergence Diagnostics: A Comparative Review, *Journal of American Statistical Association,* 91, 883–904

Dawid, A.P. (1988). The infinite regress and its conjugate analysis. In *Bayesian Statistics 3* (eds J.M. Bernardo, M.H. De Groot, D.V. Lindley and A.F.M. Smith), 95–110, Oxford: Clarendon Press

De Groot, M.E. (1970). Optimal Statistical Decisions. McGraw-Hill, New York

De Leeuw, J. and Verhelst, N.D. (1986). Maximum-likelihood estimation in generalized Rasch models, *Journal of Educational Statistics*, 11, 183–196

Dellaportas, P. (1998). Bayesian classification of neolithic tools, *Applied Statistics*, 47, 279–297

Dempster, A.P. (1969). *Elements of Continuous Multivariate Analysis.* Reading: Massachusetts, Addison Wesley

Devroye, L. (1986). *Non–Uniform Random Variate Generation.* New York: Springer–Verlag

Eagle, T. (1984). Parameter Stability in Disaggregate Retail Choice Models: Experimental Evidence, *Journal of Retailing*, 60, 101–123

Embretson, S.E., and Reise, S.P. (2000). *Item Response Theory for Psychologists,* Mahwah, NJ: Lawrence Erlbaum Associates

Ferguson, G.A. (1942). Item Selection by the Constant Process. *Psychometrika*, 7, 19–29

Fraser, C. (1988). NOHARM: A Computer Program for Fitting Both Unidimensional and Multidimensional Normal Ogive Models of Latent Trait Theory, NSW: University of New England

Gelfand, and Dey, D.K. (1994). Bayesian Model Choice: Asymptotics and Exact Calculations, *Journal of the Royal Statistical Society, Series B*, 56, 501–514

Gelfand, A.E. Hills, S.E. Racine-Poon, A. Smith, A.F.M. (1990). Bayesian Inference in Data Models Using Gibbs Sampling, *Journal of the American Statistical Association*, 85, 412, 972–985

Gelman, A. (1995). Inference and monitoring convergence, in Markov Chain Monte Carlo in Practice, Gilks W R, Richardson S and Spiegelhalter D J, pp. 131–143, London: Chapman and Hall

Gelman, A., Carlin, J.B., Stern, H.S. and Rubin, D.B. (1995). *Bayesian Data Analysis*, Chapman and Hall

Geman, S. and Geman, D. (1984). Stochastic Relaxation, Gibbs Distributions and the Bayesian Restoration of Images, *IEEE Transactions on Pattern Analysis and Machine*

Intelligence, 6, 721–741

Georgesku–Roegen, N. (1958). Threshold in Choice and the Theory of Demand, *Econometrica*, 26, 157–168

Geweke, J., Keane, M. and Runkle, D. (1994). Alternative Computational Approaches to Inference in the Multinomial Probit Model, *The Review of Economics and Statistics*, 76, 609–632

Green, P.J. (1995). Reversible Jump Markov Chain Monte Carlo computation and Bayesian Model Determination, *Biometrika*, 82, 711–732

Hajivassiliou, V. and McFadden, D. (1990). The method of simulated scores for the estimation of LDV models with an application to external debt crises, Working paper, Yale University

Hajivassiliou, V. and Ruud, P. (1994). Classical Estimation Methods for LDV Models Using Simulation. In *Handbook of Econometrics* (eds T.C. Eagle and D. McFadden) 2384–2438, Amsterdam: North Holland

Hensher, D.A. and Johnson, L.W (1979). *Applied Discrete Choice Modelling.* Halsted Press

Huber, J. and Moore, W. (1980). A Comparison of Alternative Ways to Aggregate Individual Conjoint Analysis. AMA Educators' Proceedings

Johnson, M. (1987). *Multivariate Statistical Simulation.* Wiley, New York

Johnson, R.A. and Wichern, (1988). *Applied Multivariate Statistical Analysis.* Prentice Hall, New Jersey

Krishnan, K.S. (1977). Incorporating Thresholds of Indifference in Probabilistic Choice Models, *Management Science*, 23 , 1224–1233

Lawley, D.N. (1943). On problems connected with item selection and test construction, *Proceedings of the Royal Society of Edinburgh*, 61, 273–287

Linardakis, M. (1997). A Contribution to the Bayesian Analysis of Discrete Choice Models, *M.Sc. Thesis submitted to the Department of Statistics, Athens University of Economics and Business, ISBN:960–7929–07–1*

Linardakis, M. (2001). Statistical Analysis of lexicographic data in discrete choice models, *Project funded by Athens University of Economics and Business, no E831*

Linardakis, M. and Dellaportas, P. (1998). "How much does your travel time cost"; A Bayesian Evaluation , *Proceedings of HERCMA Conference on Computer Mathematics and its Applications,* , 604–611

Linardakis, M. and Dellaportas, P. (1999). Latent variables in the multinomial probit model; A Bayesian approach, (in greek) *paper presented at the 12th Hellenic Statistical Conference, Spetses, Proceedings*

Linardakis, M. and Dellaportas, P. (2000). An Approach to Multidimensional Item Response Modelling, *in Bayesian methods with applications to science, policy and official statistics, selected papers from ISBA 2000,* 331–340

Linardakis, M. and Dellaportas, P. (2003). Assessment of Athens's metro passenger behaviour via a multiranked probit model, *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 52 (2), 185–200

Lioukas, S.K. (1984). Thresholds and Transitivity in Stochastic Consumer Choice: a Multinomial Logit Analysis, *Management Science*, 30, 110–122

Louviere, J.J. (1984). Using Discrete Choice Experiments and Multinomial Logit Choice Models to Forecast Trial in a Competitive Retail Environment: A Fast Food Restaurant Illustration, *Journal of Retailing*, 60, 81–107

Louviere, J.J. and Woodworth, G. (1983). Design and Analysis of Simulated Consumer Choice or Allocation Experiments: An Approach Based on Aggregate Data, *Journal of Marketing Research*, 20, 350–367

Luce, R. (1959). *Individual Choice Behavior: A Theoretical Analysis.* New York: Wiley

Maddala, G. S. (1983). *Limited–Dependent and Qualitative variables in Econometrics.* Cambridge University Press

Masters G.N and Wright B.D. (1997). The Partial Credit Model. In *Handbook of Modern Item response Theory* (eds var der Linden W. and Hambleton R.K), 101–121, Springer–Verlag, New York

McCullogh, R. and Rossi, P.E. (1994). An Exact Likelihood Analysis of the Multinomial Probit Model, *Journal of Econometrics*, 64, 207–240

McCullogh, R., Polson, N. and Rossi, P.E. (2000). A Bayesian Analysis of the Multivariate Probit Model with Fully Identified Parameters, *Journal of Econometrics*, 99, 173–193

McFadden, D. (1973). Conditional Analysis of Qualitative Choice Behavior, in *Frontiers in Econometrics, Zarembka P (editor).* New York, Academic Press

McFadden, D. (1978). Modelling the Choice of Residential Location, In *Karquist A et al (editors)*, Spatial Interaction Theory and Planning Models. Amsterdam: North-Holland Publishing Company

McFadden, D. (1980). Econometric Models of Probabilistic Choice, In *Manski C F and McFadden D (editors)*, Structural Analysis of Discrete Data: with Econometric Applications. MIT Press, Cambridge, Massachusetts

McFadden, D. (1989). A Method of Simulated Moments for Estimation of Discrete Response Models without Numerical Integration, *Econometrica*, 57 (5), 995–1026

McFadden, D. and Ruud, P.A. (1994). Estimation by Simulation, *The Review of Economics and Statistics*, 4, 591–608

Moore, W.L. (1980). Levels of Aggregation in Conjoint Analysis: An Empirical Comparison, *Journal of Marketing Research*, 17, 516–523

Nobile, A. (1998). A Hybrid Markov Chain for the Bayesian Analysis of the Multinomial Probit Model, *Statistics and Computing*, 8, 229–242

Nobile, A. (2000). Comment: Bayesian Multinomial Probit Models with a Normalization Constraint, *Journal of Econometrics*, **99**, 335–345

Raftery, A.E. and Lewis, S.M. (1992). How many iterations in the Gibbs Sampler?, in Bayesian Statistics 4, Bernardo J M, Berger J O, Dawid A P and Smith A F M (editors). Oxford University Press, pp. 763–773

Rasch, G. (1960). Probabilistic models for some intelligence and attainment tests, *Copenhagen, Denmark: Danish Institute for Educational Research*

Rasch, G. (1961). On General Laws and the Meaning of the Measurement in Psychology. Vol. 4, *Proceedings of the 4th Berkley Symposium on Mathematical Statistics,* Berkley: University of California Press, 321–334

Richardson, M.W. (1936). The relationship between difficulty and the differential validity of a test, *Psychometrika*, 1, 33–49

Robert, C.P. (1995). Simulation of Truncated Normal Variables, *Statistics and Computing*, 5, 121–125

Sahu, S.K. (2002). Bayesian Estimation and Model Choice in Item Response Models, *Journal of Statistical Computation and Simulation*, 72, 217–232

Smith, A.F.M. and Roberts, G.O. (1993). Bayesian Computation via the Gibbs Sampler and Related Markov Chain Monte Carlo Methods, *Journal of Royal Statistical Society, Series B*, 55, 3–25

Spanos, I., Deloukas, A. and Anastassaki, A. (1997). A Stated Choice Experiment: Value of Travel Characteristics in the Context of Attica. In *IFAC Transportation Systems* (eds Papageorgiou M. and A. Pouliezos), 427–433, Oxford: Elsevier Science Ltd

Swaminathan, H. and Gifford, J.A. (1982). Bayesian estimation in the Rasch model, *Journal of Educational Statistics*, 7, 175–192

Swaminathan, H. and Gifford, J.A. (1985). Bayesian estimation in the two-parameter logistic model, *Psychometrika*, 51, 589–601

Swaminathan, H. and Gifford, J.A. (1986). Bayesian estimation in the three-parameter logistic model, *Psychometrika*, 50, 349–364

Thissen, D. (1982). Marginal maximum-likelihood estimation for the one-parameter logistic model, *Psychometrika*, 47, 175–186

Thissen, D. and Steinberg L. (1984). A Response model for multiple choice items, *Psychometrika*, 49, 501–519

Thurstone, L. (1927). A Law of Comparative Judgement, *Psychological Review*, 34, 273–286

Train, K. (1986). Qualitative Choice Analysis, *Theory Econometrics and Application to Automobile Demand*, MIT Press

Tsutakawa, R.K. (1992). Prior distribution for item response curves, *British Journal of Mathematics and Statistical Psychology*, 45, 51–74

Tsutakawa, R.K. and Johnson, J.C. (1990). The Effect of Uncertainty of Item Parameter Estimation on Ability Estimates, *Psychometrica*, 55, 371–390

Tsutakawa, R.K. and Lin, H.Y. (1986). Bayesian estimation of item response curves, *Psychometrika*, 51, 251–267

Tsutakawa, R.K. and Soltys, M.J. (1988). Approximation for Bayesian Ability Estimation, *Journal of Educational Statistics*, 13, 117–130

van der Linden, W.J. and Hambleton, R.K. (1997). Item response theory: brief history, common models, and extensions, in *Handbook of Modern Item Response Theory, van der Linden, W.J. and Hambleton, R.K. editors*, Springer-Verlag New York Inc

Verhelst, N.D. and Molenaar, W. (1988). Logit-based parameter estimation in the Rasch model, *Statistica Neerlandica*, 42, 273–295

Wardman, M. (1988). A Comparison of Revealed Preference and Stated Preference Models of Travel Behaviour, *Journal of Transportation Economics and Policy*, 71–91

Zimmerman, D.W. and Williams R.H. (2003). A new look at the influence of guessing on the reliability of multiple choice tests, *Applied Psychological Measurement*, 27, 357–371