ATHENS UNIVERSITY
OF ECONOMICS AND BUSINESS

DEPARTMENT OF STATISTICS

POSTGRADUATE PROGRAM

Bayesian Factor Analysis: Implementation on Schizotypal Personality Disorder Data

By

Aggeliki Styl. Karatza

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

Athens, Greece
May 2006

ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ

Παραγοντική Ανάλυση κατά Bayes: Εφαρμογή σε δεδομένα Σχιζοτυπικής Διαταραχής της Προσωπικότητας

Αγγελική Στυλ. Καρατζά

ΔΙΑΤΡΙΒΗ

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

Αθήνα
Μάιος 2006

ACKNOWLEDGEMENTS

I would like to thank all my teachers who contributed to my education and especially my professors at University of Piraeus and my professors of ‘Master in Statistics’ of Athens University of Economics and Business.

I am most grateful to my supervisor, Assistant Professor Ioannis Ntzoufras, who introduced me in the interesting world of Bayesian theory and factor analysis with trust and patience.

I would also like to thank my family and friends who all these years stand by me and support me by all means.

VITA

I was born and raised in Athens, where I finished the 52nd high school. In 1996 I entered the University of Piraeus, at the Department of Statistics and Insurance. In 2003 I was accepted at the Master program of Statistics in Athens University of Economics and Business.

ABSTRACT

Aggeliki Karatza

BAYESIAN FACTOR ANALYSIS: IMPLEMENTATION ON SCHIZOTYPAL PERSONALITY DISORDER DATA

May 2006

The aim of this thesis is to reveal the latent factorial structure of schizotypal personality disorder under a set of observed schizotypal traits. We facilitate the Bayesian approach, while the classical-frequentist methodology is also implemented in an effort to compare the two approaches.

In the Bayesian approach, we combine prior information of the unknown parameters of the factor model and the data likelihood to construct the posterior distribution of the parameters. The inference is based on this posterior distribution and the corresponding descriptive measures (means or other moments). When the posterior distribution is not analytically tractable then Markov chain Monte Carlo (MCMC) methods are used to get samples from the corresponding posterior distributions.

The problem of identification is examined thoroughly as well as the general model of classical factor analysis.

Several exploratory and confirmatory factor models were used in order to examine the latent structure of the data. The aim of the analysis is to reveal the hidden dimensions of Schizotypal Personality Disorder (a disorder directly related to schizophrenia). A five-factor model was revealed through classical (non-Bayesian) exploratory factor analysis while the Bayesian analysis revealed a four factor model. Moreover, confirmatory analysis of the

schizotypic data ended in the paranoid 4-factor model (Stefanis et al., 2004) through classical analysis, while Bayesian analysis selected the Fogelson et al. (1999) 5-factor model through several information criteria.

ΠΕΡΙΛΗΨΗ

Αγγελική Καρατζά

ΠΑΡΑΓΟΝΤΙΚΗ ΑΝΑΛΥΣΗ ΚΑΤΑ ΜΠΕΫΕΣ ΣΕ ΔΕΔΟΜΕΝΑ ΣΧΙΖΟΤΥΠΙΑΣ

Μάιος 2006

Ο στόχος αυτή της διατριβής είναι να αποκαλυφθεί η λανθάνουσα παραγοντική δομή της σχιζοτυπικής διαταραχής της προσωπικότητας που υπάρχει κάτω από το σύνολο των μετρήσιμων γνωρισμάτων της σχιζοτυπίας. Η Μπεϋζιανή παραγοντική ανάλυση χρησιμοποιήθηκε για αυτό το λόγο, ενώ η κλασική μεθοδολογία εφαρμόστηκε επίσης σε μια προσπάθεια να συγκριθούν οι δύο προσεγγίσεις.

Η Μπεϋζιανή προσέγγιση θέτει εκ των προτέρων πληροφορία στις άγνωστες παραμέτρους του παραγοντικού υποδείγματος (οι συντελεστές, ο πίνακας διακύμανσης – συνδιακύμανσης των σφαλμάτων, ο πίνακας συσχετίσεων των παραγόντων) και μέσω των μεθόδων MCMC και ειδικότερα της δειγματοληψίας Gibbs μας παρέχει τις εκ των υστέρων κατανομές των παραμέτρων. Τα συμπεράσματα βασίζονται στους μέσους (ή σε άλλες παραμέτρους) των εκ των υστέρων κατανομών.

Το πρόβλημα της ταυτοποίησης μελετάται όπως επίσης το γενικό υπόδειγμα της κλασικής παραγοντικής ανάλυσης.

Στην εργασία αυτή ο σκοπός της ανάλυσης είναι να αποκαλύψει τις λανθάνουσες διαστάσεις της σχιζοτυπικής διαταραχής της προσωπικότητας που είναι μια διαταραχή άμεσα συνδεδεμένη με τη σχιζοφρένεια. Ένα υπόδειγμα πέντε παραγόντων αποκαλύφθηκε μέσω της κλασικής (μη-Μπεϋζιανής) διερευνητικής παραγοντικής ανάλυσης ενώ η Μπεϋζιανή ανάλυση αποκάλυψε ένα υπόδειγμα τεσσάρων παραγόντων. Επιπρόσθετα, η επιβεβαιωτική ανάλυση των δεδομένων σχιζοτυπίας κατάληξε στο υπόδειγμα παράνοιας τεσσάρων παραγόντων (Stefanis et al., 2004) μέσω της κλασικής

ανάλυσης ενώ η Μπεϋζιανή ανάλυση επέλεξε το υπόδειγμα 5 παραγόντων των Fogelson et al.(1999) μέσω κάποιων κριτηρίων πληροφορίας.

TABLE OF CONTENTS

1. Introduction 1

1.1 Introduction 1

1.2 Different aspects of factor analysis 2

1.3 Bayesian approach in factor analysis 3

1.4 Structure of the thesis 4

2. Factor analysis: a theoretical framework 5

2.1 Introduction 5

2.2 The factor analysis model 6

2.2.1 Introducing the model 6

2.2.2 The orthogonal model 8

2.2.3 Exploratory factor analysis (EFA) 8

2.3 Further notions 8

2.4 Confirmatory factor analysis (CFA) 10

2.5 Estimation of model parameters 11

2.5.1 Preceding of maximum likelihood methods 11

2.5.2 The method of factor analysis 14

2.5.2.1 Heywood cases 17

2.6 Goodness of fit statistics –Model selection measures 18

2.7 The problem of identification in factor analysis 20

2.7.1 Identification in EFA 21

2.7.2 Identification in CFA 22

2.7.3 Local identification 23

2.8 Rotation 25

2.9 Conclusion 26

3. Bayesian factor analysis 29

3.1 Introduction 29

3.2 Bayesian theory 29

3.3 Prior distributions 30
3.4 MCMC methods.. 32
3.5 Overrelaxation 35
3.6 Diagnosing convergence 36
3.7 Model selection 40
3.8 Bayesian approaches to factor analysis 41
3.9 Singular value decomposition 46
3.10 Comparison of Bayesian and non-Bayesian factor analysis 47

4. Application of exploratory factor analysis in schizotypic data 49

4.1 Introduction 49
4.2 Exploratory analysis – Application to schizotypic data 53
4.2.1 Classical Analysis 53
4.2.2 Bayesian analysis 55
4.3 Comparison between frequentistic and Bayesian analyses... 60

5. Application of confirmatory factor analysis in schizotypic data 61

5.1 Introduction 61
5.2 One factor confirmatory model 62
5.3 Kendler's two-factor model 63
5.4 Disorganised three-factor model 65
5.5 Paranoid four-factor model 66
5.6 Fogelson et al. five-factor model 67
5.7 Classical analysis and interpretation of best fitted model 68
5.8 Bayesian analysis 72
5.8.1 Priors 72
5.8.2 Gibbs sampling 72
5.8.3 Results 73
5.9 Comparison between frequentistic and Bayesian confirmatory analyses of data 76

6. Discussion and further research **77**

6.1 Discussion 77

6.2 Further research 78

6.2.1 Two stage factor analysis 78

6.2.2 Logit factor model 79

Appendix A **83**

Items for the nine subscales in the final 74-item version of the Schizotypal Personality Questionnaire ……………………………………………….. 83

Appendix B **89**

1. Loadings of exploratory factor analysis models…...............……..... 87
2. Factor loadings and factor correlation matrices of confirmatory factor analysis models in classical and Bayesian analysis ....….... 94
3. The code of the 4-factor model in Bayesian exploratory factor analysis ..101
4. The code of the 4-factor model in Bayesian confirmatory factor analysis..…....................…...............102

LIST OF TABLES

4.1 Chi-square values of EFA 54

4.2 Unrotated loadings of the four-factor model (Classical EFA).......................... 54

4.3 Information Criteria for Bayesian EFA 56

4.4 Posterior descriptive measure for Bayesian EFA models................................. 57

4.5 Version of AIC and BIC with $D(\bar{\theta})$ and $\bar{D}$ for EFA 58

4.6 Posterior means of factor loadings of F_4 in Bayesian EFA 58

4.7 Transformed posterior means of factor loadings of F_4 in Bayesian EFA …….. 59

5.1 Table of fitted factor models 62

5.2 Goodness of fit statistics for the fitted models 69

5.3 Factor loadings of paranoid 4-factor model 71

5.4 Correlation matrix of paranoid 4-factor model 71

5.5 Information Criteria for the five fitted models 73

5.6 Versions of AIC and BIC with $D(\bar{\theta})$ and $\bar{D}$ for CFA 73

5.7 Point estimates of deviance for Bayesian CFA 74

5.8 Factor loadings of Fogelson et al. 5-factor model with correlated factors 75

5.9 Covariance matrix of Fogelson et al. 5- factor model 74

B.1 Factor loadings and unique variance of F_1 exploratory model with classical and Bayesian analysis 89

B.2 Factor loadings and unique variance of F_2 exploratory model with classical and Bayesian analysis 90

B.3 Factor loadings and unique variance of F_3 exploratory model with classical and Bayesian analysis 91

B.4 Unrotated factor loadings and unique variance of F_4 exploratory model with classical analysis 92
B.5 Factor loadings and unique variance of F_4 in Bayesian EFA and MCMC details 92
B.6 Unrotated factor loadings and unique variance of F_5 exploratory model with classical analysis 93
B.7 Factor loadings and unique variance of F_5 in Bayesian EFA 93
B.8 Factor loadings of m_1 confirmatory model with classical and Bayesian analysis 94
B.9 Factor loadings of m_2 confirmatory model with classical and Bayesian analysis 94
B.10 Correlation matrix of latent factors of m_2 confirmatory model with classical analysis 95
B.11 Covariance matrix of latent factors of m_2 confirmatory model with Bayesian analysis 95
B.12 Factor loadings of m_3 confirmatory model with classical analysis 95
B.13 Correlation matrix of latent factors of m_3 confirmatory model with classical analysis 96
B.14 Factor loadings of m_3 confirmatory model with Bayesian analysis 96
B.15 Covariance matrix of latent factors of m_3 confirmatory model with Bayesian analysis 96
B.16 Factor loadings of paranoid 4-factor model with classical analysis 97
B.17 Correlation matrix of paranoid 4-factor model with classical analysis 97
B.18 Factor loadings of m_4 confirmatory model with Bayesian analysis 98
B.19 Covariance matrix of latent factors of m_4 confirmatory model with Bayesian analysis 98

B.20 Factor loadings of m_5 confirmatory model with classical analysis ....….. 99

B.21 Correlation matrix of latent factors of m_5 confirmatory model with classical **a**nalysis ....… 99

B.22 Factor loadings of Fogelson et al. 5-factor model with Bayesian analysis …. 100

B.23 Covariance matrix of Fogelson et al. 5- factor model with Bayesian analysis.100

LIST OF FIGURES

5.1 Path diagram for m_1 confirmatory factor model 63

5.2 Path diagram of Kendler's 2-factor model 64

5.3 Path diagram of disorganized 3-factor confirmatory model 65

5.4 Path diagram of paranoid 4-factor model 66

5.5 Path diagram of Fogelson et al. 5-factor model 68

5.6 Path diagram for fitted m_4 70

6.1 A path diagram for a two-stage factor analysis model 79

CHAPTER 1

INTRODUCTION

1.1 Introduction

The aim of this dissertation is to explore the structure of schizotypy through Bayesian factor analysis (BFA). According to DSM-III-R (American Psychiatric Association, 1987) nine symptoms reflecting cognitive, perceptual, social, interpersonal and behavioral dysfunction define the Schizotypal Personality Disorder (SPD). This disorder is considered to be genetically related to schizophrenia (Kendler et al., 1981; Kety, 1983; Bergman et al., 1996). Moreover, SPD can be examined in non-clinical populations, as well as clinical. These two facts have revealed the study of the factorial structure of SPD as an important area of research for many scientists. Among others, Raine (see Raine 1991, Raine et al., 1994) has demonstrated a significant contribution to the research in this area by constructing a 74- item self administered questionnaire, named Schizotypal Personality Questionnaire (SPQ). At the present thesis, we used the Greek version of SPQ constructed by the team of ASPIS (see Stefanis et al., 2004). The subjects participated in this study are students of Greek Technological Education Institutes and Universities. More details for SPD, the SPQ and the collection of data can be found at section 4.1 (p.47).

BFA was the subject of study for several scientists since 1972 when Press (1972) firstly introduced a basic model of BFA. For many years, the work of Press and Shigemasu (1989) was the basis of BFA until Markov chain Monte Carlo (MCMC) methodology was introduced in the statistical literature (Geman and Geman, 1984) and were finally applied in factor analysis models (see for example Rowe, 2003). This thesis is mainly based on the important work of Rowe (1998, 2000a, 2000b, 2000c, 2001, 2003) who facilitated MCMC methodology and gave an alternative version of the BFA model (see Chapter 3, p.38-44).

The main contribution of this thesis is the implementation of Bayesian methodology on schizotypic data. The Bayesian approach was not used before for the analysis of the schizotypic data according to the author's knowledge. Chapters 4 and 5 deal with the application of Bayesian and non-Bayesian schizotypic data, in an effort to compare and contrast the two approaches.

1.2 Different aspects of factor analysis

Factor analysis can be divided in two different approaches: exploratory factor analysis (EFA) and confirmatory factor analysis (CFA). EFA is used in order to explore the data concerning the number of hidden dimensions (factors) and the way they relate to the observed variables. On the contrary, CFA deals with models that have specific assumptions concerning the number of the latent factors or/and the observed variables they are related. Therefore, CFA is implemented when we wish to confirm (using a set of data) a specific scientific hypothesis, which is expressed via a corresponding model.

Since 1904, when Spearman introduced the initial form of factor analysis, a lot of research has been done concerning different methods of factor analysis. The most frequently used method of estimation is maximum likelihood where convergence of the estimated parameters is achieved through an iterative algorithm. It is the basic method of classical analysis and is easily implemented by computer programs like LISREL (see Jöreskog and Sörbom, 1996) and SPSS (SPSS User's Guide).

The problem of identification is crucial in factor analysis models. If different estimates of the parameters of the model lead to the same value of the covariance matrices, the model is not identified (Bollen, 1989a, p. 239) Different methods have been proposed in literature, either in EFA or in CFA (Bollen, 1989a, p. 238-254 and Everitt, 1984, p.16-18) The most important approaches are presented in section 2.6 (p.17).

1.3 Bayesian approach in factor analysis

According to the Bayesian theory, the parameters of the model are assumed to be random variables having a specific prior distribution. Inference is made through the posterior distributions of the parameters, which is proportional to the likelihood of the data and the prior distribution of the parameters. The Bayesian methodology has several advantages.

The most important are:

- the posterior distribution can be sequentially updated by incorporating the new available data to the model as prior information (Carlin and Louis, 2000)
- the full distributional profile (posterior distribution) of a parameter can be easily provided using MCMC methods. In this way the whole posterior information regarding the parameter of interest is available (Congdon, 2001). Recently methods have been introduced also in the frequentistic analysis but still obtaining the whole distribution of the estimator is not a standard practice and it can be estimated under specific assumptions (for example normality).
- the improvement of the precision of the parameters of interest (in comparison to the estimates through classical analysis), since extra available information can be introduced through prior information (Congdon, 2001).
- common sense interpretation of confidence intervals. Confidence intervals computed using the classical approach either contain the true unknown quantity of interest or not. On the other hand, in the Bayesian approach, the statement that the probability (conditional on the observed data) that the unknown parameter is within the 95% confidence intervals is equal to 95% is valid (for details see also Carlin and Louis, 2000, p.36)

Many scientists have contributed to the development of BFA. An important change to the implementation of BFA has begun when MCMC methods and in particular Gibbs sampling and Metropolis-Hastings algorithms were developed. Nowadays, iterative MCMC methods can be easily implemented. In this dissertation, WinBUGS

1.4 (see Spiegelhalter et al., 2003) is used in order to implement the Gibbs sampling method and estimate parameters of EFA and CFA models.

1.4 Structure of the thesis

The chapter that follows deals with classical factor analysis. A review of alternative methods of frequentistic factor analysis is presented in detail. The Bayesian theory and its implementation to factor analysis is presented at the third chapter. The two approaches of FA are also compared and their advantages and disadvantages are recorded.

In the fourth chapter we present the schizotypic data and the analysis through (classical and Bayesian) factor analysis. EFA was applicated with the use of LISREL 8.52 (see Jöreskog and Sörbom, 1996) and WinBUGS 1.4. (see Spiegelhalter et al., 2003). Five models at each category were used in order to examine the fit and determine the number of factors needed in order to have an acceptable model.

The fifth chapter deals with the application of CFA to the same data set of nine schizotypal traits. Five factor models were fitted using the Bayesian and non-Bayesian approach. A comparison of the results of the two approaches is also provided.

Finally, at chapter six, concluding remarks, as well as points of possible further research are outlined.

CHAPTER 2

FACTOR ANALYSIS: A THEORETICAL FRAMEWORK

2.1 Introduction

Factor analysis (FA) deals with the problem of revealing hidden dimensions under a set of observables. The variables that are observed are termed as *manifest* or *indicators* (Bollen, 1989a, p.16) while the unobserved are called *latent variables* or (*latent) factors*. The procedure of factor analysis takes place through the decomposition of the covariance matrix of the observed variables in terms of unknown parameters and variables. Occasionally, standardized data and hence the correlation matrix are used instead of the original data and their corresponding covariance matrix. Moreover, two different types of factor analysis models can be distinguished in literature: the EFA and CFA.

Factor analysis is mainly used in two different situations:

- as a data reduction method and
- as a method of revealing the underlying structure of the data.

Scientists are frequently asked to handle large data sets (Bartholomew et al., 2002, p.145). Therefore, they use factor analysis to "reduce" the dimension of variables of data matrix, in terms of the number of variables. Alternatively, factor analysis is used to identify one or more latent variables that are responsible for correlations among the observed variables.

The theory of FA was firstly developed by Spearman (1904). His effort was to reveal an indicator or variable which measures the mental ability of a person. This factor could be used to explain the intercorrelations between the tests of mental ability. Factor analysis is a useful tool for sciences like psychology and marketing but it is also used in scientific fields like econometrics, sociology and biometrics (Kaplan, 2000).

EFA is used when no information is available concerning the latent variables. It is used as a tool to explore the underlying structure of the data. According to Tucker and MacCallum (1997, p.132-135), EFA is used by researchers at initial stages of analysis in order to explore the data and get a picture of the number of underlying latent factors, as well as, their correlation structure and their relation with manifest variables.

On the other hand, CFA is used when either preceding information, from previous analysis, is available or subjective hypothesis is made concerning the relations among factors or between factors and manifest variables. With CFA the analyst tests his prior hypothesis that should be based on initial conclusions of EFA.

A lot of research papers as well as reading textbooks have been written concerning factor analysis. This thesis considers as a basis books of: Harman (1976), Kim and Mueller (1978), Chatfield and Collins (1980), Everitt (1984), Bollen (1989a), Basilevsky (1994), Tucker and MacCallum (1997), Bartholomew and Knott (1999), Kaplan (2000) and Bartholomew et al. (2002).

2.2 The factor analysis model

2.2.1 Introducing the model

Let us assume p observed variables denoted by $\mathbf{y}_i^T = (y_{i1}, ..., y_{ip})$ for individual $i = 1, ..., n$, where n is the number of available observations. The observed variables are assumed to be centered around their means. Moreover, we assume $q\ (\leq p)$ latent factors and their respective factor scores that are denoted by $\mathbf{f}_i^T = (f_{i1}, ..., f_{iq})$, for individual i. These factors can either be called *common* factors, in case they influence more than one manifest variable or *unique,* in case they influence only one manifest variable (Tucker and MacCallum, 1997).

In the original formulation of factor analysis, the assumed relation between factors and manifest variables is linear. It is possible to use other types of association if information of non-linearity is available (Tucker and MacCallum, 1997). The linear model, for individual i takes the following form:

$$\mathbf{y}_i = \mathbf{\Lambda f}_i + \mathbf{e}_i \ , \qquad \text{for } i = 1,...,n \qquad (2.1)$$

where $\mathbf{\Lambda}$=(λ_{ij}) is a matrix of ($p \times q$) dimension called the "loading matrix" and $\mathbf{e}_i^T = (\mathbf{e}_{i1},...,\mathbf{e}_{ip})$ is the $(p \times 1)$ vector of errors for the i-th individual. Constraints must be imposed in order to obtain an appropriate scaling for the latent variables since they are not directly measurable and their notion is often obscure. So, either the variance of the latent variable or alternatively one loading of each column of the loading matrix are constrained to be equal to one. Using the latter approach (i.e. $\lambda_{ij} = 1$, where λ_{ij} are the elements of $\mathbf{\Lambda}$) the scale of the latent variables is assumed to be the same as the scale of the observable ones (for details and an example see Bollen, 1989a, p.239).

Additionally, we assume that $\mathbf{f}_i$ and $\mathbf{e}_i$ follow multivariate normal distributions with zero mean and variance-covariance matrices $\mathbf{\Phi}$ and $\mathbf{\Psi}$, respectively. Hence

$$\mathbf{f}_i \sim N_q(\mathbf{0}, \mathbf{\Phi})$$

$$\mathbf{e}_i \sim N_p(\mathbf{0}, \mathbf{\Psi}), \qquad \text{for } i = 1,...,n \qquad (2.2)$$

In the consequence, $\mathbf{y}_i$ follows a multivariate normal distribution with zero mean and variance-covariance matrix $\mathbf{\Sigma}$ given by:

$$\mathbf{\Sigma} = \mathbf{\Lambda \Phi \Lambda}^{\mathrm{T}} + \mathbf{\Psi} \ . \qquad (2.3)$$

Equations (2.1) - (2.3) compose the basic model of factor analysis. A further assumption is that $\mathbf{\Psi}$ is diagonal, since the errors are assumed to be uncorrelated. Furthermore, the vectors of $\mathbf{e}_i$ are assumed to be uncorrelated with each factor.

2.2.2 The orthogonal model

Generally, the covariance matrix of the factors, $\mathbf{\Phi}$, is considered to be the identity matrix so that the factors themselves are not correlated. In that case, equation (2.3) takes the form:

$$\mathbf{\Sigma} = \mathbf{\Lambda\Lambda}^{\mathrm{T}} + \mathbf{\Psi} \tag{2.4}$$

The factors then share an orthogonal structure and this model is called orthogonal. However, the hypothesis of uncorrelated factors is not a realistic assumption since, in practice, latent characteristics are often correlated. Therefore, in many situations, it is more realistic to drop this assumption and use a model with correlated factors.

2.2.3 Exploratory factor analysis (EFA)

In this section EFA`s main characteristics are summarized. In EFA, there is no available information concerning the relations between the latent variables or the latent and the indicators. In addition, the number of factors is not prespecified. A third constraint is set by assuming that the errors are uncorrelated among themselves. As a consequence, the analysis starts by assuming a minimum number of factors, usually one. Then a measure of goodness of fit is calculated and if the fit is not satisfactory the analysis is made on a two-factor basis, etc.

2.3 Further notions

The model formulation of equations (2.1) - (2.3) is given in matrix notation. Alternatively, (2.1) can be written as:

$$y_{ij} = \lambda_{j1} f_{i1} + \lambda_{j2} f_{i2} + ... + \lambda_{jq} f_{iq} + e_{ij}, \quad \text{for } j = 1,...,p \text{ and } i = 1,...,n \tag{2.5}$$

Equation (2.5) aids to import further useful notions in the factor analysis content such as communality, complexity, model fit. In more detail:

- **Communality.**

The communality h_j^2 of a variable y_j, for the orthogonal factor models, is given by the sum of the squares of the common factor coefficients:

$$h_j^2 = \sum_{k=1}^{q} \lambda_{jk}^2, \qquad \text{for } j = 1,...,p \tag{2.6}$$

It is a measure of the variance of j^{th} variable accounted for by the common factors.

- **Complexity.**

As complexity we define the number of common factors influencing an observed variable. Hence, complexity in (2.6) stands for q.

- **Unique variance.**

The term unique refers to the variance of the unique factors, which usually are the error terms.

- **Factor pattern- factor structure matrices.**

The factor pattern matrix contains the coefficients of the common factors that are obtained after a factor analysis, that is the loading matrix $\mathbf{\Lambda}$, while the factor structure matrix, contains the correlations between factors and variables, that is $\mathbf{C}ov(\mathbf{f},\mathbf{y})$. The two matrices coincide in case of orthogonal factors. (Kim and Mueller, 1978a, p.77 and Kim and Mueller, 1978b, p. 84).

- **Reliability.**

The notion of reliability is used in specific parts of the following section (see section 2.7.2, p.20).

Reliability, according to Bollen (1989a, p.206), can be defined as the consistency of one measurement. It can be measured using the squared correlation coefficient between an indicator and a latent variable.

- **Model fit.**

An important aspect for each model adopted is the evaluation of its fit to the data; see for details in Tucker and Mac Callum (1997, p.142). According to them, two types of error exist in each model, the model error and the sampling error. Although in practice these two types cannot be distinguished, they must be carefully controlled. Careful choice of indicators can reduce model error. All necessary variables must be measured and considered while unnecessary variables must be excluded from the model. Furthermore, Tucker and MacCallum (1997, p.132-143) suggest the following solutions for the reduction of sampling errors are the following:

1. increase of the sample size
2. elimination of variables with high unique variances and
3. analysis of covariance rather than correlation matrices.

2.4 Confirmatory factor analysis (CFA)

Confirmatory factor analysis needs more explicit and detailed information than exploratory analysis. Concrete assumptions are made concerning the number of factors and their correlation structure. For example, we may have information from previous studies or from scientific theories or scenarios concerning which variable loads on which factor. In this case we can restrict specific loadings to be equal to zero. In addition, we can insert available information related to the relationship between factors. In CFA, the analyst has usually information available from previous studies, so he can also make specific assumptions concerning the value of factor loadings. In some cases CFA follows EFA by eliminating loadings with low values.

The orientation is quite different in CFA than in EFA, since CFA is a model validation method. As a consequence the fit of the model will be poorer since some parameters will be fixed or eliminated. What we lose in the fit of the model, we gain it as an increase in the degrees of freedom and therefore in favor of the parsimony principle (Kim and Mueller, 1978b, p.58).

2.5 Estimation of model parameters

The unknown parameters in the model of factor analysis are the elements of the loading matrix $\mathbf{\Lambda}$ as well as the variance-covariance matrices $\mathbf{\Phi}$ and $\mathbf{\Psi}$. Various methods have been proposed for factor analysis parameters estimation. According to Harman (1976) they can be divided in two general categories. The first one includes methods that require estimating the communalities while the second includes methods that require estimating the number of factors.

2.5.1 Preceding of maximum likelihood methods

At the first category of methods Harman (1976) includes the *principal factor* method, the *centroid* method and the *triangular decomposition*. Although the *principal factor* method provides a unique mathematical solution, it is not totally accepted by psychologists. It was firstly proposed by Pearson (1901) and it was further investigated by Hotelling (1933). It resembles to the principal component analysis with the difference that, in principal factor method, the analyzed correlation matrix contains estimated communalities (reduced correlation matrix).

The *centroid* method takes its name by the geometric representation of the solution produced by this approach. It was developed by Thurstone (1935, 1947). It consists of a series of residual correlation matrices with a centroid factor, extracted from each residual matrix. According to Tucker and MacCallum (1997, p.201) the centroid vector is the mean vector through which a centroid axis is passed. The factor weights are the orthogonal projections of observed vectors on the centroid factor. The procedure stops when the residual correlations and the factor weights of the resulting factors take low values.

The triangular decomposition takes advantage of a method known as "square root method" that reduces any symmetric matrix, $\mathbf{R}$ here, to a triangular matrix $\mathbf{\Lambda}$, such that $\mathbf{R} = \mathbf{\Lambda}\ \mathbf{\Lambda}^T$.

The second category includes the *maximum likelihood* method, the *MINRES* method, the *psychometric* methods, the *multiple group* methods and some factor methods that were used in the early stages of factor analysis. The last methods are going to be referred firstly, starting with the method that the pioneer of factor analysis, Spearman, introduced in 1904 (Spearman, 1904). He used a two-factor model in order to explain the intercorrelations between p observed measures of mental ability in terms of a general factor g and a specific factor s. He proved that a set of p variables can be described in terms of one general factor and p unique factors if and only if all the following tetrads vanish, that is:

$$r_{jk}r_{lm} - r_{lk}r_{jm} = 0 \qquad \text{for } j,k,l,m = 1,2,...,p; j \neq k \neq l \neq m \tag{2.7}$$

where r_{ij} denotes the correlation between the variables Y_i and Y_j. Holzinger (1930) provided a generalization of Spearman's theory in case of more than one factor, by introducing the concept of grouping of variables (Newman et al., 1937). In this model, all indicators are linearly determined by a general factor, a group factor and a unique factor. The estimation method used in this case is called the *Bi-Factor* method, which is described by Harman (1976, p.120).

The psychometric methods involve *image factor analysis* and *alpha factor analysis*. They are based on the idea that the observed variables are a sample from an assumed "universe of content", that is an infinite universe of such measures. The *image theory* was firstly developed by Guttman (1953). In this approach variables under consideration are split in two parts, the image that is the part of the variable that can be written as a linear combination of all the other variables and the anti-image that cannot be predicted by a linear combination of the other variables. The procedure of this method is based on finding the eigenvalues of a matrix that combines the observed correlation and the anti-image variance.

Concerning the method of estimation called *alpha factor* (Kaiser and Caffrey, 1965) , the following correlation matrix is used:

$$\mathbf{R}^* = \mathbf{H}^{-1}(\mathbf{R} - \mathbf{\Psi})\mathbf{H}^{-1} \tag{2.8}$$

where $\mathbf{\Psi}$ is diagonal matrix of unique components, $\mathbf{H}^{-1}$ is a diagonal matrix that contains the reciprocals of the square roots of the communalities and $\mathbf{R}$ is the observed correlation matrix. The procedure starts with initial communalities and iterates by finding the eigen solution of matrix $\mathbf{R}^*$. Then the elements of the matrix $\mathbf{H}^{-1}$ are replaced with the estimated communalities and the algorithm starts again. The final solution is obtained when communalities at subsequent steps do not differ significantly.

By the *multiple group* methods we obtain dependent factors. Such solutions are going to be discussed in a subsequent subsection that deals with the notion of “rotation”.

MINRES method stands for “minimum residuals” method. It tries to find a solution to the factor problem by minimizing the sum of squares of residuals between observed and reproduced correlations (the ones implied by the model). Eckart and Young (1936, p.211) firstly approached theoretically the method (see also Harman 1976, p.175) followed by Young and Householder (1938) and Horst (1937). The first practical implementation was provided by Harman and Jones (1966).

MINRES method is a special case of a more general approach of least squares method. In this method, as Kim and Mueller (1978b, p.21) note, firstly the number of factors is determined. Then, after calculating some initial estimates of communalities, an eigenvalue solution of the observed correlation matrix is given. The following step is to calculate communalities based on the factor pattern of previous stage. The procedure keeps on iteratively of subsequent stages/iterations until no important difference is made between communalities.

2.5.2 The method of Maximum Likelihood

In this section we present in detail the approach of *maximum likelihood* (ML) method. This is the most frequently used method for estimation in factor analysis. Tucker and MacCallum (1997) support the method by noting that it "has many desirable statistical properties, such us consistency, normality, efficiency". Another advantage of this method is that it permits statistical testing of parameters.

The method of ML presents also several drawbacks. The normality assumption is an important constraint of the method, even though ML can be used even in cases when the data does not follow multivariate normal distribution (Bartholomew et al., 2002,p.151). ML is also an iterative method and as a consequence is more compute intensive (Chatfield and Collins, 1980). Finally, a Heywood case is a problem that frequently appears when ML is implemented. (Bartholomew et al., 2002, p.172, for details see section 2.5.2.1).

The ML method will be used at the present thesis in order to compare results between the frequentist's and the Bayesian approach. According to Basilevsky (1994, p.367) the ML approach is distinguished between the unrestricted and the restricted one. The unrestricted ML can only be used in case of random factors, that is in EFA. On the other hand, when the factors are fixed, in case of CFA, restricted ML solutions should be given.

Because of the fact that the n observations from the sample follow a multivariate normal distribution, the elements of the observed covariance matrix $\mathbf{S}$ follow a Wishart distribution with $(n-1)$ degrees of freedom. Consequently, the log likelihood function L is given by the next equation.

$$L = \ln(K) - \frac{1}{2}(n-1)\left(\ln|\mathbf{\Sigma}| + \frac{1}{2}(n-p-2)[\ln|\mathbf{S}|] - \frac{1}{2}(n-1)tr(\mathbf{\Sigma}^{-1}\mathbf{S}) \right) \qquad (2.9)$$

where K is a constant involving only n and p, and $\mathbf{\Sigma}$ is the covariance of the indicators given by the model.

In case of orthogonal factors the covariance matrix is given by (2.4). In that case L is a function of the factor loadings and the variances of the errors. So, estimates of these parameters are given through maximization of L. Jöreskog (1967) proposed an alternative function that its minimization is equivalent to maximization of L. This function is the following:

$$F = \ln|\mathbf{\Sigma}| + tr(\mathbf{\Sigma}^{-1}\mathbf{S}) - \ln|\mathbf{S}| - p \tag{2.10}$$

After differentiating the function F with respect to $\mathbf{\Lambda}$ and $\mathbf{\Sigma}$ and setting the results equal to a zero matrix, we end up with the following expressions:

$$(\mathbf{I} - \mathbf{S}\hat{\mathbf{\Sigma}}^{-1})\hat{\mathbf{\Lambda}} = \mathbf{0} \tag{2.11}$$

and

$$\hat{\mathbf{\Sigma}}^{-1} = diag(\mathbf{S} - \hat{\mathbf{\Lambda}}\hat{\mathbf{\Lambda}}^{\mathbf{T}}) \tag{2.12}$$

where $\hat{\mathbf{\Lambda}}$ and $\hat{\mathbf{\Sigma}}$ are the estimates of the loading matrix and the covariance matrix of the model respectively.

Then, the algorithm starts by imposing initial values for the unknown parameters. New estimates are generated at each iteration, which substitute the old ones. This algorithm iterates until low differences appear in subsequent iterations. The above maximization procedure is performed subject to the constraint that $\mathbf{\Lambda}^{\mathbf{T}}\mathbf{\Psi}^{-1}\mathbf{\Lambda}$ is diagonal, for identifiability reasons (for details see section 2.7.1).

It is proved (Basilevsky, 1994) that, under the assumption of normality, the factor scores are estimated, conditional on the parameters of the model, by the following equation:

$$\hat{\mathbf{f}}_i = (\mathbf{I}_q + \mathbf{\Lambda}^T\mathbf{\Psi}^{-1}\mathbf{\Lambda})^{-1}\mathbf{\Lambda}^T\mathbf{\Psi}^{-1}\mathbf{x}_i, \quad \text{for } i = 1,2,...,n. \tag{2.13}$$

An alternative way of obtaining ML estimates is EM algorithm (Rubin and Thayer, 1982). It consists of an Expectation (E-) step where the expected value of the log likelihood for the factor scores given the observed data (and initial estimates of $\mathbf{\Lambda}$ and $\mathbf{\Psi}$) are obtained and a Maximization (M-) step where the expected log likelihood found in E- step is maximized. The algorithm iterates until stable values of loadings and factor scores are obtained; more details can be found in Basilevsky (1994) and Rowe (1998).

According to Liu and Rubin (1998) at the E-step of the EM algorithm we compute the expected value of the statistics

$$\mathbf{\Sigma} = \frac{1}{n}\sum_{1}^{n}(\mathbf{Y_i} - \bar{\mathbf{Y}})^T(\mathbf{Y_i} - \bar{\mathbf{Y}}), \quad \Sigma_{\mathbf{yf}} = \frac{1}{n}\sum_{1}^{n}(\mathbf{Y_i} - \bar{\mathbf{Y}})^T\mathbf{f}_i, \quad \Sigma_{\mathbf{ff}} = \frac{1}{n}\sum_{1}^{n}\mathbf{f}_i^T\mathbf{f}_i \tag{2.14}$$

At the M-step, after replacing the above statistics with their expected values we compute the maximum likelihood estimates of the unknown parameters (the factor loadings $\mathbf{\Lambda}$ and $\mathbf{\Psi}$)

Problems in the estimation procedure often are observed when negative entries of $\mathbf{\Psi}$ appear in one iteration of the above algorithm or when one or more communalities exceed the value of one. The problem is known as **Heywood case** (for more details see section 2.5.2.1, p.17).The consequence of Heywood cases is that the algorithm stops, as variances are not permitted to take negative values.

ML approach also provides a goodness of fit test:

$$c = (n-1)\min F \tag{2.15}$$

where F is the function given by equation 2.10. The test is performed to test the null hypothesis

H_0: the covariance structure is restricted to $\mathbf{\Sigma} = \mathbf{\Lambda\Lambda^T} + \mathbf{\Psi}$

against the alternative

H_a: $\mathbf{\Sigma}$ is an arbitrary positive definite matrix

Under H_0 the statistic (2.14) follows a chi-square distribution with

$$\nu = \frac{1}{2}[(p-q)^2 - (p+q)] \tag{2.16}$$

degrees of freedom, where p and q are the number of observed and latent variables respectively. In the above equation the number of common factors is assumed to be known. This is not possible in practice. Therefore, we start from a minimum assumed value of q (usually 1) and we increase the number of factors by one until the fit of the model is not rejected by the corresponding significance test.

The chi-square test is based on the assumption that the null hypothesis is true which means that the model holds for the population. This is restrictive in the sense that in reality, such a mechanism may not exist. Hence, chi-square tests frequently may lead to the rejection of the assumed model. Another important point is that the chi-square statistic is heavily influenced by the sample size. Consequently, for large samples, it takes large values, leading again to the rejection of the assumed model even if it is a good approximation of reality. In addition, the values of the chi-square test decrease as the number of the parameters are added. This fact makes the chi-square test an unreliable statistic.

For this reason, in CFA a number of other measures of fit have been developed and used in literature. These measures take into account both the parsimony principle and fitness of the model and they are briefly presented in section 2.5.

2.5.2.1 Heywood cases

A problem that frequently occurs when the iterative method of ML is implemented is called *Heywood cases.* It appears when one or more communalities exceed the value of one (Bartholomew et al., 2002, p.172). The Heywood case corresponds to zero or negative values of the variance of some errors (Chatfield and Collins, 1980, p.87).

According to SAS/STAT/User's Guide (FACTOR procedure), the Heywood cases are possibly due to:

- the inappropriateness of the model,
- the small size of the sample and
- the large or small number of the factors.

Constraining the values of the variances of errors to exceed a positive "small" value ε, that is $\psi_j \geq \varepsilon$, solves the problem. According to

‘Information Technology Estimates’ (FAQ LISREL), different methods of solutions except for ML are proposed as well as the use of appropriate starting values of the algorithm. In addition, two other approaches are available in order to avoid such cases: the gradient method and a Newton-Raphson method. Further details can be found in Tucker and MacCallum (1997, p.266-282).

2.6 Goodness of fit statistics-Model selection measures

In this section we present various statistics and measures used for the selection of the number of factors in a model. Most of them are functions of chi-square statistic and the degrees of freedom of the model. They take into account the parsimony of the model (the number of the parameters) as well as its goodness of fit.

AIC (Akaike, 1974, 1987), CAIC (consistent AIC by Bozdogan, 1987) and the single sample cross-validation index ECVI (Browne and Cudeck, 1989) are provided by LISREL and are given by:

$$AIC = -2\log L + 2d_m \tag{2.17}$$

$$CAIC = c + (1 + \ln n)d_m \tag{2.18}$$

$$ECVI = \{c/(n-1)) + 2(t/(n-1)\} \tag{2.19}$$

where $\log L$ is the log likelihood (see section, 2.5.2), c is given by $c = (n-1)F$ the chi-square measure of overall fit of the model, n the sample size, d_m the number of free parameters. Small values of these measures show a better fit of the underlying models.

In addition, two alternative goodness of fit indices, which are used as a measure of fit between different models, are the goodness of fit index (GFI) and the adjusted GFI (AGFI) given by:

$$GFI = 1 - \frac{tr\left[\left(\hat{\mathbf{\Sigma}}^{-1}\mathbf{S} - \mathbf{I}\right)^2\right]}{tr\left[\left(\hat{\mathbf{\Sigma}}^{-1}\mathbf{S}\right)^2\right]} \tag{2.20}$$

and $$AGFI = 1 - \left[\frac{q(q+1)}{2df}\right]\left[1 - GFI\right] \tag{2.21}$$

where $\hat{\mathbf{\Sigma}}$ and $\mathbf{S}$ are the estimated by the model and observed covariance matrix respectively, $\mathbf{I}$ is the identity matrix and df are the degrees of freedom of the model. They do not depend on the sample size while AGFI is a variation of GFI adjusted for degrees of freedom. They take values between zero and one, with values close to one indicating perfect fit.

Another class of indices compares the fit of the model with respect to the independence model, which is the model that assumes that no underlying structure exists concerning the manifest variables. Some of them are the Normed Fit Index (NFI):

$$NFI = \frac{\hat{F}_0 - \hat{F}_{\min}}{\hat{F}_0} \tag{2.22}$$

Non-normed Fit Index (NNFI) (Tucker and Lewis, 1973, Bentler and Bonett, 1980):

$$NNFI = \frac{\hat{F}_0 / df_0 - \hat{F}_{\min} / df_{\min}}{\hat{F}_0 / df_0 - 1 / n-1} \tag{2.23}$$

Relative Fit Index (RFI) and Incremental Fit Index (IFI, Bollen, 1986, 1989a, 1989b) and Comparative Fit Index (CFI, Bentler, 1990) given by:

$$CFI = 1 - \frac{\max\left[(n-1) \times \hat{F}_{\min} - df_{\min}, 0\right]}{\max\left[(n-1) \times \hat{F}_0 - df_0, 0\right]} \tag{2.24}$$

where $\hat{F}_{\min}$ and $\hat{F}_0$ are the function values of the fitted and the independence models respectively and $df_{\min}$ and df_0 are the degrees of freedom for the fitted and the independence model respectively.

They take values between zero (0) and one (1). James, Mulaik and Brett (1982) suggest the Parsimony Normed Fit Index (PNFI) , Mulaik et al. (1989) suggest the Parsimony Goodness of Fit Index (PGFI) given by:

$$PNFI = \frac{df_{\min}}{df_0} \frac{(\hat{F}_0 - \hat{F}_{\min})}{\hat{F}_0} \tag{2.25}$$

$$PGFI = \frac{df_{\min}}{df_0} GFI \tag{2.26}$$

with $df_{\min}$ and df_0 as above. The latter indices take into account the parsimony of the models.

Browne and Cudeck (1993) proposed a number of fit measures, which take into account the error of approximation of the assumed model in the population. They define the population discrepancy function (PDF) as

$$\hat{F}_0 = \max\left(\hat{F} - \frac{df}{n-1}, 0 \right) \tag{2.27}$$

where $\hat{F}$ is the minimum value of the fit function, df is the degrees of freedom.

Steiger (1990) defines the Root Mean Square Error of Approximation (RMSEA)

$$\varepsilon = \sqrt{\hat{F}_0 / d} \tag{2.28}$$

as a measure of discrepancy per degree of freedom. Values of ε below 0.05 indicate a close fit while significant errors of approximation in the population are represented by values of ε up to 0.08 (Driscoll et al., 2005, Browne and Cudeck, 1993)

2.7 The problem of identification in factor analysis

The problem of identification is due to the fact that $\mathbf{f}$, $\mathbf{\Lambda}$ and $\mathbf{e}$ are unobserved. Let us assume that $\boldsymbol{\theta} = (\mathbf{f}, \mathbf{\Lambda}, \mathbf{e})$ is the vector of unknown parameters of the model, that is the vector containing the unknown factor loadings, the correlations between common factors and the variances of the errors of measurement.

Let us denote by $\mathbf{\Sigma}(\boldsymbol{\theta})$ the covariance produced by a set of parameters $\boldsymbol{\theta}$. Then, the model parameters $\boldsymbol{\theta}$ are identified if no vectors $\boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_2$ exist such that $\mathbf{\Sigma}(\boldsymbol{\theta}_1) = \mathbf{\Sigma}(\boldsymbol{\theta}_2)$ (Bollen, 1989a, p.239). Alternatively, if two different vectors of the unknown parameters lead to the same value for the covariance matrices, the model is not identified. When a model is not uniquely identified then different models can be constructed with the same fit and predictive ability but with different parameter values leading to different interpretation of the relations between variables. However, in order to develop a reliable and robust scientific theory, we need to conclude in a single model. This can be solved only if we efficiently deal the problem of identification.

2.7.1 Identification in EFA

In exploratory analysis, some constraints must be imposed on the parameters in order to produce a set of identifiable parameters. As it is mentioned by Everitt (1984), if we assume orthogonal structure for the latent factors ($\mathbf{\Phi} = \mathbf{I}$), then a well-known identifiability condition arises when we restrict $\mathbf{\Lambda}^{\mathrm{T}}\mathbf{\Psi}^{-1}\mathbf{\Lambda}$ to be diagonal (where $\mathbf{\Psi}$ is the variance covariance matrix of unique factors). The constraint of diagonal $\mathbf{\Lambda}^{\mathrm{T}}\mathbf{\Psi}^{-1}\mathbf{\Lambda}$ is equivalent to restricting the first factor to have the maximum contribution to the variance in the manifest variables, the second makes a maximum contribution, subject to being uncorrelated to the first and so on (Everitt, 1984, p.17). Therefore, the ordering of the factors also corresponds to the order of contribution to the observed variables.

Using this set of constraints we impose $\frac{1}{2}q(q-1)$ restrictions on the loadings (Kaplan, 2000, p.44-45). Therefore, the number of the free parameters in the factor analysis model (assuming $\mathbf{\Phi} = \mathbf{I}$) is equal to

$$p + pq - \frac{1}{2}q(q-1) \tag{2.29}$$

Lopes and West (2004) use the assumption of full rank of the loading matrix. Moreover, Geweke and Singleton (1980) showed that if the loading matrix has a rank lower than p, then the model is not identifiable. Lopes (2003) also, apart from other constraints, refers to orthogonal $\mathbf{\Lambda}$ matrix in order to have an identified model.

2.7.2 Identification in CFA

In CFA we can use available information concerning the structure of the loading matrix. Therefore, we may constraint some loadings to be equal to fixed values, usually zero, or set specific parameters equal; see Bollen (1989a, p.239). This helps the identification problem as the number of unknown free parameters is reduced.

Bollen (1989a, p.242-246) provides some necessary but not sufficient conditions required when an identified CFA model is estimated. In the 1-factor model with two indicators, Bollen uses the value of reliability of an observed variable (see section 2.3) to solve the problem of identification. He also, concludes that the only constraint that must be imposed in order to have a one-factor model with three indicators identified is to set one of the λ_{ij} equal to 1 (or its variance $\phi_{11} = 1$). If we impose more constraints on the loadings then the model becomes overidentified.

For more complicated models, even for the three-factor ones, the algebraic computations are too tedious, so other rules are established in order to produce identifiable models. As already mentioned, these rules are necessary but not sufficient to solve the identification problem and should be applied after setting the scale of the latent variables. The rules that follow are presented at Bollen (1989a, p.242):

1. The first rule, which is called the **t-rule**, requires the number of free parameters of the unknown parameters $\boldsymbol{\theta}$ to be lower or equal to the known elements in the covariance matrix of $\mathbf{y}$:

$$d_m \leq \frac{1}{2} q(q+1) \tag{2.30}$$

2. The **three-indicator rule** requires:

- for a multifactor model three or more indicators per latent variable,
- every observed variable loads on one and only one latent variable and

- uncorrelated errors.

In this approach, factors can be correlated. It is proved that the elements of $\mathbf{\Phi}$ are identified after scaling the latent variables.

3. The **two-indicator rule**, requires:

- uncorrelated errors
- the scaled latent variables
- every indicator loads only on one latent variable (factor complexity of one)
- two indicators per latent variable and
- no zero elements in $\mathbf{\Phi}$.

Bollen (1989a, p.245-246) has generalized this rule by loosening the requirements for $\mathbf{\Phi}$. Using this generalized rule we may constraint some elements of $\mathbf{\Phi}$ to be equal to zero, without impact concerning identification. The proof is based on 'blocking' the structure and applying the not-generalized rule only in the part of the model that has correlated factors. The same approach can be implemented for all the subsets of indicators.

An alternative set of constraints that provide identifiable parameters is given by Basilevsky (1994, p.361-363). According to Basilevsky, the only assumption that is made in order to have an identifiable model is the homoscedasticity of errors between factors, which does not seem realistic for real datasets.

2.7.3 Local identification

Bollen (1989a, p.246-254) provides a set of empirical tests for model identification. He distinguishes two types of identification: global and local. His empirical tests focus on the case of local identification.

Local identification concerns a specific point of the parameter vector, say $\boldsymbol{\theta}_1$, and determines whether the implied covariance matrix changes with small changes in $\boldsymbol{\theta}_1$. It is necessary but not sufficient condition for global identification. One test for local identification (called Wald's rank rule; Wald, 1950), facilitates a vector $\boldsymbol{\sigma}(\boldsymbol{\theta})$ containing the non-redundant elements of $\boldsymbol{\Sigma}(\boldsymbol{\theta})$. The $t \times 1$ vector of unknown, unconstrained parameters $\boldsymbol{\theta}$ is locally identified at a point $\boldsymbol{\theta} = \boldsymbol{\theta}_1$ if and only if the rank of $\partial \boldsymbol{\sigma}(\boldsymbol{\theta}) / \partial \boldsymbol{\theta}$ evaluated at $\boldsymbol{\theta}_1$ is equal to t.

Another local identification test facilitates the information matrix of $\boldsymbol{\theta}$. According to this test, the vector $\boldsymbol{\theta}$ is locally identified at some point $\boldsymbol{\theta}_1$ if and only if the inverse of the information matrix exists. This test has been also recommended by Keesling (1972), Wiley (1973), Jöreskog and Sörbom, (1986), and it can be easily implemented, since the inverse of the information matrix is given by statistical programs like LISREL and EQS; for details also see Bollen (1989a, p.246-254). Of course, local identification does not imply or ensures global identification. Moreover, the issue of local identification is more complicated, since the theoretical background of local identification deals with the parameters of the population, while the tests can be implemented on the available samples. So, an unidentified point θ_1 may incorrectly pass the test or the opposite. In this case, large standard errors of specific parameters estimates indicate possible unidentified parameters.

Another empirical approach proposed by Bollen (1989a, p.251), is based on using various starting values. If the model is identified then the model parameters should always converge at the same value. Alternatively, after running once the algorithm with the sample covariance matrix, we could repeat the analysis using the predicted covariance matrix $\hat{\boldsymbol{\Sigma}}$ given by:

$$\hat{\boldsymbol{\Sigma}} = \hat{\boldsymbol{\Lambda}} \hat{\boldsymbol{\Lambda}}^{T} + \hat{\boldsymbol{\Psi}}$$

where $\hat{\boldsymbol{\Lambda}}$ and $\hat{\boldsymbol{\Psi}}$ are the estimates produced by the first analysis. Identical estimates provide an indication that the model is identified.

2.8 Rotation

Another important issue in factor analysis is rotation. The methods analyzed in section 2.4 produce mainly orthogonal factor solutions. This solution, with the exception of one-factor model, is not unique, as a consequence of the constraints' implementation (for details see section 2.7). So, an infinite number of sets of factor loadings corresponds to the same model, with the number of factors as well as the communalities unchanged. It is important that the obtained factor loadings can be easily interpreted. Therefore, if the produced loadings imply a complex structure of relations between factors and indicators, an alternative simpler and easier to interpret solution can be considered.

The process of moving from one solution to another is called rotation (Bartholomew et al., 2002, p.156). There are two categories of rotation, orthogonal and oblique rotations. In case of orthogonal, the new factors share an orthogonal structure, while in case of oblique, the factors are correlated and their structure is not orthogonal.

As it was mentioned above, the main purpose of rotation is to obtain a simpler structure describing the associations between indicator variables and factors. According to Kim and Mueller (1978b, p.31) the simplest structure matrix is obtained when the factorial complexity of each indicator is one. This does not often appear in practice.

Orthogonal rotations can be obtained mainly using three methods: Quartimax, Varimax and Equimax; for details see Kim and Mueller (1978b, p.34). In Varimax rotation the variance of the squared loadings for each factor is maximized while in Quartimax the variance of squared loadings for each variable is maximized. Equimax is a combination of these two methods. It uses a criterion that maximizes a linear combination of the quantities that are used in the two others methods. According to Tucker and MacCallum (1997, p.366) a variation of Varimax called "Raw Varimax" solution, implements the original Varimax method on the normalized factor matrix. Another combination of Quartimax and Varimax methods leads to the Orthomax method of rotation (Tucker and MacCallum, 1997, p.369). Furthermore, there is a cycling procedure of rotation that is applied on all possible pairs of factors. With the aid of a minimum quantity, the process stops when no significant transformations exist for all pairs.

When we apply the method of Quartimax rotation without the orthogonality restriction, then we produce an Oblimin rotation solution. Moreover, the oblique variation of Varimax method was termed Covarimin criterion and was proposed by Kaiser (1956). Covarimin solution did not provide satisfactorily oblique solutions in contrast to Quartimin, which gave overly oblique solutions. Due to this fact, Carroll (1957) combined these two criteria and produced the general Oblimin criterion. The Biquartimin method of rotation is a special case of this new method. Furthermore, the Oblimin procedures mentioned above, led many times to singular factor matrices, which is not desirable. The problem was dealt by Jennrich and Sampson (1966) by introducing the Direct Oblimin method, which also have important disadvantages; see Tucker and MacCallum (1997, p.377). Other oblique rotations of factors are the Promax and the Orthoblique one proposed by Harris and Kaiser (1964) that lead to orthogonal or oblique solutions, through orthogonal transformations. Finally, the Oblimax criterion maximizes the number of small and large loadings and can be used to produce either orthogonal (Quartimax) or oblique solutions.

2.9 Conclusion

Factor analysis is a multivariate methodology that is mainly used in social sciences, psychology and marketing when we assume that non observable variables exist under a set of observable ones. To reveal them we measure a set of observable symptoms or characteristics and apply factor analysis to estimate the hidden factors. Alternatively, factor analysis can be also used as a data reduction method (see section 2.1, p.5).

When fitting the factor analysis model some constraints must be implemented in order to produce a single identifiable model (see section 2.7, p.19). We can produce alternative solutions by transforming the parameters of our model using different rotation methods (see section 2.8, p.23).

The latter is the main reason why factor analysis was harshly criticized by several researchers (see for example Chatfield and Collins, 1980, p.87). BFA presented in the following chapter deals with the above problems and offers several remedies for many of them.

CHAPTER 3

BAYESIAN FACTOR ANALYSIS

3.1 Introduction

Bayesian theory was founded by Rev. Thomas Bayes (1763), an English minister and mathematician. However, his work was not really widely implemented by the scientific community until the beginning of 1950. This was due to the fact that the classical (or frequentistic) approach was easier implemented in practice and hence dominated the statistical science. Bayesian statistics started to become popular after the first development of technology and in particular the recent widespread availability of high-speed computers, which allowed the wide availablility of Bayesian methods using MCMC algorithms (Carlin and Louis, 2000, p.6 and Congdon, 2001, p.1).

Nowadays, the methods of both approaches (frequentistic and Bayesian) are equally used in practice, at least for research reasons (Carlin and Louis, 2000). The differences between Bayesian and frequentistic approach are not only methodological but also philosophical. Frequentistic inference involves uncertainty under the repetition of samples from an assumed model, which implies the probability distribution of the observed data conditional on unknown parameters to be estimated. All unknown parameters of the model are considered as fixed values. This is the gash point in the two approaches, as the same unknown parameters are considered as random variables in the Bayesian approach.

3.2 Bayesian theory

A Bayesian model consists of two main components: the sampling model (likelihood) and the prior distribution of the model parameters. The prior distribution reflects our knowledge about the unknown parameters $\boldsymbol{\theta}$ before observing the data. This knowledge can be based either on previous research or on a subjective belief of a practitioner or researcher. The information included in the prior and the likelihood is combined together to produce the posterior distribution $f(\boldsymbol{\theta} \mid \mathbf{y})$.

The mathematic equation that combined the prior and data information is based on the Bayes theorem:

$$f(\boldsymbol{\theta} \mid \mathbf{y}) = \frac{f(\mathbf{y} \mid \boldsymbol{\theta})\pi(\boldsymbol{\theta})}{m(\mathbf{y})} \tag{3.1}$$

where $\boldsymbol{\theta}$ is a $(p \times 1)$ vector of model parameters, $\mathbf{y}$ is a $(n \times 1)$ vector of data, $f(\mathbf{y} \mid \boldsymbol{\theta})$ is the conditional distribution of the observed data vector $\mathbf{y} = (y_1, \ldots, y_n)^T$ given the vector of the unknown parameters $\boldsymbol{\theta}$ (model likelihood), $\pi(\boldsymbol{\theta})$ is the prior distribution of $\boldsymbol{\theta}$, $m(\mathbf{y})$ is the marginal density of the data $\mathbf{y}$ given by:

$$m(\mathbf{y}) = \int f(\mathbf{y} \mid \boldsymbol{\theta})\pi(\boldsymbol{\theta})d\boldsymbol{\theta} \quad . \tag{3.2}$$

From (3.1) the posterior distribution is proportional to the likelihood and prior, hence

$$f(\boldsymbol{\theta} \mid \mathbf{y}) \propto f(\mathbf{y} \mid \boldsymbol{\theta})\pi(\boldsymbol{\theta}) \tag{3.3}$$

This relationship is of major importance since it simplifies the calculation of the posterior.

3.3 Prior distributions

The specification of the prior distributions directly affects the posterior distribution, which is used for the inference of the parameters.

An important category of prior distributions is the *conjugate* priors. A prior distribution is called conjugate to the likelihood $f(\mathbf{y} \mid \boldsymbol{\theta})$ when the produced posterior distribution belongs to the same family as the prior does. Using this approach, computation is considerably simplified. Morris (1983) has shown that such conjugate priors exist for exponential family models. In case of multidimensional parameter vector $\boldsymbol{\theta}$, independent conjugate priors for each conditional posterior distribution can be produced; for details see Carlin and Louis (2000, p.25-28) and Bernando and Smith (1994, p.265).

Conjugate priors can also be used in the form of a mixture. This approach is convenient when we are uncertain about the prior distribution or when we wish to handle dissimilarly the parametric space. Therefore, in the case of a one dimensional problem with $\boldsymbol{\theta}=(\theta)$ we may use a two-component mixture prior

$$\pi(\theta)=\alpha\pi_1(\theta)+(1-\alpha)\pi_2(\theta) \quad , 0\le\alpha\le 1 \tag{3.4}$$

where π_1 and π_2 are prior distributions conjugate to $f(\mathbf{y}\,|\,\theta)$ (see Carlin and Louis, 2000, p.27).

However, Bayesian theory can be also applied when no prior information is available, with the help of *non-informative* or *vague* priors. These priors allow the data to reveal themselves to the posterior distribution through the likelihood. When the parameter space is discrete and finite, that is $\theta\in\boldsymbol{\Theta}=(\theta_1,...,\theta_n)^T$, then a prior distribution of the form

$$p(\theta_i)=\frac{1}{n} \quad , \qquad \text{for } i=1,...,n \tag{3.5}$$

will represent prior ignorance since all events will be equally probable.

When, on the other hand, we have a continuous and bounded parameter space, like $\theta\in\boldsymbol{\Theta}=[\alpha,\beta]$ then a sensible non-informative prior is the uniform

$$p(\theta)=\frac{1}{\beta-\alpha} \quad . \tag{3.6}$$

When $\alpha=-\infty$ and $\beta=\infty$ we can apply a constant distribution, that is

$$p(\theta)\propto 1. \tag{3.7}$$

However, this is an improper distribution, since $\int p(\theta)d\theta=\infty$. Hence, this kind of non-informative prior should be used only in case that the marginal distribution of the data $m(\mathbf{y})$ integrated to a constant real number K. Hence only when the resulted posterior is proper.

Generally, non-informative priors prevent the researcher from making inappropriate assumptions concerning the nature of the distribution of the unknown parameters. The posterior distribution is determined only by the likelihood. In the case of a flat prior, the posterior distribution is proportional to the likelihood:

$$f(\boldsymbol{\theta} \mid \mathbf{y}) \propto f(\mathbf{y} \mid \boldsymbol{\theta}). \tag{3.8}$$

According to Carlin and Louis (2000, p.22) the choice of a prior distribution of the parameters of a model is a tedious task. If we wish to incorporate past information in our model then it can be summarised using a distributional family. Alternatively, the subjective opinion of the experts involved in the problem may be the base of selecting an appropriate prior distribution. If no or low information exists then we should choose a low or non informative prior distribution as already discussed above.

3.4 MCMC methods

When a posterior distribution cannot be calculated analytically in a closed form expression, alternative methods can be used in order to estimate it. In this section we will describe the simulation-based methods that according to Carlin and Louis (2000), can be separated in *iterative* and *non-iterative* methods. Some of the *non-iterative* Monte Carlo methods are *direct sampling, importance sampling, rejection sampling* and *weighted bootstrap*. Implementing these methods, one sample is produced in each iteration. However, they have limited applications. The alternative methods are MCMC iterative methods. The two basic methods are *Metropolis-Hastings* (M-H) and *Gibbs sampling* algorithm.

In the M-H algorithm a candidate- generating density or proposal density $q(\theta,\theta')$ needs to be specified. This proposal is used to generate candidate θ' values when the process is at state θ. The probability of moving from the state θ to the state θ' is denoted by $a(\theta,\theta')$ and equals to

$$a(\theta,\theta') = \begin{cases} \min\left(\dfrac{\boldsymbol{\pi}(\theta')q(\theta',\theta)}{\boldsymbol{\pi}(\theta)q(\theta,\theta')},1\right), & \pi(\theta)q(\theta,\theta') > 0 \\ 1, & \textit{otherwise} \end{cases} \tag{3.9}$$

where in Bayesian theory, $\boldsymbol{\pi}(\boldsymbol{\theta})$ can be substituted by $f(\boldsymbol{\theta} \mid \mathbf{y})$ which is the true joint posterior distribution of the parameters $\boldsymbol{\theta}$. The distribution $\pi(\boldsymbol{\theta})$ needs to be available only up to a constant. This is very convenient in Bayesian statistics since the posterior distributions are expressed proportional to the prior and the likelihood. Therefore, (3.9) becomes $a(\theta,\theta') = \min\left(\dfrac{f(y\mid\theta')\pi(\theta')q(\theta',\theta)}{f(y\mid\theta)\pi(\theta)q(\theta,\theta')},1\right)$ according to equation (3.3). The algorithm is implemented through the following j iterations:

Step 1. Give an initial value $\theta^{(0)}$

Step 2. Generate θ' from a proposal distribution $q(\theta^{(j)},\cdot)$

and u from $U(0,1)$

Step 3. If $u \le a(\theta^{(j)},\theta')$

- set $\theta^{(j+1)} = \theta'$

Else

-set $\theta^{(j+1)} = \theta^{(j)}$

Step 4. Set $j = j+1$

Step 5. If $j < N$ go to Step 2,

otherwise return the generated values $\{\theta^{(1)},\theta^{(2)},...,\theta^{(N)}\}$

where $U(0,1)$ is the uniform distribution at (0,1).

The above algorithm describes the generalized form of M-H algorithm suggested by Hastings (1970). In their original paper, Metropolis et al. (1953) have limited the algorithm to symmetric proposal densities with $q(\theta,\theta') = q(\theta',\theta)$. A family of such densities (proposed by Metropolis et al., 1953) is $q(\theta,\theta') = q_1(\theta'-\theta)$ with $q_1(\cdot)$ a multivariate density; usually the multivariate normal or t-density. This case is called a *random walk chain* or *random walk Metropolis algorithm*. Another family is $q(\theta,\theta') = q_2(\theta')$ that produces *independence chains* (Hastings, 1970) that do not depend on the current value of the parameter.

The specification of the scale of the proposal distribution is essential since it can lead to either large proposed jumps around the parameter space (from θ to θ') with large rejection probability or to high acceptance rate when very small jumps around the parameter space are proposed. Both situations lead to high autocorrelated chains and increase of time of convergence. For a comprehensive presentation of the M-H algorithm see Chib and Greenberg (1995).

Gibbs sampling is a special case of Metropolis-Hastings algorithm. It was firstly proposed for discrete distributions by Geman and Geman (1984). Let us assume that we are interested in drawing a sample from the unknown joint probability of K-random variables, that is $p(\mathbf{\Theta})$ with $\mathbf{\Theta} = (\theta_1,...,\theta_K)$. Instead, we can generate samples from the full or conditional distributions $p_i(\theta_i \mid \mathbf{\theta}_{-i}), \quad for\ i = 1,...,K$, where $\mathbf{\theta}_{-i} = (\theta_1...\theta_{i-1}\theta_{i+1}...\theta_K)$.

Then after providing initial values $\{\theta_1^{(0)},...,\theta_K^{(0)}\}$ steps 3-4 of M-H algorithm are substituted by the following actions:

Draw $\theta_1^{(1)} \sim f(\theta_1 \mid \theta_2^{(0)},...,\theta_K^{(0)})$

Draw $\theta_2^{(1)} \sim f(\theta_2 \mid \theta_1^{(1)},\theta_3^{(0)}...,\theta_K^{(0)})$

…

Draw $\theta_K^{(1)} \sim f(\theta_K \mid \theta_1^{(1)},...,\theta_{K-1}^{(1)})$

After t iterations we obtain $(\theta_1^{(t)},...,\theta_K^{(t)})$. It is proved (see Carlin and Louis, 2000) that

(a) $(\theta_1^{(t)},...,\theta_K^{(t)}) \xrightarrow{d} (\theta_1,...,\theta_K) \sim p(\theta_1,...,\theta_K)$, as $t \to \infty$

(b) The above convergence is exponential in t using the L_1 norm (for details see also Casella and George, 1992).

Variations of the above algorithms have been provided in literature. An example is the reversible jump MCMC (RJMCMC) that is appropriate for comparison of different models defined on a different parameter space with varying dimension. For more information on RJMCMC and other combinations of MCMC methods see Green (1995) and Carlin and Louis (2000, p.159-170).

3.5 Overrelaxation

Autocorrelation is frequently observed in the MCMC generated samples. As a result much larger samples are needed in order to attain convergence. A remedy for reducing autocorrelation was suggested by Adler (1981). The proposed method is called *overrelaxation*. According to this method, the new value in Gibbs sampling iteration is selected so as to be negatively correlated with the previous one. Let us assume that K components compose the parameter state $\mathbf{\Theta} = (\theta_1, ..., \theta_K)$ and also, all full conditional densities $\pi(\theta_i \mid \mathbf{\theta}_{-i})$ are Gaussian. In addition, the θ_i component has conditional mean μ_i and variance σ_i^2, which are functions of other components $\{\mathbf{\theta}_{-i}\}$. Then the new value θ_i' is:

$$\theta_i' = \mu_i + \alpha(\theta_i - \mu_i) + \sigma_i(1-\alpha^2)^{1/2} Z \qquad (3.10)$$

with $-1 \le \alpha \le 1$ and Z follows standard normal distribution. When α is zero, the method is equivalent to Gibbs sampling. Negative values of α are preferred because they produce values of i^{th} parameter from the other side of the mean from θ_i. Note that for $\alpha = -1$ the chain does not reach convergence.

Other methods of overrelaxation are also provided in literature. A review of these methods is presented by Neal (1998). At the same paper, Neal also introduces the method of *ordered overrelaxation*. According to this method, the new value θ_i' is chosen according to the following procedure:

1. K values are generated from the conditional distribution $\pi(\theta_i \mid \theta_{j \neq i})$
2. These K values plus the old value θ_i are arranged in non-decreasing order $\theta_i^{(0)} \le \theta_i^{(1)} \le ... \le \theta_i^{(r)} = \theta_i \le ... \le \theta_i^{(K)}$, with r the index of the ordering of the old value.
3. The new value θ_i' equals $\theta_i' = \theta_i^{(K-r)}$

With $K=1$ we get Gibbs sampling and $K \to \infty$ is equivalent to $\alpha = -1$ for Adler's method.

Neal (1998) also presents the method of ordered relaxation with respect to a uniform distribution. The values $u_i = F(\theta_i)$ follow a uniform $U(0,1)$ distribution, where $F(\cdot)$ is the cumulative distribution function for the conditional $\pi(\theta_i \mid \theta_{-i})$. Then the method of ordered overrelaxation is done in u_i, and the new value u_i' is transformed to $\theta_i' = F^{-1}(u_i')$, with F^{-1} the inverse distribution of F.

Although overrelaxation in general accelerates the convergence, it has drawbacks and should be implemented with caution; for details see Neal (1998). WinBUGS provides the possibility to produce ordered overrelaxed chains according to Neal's method (Spiegelhalter et al., 2003).

3.6 Diagnosing convergence

In general, it is difficult to detect whether a chain has reached convergence. But convergence is essential for the estimation of the posterior distribution, because inference is based on the generated sample which is assumed to be a good approximation of the target posterior distribution.

There are several convergence diagnostic tests of MCMC chains. CODA (Convergence Diagnosis and Output Analysis Software for Gibbs sampling analysis; see Best et al., 1996) is an S-plus library that provides four diagnostic tests suggested respectively by Geweke (1992), Gelman and Rubin (1992), Raftery and Lewis (1992) and Heidelberger and Welch (1983). Other tests or convergence diagnostic tools are available like Yu and Midland's CUSUM method, Liu, Lu and Rubin's L^2 convergence diagnostic, the Johnson diagnostic, Garren and Smith's convergence rate estimator; for details see Brooks and Roberts (1998). In the present thesis, convergence of chains has been tested using CODA software. For this reason the four tests available by CODA will be briefly described.

Geweke (1992) suggests that the values of a function $g(\cdot)$ of the simulated parameters can be treated as a time series. As a consequence, a spectral density $S_g(\omega)$ for this time series can be calculated, which is continuous at zero, with its value $S_g(0)$. He focuses on the mean of the function g that can be estimated by

$$\bar{g}_N = \frac{\sum_{t=1}^{N} g(\theta^{(t)})}{N} \quad , \tag{3.11}$$

with $g(\theta^{(t)})$ the values of the function g at each Gibbs sampling iteration, with N the size of the Gibbs sampling chain. He also provides an estimate of the standard error of the mean:

$$\sqrt{\frac{S_g(0)}{N}} \quad . \tag{3.12}$$

The above concept is used for two different portions of Gibbs chain: N_A and N_B, with $N_A + N_B < N$. Geweke's convergence diagnostic uses the following statistic:

$$Z_N = \frac{\bar{g}_{N_A} - \bar{g}_{N_B}}{\sqrt{\frac{1}{N_A} S_g^A(0) + \frac{1}{N_B} S_g^B(0)}} \xrightarrow{d} N(0,1), \text{ as } \quad N \to \infty \tag{3.13}$$

where $S_g^A(0)$ and $S_g^B(0)$ are spectral estimates for the two portions of the sample N_A and N_B respectively, evaluated at 0.

Geweke (1992) suggests taking $N_A = \frac{N}{10}$ and $N_B = \frac{N}{2}$, that is the first 10 % and the last 50 % of the Gibbs sample, respectively. So, Z_N is used to test the null hypothesis that the two subsamples have equal means. Values of Z_N which lie in the tails of a standard normal distribution provide an indication of non convergence (Best et al., 1996).

The second diagnostic of CODA, introduced by Gelman and Rubin (1992), can be applied on two or more parallel chains. Let us assume m generated chains with different starting points. Then the last, say n, iterations are used to reestimate the distribution of the parameters of interest as a Student t-distribution, the scale parameter of which involves both the between and within-chain variance.

Convergence is monitored by estimating the factor by which the estimated scale will shrink as $n \to \infty$, that is

$$\sqrt{R} = \sqrt{\frac{N-1}{N} + \frac{m+1}{mN}\frac{B}{W}}\sqrt{\frac{df}{df-2}} \tag{3.14}$$

where B is the variance between the means from the m chains, W is the average of the m within-chain variances and df is the degrees of freedom of t-density. “Shrink factor” for a chain that has converged is near 1.

The third diagnostic test was proposed by Raftery and Lewis (1992). It is a single chain test which focuses on the estimating a quantile (usually the 2.5^{th}) of the posterior distribution of the parameters of interest, at a given degree of precision and a required probability of attaining this degree of accuracy. Then, the program reports $N_{\min}$ - the minimum number of iterations that should be run, N the total number of iterations, B the number of burn-in iterations and k the thinning interval to be used in order to estimate the specified quantile of interest at the given precision. The above procedure is related to a binary process, subsequences of which approximate Markov chains; for details see Raftery and Lewis (1992).

Finally, Heidelberger and Welch’s convergence diagnostic (1983) is used for single chains from univariate observations. It can be generalized for multi-dimensional and multi-sample statistics. The test consists of two parts and is based on ideas from Brownian bridge theory. Using this diagnostic we test for the null hypothesis that stationarity is attained using the sampled values. This is achieved using the Cramer-Von Mises statistic (von Mises, 1931). If the null hypothesis is rejected the first 10 % of iterations is discarded and the test is repeated on the remaining sample. This procedure is repeated until 50 % at least observations or more of the chain passes the stationarity test. Half of the chain is used. If the null hypothesis is rejected for all repetitions until then we have an indication that the chain has not reached convergence. If the test is passed, the number of iterations that have been used to pass the test, the discarded iterations and the Cramer-von-Mises statistic are provided.

In the second part of the test the halfwidth test is implemented. The portion of the chain that passed the stationarity test is treated as a time series from which we estimate the spectral density at zero, $S(0)$. Then, the asymptotic standard error of the mean is equal to:

$$\sqrt{\frac{S(0)}{N_p}} \tag{3.15}$$

where N_p is the length of the retained chain. If the halfwidth of the 95 % confidence interval of the mean, evaluated with the asymptotic standard error, is less than ε times the sample mean, the halfwidth test is passed; ε is a small fraction with CODA default 0.1. In the opposite case, the halfwidth test reports failure and a longer chain should be run to achieve increased precision of the estimated parameter.

3.7 Model selection

The comparison between the models, at this thesis, is conducted through three Information Criteria: Deviance Information Criterion (DIC, Spiegelhalter et al., 2002), the Bayesian variation of Akaike Information Criterion (AIC, Akaike, 1987) and Bayesian Information Criterion (BIC, Schwarz, 1978). An alternative quantity used for model selection is Bayes factor; for details see Kass and Raftery (1995).

DIC is directly provided by WinBUGS. It is equal to:

$$DIC = \bar{D} + p_D = D(\bar{\mathbf{\theta}}) + 2p_D \tag{3.16}$$

where $\bar{D}$ is the posterior mean of the deviance, p_D is the effective number of parameters and $D(\bar{\mathbf{\theta}})$ is the point estimate of the deviance at the mean of the estimated parameters $\mathbf{\theta}$. The deviance is defined as

$$D(\mathbf{\theta}) = -2\log f(y \mid \mathbf{\theta}) \tag{3.17}$$

while $$\bar{D} = -\frac{2}{N}\sum_{t=B+1}^{N} \log f(y \mid \theta^{(t)}) . \tag{3.18}$$

The model with the smallest DIC is estimated to be the model that would best predict a replicate dataset of the same structure as that currently observed.

Akaike (1987) defines the Bayesian version of the AIC as:

$$\text{AIC}= D(\hat{\boldsymbol{\theta}}) +2\, d_m \tag{3.19}$$

where $D(\hat{\boldsymbol{\theta}})$ is the minimum value of the deviance, $\hat{\boldsymbol{\theta}}$ is the mean of the posterior distribution of the estimated parameters and d_m the number of estimated parameters.

BIC (Schwarz, 1978) is estimated by:

$$\text{BIC} = D(\hat{\boldsymbol{\theta}}) + d_m \log(n') \tag{3.20}$$

where n' is the number of observations. In case of factor analysis n'=number of individuals*number of observed variables.

Both AIC and BIC penalize for the number of parameters and in general, they tend to choose the less complex models.

3.8 Bayesian approaches to factor analysis

The unknown parameters in the model of factor analysis presented in equation (2.1) can be evaluated, using the Bayesian approach. Using this approach, we specify prior distributions on the parameters and we produce posterior distributions. Inference is based on the mean or other moments of posterior distributions.

The first attempt of BFA was made by Press (1972). After his work, many researchers have contributed with their published work, like Kaufman and Press (1973), Martin and Mac Donald (1975), Lee (1981), Press and Shigemasu (1989, revised in 1997), Rowe (1998, 2000a, 2000b, 2000c, 2001, 2003), Raftery (1993), Rowe and Press (1998), Scheines et al (1999), Hayashi and Sen (2002), D’Souza (2002), West (2003), Lopes (2003), Lopes and West (2004) and Fokoue (2004).

At this section the basic form of the model used in BFA proposed by Press and Shigemasu (1989) will be presented, as well as the alternative forms proposed by Rowe (2000a, 2000b, 2001) and Rowe and Press (1998). Moreover, a review of other models proposed in literature will be presented and the parameterization of the lower triangular matrix (proposed by Lopes and West, 2003) that will be used at the Bayesian analysis of schizotypic data in section 5.2.2.

Press and Shigemasu (1989), in the following will be denoted as PS89, present a basic form of the model used in BFA. The model is given by (2.1) (see p.7). It can also be written in matrix form as:

$$\underset{(n\times p)}{\mathbf{Y}} = \underset{(n\times q)}{\mathbf{F}}\,\underset{(q\times p)}{\mathbf{\Lambda}} + \underset{(n\times p)}{\mathbf{E}} \quad . \tag{3.21}$$

In BFA, the errors $\mathbf{e}_i$s (the elements of $\mathbf{E}$) are assumed to be normally distributed that is $\mathbf{e}_i \sim N(\mathbf{0},\mathbf{\Psi})$. But $\mathbf{\Psi}$ in that case is a positive definite matrix, not diagonal itself but diagonal on average, that is $E(\mathbf{\Psi}) > 0$ and diagonal.

The unknown quantities are ($\mathbf{\Lambda},\mathbf{F},\mathbf{\Psi}$) with $\mathbf{F}^T = (f_1,...,f_n)$. The likelihood of ($\mathbf{\Lambda},\mathbf{F},\mathbf{\Psi}$) by assuming independent y_i`s is expressed by:

$$f(\mathbf{Y} \mid \mathbf{\Lambda},\mathbf{F},\mathbf{\Psi}) \propto \left|\mathbf{\Psi}\right|^{-n/2} \exp\left(-\frac{1}{2} tr\mathbf{\Psi}^{-1}(\mathbf{Y}-\mathbf{F}\mathbf{\Lambda}^T)^T(\mathbf{Y}-\mathbf{F}\mathbf{\Lambda}^T)\right) . \tag{3.22}$$

The prior distributions used by PS89 belong to the natural conjugate family. The joint distribution of ($\mathbf{\Lambda},\mathbf{F},\mathbf{\Psi}$) has the following structure

$$f(\mathbf{\Lambda},\mathbf{F},\mathbf{\Psi}) \propto f(\mathbf{\Lambda} \mid \mathbf{\Psi}) f(\mathbf{\Psi}) f(\mathbf{F}) \tag{3.23}$$

with $\mathbf{\Lambda}$ conditional on $\mathbf{\Psi}$ has elements that are jointly normally distributed, with hyperparametrs ($\mathbf{\Lambda_0},\mathbf{H}$), $\mathbf{H} = n_0\mathbf{I}$ for some scalar n_0.Hence $f(\mathbf{\Lambda} \mid \mathbf{\Psi}) \sim N_p(\mathbf{\Lambda_0}, n_o\mathbf{I})$. The matrix $\mathbf{\Psi}$ follows an inverse Wishart distribution with hyperparameters $(v,\mathbf{B})$, with $\mathbf{B}$ diagonal, that is $\mathbf{\Psi}^{-1} \sim W(v,\mathbf{B})$. The prior distribution of $\mathbf{F}$ may be specified either by historical data or alternatively may be proportional to a constant. The factor scores of subjects are also taken to be a priori independent so as to have

$$f(\mathbf{F}) = f(f_1)...f(f_n) \tag{3.24}$$

Hence, the joint posterior density of the unknown parameters becomes with the help of Bayesian theory:

$$f(\mathbf{\Lambda},\mathbf{F},\mathbf{\Psi}\,|\,\mathbf{Y}) \propto f(\mathbf{Y}\,|\,\mathbf{\Lambda},\mathbf{F},\mathbf{\Psi})f(\mathbf{\Lambda},\mathbf{F},\mathbf{\Psi})$$

$$\propto f(\mathbf{F})\left|\mathbf{\Psi}\right|^{-\frac{n+q+v}{2}}\exp\left(-\frac{1}{2}tr\mathbf{\Psi}^{-1}\mathbf{G}\right) \tag{3.25}$$

with $\mathbf{G} = (\mathbf{Y}-\mathbf{F}\mathbf{\Lambda}^{\mathrm{T}})^{\mathrm{T}}(\mathbf{Y}-\mathbf{F}\mathbf{\Lambda}^{\mathrm{T}}) + (\mathbf{\Lambda}-\mathbf{\Lambda}_0)\mathbf{H}(\mathbf{\Lambda}-\mathbf{\Lambda}_0)^{\mathrm{T}} + \mathbf{B}$.

The marginal posteriors for $(\mathbf{\Lambda},\mathbf{F}\,|\,\mathbf{Y})$ and $(\mathbf{F}\,|\,\mathbf{Y})$ by integrating with respect to $\mathbf{\Psi}$ and $(\mathbf{\Lambda},\mathbf{\Psi})$ can be also found in PS89. For large samples and for a wide variety of priors of $\mathbf{F}$, including that proportional to a constant, it is proved that, $(f_i\,|\,\mathbf{Y})$ is distributed as a multivariate t distribution.

Due to the fact that the marginal joint posterior $(\mathbf{\Lambda},\mathbf{\Psi})$ is complicated, $\mathbf{\Lambda}$ is estimated for given $\mathbf{F}=\hat{\mathbf{F}}$. So, the authors show that any scalar element of $\mathbf{\Lambda}$ conditional on $(\hat{\mathbf{F}},\mathbf{Y})$ follows a general Student t-distribution. In addition, the mean of the distribution at $\mathbf{\Lambda}$, given the data vectors and $\hat{\mathbf{F}}$, is used as a point estimator $\hat{\mathbf{\Lambda}}$. This estimator helps to estimate the marginal distribution of $\mathbf{\Psi}$ conditional on $(\mathbf{F},\mathbf{\Lambda}) = (\hat{\mathbf{F}},\hat{\mathbf{\Lambda}})$. After algebraic manipulations, $f(\mathbf{\Psi}\,|\,\hat{\mathbf{F}},\hat{\mathbf{\Lambda}},\mathbf{Y})$ is an inverse Wishart distribution. As a point estimator of $\mathbf{\Psi}$ we can consider the mean $\bar{\mathbf{\Psi}} = E(\mathbf{\Psi}\,|\,\hat{\mathbf{F}},\hat{\mathbf{\Lambda}},\mathbf{Y})$.

Several researchers have suggested some alternative forms of priors. For example, Martin and Mac Donald (1975) use the following prior distribution

$$f(\mathbf{F},\mathbf{\Psi}) = k\exp\left(-\sum_{i=1}^{n}(\beta_i/\psi_i^2)\right) \tag{3.26}$$

where k is a normalizing constant and β_is are parameters to be assessed. The method they propose is the minimization of:

$$f(\mathbf{F},\mathbf{\Psi}\,|\,\mathbf{S}) \propto f(\mathbf{S}\,|\,\mathbf{F},\mathbf{\Psi})f(\mathbf{F},\mathbf{\Psi}) \tag{3.27}$$

where $f(\mathbf{F},\mathbf{\Psi}\,|\,\mathbf{S})$ is the posterior density of unknown parameters given the sample covariance matrix, $f(\mathbf{S}\,|\,\mathbf{F},\mathbf{\Psi})$ the conditional density function of the sample covariance matrix and $f(\mathbf{F},\mathbf{\Psi})$ the prior density function of $(\mathbf{F},\mathbf{\Psi})$. The problem is equivalent to minimizing

$$\psi = -\log(f(\mathbf{S}\,|\,\mathbf{F},\mathbf{\Psi})) - \log(f(\mathbf{\Psi})) \tag{3.28}$$

and it resembles to the ML method described in section 2.5.2. The same paper also deals with the problem of Heywood cases (see section 2.5.2.1).

Kaufman and Press (1973) refer to a non-informative prior distribution of ($\mathbf{\Lambda},\mathbf{\Psi}$) given by:

$$f(\mathbf{\Lambda},\mathbf{\Psi}) \propto \left|\mathbf{\Psi}\right|^{a-\frac{p+1}{2}}$$

or
$$f(\mathbf{\Sigma}) \propto \left|\mathbf{\Lambda}\mathbf{\Lambda}^{\mathrm{T}} + \mathbf{\Psi}\right|^{a-\frac{p+1}{2}}. \tag{3.29}$$

But they lead on the unsatisfactory result that there are no underlying factors on the average that account for the variance of $\mathbf{Y}$.

Lee (1981), used different priors for four different cases. After constraining some elements of $\mathbf{\Lambda}$ to fixed values, in order to obtain identification and easy interpretation, he assumed that the free parameters in $\mathbf{\Lambda}$ are exchangeable. They all follow the same normal distribution $N(\eta,\sigma^2)$ with η following a "locally uniform" and relatively non-informative prior and σ^2 following an inverse -χ^2 family. In the second case, the exchangeability hypothesis is dropped and λ_{ij} follows normal distributions $N(\lambda_{ij}^*,\sigma_{ij}^2)$ with mean λ_{ij}^* and variance σ_{ij}^2 as above. The prior of λ_{ij} for the third case is vague, proportional to a constant and in the fourth case it is similar to the one proposed by Martin and Mac Donald (1975).

Rowe has a major contribution in the development of BFA. In his models, the overall population mean vector $\boldsymbol{\mu}$ is also considered, that is the $\mathbf{y}_i$\`s are not subtracted by their means. So, the model takes the form:

$$\underset{(p\times 1)}{\mathbf{y}_i} = \underset{(p\times 1)}{\boldsymbol{\mu}} + \underset{(p\times q)}{\mathbf{\Lambda}}\ \underset{(q\times 1)}{\mathbf{f}_i} + \underset{(p\times 1)}{\boldsymbol{\varepsilon}_i}, \qquad \text{for } i = 1,2,...,n. \tag{3.30}$$

Rowe has used several priors. In the paper of Rowe and Press (1998) the prior distribution of $(\boldsymbol{\mu},\mathbf{\Lambda},\mathbf{F},\mathbf{\Psi})$ is given by

$$f(\boldsymbol{\mu},\mathbf{\Lambda},\mathbf{F},\mathbf{\Psi}) = f(\boldsymbol{\mu})f(\mathbf{\Lambda}\,|\,\mathbf{\Psi})f(\mathbf{F})f(\mathbf{\Psi}) \tag{3.31}$$

with $f(\boldsymbol{\mu})$ vague, $f(\mathbf{\Lambda}\,|\,\mathbf{\Psi})$ and $f(\mathbf{F})$ are normally distributed and $f(\mathbf{\Psi})$ is inverse Wishart distributed.

In an older work (Rowe 2000a) he used a normal prior distribution for $\boldsymbol{\mu}$ conditional on $\boldsymbol{\Psi}$. The prior structure for the unknown quantities is given in equation (3.33).

$$f(\boldsymbol{\mu},\boldsymbol{\Lambda},\mathbf{F},\boldsymbol{\Psi}) = f(\boldsymbol{\mu}\,|\,\boldsymbol{\Psi})f(\boldsymbol{\Lambda}\,|\,\boldsymbol{\Psi})f(\mathbf{F})f(\boldsymbol{\Psi}) \tag{3.32}$$

where $f(\boldsymbol{\mu}\,|\,\boldsymbol{\Psi})$ is set to be vague, $f(\boldsymbol{\Lambda}\,|\,\boldsymbol{\Psi})$ and $f(\mathbf{F})$ are normally distributed and $f(\boldsymbol{\Psi})$ is inverse Wishart distributed.

Rowe (2001) has also considered a normal prior for the joint distribution of $\mathbf{C} = (\boldsymbol{\mu},\boldsymbol{\Lambda})$ conditional on $\boldsymbol{\Psi}$. The relative prior for the unknown quantities is:

$$f(\mathbf{C},\mathbf{F},\boldsymbol{\Psi}) = f(\mathbf{C}\,|\,\boldsymbol{\Psi})f(\mathbf{F})f(\boldsymbol{\Psi})\,. \tag{3.33}$$

Rowe (2000b) uses normal priors for $\boldsymbol{\mu}$, $\boldsymbol{\lambda} = vec(\boldsymbol{\Lambda}^T)$, $\mathbf{F}$ and inverse Wishart for $\boldsymbol{\Psi}$. The corresponding prior structure is given by equation (3.35).

$$f(\boldsymbol{\mu},\boldsymbol{\lambda},\mathbf{F},\boldsymbol{\Psi}) = f(\boldsymbol{\mu})f(\boldsymbol{\lambda})f(\mathbf{F})f(\boldsymbol{\Psi}) \tag{3.34}$$

The conditional posterior distributions can be easily obtained. For example, Rowe (2000b) provides the conditional posterior $f(\boldsymbol{\mu}\,|\,\boldsymbol{\Lambda},\mathbf{F},\boldsymbol{\Psi},\mathbf{Y})$, $f(\mathbf{F}\,|\,\boldsymbol{\mu},\boldsymbol{\Lambda},\boldsymbol{\Psi},\mathbf{Y})$, $f(\boldsymbol{\Lambda}\,|\,\boldsymbol{\mu},\mathbf{F},\boldsymbol{\Psi},\mathbf{Y})$ as normal distributions and $f(\boldsymbol{\Psi}\,|\,\boldsymbol{\mu},\mathbf{F},\boldsymbol{\Lambda},\mathbf{Y})$ as inverse Wishart.

Lopes and West (2004) have used different constraints on the loading matrix which forms a lower triangular matrix given by (3.35).

$$\boldsymbol{\Lambda} = \begin{pmatrix} \lambda_{11} & 0 & 0 & \ldots & 0 & 0 \\ \lambda_{21} & \lambda_{22} & 0 & \cdots & 0 & 0 \\ \lambda_{31} & \lambda_{32} & \lambda_{33} & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \lambda_{q-1,1} & \lambda_{q-1,2} & \lambda_{q-1,3} & \cdots & \lambda_{q-1,q-1} & 0 \\ \lambda_{q,1} & \lambda_{q,2} & \lambda_{q,3} & \cdots & \lambda_{q,q-1} & \lambda_{q,q} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \lambda_{p,1} & \lambda_{p,2} & \lambda_{p,3} & \cdots & \lambda_{p,q-1} & \lambda_{p,q} \end{pmatrix} \tag{3.35}$$

To the loadings λ_{ij} with $i > j$ they use normal priors, that is $\lambda_{ij} \sim N(0, C_0)$. On the diagonal elements of $\boldsymbol{\Lambda}_0$ they use truncated normal distributions, that is

$\lambda_{ii} \sim N(0, C_0) I(\lambda_{ii} > 0)$ while they restrict λ_{ij} with $i < j$ to be equal to zero. In the same work, the elements of the matrix $\mathbf{F}$ a priori follow a normal distribution as usually in BFA. Moreover, they use two different approaches for specifying the number of factors. At the first approach, they use MCMC methods with prespecified number of factors q, while, on the second approach, q is considered to be an unknown parameter that has to be estimated. In the latter case, reversible jump MCMC (RJMCMC) method was used.

The main difference between the approach of PS89 and other subsequent approaches of BFA is they used closed form expressions for the posterior distributions, something that is not prerequisite when using MCMC methods. Gibbs sampling uses only the unnormalized conditional posterior distributions that are easily obtained when data and priors are available. Inference is made on the final MCMC sample from the joint posterior distribution of the parameters. Firstly, initial values for $\mathbf{F}$ and $\mathbf{\Psi}$ should be given. Then the algorithm cycles through:

$\bar{\mathbf{\Lambda}}_{(i+1)} \equiv$ a random sample from $f(\mathbf{\Lambda} \mid \bar{\mathbf{F}}_{(i)}, \bar{\mathbf{\Psi}}_{(i)}, \mathbf{Y})$

$\bar{\mathbf{\Psi}}_{(i+1)} \equiv$ a random sample from $f(\mathbf{\Psi} \mid \bar{\mathbf{F}}_{(i)}, \bar{\mathbf{\Lambda}}_{(i+1)}, \mathbf{Y})$

$\bar{\mathbf{F}}_{(i+1)} \equiv$ a random sample from $f(\mathbf{F} \mid \bar{\mathbf{\Lambda}}_{(i+1)}, \bar{\mathbf{\Psi}}_{(i+1)}, \mathbf{Y})$

When this procedure converges, we have obtained $B + N$ triplets $(\bar{\mathbf{\Lambda}}_{(1)}, \bar{\mathbf{\Psi}}_{(1)}, \bar{\mathbf{F}}_{(1)}), \ldots, (\bar{\mathbf{\Lambda}}_{(B+N)}, \bar{\mathbf{\Psi}}_{(B+N)}, \bar{\mathbf{F}}_{(B+N)})$. Observations generated in the first B iterations are discarded to avoid dependence on the initial choice of parameter. The next N observations are kept in order to evaluate the posterior distribution of the parameters. Generally, the means or, less frequently, the modes are used as estimates of the unknown parameters.

3.9 Singular Value Decomposition

At the present thesis, the constraint of the lower triangular matrix (see 3.34) was used at the exploratory analysis for reasons of parameterisation. The resulting loading matrix $\mathbf{\Lambda}$, though, can not be compared directly with the corresponding loading matrix of the classical analysis since the constraint of the diagonal product $\mathbf{\Lambda^T\Psi^{-1}\Lambda}$ (see Lawley and Maxwell, 1971, chapter 4) is not satisfied.

In order to obtain comparable estimates for the two approaches we follow the approach described bellow.

Let us denote by b the product

$$\mathbf{\Lambda^T\Psi^{-1}\Lambda = b} \tag{3.36}$$

Then it can be decomposed in an orthogonal matrix $\mathbf{U}$ and a diagonal matrix $\mathbf{V}$ with the method of *singular value decomposition* that is

$$\mathbf{b = U^T V U}\,. \tag{3.37}$$

Consequently equation (3.37) is written as:

$$\mathbf{\Lambda^T\Psi^{-1}\Lambda\ =\ U^T V U} \tag{3.38}$$

By multiplying the above equation with $\mathbf{(U^{-1})^T}$ from the left and $\mathbf{(U^{-1})}$ from the right we get:

$$\mathbf{(\Lambda U^{-1})^T\ \Psi^{1}(\Lambda\ U^{-1}) = V} \tag{3.39}$$

Equation (3.37) has been transformed into a new form that satisfies the constraint of the classical model. The transformed corresponding loading matrix $\mathbf{\Lambda_{svd}}$ is equivalent to:

$$\mathbf{\Lambda_{svd} = \Lambda U^{-1}} \tag{3.40}$$

Following the approach of Viele and Srinivasan (2000) we implement these transformations on the posterior means of the loading matrix in order to obtain estimates comparable to the ML estimates of the standard orthogonal model. The singular value decomposition is commonly used in multivariate statistics (see for example Venables and Ripley, 1999, Kateri et.al, 2005, Viele and Srinivasan, 2000).

3.10 Comparison of Bayesian and non-Bayesian factor analysis

The advantages of BFA are important. Firstly, prior information can be incorporated in BFA, something that the frequentistic factor analysis can not take advantage of. It is known that a sequence of analyses should be made in factor analysis in order to take safe results. This information, resulting from previous stages of analysis, can be used in BFA as prior information. This information may refer to the number of factors or the relationships of factors and manifest variables. Another, advantage of BFA is that correlated errors can be easily incorporated into the model without considerable theoretical or computational difficulty.

Scheines et al. (1999) make a list of the advantages of Gibbs sampling versus ML approximation. They refer to the lack of need of asymptotic normality hypothesis. They also give emphasis to the usefulness of posterior distributions. The latter give us the chance to detect multimodality and they help us to inspect the fit of the model using posterior predictive p-values. According to them, another benefit is the fact that underidentified models can give results using informative priors. Kaufman and Press (1973) have also stressed the superiority of Bayesian application in factor analysis. They support that the restrictions in classical factor analysis are very “dogmatic” and the resulting loading matrices are not unique since they can be changed by a proper rotation. This does not happen in BFA, that only needs careful specification of prior distributions.

CHAPTER 4

APPLICATION OF EXPLORATORY FACTOR ANALYSIS IN SCHIZOTYPIC DATA

4.1 Introduction

In this chapter we implement Bayesian and non Bayesian EFA methodology on schizotypic data. According to Meehl (1990), schizotypy is the fundamental construct at the level of psychism. In psychiatrical terminology, a schizotype suffers "pseudoneurotic" decompensation, with microsychotic episodes. In general, the prevalence rate of schizotypy in the general population is about 10%. Another notion that is related to schizotypy is schizotaxia. According to Meehl (1964), schizotaxia is a neural integrative defect, which is supposed to be inherited. But the imposition of social learning history upon schizotaxic individuals, results in schizotypic personalities. When a schizotype, is physically vigorous and resistant to stress, does not present symptoms of mental disease. On the other hand, there is also a subset of schizotypes that decompensate in clinical schizophrenia. The prevalence rate of shizophrenia is found increased in schizotypals in comparison to normal population.

In addition, Meehl (1964) provides four schizotypal characteristics: cognitive slippage, anhedonia, ambivalence and interpersonal aversiveness. In summary, cognitive slippage is a kind of mental thinking disorder, anhedonia is denoted as a marked, widespread and refractory defect in the pleasure capacity (person cannot find pleasure in anything), and symptoms of interpersonal aversiveness are social fear, distrust, expectation of rejection and conviction of the person's unlovability. Information for ambivalence can be found at Bleuler (1950).

In 1987, American Psychiatric Association edited the DSM-III-R Diagnostic and statistical manual of mental disorders. At this handbook, nine features defined the schizotypal personality:

1. **Ideas of reference,** which are related to the feeling that things on TV or advertisements have a special meaning for the individual or that people talk about him when talking each other.
2. **Excessive social anxiety,** which does not disappear with familiarity and is related to paranoid fears.
3. **Unusual perceptual experiences,** like the feeling of presence of shadows and in general illusions of the five senses.
4. **Odd or eccentric behaviour,** which is related to unusual mannerisms and eccentric appearance or habits.
5. **No close friends**, when the individual finds it hard or does not have interest in getting emotionally close to other people except for his family.
6. **Odd speech**, that is vague, confusing or not cohesive speech.
7. **Constricted affect**, which is a sentimental accent not in harmony with speech, idea or thoughts.
8. **Suspiciousness**, with friends or co-workers.
9. **Odd beliefs or magical thinking**, for example, experiences with supernatural, telepathy and clairvoyancy.

Since then, several scales have been constructed in order to measure the schizotypic features seperately. However, Raine (1991) constructed a 74-item self-administered questionnaire in an effort to provide an overall concept of schizotypal personality. The items of the questionnaire are binary (yes/no) and each takes the value of one (1) if the answer is positive and zero(0) if the answer is negative. The total score takes the values from 0 to 74. SPQ can be assessed in non-clinical populations, as well as clinical and provides brief subscales for the nine schizotypal features (each subscale was calculated as the sum of the questionnaire items that refer to each schizotypal subscale, see Appendix A), as well as an overall scale for schizotypy. Furthermore, nine subscales may be summarized by three groups-factors of schizotypal personality (see figure 5.3, p.65; Raine et al., 1994). SPQ is available at Raine's site.

The SPQ items were based on existing interview schedules for schizotypal personality (at a percentage of 34 %), were modeled on examples of schizotypal traits outlined in DSM-III-R (8 %), were taken from relative published questionnaires (18 %) and were also generated by the author (40 %). This version of questionnaire of Raine (1991) exhibited satisfactory internal reliability (alpha coefficient 0.90) with high correlation equal to 0.81 with STA, which is also a schizotypal personality scale based on DSM-III (American psychiatric Assocciation, 1980). The questionnaire is provided in appendix A (p.81) with its items grouped in the nine schizotypal traits.

Raine (1991) provides arguments in favor of usefulness of SPQ for the screening of schizotypes:

1. In the first assessment of the SPQ it was found that 55% of the high 10% SPQ scorers total had a DSM-III-R clinical diagnosis of SPD as assessed by the SCID. (Raine, 1991)
2. SPQ, followed by a confirmatory clinical interview, can be implemented in non-clinical populations and help the recruitment of subjects with SPD.
3. It can reduce bias created by traditional research implemented in clinical populations. The samples used in traditional studies are taken from conventional treatment centers while schizotypy is a psychological disorder that also appears in non-clinical populations.
4. Although schizotypic individuals are genetically predisposed to schizophrenia, they have protective factors against this illness. Individuals that do not feel the need of seeking out psychiatric help and belong to non-clinical populations may consist the appropriate sample of research on such protective factors.

5. SPQ may help the studies for schizophrenia. Since the control group used in schizophrenia studies or surveys may contain schizotypic individuals resulting to failure of discrimination between schizophrenic patients and control subjects. In such case, SPQ can provide a tool for screening out the schizotypic subjects from the study and reduce type II errors.

In Greece, SPQ was firstly translated by the team of Stefanis et al. (2004). An independent official translator translated it back in English, and was sent to the author for approval, which was granted. The final Greek version was produced by the comparison of the first and second draft in English. Some small changes were also necessary after administered to a test sample of 15 young employees of the University Mental Health Research Institute (UMHRI). This questionnaire was used in order to examine the covariance structure of schizotypy firstly by Stefanis et al. (2004). Since then, Βιτωράτου (2004) has analysed the reliability of the Greek SPQ, Ηλιοπούλου (2004) examined the relationship between schizotypy and the ideas of impulsive and compulsive buying used in Marketing Psychology. Finally Στάθη (2005) examined the relationship between SPD and records of knowledge.

The sample (collected by Ηλιοπούλου, 2004) is used in the present thesis to examine the latent dimensions of schizotypy using both classical and BFA. The sample consists of 167 Greek students from universities and Technological Education Institutes (TEI). In particular, the data were collected in the School of Management Sciences of the University of Aegean, in the two Universities and the TEI of Crete and in the TEI of Piraeus. It was collected during the period of exams in June 2003. After rejection of some questionnaires for reasons of validity the final sample consisted of 167 individuals, 56 % being females and 44 % males.

The mean age of the sample is 22 years old with a right-skewed age distibution, since postgraduate students also participated to the study. The participants completed a series of dichotomous items-queries. The responses in the SPQ items, which represent the nine schizotypal traits, were summed together and transformed to proportions of positive responses over the total numbers of items-questions.

LISREL 8.52 student version (see Jöreskog and Sörbom, 1996) was used in order to analyse the data with EFA as well as CFA methods. The Bayesian analysis was conducted with *WinBUGS 1.4* (see Spiegelhalter, et al., 2003).

4.2 Exploratory analysis – Application to schizotypic data

4.2.1 Classical Analysis

According to Tucker and MacCallum (1997) the parameterization that is used in frequentistic factor analysis imposes a constraint concerning the number of factors. The number of correlations of observed variables should be greater or equal to the number of free parameters of the model, that is

$$pq - q(q-1)/2 \leq p(p-1)/2 \tag{4.1}$$

where p is the number of observed variables and q the number of factors. In the present case, constraint (4.1) limits the number of factors below or equal to five. LISREL discards the missing values, so the total effective size was equal to 163.
When trying to fit three, four and five factor models Heywood cases were reported (see section 2.5.2.1, p.17). Hence, resulting loadings and the unique variances should be interpreted with caution.

Model	Degrees of freedom	Chi-square	p-value	RMSEA
$\mathbf{F_1}$	27	122.13	0.000	0.147
$\mathbf{F_2}$	19	44.14	0.001	0.090
$\mathbf{F_3}$	12	24.09	0.020	0.079
$\mathbf{F_4}$	6	11.18	0.083	0.073
$\mathbf{F_5}$	1	0.23	0.631	0.000

Table 4.1. Chi-square values of EFA

Table 4.1 presents the values of the chi-square test (see section 2.5.2), the p-values and RMSEA (see section 2.7) for the five models that were examined: F_1 (one-factor model), F_2 (two-factor model), F_3 (three-factor model), F_4 (four-factor model) and F_5 (five-factor model). According to the values of p-value we can not reject the null hypothesis that the four-factor model holds for the population in a 5% significance level. The value of RMSEA of F_4 is acceptable according to section 2.6 (p.17) and shows significant approximation to the population.

Schizotypal traits	Factor 1	Factor 2	Factor 3	Factor 4
Ideas of reference	0.572			0.355
Odd beliefs or magical thinking	0.359			0.592
Unusual perceptual experience	0.493			0.433
Odd speech	0.348		0.434	0.374
Suspiciousness	0.828	0.556		
Constricted affect	0.289		0.748	
Odd behaviour	0.828	-0.555		
No close friends	0.379		0.590	
Social anxiety			0.385	

Table 4.2. Unrotated loadings of the four-factor model (classical EFA) (loadings with absolute values ≤ 0.2 are eliminated from the table)

Table 4.2 presents the factor loadings of the four-factor model with absolute values larger than the value of 0.2. Values lower than 0.2 show negligible correlation between the variable and the factor and hence were removed to have a better picture about factor decomposition. Factor 1 can be interpreted as a general schizotypic factor as it is correlated with all schizotypal scales except "excessive social anxiety". It is strongly correlated with *suspiciousness* and *odd or eccentric behaviour*, moderately correlated with *ideas of reference*, *unusual perceptual experience*, *no close friends* and *odd beliefs or magical thinking*. Lower correlations are observed between the first factor and *constricted affect* and *odd speech.* The second factor reflects a contrast between *suspiciousness* and *odd or eccentric behaviour*. The third factor loads on *odd speech*, *constricted affect*, *no close friends* and *excessive social anxiety*. The 4th factor is correlated with *ideas of reference*, *odd beliefs or magical thinking*, *unusual perceptual experience*, *odd speech.*

The results of the other models can be found in appendix B (p.87)

4.2.2 Bayesian analysis

Bayesian EFA was conducted by imposing a lower triangular loading matrix as a constraint, for reasons of parameterization. The prior of the free loadings was chosen to be standard normal distribution, while the diagonal elements of the lower triangular loading matrix were assumed to follow a priori truncated at zero normal distribution as described in section 3.8. The error terms as well as the latent factors were assumed to be uncorrelated. For the first model (F_1), 100000 iterations were implemented after a 50000 burn-in sample. The second model (F_2) converged after 70000 iterations and a 200000 sample was used in order to have secure results. A 50000 burn-in sample was discarded for the third model (F_3). Due to high autocorrelation of the Gibbs sampling output, we kept one every 80th iteration to the final sample used for posterior inference. Furthermore, the method of overrelaxation was implemented (Neal, 1998); a total of 25000 iterations were generated. The same procedure was used for the four-factor model (F_4), with thin interval equal to 100 and the total number of 30000 iterations discarding the initial 50000 burn-in. The fifth model (F_5) converged after 100000 burn-in iterations and 30600 iterations after imposing a 100 thin interval and the method of over-relaxation (see section 3.5, p.33). The generated chains passed the four tests of CODA (see section 3.6, p.34). The resulting loading matrix was transformed according to the singular value decomposition (see section 3.9, p.45).

Table 4.3 presents the Deviance Information Criterion (DIC, Spiegelhalter et al., 2002), Akaike Information criterion (AIC, Akaike, 1987) and Bayesian Information Criterion (BIC, Schwarz, 1978) for the four models as well as, t, the number of estimated parameters (see section 3.7). The number of estimated parameters is equal to the number of the free loadings of the loading matrix and the number of unique variances at each model.

	Information Criteria			
Models	DIC	**AIC**	**BIC**	**t**
$\mathbf{F_1}$	3848	3667	3762	18
$\mathbf{F_2}$	3050	3233	3371	26
$\mathbf{F_3}$	3334	3085	3260	33
$\mathbf{F_4}$	3105	2894	3101	39
$\mathbf{F_5}$	3118	2924	3158	44

Table 4.3. Information Criteria for Bayesian EFA, (t: number of parameters); AIC and BIC have been calculated from θ corresponding to the min Deviance from the MCMC run

The four-factor model is chosen as the best fitted model with respect to AIC and BIC while DIC supports the second model to have a better fit.

An alternative way of comparing different models is the deviance. The Bayesian analysis provides the posterior density of the deviance; so different point estimates can be evaluated. Table 4.4 gives the estimates of the mean, median, 2.5%, 97.5% quantiles and minimum of the deviance for the five models as well as $D(\overline{\theta})$, that is the estimate of the deviance at the posterior mean of the stochastic nodes.

Models	$\bar{D}$	2.50%	median	97.50%	minimum	$D(\bar{\theta})$
F_1	3700 (*20.81*)	3662	3700	3743	3631	3552
F_2	3414 (*37.54)*	3338	3414	3486	3181	3777
F_3	3284 (*56.77)*	3165	3286	3390	3019	3234
F_4	3162 (*74.54)*	3009	3165	3301	2816	3221
F_5	3192 *(74.91)*	3038	3194	3333	2836	3266

Table 4.4. Posterior descriptive measure for Bayesian EFA models

At the parenthesis below the mean value of the deviance, the standard deviation of the estimate is given. The estimates are accurate since the MC errors are less than 5% of the standard deviation of the values. All measures at table 4.4 support the fourth model as the model with the smallest deviance.

Because of the small samples of Gibbs-sampling chains (25000 at F_3 and 30000 at F_4), the minimum value of deviance can not be estimated accurately. For this reason, we can calculate AIC and BIC by replacing the minimum value of deviance with $D(\bar{\theta})$ and $\bar{D}$ (that is the posterior mean of the deviance). Table 4.5 contains the estimates of the versions of AIC and BIC with $D(\bar{\theta})$ and $\bar{D}$.

models	$AIC_{D(\bar{\theta})}$	$BIC_{D(\bar{\theta})}$	$AIC_{\bar{D}}$	$BIC_{\bar{D}}$
F_1	3588	3683	3736	3831
F_2	3829	3967	3466	3604
F_3	3300	3475	3350	3525
F_4	3299	3506	3240	3447
F_5	3354	3588	3280	3513

Table 4.5. Version of AIC and BIC with $D(\bar{\theta})$ and $\bar{D}$ for EFA

We should choose the 3rd model according to the values of $BIC_{D(\bar{\theta})}$. On the other hand, the values of both $AIC_{\bar{D}}$ and $BIC_{\bar{D}}$ lead to the conclusion that F_4 shows the best fit of the five models.

Schizotypal traits	Factor 1	Factor 2	Factor 3	Factor 4
Ideas of reference	0.744			
Odd beliefs or magical thinking	0.546	0.366		
Unusual perceptual experience	0.608	0.227		
Odd speech	0.490	0.237		0.313
Suspiciousness	0.617		0.211	
Constricted affect			0.435	0.340
Odd behaviour	0.563			
No close friends			0.475	0.228
Social anxiety			0.226	

Table 4.6 Posterior means of factor loadings of F_4 in Bayesian EFA (loadings with absolute values $\leq$0.2 are eliminated from the table)

The posterior means of the factor loadings of the 4-factor model with the method of lower triangular matrix as well as their standard deviations are given in table 4.6.

Schizotypal traits	**Factor 1**	**Factor 2**	**Factor 3**	**Factor 4**
Ideas of reference	-0.679 (*0.093*)		-0.241	
Odd beliefs or magical thinking	-0.369 (*0.102*)		-0.452	-0.295
Unusual perceptual experience	-0.502 (*0.099*)		-0.400 (0.164)	
Odd speech	-0.352 (*0.1*)		-0.538 (0.160)	
Suspiciousness	-0.656 (*0.099*)			
Constricted affect		0.244 (*0.266)*	-0.292 (*0.353*)	0.388 (*0.33*)
Odd behaviour	-0.494 (*0.096*)		-0.305 (0.135)	
No close friends		0.331 (0.282)	-0.291 (*0.361*)	0.277 (*0.402*)
Social anxiety			0.300 (*0.223*)	

Table 4.7 Transformed posterior means of factor loadings of F_4 in Bayesian EFA (loadings with absolute values $\leq$0.2 are eliminated from the table)

Table 4.7 presents the posterior means of the factor loadings after the transformation of the factor loadings of F_4 with the method of singular value decomposition.

The means of the posterior distributions of the factor loadings of the rest of the models are provided in appendix B (p.87).

The first factor receives loadings from six variables: *ideas of reference, odd beliefs or magical thinking, unusual perceptual experience, odd speech, suspiciousness* and *eccentric behaviour*. The second factor is related only to *constricted affect* and *no close friends.* The third factor receives loadings from all schizotypal traits except for *suspiciousness* but the strongest loadings are those of *odd speech, odd beliefs or magical thinking* and *unusual perceptual experience*. It can be considered as a general factor of schizotypy. The fourth factor is weakly related to *odd beliefs or magical thinking*, *constricted affect* and *no close friends*.

4.3 Comparison between frequentistic and Bayesian analysis

The results of the classical analysis of the third, fourth and fifth models (see appendix B, p.87) should be interpreted with caution since Heywood cases were observed during estimation (see section 2.5.2.1, p.17). However, according to both frequentistic and Bayesian analyses the four-factor model presents the best fit. The posterior means of the Bayesian analysis with the parameterisation of the lower triangular matrix were transformed using singular value decomposition in order to satisfy the parameterisation of classical analysis (see section 3.9, p.45). The transformed loadings of the second model resemble the loadings of the corresponding classical loadings of the second model (see appendix B, p.87).

A major advantage of Bayesian computation is the avoidance of Heywood cases. This helps us obtain stable estimates of model parameters. In the Bayesian approach we propose to use the parameterization of lower triangular matrix proposed by Lopes and West (2004) and then use Singular Value Decomposition transformation to get estimates comparable to the classical orthogonal factor model which is available in standard software packages such as SPSS and LISREL.

CHAPTER 5

APPLICATION OF CONFIRMATORY FACTOR ANALYSIS IN SCHIZOTYPIC DATA

5.1 Introduction

In this chapter we examine five factor models using CFA based on psychiatric theory proposed in the related literature. The first is the standard one-factor model (m_1) that coincides with the first model of the EFA fitted in section 4.2 (p.51). The second is a 2-factor model (m_2) introduced by Kendler et al. (1991), the third one is called disorganized 3-factor model introduced by Raine et al. (1994) (denoted by m_3), the fourth is a 4-factor model (m_4) proposed by Stefanis et al. (2004) and the last model (m_5) is a 5-factor model and was introduced by Fogelson et al. (1999).

Different number of factors and underlying factor structure is prespecified in the above models by psychiatric arguments and scenarios. The four fitted models and their corresponding structure are illustrated in table 5.1.

		Schizotypal traits								
MODEL	**FACTOR**	***IR***	***MT***	***UPE***	***S***	***SA***	***NCF***	***CA***	***OB***	***OS***
1-factor	Factor 1	#	#	#	#	#	#	#	#	#
Kendler's 2-factor	Positive	#	#	#	#	#				#
	Negative				#	#	#	#	#	
Disorganized 3-factor	Cognitive/ Perceptual	#	#	#	#					
	Interpersonal				#	#	#	#		
	Disorganized								#	#
Paranoid 4-factor	Cognitive/ Perceptual		#	#						
	Negative				#	#	#	#		
	Disorganized								#	#
	Paranoid	#			#	#				
Fogelson et al.5-factor	Paranoid	#			#					
	Positive	#	#	#						
	Schizoid						#	#		#
	Avoidant	#				#				
	Disorganized				#			#	#	

Table 5.1. Table of fitted factor models

(***IR*: ideas of reference, *MT*: odd beliefs or magical thinking, *UPE*: unusual perceptual experiences, *S*: suspiciousness, *SA*: social anxiety, *NCF*: no close friends, *CA*: constricted affect, *OB*: odd behaviour, *OS*: odd speech**)

 # the factor is related to the corresponding schizotypal trait

5.2 One factor confirmatory model

Figure 5.1 presents the path diagram of the first model. All schizotypal traits are connected with the first factor. This factor can be interpreted as the general factor of schizotypy. The factor scores are a measure of schizotypy of the patient.

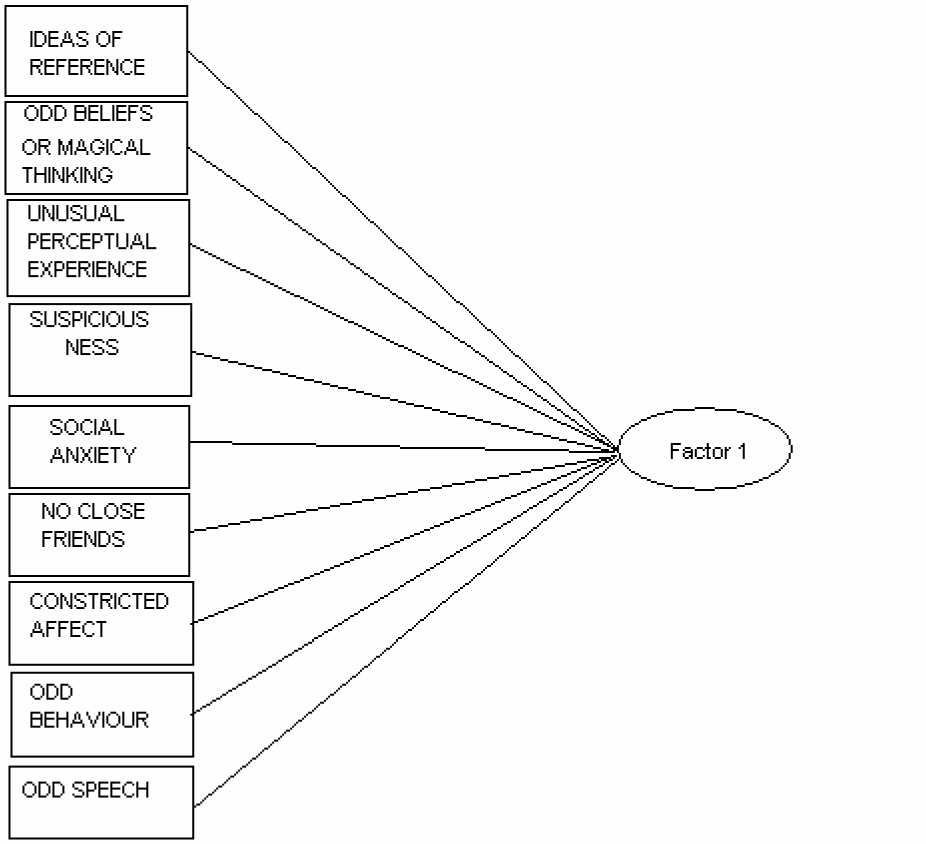

Figure 5.1. Path diagram for m_1 confirmatory factor model

5.3 Kendler's two-factor model

Kendler's 2-factor model (1991) assumes a *positive* factor related to ideas of reference, odd beliefs or magical thinking, unusual perceptual experiences, suspiciousness, social anxiety and odd speech and a *negative* factor that relates to suspiciousness, social anxiety, no close friends, constricted affect and odd behaviour. The path diagram of Kendler's model is displayed in figure 5.2.

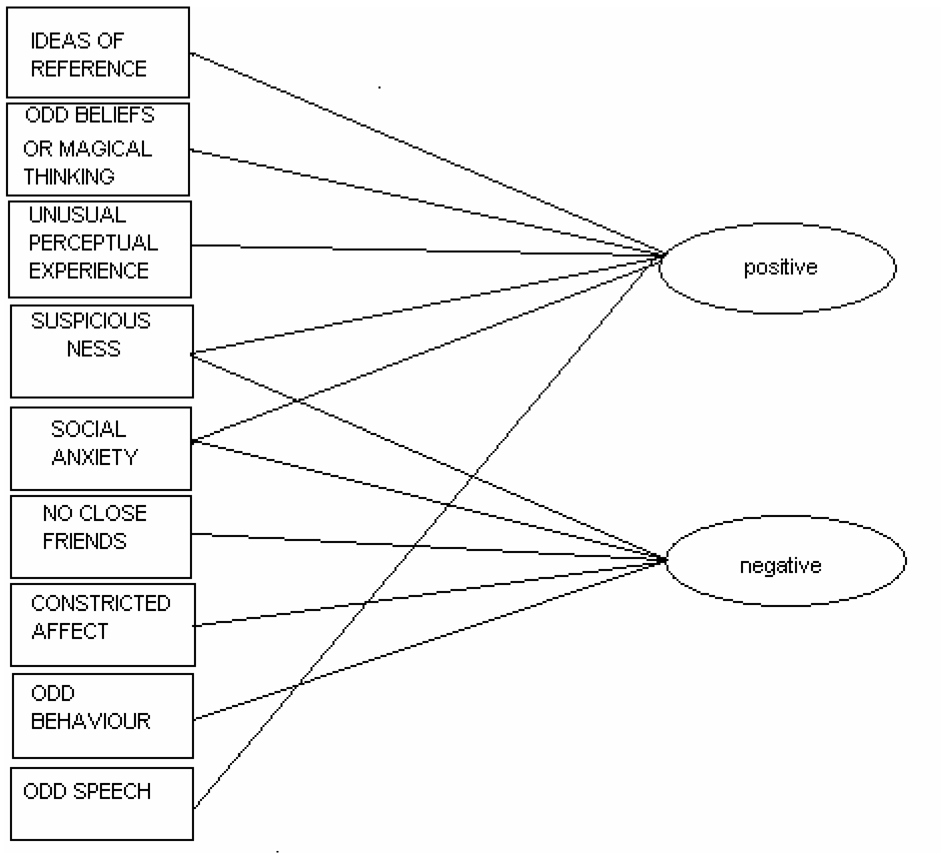

Figure 5.2. Path diagram of Kendler's 2-factor model

This model depicts the typical concept of negative-positive factors of schizotypy. The positive factor reflects aspects of cognitive-perceptual dysfunction while the negative factor is thought to reflect deficits in interpersonal functioning. However, Kendler's model differentiates with respect to the traditional 2-factor model in two points:

1. Suspiciousness and social anxiety load on both factors, while in the traditional 2-factor models they have been viewed as belonging to positive and negative factors respectively, and
2. Odd behavior belongs to the negative rather than the positive factor.

5.4 Disorganized three-factor model

The model of Raine (1994) is the most popular of all and consists of 3 factors: the *cognitive-perceptual* factor (ideas of reference, odd beliefs or magical thinking, unusual perceptual experiences and suspiciousness), the *interpersonal* factor (social anxiety, no close friends, constricted affect and suspiciousness) and the *disorganized* factor (odd behavior, odd speech). The path diagram of Raine's model is displayed in figure 5.3.

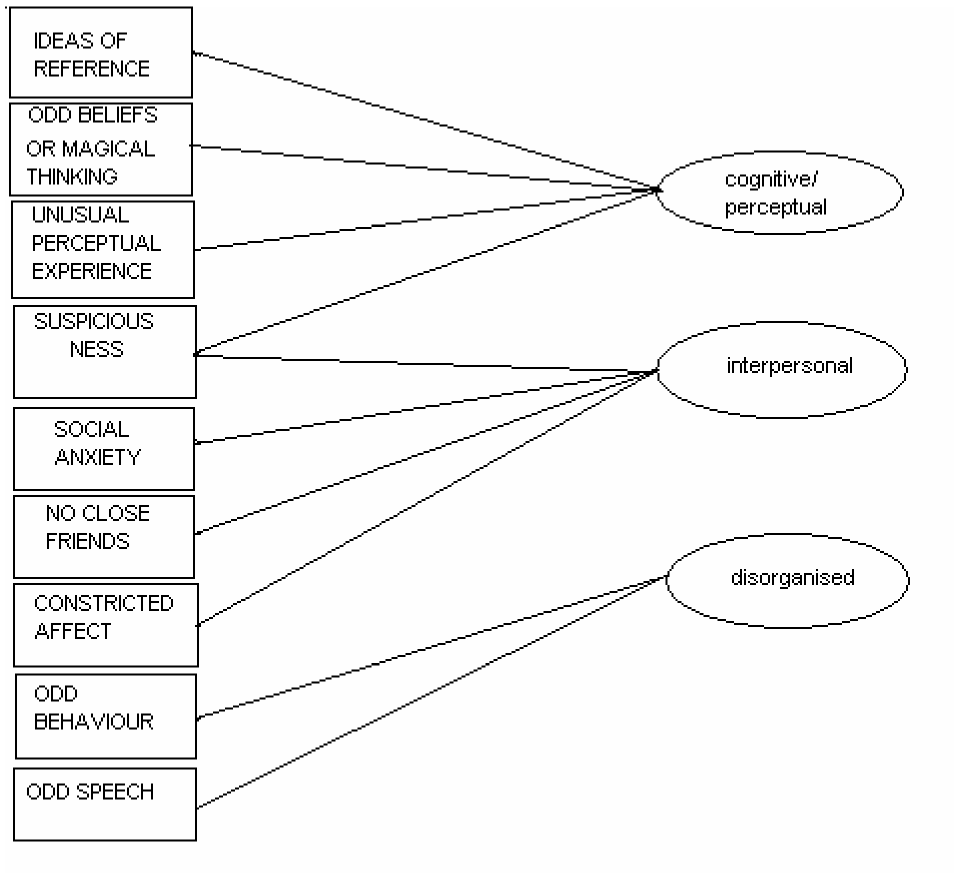

Figure 5.3. Path diagram of disorganized 3-factor confirmatory model

In Raine's model, odd behavior and odd speech form a third disorganization factor that reveals a cognitive and behavioral disorganization, while the other two factors suggest a latent trait of positive (cognitive-perceptual) deficits and negative (interpersonal) deficits respectively.

5.5 Paranoid four-factor model

The 4-factor model of Stefanis et al. (2004) assumes that four factors are related to the nine schizotypal traits: cognitive-perceptual, negative, disorganized and paranoid. The path diagram of the paranoid 4- factor model is presented in figure 5.4.

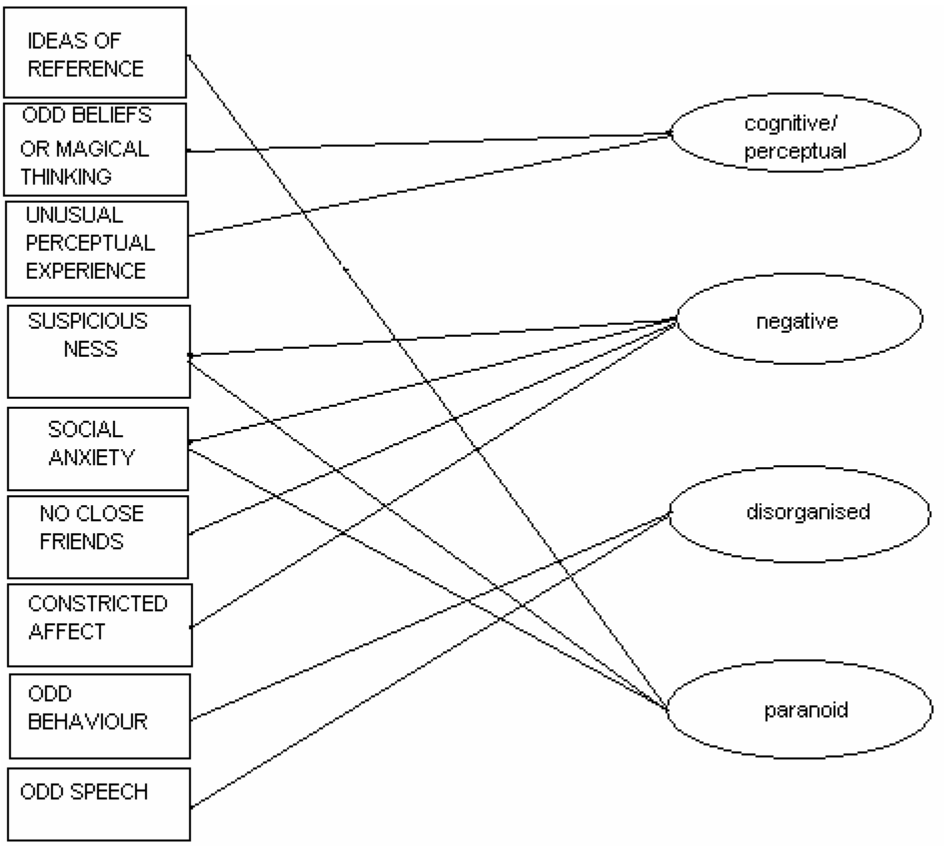

Figure 5.4. Path diagram of paranoid 4-factor model

This model was created by splitting the positive schizotypal traits (1st factor in m_2) into a paranoid and a cognitive-perceptual factor that was found to have a better fit to the data of the ASPIS study; see Stefanis et al. (2004). The *cognitive-perceptual* factor relates to odd beliefs or magical thinking, and on unusual perceptual experiences. The *negative* factor is allowed to receive loadings from suspiciousness, social anxiety, no close friends, constricted affect and it is a measure of symptoms of negative schizotypal traits. The *disorganized* factor of this model receives loadings from odd speech and odd behavior while ideas of reference, suspiciousness and social anxiety were allowed to load on the *paranoid* factor.

The existence of a separate paranoid factor is based on several studies (Stuart et al., 1995; Kay and Sevy, 1990; Bassett et al. 1994; Peralta and Cuesta 1998, 1999) that have suggested discrimination of positive traits into cognitive/perceptual and paranoid. According to Stefanis et al. (2004), because of the relative independence between cognitive-perceptual and paranoid factor, delusions and paranoia seems to have a psychological motivation rather than be created by abnormal perceptual experiences.

5.6 Fogelson et al. five-factor model

The last model (m_5) assumes that 5 factors are needed to explain the covariance between the schizotypal traits, that is paranoid, positive, schizoid, avoidant and disorganized latent factors. Figure 4.5 presents the path diagram of the Fogelson et al. (1999) 5- factor model. This model is the most complicated of all. It represents the idea of multidimensional concept in schizotypy, which is supported by many researchers.

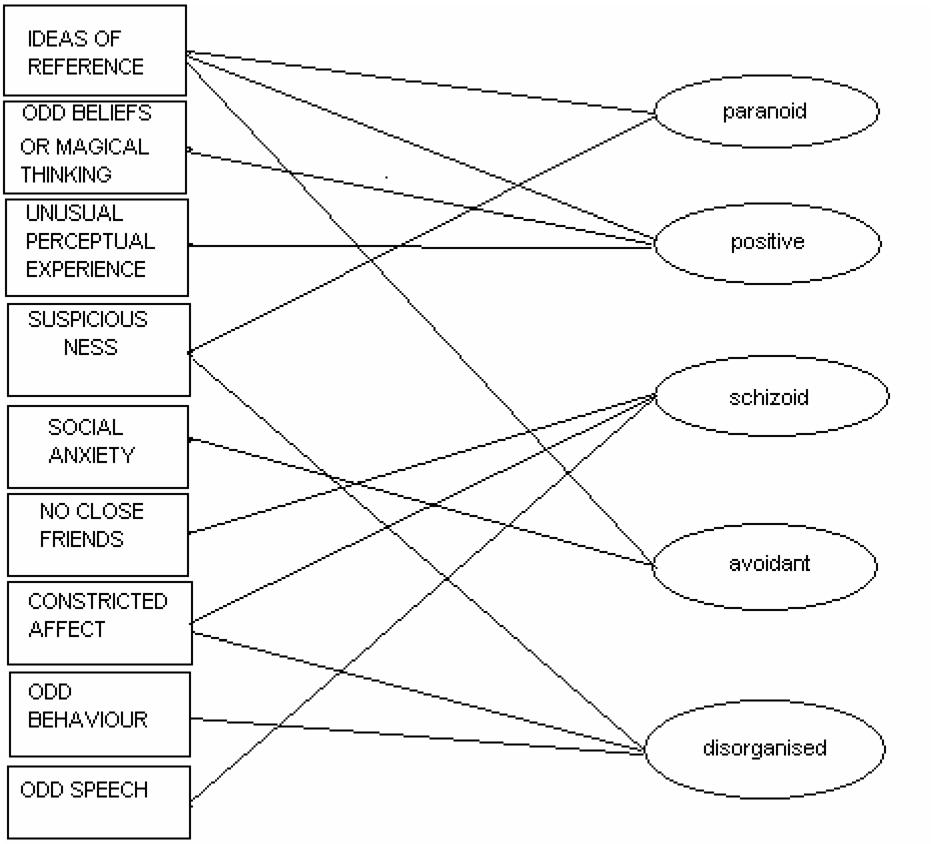

Figure 5.5. Path diagram of Fogelson et al. 5-factor model

5.7 Classical analysis and interpretation of best fitted model

The five models are compared with respect to the goodness of fit statistics (see section 2.5) provided by LISREL that are presented at table 5.2.

	FITTED MODELS				
STATISTICS OF FIT	$\mathbf{m_1}$	$\mathbf{m_2}$	$\mathbf{m_3}$	$\mathbf{m_4}$	$\mathbf{m_5}$
p-value	0.000	0.000	0.0346	0.140	0.001
AIC	163.35	135.37	94.98	87.93	111.29
CAIC	237.47	221.85	185.58	194.99	243.07
ECVI	0.984	0.815	0.572	0.530	0.670
NFI	0.788	0.832	0.910	0.938	0.920
NNFI	0.764	0.797	0.915	0.940	0.829
CFI	0.823	0.865	0.946	0.967	0.938
GFI	0.854	0.884	0.936	0.954	0.940
AGFI	0.757	0.792	0.875	0.891	0.794
PGFI	0.513	0.474	0.478	0.403	0.272
$\mathbf{F_0}$	0.609	0.418	0.169	0.102	0.207
RMSEA	0.150	0.132	0.086	0.0733	0.126

Table 5.2. Goodness of fit statistics for the fitted models

(AIC: Akaike Information Criterion, CAIC: Consistent AIC, ECVI: single sample Cross Validation Index, NFI: Normed Fit Index, NNFI: Not-Normed Fit Index, CFI: Comparative Fit Index, GFI: Goodness of Fit Index, AGFI: Adjusted GFI, PGFI: Parsimony Goodness of Fit Index, $\mathbf{F_0}$: population discrepancy function, RMSEA: Root Mean Square Error of Approximation)

The indices that are based on chi-square value can be used to compare the fitted models. The fourth model (m_4) gives the lowest values of ECVI and AIC, while CAIC indicates that m_3 seems to present a better fit compared to m_4. On the other hand, most of the fit indices (NFI, NNFI, CFI, PGFI) show that the best fit is accomplished when m_4 is fitted. The values of the other goodness of fit indices (GFI, AGFI) of m_4 take values up to 0.9 with the exception of AGFI that takes the value of 0.891. The value of RMSEA for m_4 is below 0.08 and indicates a moderate fit. In addition F_0 , decreases significantly from 0.169 of m_3 to 0.102 of m_4 indicating a decrease in the error of approximation in the population when m_4 is fitted.

After the comparison of the goodness of fit statistics of the five fitted models, we conclude that m_4 describes better the structure of the schizotypal traits and the underlying factors. The path diagram of m_4 is presented at figure 5.6.

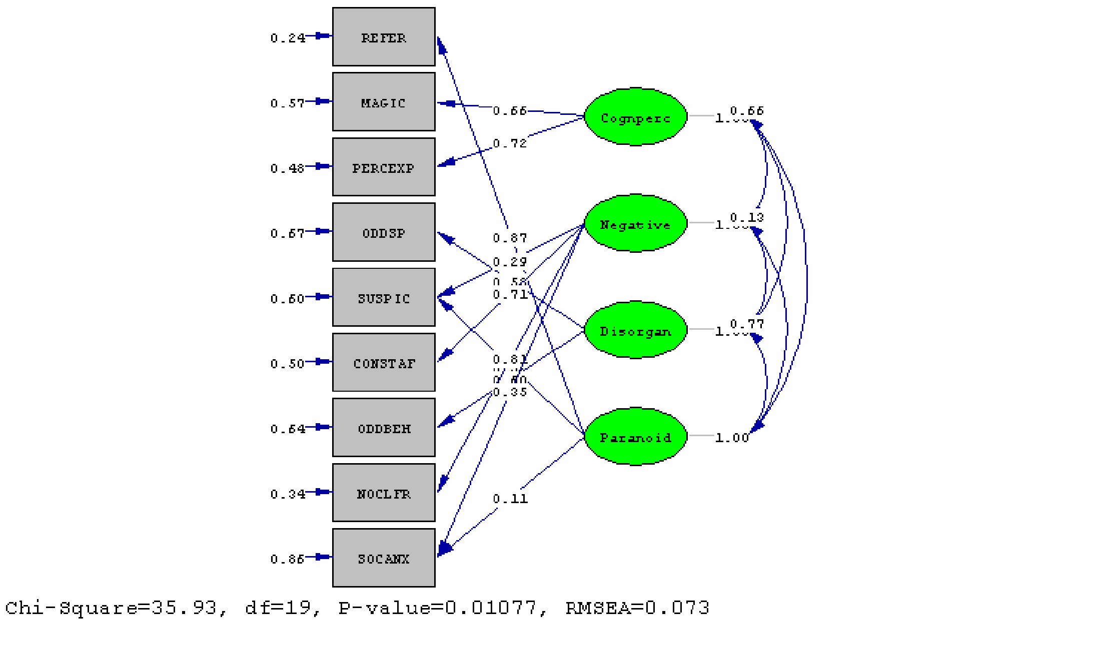

Figure 5.6. Path diagram for fitted m_4

The loadings of the fourth model are presented at table 5.3. the loadings of the other four models as well as the correlations of the factors are presented in appendix B (p.87).

	Factors			
Schizotypal traits	**Paranoid**	**Cognitive/ perceptual**	**Disorganized**	**Negative**
Ideas of reference	0.874			
Odd beliefs or magical thinking		0.655		
Unusual perceptual experiences		0.722		
Odd speech			0.577	
Suspiciousness	0.523			0.29
Constricted affect				0.709
Odd behaviour			0.596	
No close friends				0.815
Social anxiety	0.109			0.349

Table 5.3. Factor loadings of paranoid 4-factor model

The schizotypal traits share from moderate to high factor loadings with the factors except for the loading of social anxiety with the paranoid factor that is low: 0.109.

Table 5.4 presents the correlation matrix of the four factors, as well as their standard errors in parenthesis.

	Cognitive/ perceptual	**Negative**	**Disorganized**	**Paranoid**
Cognitive/ perceptual	1			
Negative	0.304 (*0.105*)	1		
Disorganized	0.910 (*0.105*)	0.650 (*0.105*)	1	
Paranoid	0.663 (*0.093*)	0.132 (*0.102*)	0.771 (*0.110*)	1

Table 5.4. Correlation matrix of paranoid 4-factor model

5.8 Bayesian analysis

5.8.1 Priors

Five models were fitted through Bayesian CFA: m_1, m_2, m_3, m_4, m_5. Their structure is presented at table 5.1. In the models under consideration we assume correlated factors. The prior for the precision matrix of the factors was selected to be the Wishart distribution with 100 degrees of freedom for m_2, m_3, m_4 and 110 for m_5. The degrees of freedom of the Wishart distribution were chosen so as to have the posterior variances of the factors approximate one, for identifiability reasons (see section 2.2.1, p.6). The priors of the loadings were chosen to be univariate standard normal distribution while, the priors of some loadings were chosen to be normal distributions truncated at zero (without this constraint, different chains converged at the same loadings with opposite signs. In addition, the errors were assumed to be uncorrelated.

5.8.2 Gibbs sampling

Since the first model of confirmatory analysis (m_1) is the same with the 1-factor model of exploratory analysis, results for m_1 can be obtained from the analysis of section 3.2. For the second model of Bayesian CFA a sample of 100000 values was enough after a 50000 sample of burn-in iterations in order to obtain convergence. Two chains were created for every model and the convergence was tested through CODA of S-plus. The third model m_3, needed 50000 burn-in sample and 150000 subsequent iterations in order to converge. In addition, the fourth model m_4 converged after a 70000 burn-in sample and 150000 iterations and the convergence of the fifth model was obtained after 70000 burn-in sample and a sample of 200000 iterations from the posterior distribution of the estimated parameters.

5.8.3 Results

Model comparison was based on AIC, BIC and DIC that were presented in section 3.7. The number of free parameters is assumed to be the number of estimated factor loadings, plus the number of the variances of the errors, plus the number of the correlations of the latent factors.

The minimum value of the deviance was used in order to calculate the AIC and BIC. The results of the calculation of AIC and BIC as well as the DIC that are provided by WinBUGS are presented at table 5.5.

	t	DIC	AIC	BIC
$\mathbf{m_1}$	18	3847	3667	3762
$\mathbf{m_2}$	23	3775	3325	3447
$\mathbf{m_3}$	25	3708	3117	3250
$\mathbf{m_4}$	30	3629	2822	2981
$\mathbf{m_5}$	37	3576	2616	2812

Table 5.5. Information Criteria for the five fitted models (DIC: Deviance Information Criterion, AIC: Akaike Information Criterion, BIC: Bayesian Information Criterion, t: number of parameters)

models	$\mathbf{AIC_{D(\bar{\theta})}}$	$\mathbf{BIC_{D(\bar{\theta})}}$	$\mathbf{AIC_{\bar{D}}}$	$\mathbf{BIC_{\bar{D}}}$
$\mathbf{m_1}$	3588	3683	3736	3831
$\mathbf{m_2}$	3295	3417	3558	3680
$\mathbf{m_3}$	3113	3246	3435	3568
$\mathbf{m_4}$	2837	3996	3563	3422
$\mathbf{m_5}$	2678	2874	3169	3365

Table 5.6. Versions of AIC and BIC with $D(\bar{\theta})$ and $\bar{D}$ for CFA

Table 5.6 presents the values of Akaike information Criterion and Bayesian Information criterion calculated with the values of the mean of deviance ($\bar{D}$) and the values of deviance estimated at the mean of the parameters ($D(\bar{\theta})$).

Models	**mean**	**2,50%**	**median**	**97,50%**	**minimum**	**$D(\bar{\theta})$**
m_1	3700 (*20.81*) *(0.118)*	3662	3700	3743	3631	3552
m_2	3512 *(39.65)* *(0.367)*	3431	3513	3588	3279	3249
m_3	3385 *(52.52)* *(0.53)*	3270	3389	3479	3067	3063
m_4	3203 *(72.38)* *(0.749)*	3050	3207	3334	2762	2777
m_5	3094 *(109.4)* *(1.39)*	2868	3098	3297	2542	2604

Table 5.7. Point estimates of deviance for Bayesian CFA

Table 5.7. presents point estimates of deviance. The first parenthesis below the mean of the deviance contains the standard deviation of the estimate and the second parenthesis contains the MC error of the estimate. The minimum values of the deviance at the five models do not differ significantly from the values of $D(\bar{\theta})$. This is an indication of a satisfactory approximation of the deviance.

According to the values of all the calculated Information criteria presented at tables 5.5 and 5.6 m_5 provides the best fit among the fitted models. As a consequence, m_5 is selected to be the most appropriate model to explain the covariance structure of the nine schizotypal traits.

	FACTORS				
Schizotypal traits	***Paranoid***	***Positive***	***Schizoid***	***Avoidant***	***Disorganized***
Ideas of reference	0.456 (*0.144*)	0.3816 (*0.111*)		-0.08 (*0.125*)	
Odd beliefs or magical thinking		0.604 (*0.098)*			
Unusual perceptual experiences		0.675 (*0.098*)			
Odd speech			0.466 (*0.085*)		
Suspiciousness	0.602 (*0.152)*				0.255 (*0.121*)
Constricted affect			0.808 (*0.107*)		-0.14 (*0.091*)
Odd behaviour					0.738 (*0.110*)
No close friends			0.662 (*0.091)*		
Social anxiety				0.699 (*0.18*)	

Table 5.8. Factor loadings of Fogelson et al. 5-factor model with correlated factors

The loadings of the fifth model as well as their standard deviations are presented at table 5.8. The loadings of the other four models as well as the correlations of the corresponding latent factors are provided in appendix B (p.87). Ideas of reference loads weekly on the avoidant factor as well as the schizotypal trait of constricted affect that loads weekly on the disorganized factor. In this way, the avoidant factor receives loadings only from social anxiety and odd behaviour is the only schizotypal trait that loads on the disorganized factor. The remaining factor loadings appear from moderate (0.255) to high (0.808) values.

	Paranoid	Positive	Schizoid	Avoidant	Disorganized
Paranoid	1.029 (*0.155*)				
Positive	0.223 (*0.109*)	1.102 (*0.166*)			
Schizoid	0.197 (*0.098*)	0.23 (*0.093*)	1.079 (*0.161*)		
Avoidant	0.129 (*0.095*)	0.178 (*0.095*)	0.223 (*0.096*)	1.008 (*0.146*)	
Disorganized	0.211 (*0.112*)	0.365 (*0.105*)	0.306 (*0.102*)	0.105 (*0.094*)	1.118 (*0.171*)

Table 5.9. Covariance matrix of Fogelson et al. 5- factor model

The above table presents the estimates of covariance between the factors of the fitted 5- factor model with their standard errors in parenthesis. The disorganized factor shares the highest value of covariance with the positive factor, that is 0.365. In addition, the values of covariance of all the other combinations of factor remain in low to moderate values.

5.9 Comparison between frequentistic and Bayesian confirmatory analyses of data

The classical analysis is based on the iterative method of ML. In Stefanis et al. (2004) the 4-factor was the best fitted model according to the results of classical analysis while we conclude that Fogelson et al 5- factor model presented the lowest values of fit indices among the five models that were fitted through Bayesian analysis.

CHAPTER 6

DISCUSSION AND FURTHER RESEARCH

6.1 Discussion

In this thesis we have examined the dimensionality of schizotypy and SPD using Bayesian theory. At the early stages of research schizotypic disorder was assumed to be one or two-dimensional. But further and recent research revealed a multidimensional structure of SPD. The three factors of Raine et al. (1994) were not confirmed by subsequent studies (see Bergman, 1996; Stefanis et al., 2004).

Stefanis et al. (2004) introduced the paranoid 4-factor model which was also validated by the data of the present thesis using classical factor analytic methods. However, it was not confirmed by Bayesian factor analytic methods used in this thesis. Instead, the 5-factor model of Fogelson et al. (1999) seems to provide a better fit among the five fitted, according to specific information criteria (see section 5.7, p.66). The estimated factor loadings of models m_1 - m_5, though, were approximately, identical with the two approaches (see appendix B, p. 92-98).

In the Bayesian approach of exploratory analysis, the factor loadings were assumed to form a lower triangular matrix with positive diagonal elements. (see section 3.8, p.39). The estimated parameters presented high autocorrelation. Therefore, the method of overrelaxation and selection of a subsample of the output was used, a fact that retarded the convergence a lot.

The use of singular value decompositition allowed comparing the two different parameterizations of classical and Bayesian analyses since the loadings have been transformed according to the classical constraint (see section 3.9, p.45). However, the classical analysis presented Heywood cases (see section 2.5.2.1, p.17), consequently the values of the loadings and the unique variances should be interpreted with caution. In fact, the classical analysis of the second model (that did not presented Heywood cases) gave almost identical loadings with the Bayesian analysis. This is an indication that the Bayesian analysis of the 3rd, 4th and 5th model provide stable estimates of factor loadings without being distorted by Heywood cases. Nevertheless, Bayesian and non-Bayesian analysis revealed a multidimensional factorial structure, a four and a 5-factor model respectively. The two models should be further examined in a confirmatory way on different data.

6.2 Further research

6.2.1 Two-stage factor analysis

The nature of the schizotypic data used at the present dissertation is such that a two-stage factor analysis may be easily implemented. The questionnaire consists of 74 items with dichotomous yes/no responses. Therefore, a reasonable assumption is that each variable follows Bernoulli distribution.

In a two-stage analysis CFA model, is implemented in two levels. At the first stage, for schizotypic data and SPQ (see section 4.1), each item is associated to one of the nine schizotypal traits they belong to. At the second stage, these schizotypic traits load on specific factors under a specific model.

In figure (6.1) the model of two-stage factor analysis is presented, where with y_i , $i = 1,...,15$ the observed variables, with k_j, $j = 1,2,3$ the first stage factors and with f_1 the second stage one factor representing an overall schizotypy score.

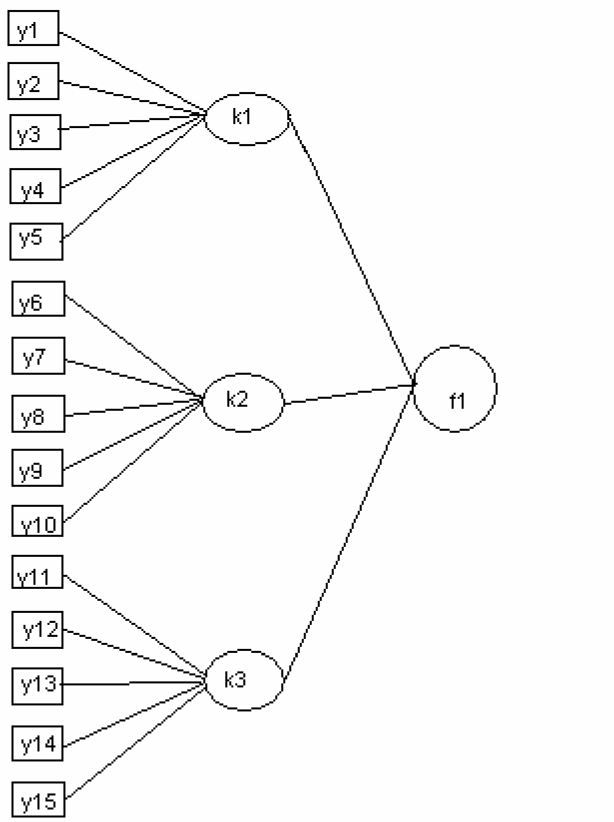

Figure 6.1. A path diagram for a two-stage factor analysis model

Such models can be constructed using higher-order factor analysis in its general set up; for details see Bollen (1989a, p.313). Ansari and Jedidi (2000) describe the use of MCMC procedures for performing factor analysis of multilevel binary data.

6.2.2 Logit factor model

At the present thesis, the responses used in the factor analysis model that is presented in equation (2.1) consist of the values of nine schizotypal variables. These values were created by summing the binary responses of specific answers of SPQ (see section 4.1, p.47) and they were transformed into proportions percent.

The factor analysis model used is based on the assumption that the errors and as a consequence the dependent variables follow the Normal distribution. Generally, the nine schizotypal subscales are skewed and often do not meet this assumption. However, the analysis was made under the assumption of normality, since this is the standard approach in psychiatric research.

A more realistic model, called the *logit factor model*, can be constructed assuming that all variables follow the binomial distribution. The nine schizotypal traits, though, are in reality, binomial variables since they are the sums of specific binary responses.

Let us assume n observations of p manifest variables Y_i each following binomial distribution, that is

$$Y_i \sim B(N_i, \pi_i), \text{ for } i = 1,...,p \qquad (6.1)$$

where N_i is the number of independent Bernoulli coordinates of each binomial and π_i the probability of "success" of the Bernoulli distributions.

In our case, N_i is the number of the items-questions of each of schizotypic scale and π_i is the probability of a positive response to each of question or item. In addition, the responses of each item are supposed to be independent.

The logit model is defined as:

$$\log it_{\pi_i}(f) = \log \frac{\pi_i(f)}{1-\pi_i(f)} = a_{i0} + \sum_{j=1}^{q} a_{ij} f_j \qquad , \text{ for } i = 1,...,p \qquad (6.2)$$

where a_{i0} is the overall mean, a_{ij} are unknown coefficients and f_j are the q assumed latent factors. The factors are also assumed to be continuous and follow the standard normal distribution, that is:

$$f_j \sim N(0,1) \ , \ \text{for } j = 1,...,q. \qquad (6.3)$$

In addition, the p observed variables are assumed to be independent conditional on the latent variables.

The interpretation of equation (6.2) is quite complicated since it is not the familiar linear model. The effect of a unit change in f_1, for instance, is to increase the log odds $\left(\log\frac{\pi_i}{1-\pi_i}\right)$ by an amount of a_{i1}.

Alternatively, the probability of a positive response is:

$$\pi_i = \frac{\exp(a_{i0} + a_{i1}f_1 + a_{i2}f_2 + \ldots + a_{iq}f_q)}{1+\exp\left(a_{i0} + a_{i1}f_1 + a_{i2}f_2 + \ldots + a_{iq}f_q\right)} \quad . \tag{6.4}$$

McCullagh and Nelder (1983) refer to the generalized linear models with logit link while Bartholomew et al. (2002) present the logit factor model with binary observations.

The implementation of the model (6.2) can be done with programs like LISREL (see Jöreskog and Sörbom, 1996) and in Bayesian approach using WinBUGS (see Spiegelhalter et al., 2003).

APPENDIX A

Items for the nine subscales in the final 74-item version of the Schizotypal Personality Questionnaire

Ideas of reference

1. Do you sometimes feel that things you see on the TV or read in the newspaper have a special meaning for you?

10. I am aware that people notice me when I go out for a meal or to see a film.

19. Do some people drop hints about you or say things with a double meaning?

28. Have you ever noticed a common event or object that seemed to be a special sign for you?

37. Do you sometimes see special meanings in advertisements, shop windows, or in the way things are arranged around you?

45. When shopping do you get the feeling that other people are taking notice of you?

53. When you see people talking to each other, do you often wonder if they are talking about you?

60. Do you sometimes feel that other people are watching you?

63. Do you sometimes feel that people are talking about you?

Excessive Social Anxiety

2. I sometimes avoid going to places where there will be many people because I will get anxious

11. I get very nervous when I have to make polite conversation.

20. Do you ever get nervous when someone is walking behind you?

29. I get anxious when meeting people for the first time.

38. Do you often feel nervous when you are in a group of unfamiliar people?

46. I feel very uncomfortable in social situations involving unfamiliar people.

54. I would feel very anxious if I had to give a speech in front of a large group of people.

71. I feel very uneasy talking to people I do not know well.

Odd beliefs or Magical Thinking

4.Have you had experiences with the supernatural?

12. Do you believe in telepathy (mind-reading)?

21. Are you sometimes sure that other people can tell what you are thinking?

30. Do you believe in clairvoyance (psychic forces, fortune telling)?

39. Can other people feel your feelings when they are not there?

47. Have you had experiences with astrology, seeing the future, UFOs, ESP or a sixth sense?

55. Have you ever felt that you are communicating with another person telepathically (by mind-reading)?

Unusual Perceptual Experiences

4. Have you often mistaken objects or shadows for people, or noises for voices?

13. Have you ever had the sense that some person or force is around you, even though you cannot see anyone?

22. When you look at a person or yourself in a mirror, have you ever seen the face change right before your eyes?

31. I often hear a voice speaking my thoughts aloud.

40. Have you ever seen things invisible to other people?

48. Do everyday things seem unusually large or small?

56. Does your sense of smell sometimes become unusually strong?

61. Do you ever suddenly feel distracted by distant sounds that you are not normally aware of?

64. Are your thoughts sometimes so strong that you can almost hear them?

Odd or Eccentric Behavior

5. Other people see me as slightly eccentric (odd).

14. People sometimes comment on my unusual mannerisms and habits.

23. Sometimes other people think that I am a little strange.

32. Some people think that I am a very bizarre person.

67. I am an odd, unusual person.

70. I have some eccentric (odd) habits.

74. People sometimes stare at me because of my odd appearance.

No Close Friends

6. I have little interest in getting to know other people.

15. I prefer to keep to myself.

24. I am mostly quiet when with other people

33. I find it hard to be emotionally close to other people.

41. Do you feel that there is no one you are really close to outside of your immediate family or people you can confide in or talk to about personal problems?

49. Writing letters to friends is more trouble than it is worth.

57. I tend to keep in the background on social occasions.

62. I attach little importance to having close friends.

66. Do you feel that you are unable to get "close" to people?

Odd Speech

7. People sometimes find it hard to understand what I am saying.

16. I sometimes jump quickly from one topic to another when speaking.

25. I sometimes forget what I am trying to say.

34. I often ramble on too much when speaking.

42. Some people find me a bit vague and elusive during a conversation.

50. I sometimes use words in unusual ways.

58. Do you tend to wander off the topic when having a conversation?

69. I find it hard to communicate clearly what I want to say to people.

72. People occasionally comment that my conversation is confusing.

Constricted Affect

8. People sometimes find me aloof and distant.

17. I am poor at expressing my true feelings by the way I talk and look.

26. I rarely laugh and smile.

35. My "non-verbal" communication (smiling and nodding during a Y N conversation) is poor.

43. I am poor at returning social courtesies and gestures.

51. I tend to avoid eye contact when conversing with others.

68. I do not have an expressive and lively way of speaking.

73. I tend to keep my feelings to myself.

Suspiciousness

9. I am sure I am being talked about behind my back.

18. Do you often feel that other people have got it in for you?

27. Do you sometimes get concerned that friends or co-workers are not really loyal or trustworthy?

36. I feel I have to be on my guard even with friends.

44. Do you often pick up hidden threats or put-downs from what people say or do?

52. Have you found that it is best not to let other people know too much about you?

59. I often feel that others have it in for me.

65. Do you often have to keep an eye out to stop people from taking advantage of you?

APPENDIX B

1. Loadings of exploratory factor analysis models.

Model 1

F_1	Factor 1 *(classical)*	Unique variance *(classical)*	Factor 1 *(Bayesian)*	Unique variance *(Bayesian)*
Ideas of reference	0.631	*0.602*	0.622	*0.083*
Odd beliefs or magical thinking	0.542	*0.706*	0.548	*0.084*
Unusual perceptual experience	0.663	*0.561*	0.668	*0.080*
Odd speech	0.594	*0.647*	0.601	*0.082*
Suspiciousness	0.600	*0.640*	0.605	*0.082*
Constricted affect	0.410	*0.832*	0.417	*0.089*
Odd behaviour	0.615	*0.621*	0.620	*0.081*
No close friends	0.433	*0.812*	0.436	*0.088*
Social anxiety	0.334	*0.888*	0.347	*0.087*

Table B.1. Factor loadings and unique variance of F_1 exploratory model with classical and Bayesian analysis

Model 2

F_2	**Factor 1 (*classical*)**	**Factor 2 (*classical*)**	**Unique Variance (*classical*)**	**Factor 1 (*Bayesian*)**	**Factor 2 (*Bayesian*)**	**Unique Variance (*Bayesian*)**
Ideas of reference	0.553	-0.436	*0.504*	-0.544	-0.464	*0.521*
Odd beliefs or magical thinking	0.456	-0.424	*0.612*	-0.509	-0.337	*0.654*
Unusual perceptual experience	0.6	-0.303	*0.549*	-0.617	-0.277	*0.568*
Odd speech	0.602		*0.634*	-0.603		*0.657*
Suspiciousness	0.575		*0.642*	-0.587		*0.652*
Constricted affect	0.613	0.532	*0.34*	-0.569	0.501	*0.451*
Odd behaviour	0.574	-0.227	*0.619*	-0.595		*0.638*
No close friends	0.589	0.421	*0.475*	-0.591	0.504	*0.431*
Social anxiety	0.377		*0.842*	-0.381		*0.854*

Table B.2. Factor loadings and unique variance of F_2 exploratory model with classical and Bayesian analysis

Model 3

F_3	**Factor 1** *classical*	**Factor 2** *classical*	**Factor 3** *classical*	**Unique Variance** *classical*	**Factor 1** *Bayesian*	**Factor 2** *Bayesian*	**Factor 3** *Bayesian*	**Unique Variance** *Bayesian*
Ideas of reference	0.5		0.419	*0.54*	-0.66		-0.291	*0.500*
Odd beliefs or magical thinking	0.244	0.286	0.589	*0.511*	-0.639			*0.580*
Unusual perceptual experience	0.426	0.342	0.401	*0.541*	-0.657			*0.557*
Odd speech	0.237	0.57	0.26	*0.551*	-0.568			*0.563*
Suspiciousnes s	0.997			*0*	-0.519		-0.380	*0.530*
Constricted affect	0.271	0.693	-0.375	*0.305*	-0.213	0.548		*0.431*
Odd behaviour	0.384	0.337	0.299	*0.65*	-0.560			*0.615*
No close friends	0.306	0.574	-0.266	*0.506*	-0.225	0.536		*0.429*
Social anxiety	0.195	0.352		*0.838*	-0.271	0.219		*0.815*

Table B.3. Factor loadings and unique variance of F_3 exploratory model with classical and Bayesian analysis

Model 4

F_4	Factor 1	Factor 2	Factor 3	Factor 4	Unique Variance
Ideas of reference	0.572			0.355	*0.545*
Odd beliefs or magical thinking	0.359			0.592	*0.506*
Unusual perceptual experience	0.493			0.433	*0.539*
Odd speech	0.348		0.434	0.374	*0.541*
Suspiciousness	0.828	0.556			*0*
Constricted affect	0.289		0.748		*0.328*
Odd behaviour	0.828	-0.555			*0*
No close friends	0.379		0.59		*0.474*
Social anxiety			0.385		*0.785*

Table B.4. Unrotated factor loadings and unique variance of F_4 exploratory model with classical analysis

Schizotypal traits	Factor 1	Factor 2	Factor 3	Factor 4	Unique Variance
Ideas of reference	-0.679 (*0.093*)		-0.241		*0.472*
Odd beliefs or magical thinking	-0.369 (*0.102*)		-0.452	-0.295	*0.559*
Unusual perceptual experience	-0.502 (*0.099*)		-0.400 (0.164)		*0.542*
Odd speech	-0.352 (*0.1*)		-0.538 (0.160)		*0.494*
Suspiciousness	-0.656 (*0.099*)				*0.462*
Constricted affect		0.244 (*0.266)*	-0.292 (*0.353*)	0.388 (*0.33*)	*0.417*
Odd behaviour	-0.494 (*0.096*)		-0.305 (0.135)		*0.559*
No close friends		0.331 (0.282)	-0.291 (*0.361*)	0.277 (*0.402*)	*0.386*
Social anxiety			0.300 (*0.223*)		*0.779*

Table B.5. Factor loadings and unique variance of F_4 in Bayesian EFA and MCMC details

Model 5

Schizotypal traits	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Unique Variance
Ideas of reference	0.572			0.340		*0.523*
Odd beliefs or magical thinking	0.361			0.681	-0.273	*0.287*
Unusual perceptual experience	0.494		0.224	0.365		*0.569*
Odd speech	0.350		0.503	0.343	0.383	*0.351*
Suspiciousness	0.828	0.556				*0*
Constricted affect	0.289		0.630	-0.223		*0.452*
Odd behaviour	0.829	-0.555				*0*
No close friends	0.381		0.670	-0.323	-0.205	*0.259*
Social anxiety			0.405			*0.789*

Table B.6. Unrotated factor loadings and unique variance of F_5 exploratory model with classical analysis

Schizotypal traits	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Unique Variance
Ideas of reference	-0.682		-0.221			*0.478*
Odd beliefs or magical thinking	-0.378		0.415	-0.320		*0.572*
Unusual perceptual experience	-0.490		-0.372	-0.217		*0.549*
Odd speech	-0.437	0.220	-0.228	-0.343		*0.490*
Suspiciousness	-0.636		-0.299	0.23		*0.473*
Constricted affect	-0.259	0.492				*0.417*
Odd behaviour	-0.492		-0.257			*0.552*
No close friends		0.462				*0.385*
Social anxiety		0.239				*0.782*

Table B.7. Factor loadings and unique variance of F_5 in Bayesian EFA

2. Factor loadings and factor correlation matrices of CFA models in classical and Bayesian analysis

m_1	Factor 1 (*classical*)	Factor 1 (*Bayesian*)
Ideas of reference	0.617	0.622
Odd beliefs or magical thinking	0.541	0.548
Unusual perceptual experience	0.656	0.668
Odd speech	0.592	0.601
Suspiciousness	0.605	0.605
Constricted affect	0.409	0.417
Odd behaviour	0.616	0.620
No close friends	0.430	0.436
Social anxiety	0.339	0.347

Table B.8. Factor loadings of m_1 confirmatory model with classical and Bayesian analysis

<u>**Model 2**</u>

m_2	Positive (*classical*)	Negative (*classical*)	Positive (*Bayesian*)	Negative (*Bayesian*)
Ideas of reference	0.639		0.646	
Odd beliefs or magical thinking	0.594		0.600	
Unusual perceptual experience	0.699		0.691	
Odd speech	0.569		0.542	
Suspiciousness	0.479	0.204	0.477	0.232
Constricted affect		0.696		0.690
Odd behaviour		0.422		0.377
No close friends		0.792		0.800
Social anxiety	0.197	0.267	0.212	0.274

Table B.9. Factor loadings of m_2 confirmatory model with classical and Bayesian analysis

	Positive	Negative
Positive	1.000	
Negative	0.440	1.000

Table B. 10. Correlation matrix of latent factors of m_2 confirmatory model with classical analysis

	Positive	Negative
Positive	1.021	
Negative	0.198	1.029

Table B. 11. Covariance matrix of latent factors of m_2 confirmatory model with Bayesian analysis

Model 3

m_3	Cognitive/ perceptual	Interpersonal	Disorganized
Ideas of reference	0.692		
Odd beliefs or magical thinking	0.602		
Unusual perceptual experience	0.669		
Odd speech			0.575
Suspiciousness	0.501	0.221	
Constricted affect		0.733	
Odd behaviour			0.599
No close friends		0.788	
Social anxiety		0.372	

Table B.12. Factor loadings of m_3 confirmatory model with classical analysis

	Cognitive/ perceptual	Interpersonal	Disorganized
Cognitive/ perceptual	1.000		

Interpersonal	0.285	1.000	
Disorganized	0.957	0.665	1.000

Table B. 13. Correlation matrix of latent factors of m_3 confirmatory model with classical analysis

m_3	Cognitive/ perceptual	Interpersonal	Disorganized
Ideas of reference	0.614		
Odd beliefs or magical thinking	0.524		
Unusual perceptual experience	0.615		
Odd speech			0.529
Suspiciousness	0.483	0.223	
Constricted affect		0.701	
Odd behaviour			0.567
No close friends		0.738	
Social anxiety		0.352	

Table B.14. Factor loadings of m_3 confirmatory model with Bayesian analysis

	Cognitive/ perceptual	Interpersonal	Disorganized
Cognitive/ perceptual	1.168		
Interpersonal	0.181	1.088	
Disorganized	0.442	0.313	1.245

Table B. 15. Covariance matrix of latent factors of m_3 confirmatory model with Bayesian analysis

Model 4

	Factors			
Schizotypal traits	**Cognitive/ perceptual**	**Negative**	**Disorganized**	**Paranoid**
Ideas of reference				0.874
Odd beliefs or magical thinking	0.655			
Unusual perceptual experiences	0.722			
Odd speech			0.577	
Suspiciousness		0.29		0.523
Constricted affect		0.709		
Odd behaviour			0.596	
No close friends		0.815		
Social anxiety		0.349		0.109

Table B.16. Factor loadings of paranoid 4-factor model with classical analysis

	Cognitive/ perceptual	**Negative**	**Disorganized**	**Paranoid**
Cognitive/ perceptual	1			
Negative	0.304 (*0.105*)	1		
Disorganized	0.910 (*0.105*)	0.650 (*0.105*)	1	
Paranoid	0.663 (*0.093*)	0.132 (*0.102*)	0.771 (*0.110*)	1

Table B.17. Correlation matrix of paranoid 4-factor model with classical analysis

$\mathbf{m_4}$	**Cognitive/ perceptual**	**Negative**	**Disorganized**	**Paranoid**
Ideas of reference				0.716
Odd beliefs or magical	0.536			

thinking				
Unusual perceptual experience	0.661			
Odd speech			0.501	
Suspiciousness		-0.262		0.490
Constricted affect		-0.686		
Odd behaviour			0.555	
No close friends		-0.275		
Social anxiety		-0.331		0.097

Table B.18. Factor loadings of m_4 confirmatory model with Bayesian analysis

	Cognitive/ perceptual	**Negative**	**Disorganized**	**Paranoid**
Cognitive/ perceptual	1.281			
Negative	-0.207	1.097		
Disorganized	0.483	-0.325	1.334	
Paranoid	0.437	-0.122	0.444	1.236

Table B. 19. Covariance matrix of latent factors of m_4 confirmatory model with Bayesian analysis

Model 5

m_5	**Paranoid**	**Positive**	**Schizoid**	**Avoidant**	**Disorganized**

Ideas of reference	0.570	0.322		0.250	
Odd beliefs or magical thinking		0.626			
Unusual perceptual experience		0.755			
Odd speech			0.524		
Suspiciousness	0.769				-0.027
Constricted affect			0.986		-0.346
Odd behaviour					0.896
No close friends			0.692		
Social anxiety				-0.497	

Table B.20. Factor loadings of m_5 confirmatory model with classical analysis

	Paranoid	**Positive**	**Schizoid**	**Avoidant**	**Disorganized**
Paranoid	1.000				
Positive	0.671	1.000			
Schizoid	0.572	0.44	1.000		
Avoidant	-0.515	-0.614	-0.705	1.000	
Disorganised	0.588	0.605	0.581	-0.193	1.000

Table B.21. Correlation matrix of latent factors of m_5 confirmatory model with classical analysis

	FACTORS				
Schizotypal traits	**Paranoid**	**Positive**	**Schizoid**	**Avoidant**	**Disorganized**
Ideas of reference	0.456 (*0.144*)	0.3816 (*0.111*)		-0.08 (*0.125*)	
Odd beliefs or magical thinking		0.604 (*0.098)*			
Unusual perceptual experiences		0.675 (*0.098*)			
Odd speech			0.466 (*0.085*)		
Suspiciousness	0.602 (*0.152)*				0.255 (*0.121*)
Constricted affect			0.808 (*0.107*)		-0.14 (*0.091*)
Odd behaviour					0.738 (*0.110*)
No close friends			0.662 (*0.091)*		
Social anxiety				0.699 (*0.18*)	

Table B.22. Factor loadings of Fogelson et al. 5-factor model with Bayesian analysis

	Paranoid	**Positive**	**Schizoid**	**Avoidant**	**Disorganized**
Paranoid	1.029 (*0.155*)				
Positive	0.223 (*0.109*)	1.102 (*0.166*)			
Schizoid	0.197 (*0.098*)	0.23 (*0.093*)	1.079 (*0.161*)		
Avoidant	0.129 (*0.095*)	0.178 (*0.095*)	0.223 (*0.096*)	1.008 (*0.146*)	
Disorganized	0.211 (*0.112*)	0.365 (*0.105*)	0.306 (*0.102*)	0.105 (*0.094*)	1.118 (*0.171*)

Table B.23. Covariance matrix of Fogelson et al. 5- factor model with Bayesian analysis

3. The code of the 4-factor model in Bayesian exploratory factor analysis.

```
model;
{
   for (i in 1:n){
  for(j in 1:p){
    y[i,j]~dnorm(mu[i,j],tau[j])
mu[i,j]<-l1[j]*f1[i]+l2[j]*f2[i]+l3[j]*f3[i]+l4[j]*f4[i]
      }
           }
l2[1]<-0.0
l3[1]<-0.0
l3[2]<-0.0
l4[1]<-0.0
l4[2]<-0.0
l4[3]<-0.0
l1[1]~dnorm(0,1)I(0,)
l2[2]~dnorm(0,1)I(0,)
l3[3]~dnorm(0,1)I(0,)
l4[4]~dnorm(0,1)I(0,)
for(j in 2:p){
l1[j]~dnorm(0,1)
                }
for(j in 3:p){
l2[j]~dnorm(0,1)
                }
for(j in 4:p){
l3[j]~dnorm(0,1)
                }
 for(j in 5:p){
 l4[j]~dnorm(0,1)
                 }
for(j in 1:p){
       tau[j]~dgamma(1,1)
        s[j]<-1/tau[j]
                       }
      for(i in 1:n){
          f1[i]~dnorm(0,1)
         f2[i]~dnorm(0,1)
         f3[i]~dnorm(0,1)
         f4[i]~dnorm(0,1)
                    }
                }
```

4. The code of the 4-factor model in Bayesian confirmatory factor analysis.

```
model;
{
  for (i in 1:n){
 for(j in 1:p){
   y[i,j]~dnorm(mu[i,j],tau[j])
mu[i,j]<-l[1,j]*f[i,1]+l[2,j]*f[i,2]+l[3,j]*f[i,3]+l[4,j]*f[i,4]
     }
          }
l[1,1]<-0
l[1,2]~dnorm(0,1)I(0,)
l[1,3]~dnorm(0,1)
l[1,4]<-0
l[1,5]<-0
l[1,6]<-0
l[1,7]<-0
l[1,8]<-0
l[1,9]<-0
l[2,1]<-0
l[2,2]<-0
l[2,3]<-0
l[2,4]<-0
l[2,5]~dnorm(0,1)
l[2,6]~dnorm(0,1)
l[2,7]<-0
l[2,8]~dnorm(0,1)
l[2,9]~dnorm(0,1)
l[3,1]<-0
l[3,2]<-0
l[3,3]<-0
l[3,4]~dnorm(0,1)I(0,)
l[3,5]<-0
l[3,6]<-0
l[3,7]~dnorm(0,1)
l[3,8]<-0
l[3,9]<-0
l[4,1]~dnorm(0,1)I(0,)
l[4,2]<-0
l[4,3]<-0
l[4,4]<-0
l[4,5]~dnorm(0,1)
l[4,6]<-0
l[4,7]<-0
l[4,8]<-0
```

```
l[4,9]~dnorm(0,1)
for(j in 1:p){
        tau[j]~dgamma(1,1)
         s[j]<-1/tau[j]
                         }
for (i in 1:n){
      f[i,1:4]~dmnorm(miden[1:4],prec[1:4,1:4])
                    }
 prec[1:4,1:4]~dwish(R[1:4,1:4],100)
sigma[1:4,1:4]<-inverse( prec[1:4,1:4])
                }
  }
```

REFERENCES

Adler, S.L. (1981). Over-relaxation method for the Monte Carlo evaluation of the partition function for multiquadratic actions. Physical Review D, 23, 2901-2904

Akaike, H. (1974). A new look at the Statistical Model Identification. *IEEE Transactions on Automatic Control.* AC-19, 716-723

Akaike, H. (1987). Factor analysis and AIC. *Psychometrika,* 52, 317-332

American Psychiatric Association (1980). *DSM-III-R: Diagnostic and Statistical Manual of Mental Disorders. 3rd ed.* Washington, DC: The Association

American Psychiatric Association (1987). *DSM-III-R: Diagnostic and Statistical Manual of Mental Disorders. 3rd ed., revised.* Washington, DC: The Association

Ansari, A. and Jedidi, K. (2000). Bayesian factor analysis for multilevel binary observations. *Psychometrika,* 65(4), 475-496

Bartholomew, D.J. and Knott, M. (1999). *Latent variable models and factor analysis.* Arnold Kendall's Library of Statistics 7

Bartholomew, D.J., Steele, F., Moustaki, I. and Galbraith, J.I. (2002). *The analysis and interpretation of multivariate data for social scientists.* Chapman and Hall/ CRC

Basilevsky, A. (1994). *Statistical Factor Analysis and Related Methods: Theory and Applications.* John Wiley and Sons, Canada

Bassett, A. S., Bury, A. and Honer, W.G. (1994). Testing Liddle's three syndrome model in families with schizophrenia. *Shizophrenia Research,* 12, 213-221

Bayes, R.T. (1763). An essay towards solving a problem in the doctrine of chances. *Philosophical Transactions of the Royal Society of London,* 53, 370-418

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin,* 107, 238-246

Bentler, P. M. and Bonnet, D.G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin*, 88, 588-606

Bergman, A.J., Harvey, P.D., Mitropoulou, V., Aronson, A., Marder, D., Silverman, J., Trestman, R. and Siever, L.J. (1996). The factor structure of

Schizotypal symptoms in a clinical population. *Shizophrenia Bulletin*, 22(3), 501-509

Bernando, J.M. and Smith, A.F.M. (1994). *Bayesian Theory.* John Wiley and Sons, England

Best, N., Cowles, M.K. and Vines, K. (1996). CODA, Convergence Diagnosis and Output Analysis Software for Gibbs sampling output, Version 0.30 (manual). Available at http://www.mrc-bcu.cam.ac.uk/bugs/ documentation/Download/ cdaman03.pdf

Bleuler, E. (1950). *Dementia praecox*. New York: International Universities Press

Bollen, K.A. (1986). Sample size and Bentler's and Bonett's nonnormed fit index. *Psychometrika,* 51,375-377

Bollen, K.A. (1989a). *Structural Equations with Latent Variables.* John Wiley and Sons, Canada

Bollen, K.A. (1989b). A new incremental fit index for general structural equation models. *Sociological methods and Research,* 17, 303-316

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

Brooks, S.P. and Roberts, G.O. (1998). Convergence assessment techniques for Markov chain Monte Carlo. *Statistics and Computing,* 8, 319-335

Browne, M.W. and Cudeck, R. (1989). Single sample cross-validation indexes for covariance structures. *Multivariate Behavioral Research,* 24, 45-55

Browne, M.W. and Cudeck, R. (1993). Alternative ways of assessing model fit. *Testing Structural Equations Models (Bollen, K.A. and Long, J.S. ,eds.).* Newbury Park, CA: Sage Publications, 136-162

Carlin, B.P. and Louis, T.A. (2000). *Bayes and Empirical Bayes Methods for Data Analysis (2nd edition).* Chapman and Hall/CRC, United states of America

Carroll, J.B. (1957). Biquartimin criterion for rotation to oblique simple structure in factor analysis. *Science,* 125, 1114-1115

Casella, J.B. and George, E. (1992). Explaining the Gibbs sampler. *The American Statistician,* 46, 167-174

Chatfield, C. and Collins, A.J. (1980). *Introduction to Multivariate Analysis.* Chapman and Hall, London

Chib, S. and Greenberg, E. (1995). Understanding the Metropolis-Hastings algorithm. *The American Statistician,* 49, 327-335

Congdon, P. (2001). *Bayesian Statistical Modelling.* John Wiley and Sons, England

Driscoll, H., Campbell A. and Muncer, S. (2005). Confirming the Structure of a Ten-Item expagg scale using confirmatory factor analysis. *Current Research In Social Psychology,* 10 (15)

D'Souza, A. (2002). Bayesian factor analysis. Draft Preprint. University of southern California. Computation learning and motor control lab. Available at: http://www-clmc.usc.edu/~adsouza/notes/ bayesian _fa.pdf

Eckart, C. and Young, G. (1936). The approximation of one matrix by another of lower rank. *Psychometrika*, 1, 211-218

Everitt, B.C. (1984). *An Introduction to Latent Variable Models.* Chapman and Hall, London and New York

Fogelson, D.L., Nuechterlein, K.H., Asarnow, R.F., Payne, D.L., Subotnik, K.L. and Giannini, C.A.(1999). The factor structure of schizophrenia spectrum personality disorders: Signs and symptoms in relatives of psychotic patients from the UCLA family members study. *Psychiatry Research,* 87(2-3), 137-146

Fokoue, E., (2004). Stochastic determination of the intrinsic structure in Bayesian factor analysis. *Technical Report No. 2004-17,* Statistical and Applied Mathematical Sciences Institute

Gelman, A. and Rubin, D.B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science,* 7, 457-511

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

Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments (with discussion). *Bayesian Statistics 4, Oxford: Oxford University Press,* 169-193

Geweke, I. and Singleton, K.J. (1980). Interpreting the likelihood ratio statistic in factor models when sample size is small. *Journal of the American Statsitical Association,* 75, 133-137

Green, P.J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrica,* 82, 711-732

Guttman, L. (1953). Image theory for the structure of quantitative variates. *Psychometrika,* 18, 277-296

Harman, H.H. (1976). *Modern Factor Analysis.* Chicago University Press, Chicago

Harman, H.H. and Jones, W.H. (1966). Factor analysis by minimizing residuals (Minres). *Psychometrika,* 31, 351-369

Harris, C.W. and Kaiser, H.F. (1964). Oblique factor analysis solutions by orthogonal transformations. *Psychometrika,* 29, 347-362

Hastings, W.K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika,* 57, 97-109

Hayashi, K. and Sen, P.K. (2002). Bayesian factor analysis: Bias-corrected estimator of factor loadings in Bayesian factor analysis. *Educational and Psychological Measurement,* 62(6), 944-959

Heidelberger, P. and Welch, P.D. (1983). Simulation run length control in the presence of ana initial transient. *Operations Research,* 31, 1109-1144

Holzinger, K. (1930). Statistical resume of the Spearman Two-Factor Theory. *University of Chicago,* Chicago

Horst, P. (1937). A method of factor analysis by means of which all coordinates of the factor matrix are given simultaneously. *Psychometrika,* 2, 225-236

Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of Education Technology,* 24, 417-441, 498-520

James, L.R., Mulaik, S.A. and Brett, J.M. (1982). *Causal analysis: Assumptions, models and data.* Beverly Hills: Sage

Jennrich, R.I. and Sampson, P.F. (1966). Rotation for simple loadings. *Psychometrika,* 31, 313-323

Jöreskog, K.G. (1967). Some contributions to maximum likelihood factor analysis. *Psychometrika,* 32, 443-482

Jöreskog, K.G. and Sörbom, D. (1986). *LISREL VI: Analysis of linear structural relationships by maximum likelihood, instrumental variables, and least square methods (User's guide, 4th ed.).* Mooresville, IN: Scientific software

Jöreskog, K.G. and Sörbom, D. (1996). *LISREL 8: User's Reference Guide.* Scientific Software International

Information Technology Estimates: Research Consulting. LISREL FAQ. Available at: http://www.utexas.edu/its/rc/answers/lisrel/lisrel2.html

Kaiser, H.F. (1956). Note on Carroll's analytic simple structure. *Psychometrika,* 21, 89-92

Kaiser, H.F. and Caffrey, J. (1965). Alpha factor analysis. *Psychometrica,* 30, (1)

Kaplan, D. (2000). *Structural Equation Modeling, Foundations and Extensions.* Sage Publications

Kass, R.E. and Raftery, A.E. (1995). Bayes factors. *Journal of American Statistical Association,* 90, 773-795

Kateri, M., Nikolaou, A. and Ntzoufras, I. (2005). Bayesian inference for the RC(m) association model. *Journal of Computational and Graphical Statistics,* 14, 116-138.

Kay, S.R. and Sevy, S. (1990). Pyramidal model of Schizophrenia. *Schizophrenia Bulletin,* 16(3), 537-545

Kaufman, G. M. and Press, S.J. (1973). Technical Report No. 7322. School of Business, Department of Economics and business, Center for Mathematical Studies in Business and Economics

Keesling, W. (1972). Maximum likelihood approaches to causal flow analysis. Ph. D. thesis, University of Chicago

Kendler, K.S., Gruenberg, A.M. and Strauss, J.S. (1981). An independent analysis of the Copenhagen sample of the Danish adoption study of schizophrenia: II. The relationship between schizotypal personality disorder and schizophrenia. *Archives of General Psychiatry,* 38, 982-984

Kendler, K. S., Ochs, A. L., Gorman, A.M., Hewitt, J.K., Ross, D.E. and Mirsky, A.F. (1991). The structure of schizotypy: A pilot multitrait twin study. *Psychiatry Research,* 36(1), 19-36

Kety, S. S. (1983). Mental illness in the biological and adoptive relatives of schizophrenic adoptees: Findings relevant to genetic and environmental factors in etiology. *American Journal of Psychiatry,* 140, 720-727

Kim, J. and Mueller, C.W. (1978a). *Introduction to Factor Analysis: what it is and how to do it.* Sage Publications, Iowa

Kim, J. and Mueller, C.W. (1978b). *Introduction to Factor Analysis: Statistical Methods and Practical Issues.* Sage Publications, Iowa

Lawley, D.N. and Maxwell, A.E. (1971). *Factor analysis as a statistical method.* Butterworths, London

Lee, S.Y. (1981). A Bayesian approach to confirmatory factor analysis. *Psychometrika,* 46, 153-160

Liu, C. and Rubin, D.B. (1998). Maximum likelihood estimation of factor analysis using the ECME algorithm for complete and incomplete data. *Statistica Sinica,* 8, 729-747

Lopes, H.F. (2003). Expected posterior priors in factor analysis. *Brazilian Journal of Probability and Statistics,* 17, 91-105

Lopes, H.F. and West, M. (2004). Bayesian model assessment in factor analysis. *Statistica Sinica,* 14(1), 41-67

Martin, J.M. and McDonald, R.P. (1975). Bayesian estimation in unrestricted factor analysis: a treatment for Heywood cases. *Psychometrika,* 40(4), 505-517

McCullagh, P. and Nelder, J.A. (1983, 1989). *Generalized Linear Models.* Chapman and Hall, Cambridge

Meehl, P.E. (1964). Manual for use with checklist of schizotypic signs. No. PR-73-5. *University of Minnesota research Laboratories of the Department of Psychiatry*

Meehl, P.E. (1990). Toward an integrated theory of schizotaxia, schizotypy and shizophrenia. *Journal of Personality Disorders,* 4, 1-99

Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H. and Teller, E. (1953). Equations of state calculations by fast computing machines. *Journal of Chemical Physics,* 21, 1087-1091

Morris, C.N. (1983). Natural exponential families with quadratic variance functions: Statistical theory. *Annals of Statistics,* 11, 515-529

Mulaik, S., James, L., Van Astine, J., Bennet, N., Lind, S. and Stilwell, C. (1989). Evaluation of goodness-of-fit indices for structural equation models. *Psychological Bulletin,* 105, 430-460

Neal, R.M. (1998). Suppressing random walks in Markov Chains Monte Carlo using ordered overrelaxation. *Learning in Graphical Models, M.I.*

Newman, H.H., Freeman, F.N. and Holzinger, K.J. (1937). Twins: A study of herediddty and environment. *University of Chicago Press, Chicago*

Pearson, K. (1901). On lines and planes of closest fit to systems of points in space. *The London, Edinburgh and Dublin Philosophical Magazine and Journal of Science*, 2, 559-572

Peralta, V. and Cuesta, M.J.(1998). Factor structure and clinical validity of competing models of positive symptoms in schizophrenia. *Biological Psychiatry,* 44(2), 107-114

Peralta, V. and Cuesta, M.J.(1999). Dimensional structure of psychotic symptoms: an item-level analysis of SAPS and SANS symptoms in psychotic disorders. *Schizophrenia Research,* 38(1), 13-26

Press, S.J. (1972). *Applied multivariate analysis: using Bayesian and frequentist methods of inference.* Melbourne, FL; Krieger

Press, S.J. (1982). *Applied multivariate analysis.* Holt, Rinehart and Winston, New York

Press, S.J. and Shigemasu, K. (1989). Bayesian inference in factor analysis. *Contributions to Probability and Statistics, chapter 15*. Springer-Verlag

Press, S.J. and Shigemasu, K. (1997). Bayesian inference in factor analysis-Revised. *Technical Report No. 243,* Department of Statistics, University of California, Riverside

Raftery, A.E. (1993). Bayesian model selection in structural equation models. *Testing Structural Equation Models (Bollen, K.A. and Long, J.S., eds.).* Beverly Hills: Sage, 163-180

Raftery, A.E. and Lewis, S. (1992). How many iterations in the Gibbs sampler? *Bayesian Statistics 4, Oxford: Oxford University Press,* 763-773

Raine A. (1991). *SPQ***.** Available at**:** http://www-rcf.usc.edu/~raine

Raine, A. (1991). The SPQ: A scale for the assessment of schizotypal personality based on DSM-III-R criteria. *Schizophrenia Bulletin,* 17, (4), 555-564

Raine, A., Reynolds, C.A., Lencz, T., Scerbo, A., Triphon, N. and Kim, D. (1994). Cognitive perceptual interpersonal and disorganised features of schizotypal personality. *Schizophrenia Bulletin,* 20(1), 191-201

Rowe, D.B. (1998). Correlated Bayesian factor analysis. Ph.D. Thesis, *Department of Statistics, University of California, Riverside CA 92521*

Rowe, D.B. (2000a). On estimating the mean in Bayesian factor analysis. *Social Science Working Paper 1096*, Division of Humanities and Social Sciences, Caltech, Pasadena

Rowe, D.B. (2000b). Incorporating prior knowledge regarding the mean in Bayesian factor analysis. *Social Science Working Paper 1097,* Division of Humanities and Social Sciences, Caltech, Pasadena

Rowe, D.B. (2000c). A Bayesian factor analysis model with generalized prior information. *Social Science Working Paper 1099,* Division of Humanities and Social Sciences, Caltech, Pasadena

Rowe, D.B. (2001). A model for Bayesian factor analysis with jointly distributed means and loadings. *Social Science Working Paper 1108,* Division of Humanities and Social Sciences, Caltech, Pasadena

Rowe, D.B. (2003). On using the sample mean in Bayesian factor analysis. *Journal of Iinterdisciplinary Mathematics,* 6(3), 319-329

Rowe, D.B. and Press, S.J. (1998). Gibbs sampling and hill climbing in Bayesian factor analysis. *Technical Report No. 255*, Department of statistics, University of California, Riverside CA 92521

Rubin, D. and Thayer, D. (1982). EM algorithms for ML factor analysis. *Psychometrika,* 47(1), 69-76

SAS / STAT/ User's Guide. Available at: http://www.id.unizh.ch/software/unix/ statmath /sas/sasdoc/stat/chap26

Scheines, R., Hoijtink, H. and Boomsma, A. (1999). Bayesian estimation and testing of structural equation models. *Psychometrika*, 64(1), 37-52

Schwartz, G. (1978). Estimating the dimension of a model, *Annals of Statistics,* 6, 461-464

Spearman, C. (1904). 'General Intelligence' objectively determined and measured. *American Journal of Psychology,* 15, 201-93

Spiegelhalter, D. J., Best, N.G., Carlin, B.P. and van der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of Royal Statistical Society B,* 64, part 4, 583-639

Spiegelhalter, D. J., Best, N.G, Thomas, A. and Lunn, D. (2003). *WinBugs User Manual.* Available at http://www.mrcbsu.cam.ac.uk /bugs/winbugs/ manual14.pdf

SPSS Base 13.0 User's Guide. SPSS Inc

Steiger, J.H. (1990). Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research,* 25, 173-180

Stefanis, N.C., Smyrnis, N., Avramopoulos, D., Evdokimidis, I., Ntzoufras, I. and Stefanis, C. N. (2004). Factorial Composition of self-rated schizotypal traits amongst young males undergoing military training. *Schizophrenia Bulletin,* 30(2), 335-350

Stuart, G.W., Malone, V., Currie, J. Klimidis, S. and Minas, I.H. (1995). Positive and negative symptoms in neuroleptic-free psychotic inpatients. *Schizophrenia Research,* 16(3), 175-188

Tucker, L. R. and Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika,* 38,1-10

Tucker, L. R. and MacCallum, R.C. (1997). *Exploratory Factor Analysis.* Available at: http://www.unc.edu/~rcm/book/factor.pdf

Thurstone, L. L. (1935/1947). *The vectors of mind/Multiple Factor Analysis: A development and expansion of the vectors of mind.* Chicago: University of Chicago Press

Von Mises , R. (1931). *Wahrscheinlichkeitschrechnung.* Deutiche, Vienna

Venables, W. N. and Ripley, B. D. (1999). *Modern Applied Statistics with S-Plus.* Springer

Viele, K. and Srinivasan, C. (2000). **Parsimonious estimation of multiplicative interaction in analysis of variance using Kullback-Leibler Information.** ***Journal of Statistical Planning and Inference,*** **84, 201-209**

Wald, A. (1950). *Statistical Decision Function*. Wiley

West, M. (2003). Bayesian factor regression models in the 'Large p, small n' paradigm. *Bayesian Statistics 7, Oxford University Press*

Wiley, D. E. (1973). The identification problem for structural equation models with unmeasured variables. *Structural Equation Models in the Social Sciences, Seminar Press, New York*

Young, G. and Householder, A. (1938). Discussion of a set of points in terms of their mutual distances. *Psychometrika*, 3, 19-22

Βιτωράτου, Σ. (2004). Τυποποίηση και ανάλυση ψυχομετρικών χαρακτηριστικών νεοσύλλεκτων εφέδρων οπλιτών. *Διπλωματική εργασία,* Εθνικό και Καποδιστριακό Πανεπιστήμιο Αθηνών: Ιατρική σχολή, τμήμα Μαθηματικών και Πανεπιστήμιο Ιωαννίνων, τμήμα Μαθηματικών

Ηλιοπούλου (2004). Σχιζοτυπία και Συμπεριφορά Καταναλωτή. *Διπλωματική εργασία.* Μεταπτυχιακό δίπλωμα στη Διοίκηση Επιχειρήσεων. Τμήμα Διοίκησης Επιχειρήσεων. Πανεπιστήμιο Αιγαίου

Στάθη, Χ. (2005). Διερεύνηση των σχέσεων μεταξύ σχιζοτυπικών χαρακτηριστικών της προσωπικότητας και γνωσιακών επιδόσεων. *Διπλωματική εργασία.* Μεταπτυχιακό δίπλωμα στη Βιοστατιστική. Εθνικό και Καποδιστριακό Πανεπιστήμιο Αθηνών: Ιατρική σχολή, τμήμα Μαθηματικών και Πανεπιστήμιο Ιωαννίνων, τμήμα Μαθηματικών