Carel F.W. Peeters
Utrecht University

Bayesian Exploratory and Confirmatory Factor Analysis

Perspectives on Constrained-Model Selection

Peeters, Carel Frederik Wilhelm
Bayesian Exploratory and Confirmatory Factor Analysis: Perspectives on Constrained-Model Selection
Proefschrift Universiteit Utrecht, Utrecht. - Met lit. opg. - Met samenvatting in het Nederlands.
ISBN: 978-90-393-5787-3
Druk: Drukkerij Mostert & Van Onderen!, Leiden

This research was partly supported by grant NWO-VICI-453-05-002 of the Netherlands Organization for Scientific Research (NWO) and was performed under the scientific umbrella of the Interuniversity Graduate School of Psychometrics and Sociometrics (IOPS).

Cover design by Floryt Peeters-van Wesel and the author. Inspired by Pieter A.H. Hofman's design for *Elegie sur la mort d'une princesse enchantée* by Ignacy Lilien. Typeset in LaTeX 2_ε by the author using Springer Verlag's svmono.cls document class. The snippet of text on the very last page of this work is taken from 'Invictus' by William Ernest Henley.

BAYESIAN EXPLORATORY AND CONFIRMATORY FACTOR ANALYSIS: PERSPECTIVES ON CONSTRAINED-MODEL SELECTION

Bayesiaanse exploratieve en confirmatieve factoranalyse:
perspectieven op de selectie van gerestricteerde modellen

(met een samenvatting in het Nederlands)

PROEFSCHRIFT

ter verkrijging van de graad van doctor aan de Universiteit Utrecht
op gezag van de rector magnificus, prof. dr. G.J. van der Zwaan,
ingevolge het besluit van het college voor promoties
in het openbaar te verdedigen op

maandag 4 juni 2012 des ochtends te 10.30 uur

door

Carel Frederik Wilhelm Peeters

geboren op 16 maart 1982
te Nijmegen

Promotor: Prof. dr. P.G.M. van der Heijden

Samenstelling leescommissie:

Leden: Prof. dr. ing. P.H.C. Eilers
Department of Biostatistics, Erasmus MC
Prof. dr. J.J.C.M. Hox
Department of Methodology & Statistics, Utrecht University
Prof. dr. J.M.F. ten Berge
Department of Psychometrics & Statistics, University of Groningen
Prof. dr. S. van Buuren
Department of Statistics, TNO Quality of Life

To my parents
Carel & Annelies

(. . .) we buy information
with assumptions

-Clyde H. Coombs

We must not suppose that science
teaches us that every *thing* anyone
would ever want to take seriously is
identifiable as a collection of particles
moving about in space and time

-Daniel C. Dennett

Preface

Statistical science is a discipline in perpetual transition. The ongoing transition is due to both its outer and inner world. The outer world provides the discipline with the ever changing realities of data acquisition in the sciences as fields of enquiry emerge and evolve. The inner world is shaped by the search for a unified philosophical and mathematical foundation. This dissertation focuses on a statistical topic that bears elements of both the outer and the inner world. A topic that I perceive to be statistical heartland: Data based parametric-model selection.

The connotation of 'model selection' as used in this dissertation is that of selection of a statistical model for the data, possibly to be used as a prediction function. The goal then is to select, in settings in which such is deemed appropriate, from a set of *a priori* specified models, the one model that best balances model fit and model complexity. The practice of model selection is viewed as having two main *foci*, both placed in the perspective of functional constraints, being: (I) variable selection, which is viewed in terms of selection of the dimension of a model or the selection of exclusion constraints in the design matrix; and (II) the selection of appropriate truncations of the parameter space, which essentially entails placement of inequality constraints on the model parameters of interest. The underlying work will embark on both.

Both the frequentist (including information theory) and the machine learning approach towards model selection have much to add to the endeavor. The approach here, however, is Bayesian. The focus is on usage of the marginal likelihood as a prior predictive density in Type I and Type II model selection settings. The stance will be that of the objective Bayesian, in the sense that prior choice is automated through default or noninformative prior usage while retaining propriety of the main Bayesian model selection criterion: The Bayes factor as a ratio of marginal likelihoods. The Bayes factor takes preference as it provides a conceptually consistent quantity for the comparison of a multitude of competing models. Additionally, it penalizes model complexity in a natural manner. Moreover, it does not require competing models to be nested for model comparison to be meaningful. A drawback of the Bayes factor is that it is often analytically untractable and hard to compute. The dissertation shows that, under mild assumptions, computation of analytically untractable Bayes factors may be simplified for both Type I and Type II model selection settings.

The subjective factor, as with any approach to model selection, is mainly entrenched within the determination of the model set. The underlying assumption is that the substantive researcher makes an informed choice, preferably based on their principled use, of the models to be included in the model set. I note, however, that I deem the practice of model selection *not* to be conceptually or fundamentally different when the true model is not contained in the model set under consideration.

The parametric model I focus on is the normal-theory factor analytic model. Factor analysis (FA) assumes that a random p-dimensional vector X consists of correlated variables that can be grouped, using their covariances, into a m-dimensional vector of latent factors Ξ, with m smaller than p. Although this model of psychometric origin has originally been met with some disdain by statisticians, it now is heavily utilized in such diverse fields as medical research, social and behavioral science, and natural science. It is used as a data-reduction technique in a form usually termed exploratory FA (EFA) and as a data-analytic technique in a form called confirmatory FA (CFA). The model essentially entails a generalization of multivariate regression and forms the base on which general models, such as structural equation models, rest.

The model sets considered in this dissertation then contain factor analytic models of differing dimension m, or competing inequality constrained factor structures given m. The considered statistical model and accompanying model sets connect to Type I and Type II model selection problems and express my conviction that in latent variable modeling the intrinsic dimensionality of the model should be determined before competing structures (possibly containing functional restrictions on the parameters retained) are to be tested. Another reason to consider the FA model is that the many indeterminacies inherently linked to the model forces one to think in rather general terms about Bayesian procedures. Moreover, reviewing these indeterminacies spurs learning on the peculiarities of both the FA model and Bayesian computation.

Essentially, the dissertation strives to be the scientific analogue of a layer cake. The base layer is formed by the development of Bayes factors and accompanying computation strategies for constrained-model selection. These strategies are subsequently utilized in the second layer: The reformulation of EFA and CFA in the Bayesian framework from a constrained model perspective. The dissertation provides a true Bayesian EFA which may be used for the selection of the intrinsic dimension m. Moreover, in the reformulation of CFA structure is imposed on the model through inequality constraints on the parameters of interest, rather than through exclusion constraints. The conjunction of these two layers, which is essentially a coming together of statistics and psychometrics, will be exemplified in the final layer: The factor analytic analysis of real data from various fields outside the immediate behavioral science realm that brought about the FA model.

The appeal of statistics as a mathematical science is that it inherently has elements of both the pure and the applied. Whenever choices had to be made with respect to the (overly) formal or informed applicability and practicability, the preference was, conform the nature of statistics, with the latter. For example, although the developed procedures in this dissertation enjoy theoretical justification, they have not been analyzed from a strict decision-theoretic perspective. The semi-formal approach taken, I believe, connects to the reality of statistics as an art and science between inductive logic and empirical research (cf. Sprenger, 2008). I justify such an approach, next to practicability, by way of Alfred North Whitehead's *adagium*: "It is more important that a proposition be interesting than that it be true". I hope you, the reader, will find something interesting among the very humble contributions.

Bussum,
March 2012

Carel F.W. Peeters

Contents

List of Figures and Tables

Figures

Tables

1

Introduction: Bayesian Constrained-Model Selection in Factor Analytic Modeling

1.1 Preamble

This dissertation revolves around the practice of statistical model selection. The connotation of 'model selection' as used in this dissertation is that of selection of a statistical model for given data. The goal then is to select, in settings in which such is deemed appropriate, from a set of *a priori* specified models, the one model that best balances model fit and model complexity. The practice of model selection is viewed as having two main *foci*, both placed in the perspective of (functional) restrictions, being: Determination of the dimension of a model and the selection of appropriate inequality constraints on the model parameters of interest.

The parametric model reviewed in the model selection efforts is the factor analytic model. Factor analysis (FA) assumes that a random p-dimensional vector X consists of correlated variables that can be grouped into a random m-dimensional vector of latent factors Ξ, with $m < p$. It compromises a class of techniques for the (structural) analysis of the sample covariance matrix of (respondent) scores on a test battery. In addition to being a data-reductional tool, FA is also used as a data-analytic tool, aimed at validating a certain hypothetical structure deemed to underlie the data. Factor analytic efforts have also become appreciated aids in techniques such as cluster and discriminant analysis, and regression in fields as diverse as medical research, social and behavioral science, and natural science.

The factor analytic model naturally connects with the restrictions-based view on model selection. Considering FA as a data-reductional tool a key question becomes which dimensionality m of the vector of latent factors is optimal in the sense of balancing parsimony and fit. Emphasizing FA as a data-analytic technique the key question is which hypothetical parameter-based covariance structure best represents the sample covariance matrix. The model sets considered in this dissertation then contain factor analytic models of differing dimension m, or competing inequality constrained factor structures given m. The considered model sets express the conviction that in latent variable modeling the intrinsic dimensionality of the model should be determined before competing structures (possibly containing functional restrictions on the parameters retained) are to be tested.

The approach towards the FA model and model selection is Bayesian. In this dissertation the Bayesian approach, constrained-model selection, and FA are intrinsically interwoven. The development of Bayes factors and accompanying computation strategies for constrained-model selection forms the red thread. These strategies are subsequently utilized in the reformulation of data-reductional and data-analytic FA in the Bayesian framework from a constrained-model perspective. The dissertation provides a Bayesian data-reductional FA model which may be used for the selection of the intrinsic dimension m. Moreover, in the reformulation of data-analytic FA, structure is imposed on the model through inequality constraints on the parameters of interest, rather than through standard exclusion restrictions.

In the remainder the recurring elements in this dissertation are primed by shortly reviewing the FA model, constrained-model inference, and the Bayesian approach towards model selection (Section 1.2). Next, the perceived lacunae in their interrelations provide the guiding research aims (Section 1.3). Section 1.4 previews the developments and contributions of the dissertation in light of these research aims and simultaneously provides an overview of the dissertation.

1.2 Priming the Elements

1.2.1 The Factor Analytic Model

The work of Spearman (Spearman, 1904) marked the beginning of the quantitative treatment of latent variables. Latent variables are constructs that are not directly measurable in the sense that there are no independent criteria for their numerical determination. Examples include concepts such as ‘intelligence’ in psychological research or ‘impaired glucose metabolism’ in epidemiology. Spearman’s seminal ideas came to be extended in a class of techniques denoted by ‘factor analysis’, the driving idea of which is that the dependence in a battery of variables can be explained by underlying latent factors common to (subgroupings in) the test battery. Factor analysis is thus a multivariate technique, related to classic topics such as correlation and regression. In its modern form, factor analysis may be generally seen as a class of techniques or tools for the sparse modeling of a covariance matrix. The reader is referred to Bollen (1989), Basilevsky (1994), and Mulaik (2010) for an overview of FA, its foundations, and related methods.

Here, focus will be with the normal-theory linear common factor model. Let $\mathbf{X}^{\mathrm{T}} \equiv [\mathbf{x}_1, \ldots, \mathbf{x}_n]$ define p-variate observation vectors on $i = 1, \ldots, n$ subjects, such that $\mathbf{x}_i^{\mathrm{T}} \equiv [\mathrm{x}_{i1}, \ldots, \mathrm{x}_{ip}] \in \mathbb{R}^p$ denotes a realization of the random vector $X_i^{\mathrm{T}} \equiv [X_{i1}, \ldots, X_{ip}] \in \mathbb{R}^p$. Also, let $\boldsymbol{\Xi}^{\mathrm{T}} \equiv [\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_n]$ define m-variate vectors of latent factor scores on n subjects with $\boldsymbol{\xi}_i^{\mathrm{T}} \equiv [\xi_{i1}, \ldots, \xi_{im}] \in \mathbb{R}^m$. The $\boldsymbol{\xi}_i$ thus represent a latent variable of dimension m, whose elements are referred to as *common factors*. The factor analysis model states that each random variable X_i is a linear combination of the latent random variables, such that:

$$\underset{(p \times 1)}{\mathbf{x}_i} = \underset{(p \times 1)}{\boldsymbol{\mu}} + \underset{(p \times m)}{\boldsymbol{\Lambda}} \cdot \underset{(m \times 1)}{\boldsymbol{\xi}_i} + \underset{(p \times 1)}{\boldsymbol{\epsilon}_i} \tag{1.1}$$

with $m < p$. In (1.1) $\boldsymbol{\mu} \in \mathbb{R}^p$ denotes an overall mean vector, the $\boldsymbol{\epsilon}_i \in \mathbb{R}^p$ denote the error measurements, and $\boldsymbol{\Lambda} \in \mathbb{R}^{p \times m}$ is a matrix of factor loadings in which each element λ_{jk} is the loading of the jth variable on the kth factor, $j = 1, \ldots, p$, $k = 1, \ldots, m$.

Throughout, the model maintains the following assumptions: (i) $\mathbf{x}_i \perp\!\!\!\perp \mathbf{x}_{i'}, \forall i \neq i'$; (ii) $\text{rank}(\boldsymbol{\Lambda}) = m$; (iii) $\boldsymbol{\epsilon}_i \sim \mathcal{N}_p(\mathbf{0}, \boldsymbol{\Psi})$, with $\boldsymbol{\Psi} \equiv \text{diag}(\psi_{11}, \ldots, \psi_{pp})$, and $\psi_{jj} > 0$; (iv) $\boldsymbol{\xi}_i \sim \mathcal{N}_m(\mathbf{0}, \boldsymbol{\Phi})$; and (v) $\boldsymbol{\xi}_i \perp\!\!\!\perp \boldsymbol{\epsilon}_{i'}, \forall i, i'$. With these assumptions in place, covariance algebra can be used to infer that the model-implied covariance structure of the $\mathbf{x}_i$ is

$$\begin{aligned}
\mathbb{E}(\mathbf{x}_i - \boldsymbol{\mu})(\mathbf{x}_i - \boldsymbol{\mu})^{\mathrm{T}} &= \mathbb{E}[(\boldsymbol{\Lambda}\boldsymbol{\xi}_i + \boldsymbol{\epsilon}_i)(\boldsymbol{\Lambda}\boldsymbol{\xi}_i + \boldsymbol{\epsilon}_i)^{\mathrm{T}}] \\
&= \boldsymbol{\Lambda}\mathbb{E}(\boldsymbol{\xi}_i\boldsymbol{\xi}_i^{\mathrm{T}})\boldsymbol{\Lambda}^{\mathrm{T}} + 2\boldsymbol{\Lambda}\mathbb{E}(\boldsymbol{\xi}_i\boldsymbol{\epsilon}_i^{\mathrm{T}}) + \mathbb{E}(\boldsymbol{\epsilon}_i\boldsymbol{\epsilon}_i^{\mathrm{T}}) \\
&= \boldsymbol{\Lambda}\boldsymbol{\Phi}\boldsymbol{\Lambda}^{\mathrm{T}} + \boldsymbol{\Psi} \equiv \boldsymbol{\Sigma}. \qquad (1.2)
\end{aligned}$$

Then, for existence (vi), generally $(p - m)^2 - p - m \geqslant 0$, simply stating that the number of nonredundant elements in the sample covariance matrix $\mathbf{S}$ must be greater than or equal to the number of freely estimable parameters in $\boldsymbol{\Sigma}$, which places an upper bound on m.

Steiger (1994) identifies several interrelated rationales for usage of this model. The first is the *partial correlation-explanation rationale*. In this view the common factors are seen as the explanatory concepts underlying observed covariances in a set of measures. This rationale can be found in fields such as psychology and sociology (e.g., McDonald, 1985). The second is the *random noise rationale*. The observed variables are then thought to represent measurements of some physical process degraded by random noise. The common factors may then be seen as the consistent underlying sources while the error measurements represent the noise. This rationale can be found in the realm of pattern recognition and signal unmixing (e.g., Rowe, 2003), and natural science applications (e.g., Bagoly, Borgonovo, Mészáros, Balázs, & Horváth, 2009). The third rationale is the *true score rationale* found in psychometrics and education research (e.g., Waugh & Chapman, 2005). It is based on classical true score theory which assumes measurements of an ability item to be composed of a true score and an error component. Factor analysis is then utilized to lessen the attenuation by unreliability in order to assess covariances or correlations among ability items. The rationale in use by all fields that use FA is the *data-reduction rationale*. The projection of the p observed items onto the lower-dimensional factor space, is then thought to be a compressed representation of the data with which one can identify the major underlying sources of variation.

FA generally knows two *modi operandi*: Exploratory and confirmatory. In exploratory FA (EFA) both m and the meaning of latent factors are unknown. In the exploratory sense, FA is a theory-generating technique used for the identification of meaningful latent factors. Confirmatory FA (CFA) is a theory-testing technique. An *a priori* factor structure is assumed, with given m, with a pre-specified loadings matrix in which exclusion constraints indicate which variables are indicators of which latent factor(s), and with possibly correlated factors and error variances.

Parameter restriction is an important topic in both EFA and CFA as the factor model is inherently underidentified. The latent factors are in need of a metric. Moreover, from (1.2) it may be easily seen that for any nonsingular $\mathbf{R} \in \mathbb{R}^{m \times m}$:

$$\mathbf{\Sigma} = \mathbf{\Lambda}\mathbf{\Phi}\mathbf{\Lambda}^{\mathrm{T}} + \mathbf{\Psi} = (\mathbf{\Lambda}\mathbf{R})[\mathbf{R}^{-1}\mathbf{\Phi}(\mathbf{R}^{\mathrm{T}})^{-1}](\mathbf{\Lambda}\mathbf{R})^{\mathrm{T}} + \mathbf{\Psi}.$$

This implies that there is an infinite number of alternative matrices $\mathbf{\Lambda}^{\ddagger} = \mathbf{\Lambda}\mathbf{R}$ and $\mathbf{\Phi}^{\ddagger} = \mathbf{R}^{-1}\mathbf{\Phi}(\mathbf{R}^{\mathrm{T}})^{-1}$ that generate the same covariance structure $\mathbf{\Sigma}$. The operation $\mathbf{\Lambda} \mapsto \mathbf{\Lambda}\mathbf{R}$ is termed 'rotation'. Thus, in any solution, $\mathbf{\Lambda}$ can be made to satisfy m^2 additional conditions, which is naturally equivalent to the number of independent elements of $\mathbf{R}$. Restriction of parameters is required for identification purposes, i.e., in setting the metric of the latent factors and for providing general identification in ensuring, at a minimum, existence of a solution. From the above it is clear that any method of estimation requires at a minimum m^2 restrictions on $\mathbf{\Lambda}$ and $\mathbf{\Phi}$. The EFA tradition usually achieves this by requiring that $\mathbf{\Phi} = \mathbf{I}_m$ and $\mathbf{\Lambda}^{\mathrm{T}}\mathbf{\Psi}^{-1}\mathbf{\Lambda}$ be diagonal accompanied by an order condition on the diagonal elements. These restrictions are arbitrary such that whence estimation is settled EFA traditionally endeavors on applying a rotation that satisfies certain criteria for interpretation purposes. The confirmatory FA (CFA) tradition seeks to incorporate theory into the factor model while unique solutions are sought after by imposing restrictions such that rotation is deemed unnecessary. These restrictions emulate sparse structure and are of the fixed-value equality kind, with zero (exclusion) and unity being the most common values (see, for example, Bollen, 1989). Equality restrictions are also used in testing measurement invariance across groups. Provisions for inequality restrictions on parameters in the context of FA are mainly used for the prevention of impermissible estimates, most notably in combating Heywood cases (Heywood, 1931), i.e., $\hat{\psi}_{jj} \leqslant 0$, or $\hat{\phi}_{mm} \leqslant 0$.

The topic of parameter restrictions is thus heavily intertwined with FA. The exploratory and confirmatory modes of factor analysis are intricately related to the view on constrained-model selection as developed here.

1.2.2 Constrained-Model Inference

Type I and Type II Constrained-Model Selection

Constrained statistical inference is usually taken to mean the conduct of making inferences when there are inequality, order, or shape restrictions on either the decision, parameter or sample space (e.g., Silvapulle & Sen, 2005). The approach taken here will consider restrictions on the parameter space and takes a somewhat different stance in reviewing what are considered two interconnected types of restricted-model selection. These types have a natural connection to selection problems in the EFA and CFA model and, for lack of better terminology, will be termed Type I and Type II throughout.

Type I (Dimensionality determination). Type I constrained-model selection is taken to mean the determination of the appropriate dimensionality of a model. The problem of choosing an appropriate dimensionality is found in many applied contexts. Typical examples include the choice of variables to include as predictors in regression, the choice of degree for polynomial regression, or the choice of the rank constraint on the coefficient matrix in reduced-rank regression. Such choices are viewed as restricting the dimension of a model (relative to higher- or lower-dimensional representations) and for many linear model types can be expressed as the selection of exclusion restrictions in a design matrix. This type of constrained-model selection connects with EFA in the sense of selecting the optimal dimensionality m.

Example 1.1. The conception metabolic syndrome (MBS) refers to a clustering of interrelated risk factors of metabolic origin (Eckel, Grundy, & Zimmet, 2005; Unwin, 2006). MBS is thought to be a precursor for the development of type 2 diabetes mellitus and atherosclerotic cardiovascular disease. The syndrome has, next to an epidemiological status, social, behavioral and psychological components, as MBS might result from maladaptive human metabolism in the face of food energy abundance in combination with a sedentary lifestyle (Wilkin & Voss, 2004; Miranda, DeFronzo, Califf, & Gruyton, 2005). The syndrome is considered to be a major threat to current and future public health.

The pathophysiologic constellation of the MBS has become an important subject of research and latent factors extracted from factor analytic modeling efforts are taken as indicative of physiologic processes that underlie the syndrome. Suppose data from the Yale University School of Medicine on 464 obese and overweight children and adolescents first published in Weiss, *et al.* (2004) are available. The data contain measurements on the body mass index (BMI), blood glucose level at (fasting) baseline (GB) and two hours after (G2) oral glucose intake (both in mg/dl), fasting levels of triglycerides (trig.; mg/dl) and high-density lipoprotein (HDL) cholesterol (mg/dl), systolic and diastolic blood pressure (SBP, DBP; both in mm Hg), and insulin resistance (IR). These measurements are related to what are believed to be phenotypic expressions of MBS (e.g., Einhorn et al., 2003).

What FA can do is, through the latent factors, give indications of pathophysiologic domains that underlie phenotypic expressions of MBS. Suppose that, as in Weiss, *et al.* (2004) the natural logarithm is taken of the glucose, insulin resistance and triglycerides measurements to abide the normality assumption underlying the common factor model. In an EFA modeling effort the key model question becomes which choice of m, such that

$$\boldsymbol{\Lambda} = \begin{bmatrix} \lambda_{11} & \cdots & \lambda_{1m} \\ \lambda_{21} & \cdots & \lambda_{2m} \\ \lambda_{31} & \cdots & \lambda_{3m} \\ \lambda_{41} & \cdots & \lambda_{4m} \\ \lambda_{51} & \cdots & \lambda_{5m} \\ \lambda_{61} & \cdots & \lambda_{6m} \\ \lambda_{71} & \cdots & \lambda_{7m} \\ \lambda_{81} & \cdots & \lambda_{8m} \end{bmatrix} \begin{matrix} \text{BMI} \\ \log_e\{\text{trig.}\} \\ \text{HDL chol.} \\ \log_e\{\text{IR}\} \\ \log_e\{\text{GB}\} \\ \log_e\{\text{G2}\} \\ \text{SBP} \\ \text{DBP} \end{matrix},$$

best balances parsimony with the adequate representation of the information contained in the original covariance matrix. The choice of m is bound by the existence condition given in Section 1.2.1. The MBS forms a recurrent example throughout this dissertation. ■

Type II (Inequality, order, or shape restriction determination). Type II model selection is taken to mean the determination of appropriate inequality, order or shape restrictions on the parameter space. This type of constrained-model selection connects with FA in general through the impermissableness of Heywood cases expressed in model assumption (iii) from Section 1.2.1. This dissertation will connect Type II constrained-model selection explicitly to CFA by focussing on determination of linear inequality constraints of the following form:

$$\boldsymbol{\Omega}_b \boldsymbol{\lambda}_f - \boldsymbol{\alpha}_b > \mathbf{0}. \tag{1.3}$$

In (1.3) $\boldsymbol{\lambda}_f$ is a vectorization of the free parameters in $\boldsymbol{\Lambda}$ of known factor dimensionality m, $\boldsymbol{\Omega}_b$ is a constraints matrix representing a system of $l = 1, \ldots, L_b$ linear restrictions and $\boldsymbol{\alpha}_b$ denotes a real fixed-value vector of length L_b. The conjunction of $\boldsymbol{\Omega}_b$ and $\boldsymbol{\alpha}_b$ on $\boldsymbol{\lambda}_f$ then defines inequality constrained model M_b. The full set of reasons for reviewing constraints of the form (1.3) will be given in Section 1.3.1, but importantly, they can be used to alternatively express factor structure. Interest lies especially with constraints of form (1.3) such that each row of $\boldsymbol{\Omega}_b$ will be a permutation of either $(1, 0, \ldots, 0)$, $(-1, 0, \ldots, 0)$, $(1, 1, 0, \ldots, 0)$, $(-1, 1, 0, \ldots, 0)$, or $(-1, -1, 0, \ldots, 0)$. These constraints prove to be substantive from a theoretical point of view as they directly allow to express the direction and (relative) strength of factor loadings.

Example 1.2. Now, suppose that previous research has indicated a factor solution of $m = 2$ as optimal for the MBS data mentioned in Example 1.1. Or that one has settled on $m = 2$ by way of a previous EFA. Also suppose that one is interested in performing a CFA on the data, in which the hypothesized factor structure revolves around two pathophysiologic constellations indicating abnormalities in the glucose and lipid metabolism, respectively. A classical factor analyst might specify such a structure as follows:

$$\mathbf{\Lambda} = \begin{bmatrix} 0 & \lambda_{12} \\ 0 & \lambda_{22} \\ 0 & \lambda_{32} \\ \lambda_{41} & \lambda_{42} \\ \lambda_{51} & 0 \\ \lambda_{61} & 0 \\ 0 & \lambda_{72} \\ 0 & \lambda_{82} \end{bmatrix} \begin{matrix} \text{BMI} \\ \log_e\{\text{trig.}\} \\ \text{HDL chol.} \\ \log_e\{\text{IR}\} \\ \log_e\{\text{GB}\} \\ \log_e\{\text{G2}\} \\ \text{SBP} \\ \text{DBP} \end{matrix} .$$

There are several perceived problems with such formulations of confirmatory factor structure (see Section 1.3.1), one of which is that it does not allow for much specificity. An interesting extension of CFA would be to use only minimal restrictions for (rotational) identification and subsequently express factor structure using a system of inequality constraints of the form (1.3). Factor structure for the MBS data can then be formulated, for example, as

$$\mathbf{\Lambda} = \begin{bmatrix} \lambda_{11} & > & |\lambda_{12}| \\ |\lambda_{21}| & < & -\lambda_{22} \\ \lambda_{31} = 0 & & \lambda_{32} > 0 \\ \lambda_{41} & > & |\lambda_{42}| \\ \lambda_{51} > 0 & & \lambda_{52} = 0 \\ \lambda_{61} & > & |\lambda_{62}| \\ |\lambda_{71}| < .3 & & |\lambda_{72}| < .3 \\ |\lambda_{81}| < .3 & & |\lambda_{82}| < .3 \end{bmatrix} \begin{matrix} \text{BMI} \\ \log_e\{\text{trig.}\} \\ \text{HDL chol.} \\ \log_e\{\text{IR}\} \\ \log_e\{\text{GB}\} \\ \log_e\{\text{G2}\} \\ \text{SBP} \\ \text{DBP} \end{matrix} .$$

A formulation like $|\lambda_{21}| < -\lambda_{22}$ states that the negative of λ_{22} is believed to be larger than λ_{21}, irrespective of the latter's sign. Note that this is a much more informative formulation than the usual strategy of setting $\lambda_{21} = 0$ and letting λ_{22} be free to be estimated in order to express the belief that $\log_e\{\text{trig.}\}$ is an indicator for the second latent factor rather than the first one. A statement like $|\lambda_{82}| < .3$ indicates the belief that λ_{82} will take a value in the interval $[-.3, .3]$. Usage of inequalities directly allows to express the direction and (relative) strength of factor loadings and as such brings more specificity to competing models of factor structure. ■

Classical and Information Theoretic Approaches

Both classical and information theoretic approaches toward Type I and Type II constrained-model selection have added much to the endeavor. These approaches are however, especially for factor analytic model considerations, not without problems. Below these approaches are discussed, with a focus on likelihood ratio (LR) testing procedures and information criteria.

Dimensionality selection in EFA is in practice often based on heuristical procedures such as eyeballing a scree plot (Cattell, 1966) or utilization of the Guttman-Kaiser rule (Guttman, 1954; Kaiser, 1960). Formal procedures for dimensionality selection revolve around the likelihood ratio. LR testing procedures have become available for EFA with the advent of computationally feasible maximum likelihood (ML) estimation for factor analysis (Jöreskog, 1967). Standard statistical theory states that, pending certain regularity conditions, the LR test statistic is asymptotically distributed as a χ^2 variable with degrees of freedom equal to the difference in the number of free parameters between the two nested models under consideration. In practice LR testing for EFA then proceeds by sequences of tests comparing the model with at most m factors against either the saturated model or the model with at least $m+1$ factors. However, standard statistical theory often cannot be applied in this situation due to violation of regularity conditions.

Assume that the obvious regularity conditions are abided: Sample size is moderate, the observables are multivariate normally-distributed, and the problem considered has nonnegative degrees of freedom. Two regularity conditions intricately related to the nature of the FA model are then pivotal in the validity of the LR test, being assumptions (ii) and (iii) from Section 1.2.1 indicating that the matrix of factor loadings and the diagonal matrix of error variances should be of full rank. Rank deficiencies in mentioned matrices are however endemic when retaining too many factors (Chapter 2). Due to problems of rank deficiency and subsequent nonidentifiability of model parameters the LR test that employs an exploration into numbers of factors that exceed the true factor dimensionality no longer follows the χ^2 distribution (Takane, van der Heijden, & Browne, 2003; Hayashi, Bentler, & Yuan, 2007) rendering only the difference test between the model of correct (intrinsic) dimensionality m and the saturated model meaningful. The intrinsic dimensionality is however typically unknown when performing an EFA. The mentioned problems will occur in all asymptotically equivalent procedures, such as the Wald test and the Langrangian multiplier test, and in practice leads to overfactoring (retaining too many latent factors; e.g., Hayashi et al., 2007).

Information criteria such as the AIC (Akaike, 1973, 1987), the BIC (Akaike, 1978; Schwarz, 1978), and their variants (e.g., Bozdogan & Shigemasu, 1998) improve on the LR test by the inclusion of a penalty for model complexity, making non-nested model comparisons possible. However, as the heart of these criteria also consists of evaluations of the model likelihood function at maximum likelihood estimates, they too may be affected by the regularity conditions on the rank of certain matrices mentioned above (see Chapter 2). This may account for the oft observed fact that many information criteria have a bias towards more complex models, resulting in a tendency to overestimate model size in EFA as, among others, Lopes and West (2004) have demonstrated. Moreover, it is still unclear if the penalty term should also include the number of (realizations of) latent factors. As the complexity correction criterion is explicit it is hard to establish one that is consistent across a range of data and model specifications.

Type II constrained inference for FA has focussed mainly on (prevention of) Heywood cases. A common mode of inference under restrictions of the inequality type is likelihood-based inference. In this respect it is important if non-negativity of variances are explicitly parameterized and if one considers the LR test as an overall test of goodness-of-fit or as a test for evaluating one or several boundary parameters. In case of explicit parametrization, parameter optimization occurs over a restricted parameter space. The LR test then cannot be used as an overall test of goodness-of-fit as its distribution cannot be characterized: The parameter values under the null hypothesis no longer are interior points of the maintained hypothesis (the null model posits a model which assumes that no variance parameter is negative) and the null hypothesis has unidentified parameters. The LR statistic is unaffected if the space over which it is optimized is unrestricted. It will follow the χ^2 distribution if other regularity and distributional assumptions are met. However, impermissible solutions are then possible and interpretation of the LR test may be problematic (the model allows for negative variances).

Much research has thus focussed on the LR test for the evaluation of one or several boundary variance parameters. In such situations the null hypothesis has parameter values on the boundary of the maintained hypothesis but unidentified parameters under the null hypothesis are avoided. The LR test then behaves as a mixture of χ^2 distributions (cf. Chernoff, 1954; Shapiro, 1985; Self & Liang, 1987; Stoel, Garre, Dolan, & van den Wittenboer, 2006; Savalei & Kolenikov, 2008). While most standard packages performing FA and covariance structure modeling in general (as for example EQS®, M*plus*®, LISREL®, IBM® SPSS® Amos) do produce constrained test statistics, referencing these statistics to the right mixture distribution seems to be rare (see Savalei & Kolenikov, 2008).

In a broader sense, Type II restricted-model inference has been studied in the literature under various names, such as one-sided testing (Kudô, 1963), isotonic regression (Barlow, Bartholomew, Bremner, & Brunk, 1972), order-restricted inference (Robertson, Wright, & Dijkstra, 1988), and contrasts (Rosenthal, Rosnow, & Rubin, 2000). Consult Dijkstra, Robertson, and Silvapulle (2002) and Silvapulle and Sen (2005) for a more recent overview and developments. Most often, proposed solutions imply complex, problem-specific machinery that does not extend well beyond simple order-restrictions (e.g., Cohen, Kemperman, & Sackrowitz, 2000). Subsequently, more elaborate systems of inequality constraints on parameters other than variance components have been studied from an estimation perspective mostly (e.g., S. Y. Lee, 1980, 1981; Jamshidian & Bentler, 1993; Tsonaka & Moustaki, 2007), leaving the development of inferential procedures for such situations out of focus. There is thus room in the factor analytic realm for more generic procedures of model inference under systems of inequality constraints more elaborate than the simple (transitive) order restriction.

Information criteria cannot be used to circumvent the LR Type II testing problems mentioned above. Standard information criteria assume a parameter either to be part or not to be part of the maintained model. They thus have to be adapted to deal with restricted-model selection. These adaptations are connected to the appropriateness of usage of the restricted ML estimator and finding the configuration of distinct parameters in a restricted parameter space. Anraku (1999) developed an information criterion for restricted-model selection, but it is only appropriate for simple order restrictions in ANOVA-type models.

In this dissertation a turn towards the Bayesian road is taken in the search for more general solutions for Type I and Type II model inference, with the focal point on model selection. Adequate prior information will ensure $\psi_{jj} > 0$ and $\phi_{mm} > 0$, such that focus may be with selection of optimal dimension m (Chapter 2) and, given m, selection amongst competing models formulated with inequalities of the general form (1.3) (Chapter 4). Next, Bayesian model selection is shortly reviewed before delving deeper into the motivations for this dissertation.

1.2.3 Bayesian Model Selection

The Bayesian viewpoint is distinct from the classical approach to statistics. Let $\boldsymbol{\Theta}$ denote a model-specific collection of unknown parameters of continuous metric. The frequentist approach solely utilizes the likelihood of the observed data $L(\boldsymbol{\Theta}; \mathbf{X})$, in that a retrospective evaluation is made of a certain statistic used to estimate $\boldsymbol{\Theta}$ over all possible $\mathbf{X}$ values conditional on the true unknown $\boldsymbol{\Theta}$ which is deemed fixed. The Bayesian approach views $\boldsymbol{\Theta}$ as random. This allows for probability statements about $\pi(\boldsymbol{\Theta}|\mathbf{X})$, the distribution of model parameters conditioned on the observed data. To provide the mentioned conditional probabilities, a joint probability function for $\boldsymbol{\Theta}$ and $\mathbf{X}$ must be provided for. To this purpose a prior distribution $\pi(\boldsymbol{\Theta})$ must be specified, which reflects the formalized knowledge or uncertainty about the parameters before observation of the data. Using a basic property of conditional probability known as Bayes' rule (cf. Bayes, 1958; Laplace, 1986), one obtains the posterior distribution as:

$$\pi(\boldsymbol{\Theta}|\mathbf{X}) = \frac{L(\boldsymbol{\Theta}; \mathbf{X})\pi(\boldsymbol{\Theta})}{\displaystyle\int L(\boldsymbol{\Theta}; \mathbf{X})\pi(\boldsymbol{\Theta})\, \partial\boldsymbol{\Theta}}. \tag{1.4}$$

Expression (1.4) encapsulates the core machinery of Bayesian statistics, whose flexibility has proven to extend to complex problems (consult, for example, Press, 2003; Gelman, Carlin, Stern, & Rubin, 2004). The denominator in (1.4) is called the prior predictive density or marginal likelihood and is key in Bayesian model selection.

Let us shortly review Bayesian model selection for latent variable models. Let $\boldsymbol{\vartheta}$ denote latent data. For the factor model described in Section 1.2.1 $\boldsymbol{\Theta} = \{\boldsymbol{\mu}, \boldsymbol{\Lambda}, \boldsymbol{\Psi}, \boldsymbol{\Phi}\}$ and $\boldsymbol{\vartheta} = \boldsymbol{\Xi}$. Now, let $g(\boldsymbol{\vartheta}|\boldsymbol{\Theta})$ denote the density of latent data $\boldsymbol{\vartheta}$ given $\boldsymbol{\Theta}$ and assume that the complete data likelihood consists of $L(\boldsymbol{\Theta}, \boldsymbol{\vartheta}; \mathbf{X})g(\boldsymbol{\vartheta}|\boldsymbol{\Theta})$. Suppose also that the prior $\pi(\boldsymbol{\Theta})$ is available for the unknown model parameters $\boldsymbol{\Theta}$. The marginal likelihood is then expressed as:

$$m(\mathbf{X}) = \int L(\boldsymbol{\Theta}, \boldsymbol{\vartheta}; \mathbf{X})\pi(\boldsymbol{\Theta})g(\boldsymbol{\vartheta}|\boldsymbol{\Theta})\ \partial(\boldsymbol{\Theta}, \boldsymbol{\vartheta}). \tag{1.5}$$

The marginal likelihood expresses the likelihood of the data conditional on the model entertained. It is the pivotal quantity in the construction of the Bayes factor, the main Bayesian model selection criterion. Suppose that S competing models M_s are under consideration, for $s = 1, \ldots, S$. The Bayes factor of M_s to $M_{s'}$ is then expressed as (Jeffreys, 1935, 1961):

$$B_{ss'} = \frac{m_s(\mathbf{X})}{m_{s'}(\mathbf{X})} = \frac{\displaystyle\int L_s(\boldsymbol{\Theta}_s, \boldsymbol{\vartheta}_s; \mathbf{X})\pi_s(\boldsymbol{\Theta}_s)g_s(\boldsymbol{\vartheta}_s|\boldsymbol{\Theta}_s)\ \partial(\boldsymbol{\Theta}_s, \boldsymbol{\vartheta}_s)}{\displaystyle\int L_{s'}(\boldsymbol{\Theta}_{s'}, \boldsymbol{\vartheta}_{s'}; \mathbf{X})\pi_{s'}(\boldsymbol{\Theta}_{s'})g_{s'}(\boldsymbol{\vartheta}_{s'}|\boldsymbol{\Theta}_{s'})\ \partial(\boldsymbol{\Theta}_{s'}, \boldsymbol{\vartheta}_{s'})}. \tag{1.6}$$

The Bayes factor embodies the ratio of posterior odds to prior odds for the models under consideration. The expression in (1.6) resembles a likelihood ratio. But instead of evaluating the respective likelihoods at the maximum likelihood estimates, the parameters are integrated out with respect to the respective priors. The BF thus can be viewed as representing a 'weighted' likelihood ratio that provides a measure "of the evidence provided by the data in favor of one scientific theory, represented by a statistical model, as opposed to another" (Kass & Raftery, 1995) which does not need referencing to an asymptotic distribution. The Bayes factor behaves like a natural Occam's razor, as model fit and complexity are accounted for in the marginal likelihood (Spiegelhalter & Smith, 1982; Jefferys & Berger, 1992). For interpretation the quantity (1.6) can be referred to half-units on the $\log_{10}$ scale (Jeffreys, 1961, Appendix B) or one can consider $2\log_e B_{ss'}$ (Kass & Raftery, 1995), which is on the same scale as likelihood ratio statistics. Another interpretational aid might be the posterior model probability, defined as:

$$P(M_{s'}|\mathbf{X}) = \left(\sum_{s=1}^{S} \frac{p_s}{p_{s'}} \cdot B_{ss'} \right)^{-1}. \tag{1.7}$$

In (1.7) p_s denotes the prior probability one assigns to model M_s being best, $\sum_{s=1}^{S} p_s = 1$. The posterior model probability $P(M_{s'}|\mathbf{X})$ gives the posterior probability, given the batch of models under consideration, that model $M_{s'}$ is the correct model for the data at hand. A normalization of the Bayes factor ensues when letting $p_s = S^{-1}\ \forall s$.

The Bayes factor as a model selection criterion has the following advantages (cf. Kass & Raftery, 1995; S. Y. Lee, 2007, Chapter 5): (i) It provides both a measure of evidence against a competing model and a measure of support for the alternative model; (ii) It will not by default favor the alternative model in (very) large samples; (iii) The comparison of any two models does not depend on the assumption that either one is 'true'; (iv) It allows one to take model uncertainty into account, thus providing a consistent quantity for the comparison of a multitude of competing models; (v) It can handle the comparison of both nested and nonnested models. There are also some disadvantages to utilization of the Bayes factor. The aims and motivations for this work lie in the interrelations between those disadvantages and constrained-model selection in the factor analytic setting.

1.3 Guiding Research Aims

1.3.1 Motivations

The Bayes factor and factor analytic modeling know certain interrelated problems and lacunae in their relation to Type I and II constrained-model selection. These drawbacks are embedded within two general difficulties with (1.5) and thus (1.6). First, the marginal likelihood will generally not be analytically tractable and especially when the parameter spaces are of high dimension, computation strategies can be challenging (see Kass & Raftery, 1995 and Han & Carlin, 2001). Second, prior distribution assessment, especially in multivariate cases, may display considerable difficulty when not using default (noninformative) options and may prove influential in the sense that differing specifications may render differing outcomes.

The main problem of Bayes factor utilization for Type I model selection is that both the use of improper noninformative and proper but vague priors yield indeterminate answers for (1.6) when the models to be compared are of differing dimension (Jeffreys, 1961). Such priors are (in the limit) only defined up to arbitrary constants, leading the Bayes factor to be defined only up to a ratio of arbitrary constants. This is undesirable as especially under default prior choices the interpretation of (1.6) as a weighted likelihood ratio is warranted (Berger & Pericchi, 2004). This problem reflects on Bayesian approaches towards EFA and the selection of m. Most approaches use conjugate priors, usually in conjunction with Anderson and Rubin's triangularity condition (1956, p. 121) on $\boldsymbol{\Lambda}$ for rotational determination. But even when the conjugate priors are vague, the minimal restrictions induce strong prior information. The block lower triangular form of $\boldsymbol{\Lambda}$ implies a preferred ordering by giving weight to the first m variables, thus determining both model interpretation and marginal computation strategies for the choice of m (Lopes & West, 2004). This calls into question the exploratory nature of Bayesian EFA. Moreover, the nature of regularity conditions tied to the rank of certain matrices, such as assumption (ii) from Section 1.2.1, might also have a determining effect on Bayesian strategies towards factor analytic dimensionality selection in general. A topic that has received little attention (see Lopes & West, 2004 for a notable exception and tentative exploration of this issue).

Quite recently a body of work emerged on using the Bayes factor for Type II model selection, especially for situations in which inequalities are placed on the parameter space (Kato & Hoijtink, 2006; Klugkist & Hoijtink, 2007; Laudy & Hoijtink, 2007; Hoijtink, Klugkist, & Boelen, 2008; Kato & Peeters, 2008; Mulder, Hoijtink, & Klugkist, 2010; M.-S. Oh & Shin, 2011). These works, however, only focus on (in)equalities on mean-type parameters and do not consider constraints on regression-type parameters. Inequality-constrained inference on regression-type parameters still has received relatively little attention in the literature. Moreover, many of these methods imply examination of the sample variance for prior specification (e.g., Klugkist & Hoijtink, 2007) or usage of very elaborate sample-based priors that need manual adjustment in order to avoid the implicit and automatic penalty for model complexity accounted for in (1.5) from defecting for restricted parameter spaces (e.g., Mulder et al., 2010).

In this dissertation Type II constrained-model selection is connected to CFA through the idea of expressing factor structure using inequalities on and between the λ_{jk}, which are essentially regression parameters. It is felt that restrainment of parameters in the traditional CFA context is in itself restrictive. As elaborated in Section 1.2.1, CFA seeks to incorporate theory into the factor model by imposing certain restrictions. These restrictions are meant to emulate sparse structure and are of the fixed-value equality kind with a focus on exclusion restrictions (e.g., $\lambda_{jk} = 0$). Such rigid character of parameter specification can be termed an unrealistic feature, as only the existence or non-existence of relationships between (observed and latent) parameters are usually taken into account. There are several perceived problems connected to such rigidity. First, it implies a loss of information in the sense that, to express an *a priori* factor structure, more exclusion restrictions are usually applied than is necessary for identification of the overparameterized FA model. Also, exclusion restrictions may amount to errors of omission, may make the unrealistic assumption that items are factorially pure (in the population), and may induce bias in estimates of the free parameters (cf. Ferrando & Lorenzo-Seva, 2000; van Prooijen & van der Kloot, 2001). These issues are intricately connected to the well-known and widespread situation of exploratively obtained factor structures not being confirmed by CFA (e.g., see McCrae, Zonderman, Costa, Bond, & Paunonen, 1996; Lonigan, Hooe, David, & Kistner, 1999). Moreover, there is little provision to let restricted parameters express substantive theoretical ideas regarding magnitude and direction of parameter effects. Researchers in substantive fields usually have informed ideas regarding direction and magnitude of parameter effects that cannot be expressed using exclusion restrictions. What is wished for then, is expressions of factor structure, not through usage of exclusion restrictions in $\mathbf{\Lambda}$, but by the imposition of inequality constraints, hence, through a system as (1.3).

A gap in FA practice is that there is no unequivocal strategy for integrating EFA and CFA. While the modi operandi are often viewed as distinct, it might be fruitful to view EFA and CFA as complementary techniques. For example, in the CFA model all attention regarding misspecification is geared towards the pre-specified pattern of factor loadings. The evaluation of model fit in CFA is then essentially the evaluation of a diffuse hypothesis (Hoyle & Duvall, 2004), as it is unclear in case of misspecification if the pattern of loadings or the factor dimensionality is to blame. A simple proposal stressing complementarity would be to separate dimensionality and pattern selection: Use EFA to determine the optimal choice of m only, then formulate, for given m, competing factor analytic structures for confirmatory statistical scrutiny. The topic of integrating EFA and CFA deserves further consideration.

1.3.2 Research Aims

The motivations for this research are connected to the desire to develop provisions for the mentioned ailments. There are then three research aims that guide the dissertation.

Research Aim 1 *To construct a conceptually and computationally simple Bayes factor for Type I constrained-model selection that is determinate under usage of improper priors. Subsequently, this Bayes factor is to be embedded within a strategy towards a truly Bayesian EFA concerned with the selection of an optimal dimensionality for* m.

Aim 1 intends to allow for the use of improper priors in Bayes factor computation for Type I constrained-model selection. While certain arrangements for improper prior usage have been developed previously (e.g., Berger & Pericchi, 1996), these sometimes rely on analytic tractability and generally do not have numerical implementations for complex problems. Such numerical extensions are aimed for. This subsequently allows for usage of improper prior information in EFA such that possible order preferences can be circumvented. Moreover, Bayesian computation proceeds by assessing the full posterior space, thus providing more information than ML point estimation, and giving a stronger stance in evaluating violation of regularity conditions. As such, more mature factor analytic dimensionality selection may ensue.

Research Aim 2 *To construct a conceptually and computationally simple Bayes factor for Type II constrained-model selection that is geared towards inequalities on regression-type parameters. Subsequently, this Bayes factor is to be embedded within a strategy which allows one to express factor analytic structure using inequality constraints rather than through exclusion restrictions.*

Most generally, fulfillment of Aim 2 will extend Bayesian model selection efforts regarding type II model selection and will add to the body of literature regarding inequality-constrained inference on regression-type parameters. More specifically, this research aim holds importance for the theory and technique of CFA. The formulation and development of inequality constrained models allows for theoretically meaningful constrained coefficients beyond those needed simply to identify a model. This will allow for more realistic specification of model translated theories as substantive ideas regarding the strength of relationships and associations can then be explicitly taken into account. In doing so, more information can be subtracted from a single analysis and a new take on CFA can be rendered.

Research Aim 3 *To let the provisions from Aims 1 and 2 conjoin in order to develop an integrative factor analytic strategy that proposes a bridge crossing the divide between EFA and CFA and to bring this strategy to bear on substantive fields of study outside the direct realm that brought about the FA model.*

Aim 3 reveals that EFA and CFA are seen as complementary techniques. EFA is not only seen as an abductive method (Haig, 2005), but also as an inferential technique regarding choice of m. As such, EFA and CFA can potentially be used in conjunction. Aim 3 answers the call for thinking of ways of "integrating exploratory and confirmatory approaches to factor analysis" (Steiger, 1994). A factor analytic strategy is proposed that aims to integrate EFA and CFA without spurring concerns regarding cross-validation and *post hoc* selection. The reason for seeking applications outside the boundaries of psychometrics is that, outside this realm, FA seems less immediate, but is potentially very useful and viable. Moreover, the trickle down effect for new factor analytic developments will most probably meet more obstacles outside the psychometric field.

The first two focal points are of a technical nature, while the last envisioned endeavor has a more applied nature. It will be the technical part that will be given the initial weight. The boundaries that guide the research aims are reviewed before giving an extensive outline of the dissertation.

1.3.3 Assumptions and Boundaries

It seems prudent to explicitly state the assumptions and boundaries of present work. First, the underlying assumption in this work is that the common factor analytic model is an appropriate model for the data at hand. Also, the selection of manifest variables is not considered (for this topic, see e.g., Kano & Harada, 2000; Kano, 2007). It is assumed that a meaningful set of variables, representing a non-singular sample covariance matrix, is available.

Focus is thus with the common factor analytic model, although its properties are thought of as well-studied. However, this dissertation will show that certain (relatively) unknown sources of indeterminacy cast doubt on the regularity conditions of heavily used test statistics. Moreover, usage of elaborate systems of inequality constraints as an expression of factor structure is seen as an important extension of the common factor model.

Another boundary of the work is that it will focus on moderately-sized data (data collected by the individual scientist). The provisions that are developed are computationally intensive and not directly aimed at handling very large data sets. Although interesting and important, focus will thus not be with the massive amounts of data currently sweeping many applications, such as those in genomics (e.g., Carvalho et al., 2008). However, studying Type I and Type II constrained-model selection for the classic common factor model geared towards moderately-sized data is prone to give pointers for the directions and pitfalls of factor analysis for massive (high-throughput) data.

Lastly, the take on constrained-model selection expresses a conviction regarding model selection and the FA model. The conviction is that in latent variable modeling one is to determine the intrinsic dimensionality of the model (reflected here through the choice of m), before competing structures (possibly containing inequalities or other functional parameter restrictions) are to be assessed, in order to avoid embarking on a diffuse model question.

1.4 Overview

1.4.1 Bird's-eye View

Part I comprises Chapters 2 to 4 and covers Research Aims 1 and 2. Focus is on statistical and computational issues regarding formulation, prior selection, parameter estimation and model selection for constrained Bayesian factor models for the multivariate normal covariance matrix. These issues will be elaborated with simulated and real-data examples. The illustrations serve as the precursor for Part II which comprises Research Aim 3 and which purports that researchers often have competing theories that can be translated into inequality-constrained factor analytic models. It will be shown that after this translation an evaluation of these theories is rather straightforward. Below an overview is provided of the developments accompanying the stated aims. Table 1.1 provides a reading guide to the dissertation. This dissertation is constructed of research articles. This leads to each chapter being self-contained and some inevitable redundancy in presentation. Notation, however, is consistent and largely self-explanatory throughout. Fortran 90 programs implementing the methods proposed in the technical chapters can be obtained from the author.

Table 1.1. Reading Guide

	Elements					
	Factor analysis		Constrained inference		Bayesian model	
Chapter	EFA	CFA	Type I	Type II	selection	Aim
2	✓		✓		✓	1
3	✓	✓				2
4		✓		✓	✓	2
5	✓	✓	✓	✓	✓	3
6	✓	✓	✓	✓	✓	3

1.4.2 Part I

Chapter 2 (Peeters, invited revision a) reviews Markov chain Monte Carlo (MCMC) computation of the marginal likelihood. The candidate estimator method for marginal likelihood computation (Besag, 1989; Chib, 1995) is adapted to deal with (i) improper noninformative priors and (ii) the existence of (well-separated) symmetric posterior modes due to permutative invariance over the parameter indices, such that the ensuing Bayes factor is still determinate. Pending certain conditions, the provisions provide for what can be seen as a simulation consistent MCMC implementation of well-known default Bayes factors (Berger & Pericchi, 1996). This automated candidate estimator is subsequently applied to latent factor dimensionality selection in EFA. This application is inspired by (i) a desire to stringently test the provisions and (ii) to spur learning on some lesser known indeterminacies in the factor model and their interrelationships with computational approaches towards dimensionality selection. It will be shown that a failure to abide regularity condition (ii) from Section 1.2.1 may result in violation of a crucial regularity condition for simulation consistency of estimates stemming from MCMC sampling. This implies that Bayesian approaches towards factor analytic dimensionality selection may be haunted by some of the same regularity conditions that hamper classical approaches. Embedding the automated candidate estimator in a certain assessment strategy that keeps check of the regularity condition for simulation consistency provides an appropriate stopping rule for factor analytic data compression. In passing, a truly Bayesian EFA is proposed. These results also hold importance for non-Bayesian approaches towards factor analytic dimensionality selection. They imply that for informed decisions regarding factor dimensionality, LR and information theoretic approaches benefit from a complete exploration of the likelihood, which can be achieved by objective Bayesian methods.

Chapter 3 (Peeters, 2012) deals with two sets of conditions for rotational identification of the oblique factor solution under utilization of fixed zero elements in the factor loadings matrix (Jöreskog, 1979). These condition sets, formulated under factor correlation and factor covariance metric respectively, were claimed to be equivalent and to lead to global rotational uniqueness of the factor solution. It is shown that the conditions for the oblique factor correlation structure need to be amended for global rotational uniqueness, and hence, that the condition sets are not equivalent. The amended condition set provides a way to design an unrestricted solution to the (Bayesian) CFA model. Unrestricted solutions correspond to exploratory factor analysis (EFA) in the sense that only minimal restrictions are placed on the model to achieve a (global) rotationally unique solution for m factors. As such, an unrestricted solution for m common factors does not restrict the factor space and will yield an optimal fit for any model with m factors (Mulaik, 2010, Section 15.4). As indicated above, in the EFA tradition this is usually achieved by requiring $\mathbf{\Phi} = \mathbf{I}_m$ and $\mathbf{\Lambda}^{\mathrm{T}}\mathbf{\Psi}^{-1}\mathbf{\Lambda}$ be diagonal accompanied by an order condition on the diagonal elements. But such restrictions are mere convenience for estimation purposes and have no interpretational meaning. An unrestricted confirmatory factor model (UCFM) is a FA model that places only minimal restrictions on $\mathbf{\Lambda}$ and $\mathbf{\Phi}$ for achieving global rotational uniqueness of the factor solution, with the restrictions chosen such that they convey preconceived theoretical meaning and thus render unnecessary post-hoc rotation of the solution for interpretation purposes. The UCFM is pivotal in designing a strategy for inequality-constrained CFA in Chapter 4.

In Chapter 4 (Peeters, invited revision b) a Bayesian framework is proposed which takes parameter restrictions in the context of CFA beyond exclusion restrictions and the prevention of impermissible estimates by allowing inequality and approximate equality constraints to express substantive theoretical ideas regarding direction and magnitude of effect of factor loadings. It focuses first on the development of a Bayes factor for Type II constrained-model selection. Interest then lies with the demarcation of competing inequality constrained formulations of factor analytic correlation structure. The strategy consists of choosing as a base model a UCFM. Substantive theory will then not be represented by structural exclusions to express a pre-specified loading pattern, but by imposing inequalities on and between the free parameters in $\mathbf{\Lambda}$. It is then shown that when (i) proper but noninformative priors are chosen that are flat on the parameter space of the parameters on which inequalities are placed; and (ii) all competing inequality-constrained models are subsets of the UCFM; then the ensuing Bayes factor is determinate, its complexity is well-defined, and its computation is strongly simplified as under this framework model fit and complexity are explicitly connected to, respectively, the posterior and prior probability mass satisfying the constraints defining the constrained model. Figure 1.1 gives a schematic geometric depiction of the workings

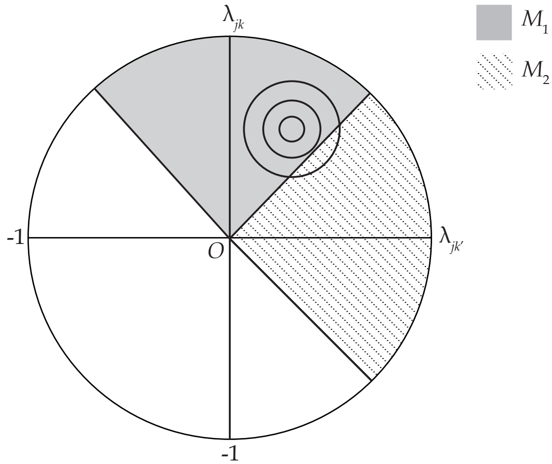

Fig. 1.1. Schematic geometric depiction of the Bayes factor for inequality-constrained model selection in the factor analytic setting. The unit circle represents the prior space (probability mass) which is bound by the communality of the standardized FA solution. Model M_1 represents the feasible space defined by $\lambda_{jk} > |\lambda_{jk'}|$ whereas competing model M_2 represents the feasible space defined by $\lambda_{jk'} > |\lambda_{jk}|$, both of which are expressible through (1.3). The concentric circles represent the posterior probability mass. Both models are of the same complexity (the feasible space comprises 1/4 of the total space), but the posterior mass is mostly located in the feasible region defined by the constraints connected to model M_1. Say 7/8 of the posterior mass is located in the feasible region of M_1 and that 1/8 of the posterior mass is located in the feasible region of M_2. The bayes factor of M_1 to the unconstrained model, $B_{10} = (7/8) \div (1/4) = 7/2$. The analogous Bayes factor for M_2, $B_{20} = (1/8) \div (1/4) = 1/2$. The Bayes factor of model M_1 to M_2 is then found as $B_{12} = (7/2) \div (1/2) = 7$. In this toy example there is thus positive evidence of M_1 against M_2: Loosely speaking, model M_1 is seven times as likely for the data at hand than model M_2.

of the Bayes factor for inequality-constrained-model selection in the factor analytic setting.

The inequality constrained approach emphasizes a break with simple structure models. Constraining factor loadings to zero may be unrealistic and may push the outcomes in a certain direction, as it leaves little possibility for the substantive evaluation of the implicit claim that certain relations are non-existent, and gives no inroads for the specification of the relative strengths of relationships. With inequality constraints one can more realistically accommodate theoretical ideas so that one can determine the model with the highest posterior probability when initially a multitude of plausible explanations is available. The developments may be viewed as constituting a new take on CFA. Chapter 4 ends with a proposal for a factor analytic strategy for bridging the EFA-CFA divide. This proposal forms the heart of Aim 3 which is deepened in Part II.

1.4.3 Part II

Part II conjoins the developments from Part I in an alternative factor analytic strategy merging EFA and CFA. This strategy lets EFA prelude inequality-constrained CFA efforts by making it part of a total inferential procedure involving the selection of an optimal dimension m before competing confirmatory structures are assessed. The strategy consists of the following steps: (1) Embark on evaluating a series of unrestricted (EFA) models with respect to their factor dimensionality; (2) Whence settled on latent factor dimensionality m, specify a UCFM; (3) Formulate, using the UCFM as a base model, competing inequality-constrained factor structures making use of a system of inequality constraints on and between the free parameters in the loadings matrix (i.e., making use of (1.3)); (4) Compute for each constrained model the Type II constrained-model selection Bayes factor and determine the constrained model most supported by the data.

In Chapter 5 (Peeters, Dziura, & van Wesel, under review) the alternative strategy is brought to bear on research regarding MBS. As stated previously, MBS data form a recurring example throughout Part I. In this chapter the example is deepened. FA has become part and parcel in MBS research. Both exploration- and confirmation-driven factor analyzes are rampant. However, factor analytic results on MBS differ widely. A situation that is at least in part attributable to misapplication of FA. Here, factor analytic efforts in the study of MBS are reviewed with emphasis on misusage of the FA model. The alternative factor analytic strategy is proposed to confronts weaknesses in application of the FA model in the MBS realm. A high-profile MBS data set with anthropometric measurements on overweight and obese children and adolescents is reanalyzed using the alternative strategy. The findings may give renewed cachet to both FA and its connection to MBS research.

In Chapter 6 (Peeters & van Wesel, under review) a framework concerning Bayesian inequality-constrained statistical inference in the general linear model is depicted. Focus is on the ANOVA and FA model. In the political sciences a confirmatory-oriented approach towards the incorporation of inequalities on parameters of interest is often appropriate. Classical approaches towards analyzing (multiple) inequality-constrained hypotheses are often difficult. Type II constrained Bayesian model selection as developed in Chapter 4 is then illustrated with two examples from political science. The first example concerns an analysis of variance context while the second concerns a confirmatory factor analysis that makes use of the alternative factor analytic strategy. Both examples are meant to show the possibilities of the methods shaping the constrained-model selection framework.

Chapter 7 concludes with a discussion. It is a reflexive chapter as it explores the limitations of present work as well as how these limitations can provide inroads for further research. Moreover, it states how the methods developed in this dissertation are perceived to be connected to a philosophy of method.

Part I

Statistical & Computational Modeling

2

A Default Candidate Estimator for Marginal Likelihood Estimation with Application to Factor Analytic Dimensionality Selection

Peeters, C.F.W. (currently under revision for *Psychometrika*)

2.1 Introduction

Suppose that the models in $\mathcal{A} = \{M_1, \ldots, M_A\}$ are under consideration for data $\mathbf{X}$. Assume that each model M_a is characterized by the model-specific collection of unknown parameters $\mathbf{\Theta}_a$ and likelihood $L_a(\mathbf{\Theta}_a; \mathbf{X})$. Also assume that we have $\pi_a(\mathbf{\Theta}_a)$, $a = 1, \ldots, A$, as the available prior distributions for the unknown parameters. The key quantity in Bayesian model selection is the marginal likelihood of the data

$$m_a(\mathbf{X}) = \int L_a(\mathbf{\Theta}_a; \mathbf{X})\pi_a(\mathbf{\Theta}_a)\partial\mathbf{\Theta}_a, \tag{2.1}$$

which constitutes the normalizing constant for the posterior distribution $\pi_a(\mathbf{\Theta}_a|\mathbf{X})$ and is the pivotal quantity in the construction of the Bayes factor (Jeffreys, 1935, 1961):

$$B_{aa'} = \frac{m_a(\mathbf{X})}{m_{a'}(\mathbf{X})} = \frac{\int L_a(\mathbf{\Theta}_a; \mathbf{X})\pi_a(\mathbf{\Theta}_a)\partial\mathbf{\Theta}_a}{\int L_{a'}(\mathbf{\Theta}_{a'}; \mathbf{X})\pi_{a'}(\mathbf{\Theta}_{a'})\partial\mathbf{\Theta}_{a'}}. \tag{2.2}$$

The Bayes factor as a model selection criterion has the following advantages (cf. Kass & Raftery, 1995; S. Y. Lee, 2007, Chapter 5): (i) It provides both a measure of evidence against a competing model and a measure of support for the alternative model; (ii) It will not by default favor the alternative model in (very) large samples; (iii) The comparison of any two models does not depend on the assumption that either one is 'true'; (iv) It allows one to take model uncertainty into account, thus providing a consistent quantity for the comparison of a multitude of competing models; (v) It can handle the comparison of both nested and nonnested models; (vi) It can be easily converted into an interpretable scale.

Notwithstanding, there are several difficulties with (2.1) and thus (2.2): (i) The marginal likelihood will generally not be analytically tractable and especially when the parameter spaces are of high dimension, computation strategies can be challenging (cf. Kass & Raftery, 1995; Han & Carlin, 2001); (ii) Both the use of improper noninformative and proper but vague priors yield indeterminate answers for (2.2) (Jeffreys, 1961). This is undesirable as especially under default prior choices the interpretation of (2.2) as a weighted likelihood ratio is warranted (Berger & Pericchi, 2004); (iii) When not using default options for the priors $\pi_a(\mathbf{\Theta}_a)$, their assessment, especially in multivariate cases, may display considerable difficulty and may prove influential in the sense that differing specifications may render differing outcomes.

This paper has two main goals. First, we seek to develop an automated marginal likelihood computation strategy aimed at providing indirect estimation of the marginal likelihood in enumerable model spaces for the general non-nested setting. Automated is taken to mean determinate under usage of improper (limiting) priors and possible permutative invariance over the parameter indices. Second, we seek to apply this strategy to the problem of dimensionality selection in factor analysis (FA). The remainder of the introduction provides the backdrop for these motivations.

2.1.1 Factor Analysis and Dimensionality Selection

FA assumes that a random vector $X \in \mathbb{R}^p$ consists of correlated variables that can be grouped by their covariances into a random vector of latent factors $\Xi \in \mathbb{R}^m$, with $m < p$. In addition to being a data-reductional tool, FA as a data-analytic tool has become an appreciated aid in such diverse fields as medical research, social and behavioral science, and natural science.

Common exploratory FA (EFA) is a staple in the exploratory spectrum of the FA tradition. While many analyses conducted using this technique preassign m or assume m to be known in advance, such an approach may not be tenable or appropriate in an exploratory analysis. The selection of m reflects the requirement of balancing parsimony with the adequate representation of the information contained in the original covariance matrix. A poor choice of latent dimensionality has inappreciable consequences for the reduction and interpretation of data. Retaining too few latent factors implies loss of information, while retaining too many factors implies a violation of the parsimony principle.

2.1.2 Classical Stopping Rules

Many stopping rules have been developed for assessing multivariate data compression in the FA setting. These procedures can be generally classified as either heuristical or statistical, where the statistical methods may be further partitioned into analytical and parametrical procedures (cf. Bartholomew, 1987; Jackson, 1993; Basilevksy, 1994; Fabrigar, Wegener, MacCallum, & Strahan, 1999; Hong, Mitchell, & Harshman, 2006; Cangelosi & Goriely, 2007).

The simplest heuristical methods are to examine the residual covariance matrix or the proportion of total variance explained. The researcher decides on the number of factors which lead to a satisfactory residual matrix or proportion of explained variance. Two of the most well-known and widely used heuristical procedures are based on a direct examination of the eigenvalues extracted from the observed covariance matrix. The first is the Guttman-Kaiser rule (Guttman, 1954; Kaiser, 1960) which states that one should retain at most those factors associated with eigenvalues whose magnitude exceeds the average eigenvalue. The second is the scree plot (Cattell, 1966) which plots the successive eigenvalues against the rank order. The number of factors is decided upon based on the sharpest drop of the eigenvalue. Drawbacks of these methods are that they do not provide a formal test but rely on subjective judgment regarding the number of factors to retain as well as the possibility that they may spur arbitrariness of decision when ambiguous patterns emerge from the data. As such they have been prone to serious under- and overfactoring.

The analytical procedures seek to propose more objective variations of the heuristical methods listed above. The minimum average partial correlation procedure (Velicer, 1976) proposes a more stringent look at the residual matrix. Parallel analysis (Horn, 1965) poses a sampling-based alternative to the Guttman-Kaiser rule. Various regression-based procedures have been developed to form an objective alternative for the scree plot (Gorsuch & Nelson, 1981; Jurs, Zoski, & Mueller, 1993; Zoski & Jurs, 1993). Additionally, various resampling procedures have been proposed for both the Guttman-Kaiser rule and scree plot (Z. V. Lambert, Wildt, & Durand, 1990; Jackson, 1993; Hong et al., 2006). While analytical procedures improve the accuracy of heuristical methods, they are subject to the same caveats.

Likelihood ratio (LR) testing procedures have become available for EFA with the advent of computationally feasible maximum likelihood (ML) estimation for factor analysis (Jöreskog, 1967). As is well-known, standard statistical theory states that the LR test statistic is asymptotically distributed as a (noncentral) χ^2 variable with degrees of freedom equal to the difference in the number of free parameters between the two nested models under consideration. LR testing in EFA proceeds by sequences of tests comparing the model with at most m factors against either the saturated model or the model with at least $m+1$ factors. While this procedure has a formal statistical foundation, it is not without problems.

Generally, regularity conditions (Wilks, 1938) are quite easily violated. For example, the LR test is very sensitive to sample size. The asymptotic χ^2 approximation is invalid with small sample sizes (see Geweke & Singleton, 1980, for an exploration of this issue). On the other hand, by the consistency of the testing procedure the rejection of the null hypothesis becomes the sure event for a sample size sufficiently large. Say, however, that we abide the obvious regularity conditions: sample size is moderate, the observables are multivariate normally-distributed, and the problem considered has nonnegative degrees of freedom. Two regularity conditions intricately related to the nature of the FA model are then pivotal in the validity of the LR test: (a) the matrix of factor loadings is of full rank, and (b) the diagonal matrix of error variances is of full rank (the mentioned matrices will be defined in Section 2.2.1). Rank deficiencies in mentioned matrices are however endemic when retaining too many factors (Section 2.2.2). Due to problems of rank deficiency and subsequent nonidentifiability of model parameters the LR test that employs an exploration into numbers of factors that exceed the true factor dimensionality no longer follows the χ^2 distribution (Takane et al., 2003; Hayashi et al., 2007). This may render only the difference test between the model of correct dimensionality and the saturated model meaningful and in practice leads to overfactoring.

Related to the LR test is a plethora of (variants on) information criteria such as the AIC (Akaike, 1973, 1987), BIC (Akaike, 1978; Schwarz, 1978), and ICOMP (Bozdogan & Shigemasu, 1998). These are usually thought of as improvements over the LR test. However, since their complexity correction criterion is explicit it is hard to establish one that is consistent across a range of data and model specifications. Moreover, as the heart of these criteria also consists of evaluations of the model likelihood function at the maximum likelihood estimates, they too are affected by the regularity conditions on the rank of certain matrices mentioned above, possibly accounting for the oft observed fact that many information criteria have a bias towards more complex models, resulting in a tendency to overestimate model size in FA (e.g., Lopes & West, 2004).

The marginal likelihood implicitly weighs model fit and model complexity. Also, the Bayesian way of life is usually thought of as being less amenable to regularity conditions. Hence the attractiveness of the Bayesian approach.

2.1.3 Bayesian Approaches

Several strategies have been employed to compute Bayes factors for the sake of Bayesian dimensionality selection in FA. Press and Shigemasu (1999) employ large sample approximations. Many current approaches however thrive on Markov chain Monte Carlo (MCMC) techniques, which provide for a more general recipe that translates to a much larger set of problems and is much more robust against sample sizes. Polasek (2000) uses the candidate estimator method (Besag, 1989; Chib, 1995), which aims to give an estimate of the marginal likelihood of each model under consideration. Lee and Song (2002) utilize path sampling (Gelman & Meng, 1998) which is an extension of thermodynamic integration that sets out to compute a ratio of normalizing constants. Fokoué (2009) and Lopes and West (2004) sample both parameter and model space with birth-and-death MCMC (Stephens, 2000) and reversible jump MCMC (Green, 1995) algorithms respectively.

While useful, these techniques suffer from several drawbacks related to the general difficulties of marginal computation and the specifics of the MCMC procedures entertained (see Han & Carlin, 2001 for a comparison of various MCMC strategies for marginal likelihood and Bayes factor computation). Computing the marginal for each model under consideration using MCMC can become essentially an exercise in bookkeeping when the model set is large. MCMC methods that sample both model and parameter space often imply serious and difficult tuning efforts to promote mixing in high dimensional space. Moreover, all methods require proper informative priors whose assessment may be hard in multivariate situations or inappropriate when one wishes to induce little prior information. Specifically for the FA situation, implementation of these methods has generally required identification constraints which imply a preferred ordering, and as such induce unintended prior information (Section 2.2.3).

An interesting strategy is the candidate estimator method, first proposed by Besag (1989) and fully explored by Chib (1995) and Chib and Jeliazkov (2001). It is based on the following identity implied by Bayes' theorem:

$$m_a(\mathbf{X}) = \frac{L_a(\boldsymbol{\Theta}_a; \mathbf{X})\pi_a(\boldsymbol{\Theta}_a)}{\pi_a(\boldsymbol{\Theta}_a|\mathbf{X})}, \tag{2.3}$$

whose evaluation at any candidate estimate $\boldsymbol{\Theta}_a^*$ of $\boldsymbol{\Theta}_a$ provides for an estimate of the marginal likelihood. As $L_a(\boldsymbol{\Theta}_a; \mathbf{X})$ and $\pi_a(\boldsymbol{\Theta}_a)$ are generally available in closed form they can be evaluated directly at $\boldsymbol{\Theta}_a^*$. The estimate of the posterior ordinate can be obtained through MCMC sampling. This strategy lends its attractiveness from its conceptual simplicity and the fact that its implementation needs little additional programming beyond coding for standard Bayesian analyses. However, besides being amenable to some of the drawbacks stated above, the method may produce flawed estimates when multimodal problems are considered (Neal, 1999) (Section 2.3.2).

Ideally, what would be desired is a marginal likelihood computation strategy that is conceptually simple, that is not detrimented by multimodality due to permutative invariance, and that gives determinate (pseudo) Bayes factors when incorporating improper noninformative prior information. Here we set out to provide for such a strategy. Basically, the candidate estimator method will be integrated with the use of training samples for the construction of posterior priors such as (i) to cope with the indeterminacy in (2.2) when using noninformative priors and (ii) to provide, conditional on the possibility to restrict draws from an MCMC scheme to exactly one permutative mapping, an automatic correction for the existence of multiple symmetrical modes often encountered in latent variable models.

By means of demonstration we will have special applicatory attention for the problem of factor analytic dimensionality selection in the common EFA setting. The applicatory attention is inspired by a desire to stringently test the provisions and to spur learning on some lesser known indeterminacies in the factor model and their interrelationships with (MCMC-based) computational approaches towards dimensionality selection. We will especially show that a failure to abide regularity condition (a) mentioned in previous section may result in violation of a crucial regularity condition for simulation consistency of estimates stemming from MCMC sampling (Section 2.4.5). This implies that Bayesian approaches towards factor analytic dimensionality selection may be haunted by some of the same regularity conditions that hamper classical approaches. It also forces one to embed a marginal likelihood computation strategy in an assessment strategy in order to come to an appropriate stopping rule for factor analytic data compression.

2.1.4 Overview

Section 2.2 provides the theoretical framework of the common EFA model. The identifiability and indeterminacy issues inherent to this model motivate some extensions of the candidate estimator method in Section 2.3. In Section 2.4 this automated marginal likelihood computation strategy is brought to bear on the problem of latent dimensionality selection in common EFA. This section also contains some simulation efforts. Section 2.5 illustrates the conjunction of developed techniques and FA through exploring a substantive application from the field of medicine. Section 2.6 concludes with a discussion.

2.2 The Factor Analytic Model

The model we consider here is the common FA model. Let $M_a \in \mathcal{A}$ express such a model of given factor dimensionality m. Model structure and assumptions are shown to be intricately related to the topics of 'identifiability' and 'indeterminacy'. The approach of which motivates the specificities of the marginal likelihood computation strategy to be developed in following sections. To aid notational simplicity, the model index a will be suppressed where appropriate for the remainder of the text.

2.2.1 Model Structure and Assumptions

Let $\mathbf{X}^{\mathrm{T}} \equiv [\mathbf{x}_1, \ldots, \mathbf{x}_n]$ define p-variate observation vectors on $i = 1, \ldots, n$ subjects, such that $\mathbf{x}_i^{\mathrm{T}} \equiv [\mathrm{x}_{i1}, \ldots, \mathrm{x}_{ip}] \in \mathbb{R}^p$ denotes a realization of the random vector $X_i^{\mathrm{T}} \equiv [X_{i1}, \ldots, X_{ip}] \in \mathbb{R}^p$. Also, let $\boldsymbol{\Xi}^{\mathrm{T}} \equiv [\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_n]$ define m-variate vectors of latent factor scores on n subjects with $\boldsymbol{\xi}_i^{\mathrm{T}} \equiv [\xi_{i1}, \ldots, \xi_{im}] \in \mathbb{R}^m$. The factor analysis model states that each random variable X_i is a linear combination of the latent random variables, such that:

$$\underset{(p \times 1)}{\mathbf{x}_i} = \underset{(p \times 1)}{\boldsymbol{\mu}} + \underset{(p \times m)}{\boldsymbol{\Lambda}} \cdot \underset{(m \times 1)}{\boldsymbol{\xi}_i} + \underset{(p \times 1)}{\boldsymbol{\epsilon}_i} \tag{2.4}$$

with $m < p$. In (2.4) $\boldsymbol{\mu} \in \mathbb{R}^p$ denotes an overall mean vector, the $\boldsymbol{\epsilon}_i \in \mathbb{R}^p$ denote the error measurements, and $\boldsymbol{\Lambda} \in \mathbb{R}^{p \times m}$ is a matrix of factor loadings in which each element λ_{jk} is the loading of the jth variable on the kth factor, $j = 1, \ldots, p$, $k = 1, \ldots, m$.

The model maintains the following assumptions: (i) $\mathrm{x}_{ij} \perp\!\!\!\perp \mathrm{x}_{i'j'} | \{\xi_{i1}, \ldots, \xi_{im}\}, \forall i, j \neq i', j'$; (ii) $\mathrm{rank}(\boldsymbol{\Lambda}) = m$; (iii) $\boldsymbol{\epsilon}_i \sim \mathcal{N}_p(\mathbf{0}, \boldsymbol{\Psi})$, with $\boldsymbol{\Psi} \equiv \mathrm{diag}[\psi_{11}, \ldots, \psi_{pp}]$, and $\psi_{jj} > 0$; (iv) $\boldsymbol{\xi}_i \sim \mathcal{N}_m(\mathbf{0}, \mathbf{I}_m)$; and (v) $\boldsymbol{\xi}_i \perp\!\!\!\perp \boldsymbol{\epsilon}_{i'}, \forall i, i'$. The likelihood for the observations conditional on the realization of $\boldsymbol{\Xi}$ can then be expressed as:

$$L(\boldsymbol{\mu}, \boldsymbol{\Lambda}, \boldsymbol{\Xi}, \boldsymbol{\Psi}; \mathbf{X}) = \prod_{i=1}^{n} f(\mathbf{x}_i | \boldsymbol{\mu}, \boldsymbol{\Lambda}, \boldsymbol{\xi}_i, \boldsymbol{\Psi}) = \prod_{i=1}^{n} (2\pi)^{-\frac{p}{2}} |\boldsymbol{\Psi}|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\boldsymbol{\epsilon}_i^{\mathrm{T}} \boldsymbol{\Psi}^{-1} \boldsymbol{\epsilon}_i\right\}, \tag{2.5}$$

where $\boldsymbol{\epsilon}_i = \mathbf{x}_i - \boldsymbol{\mu} - \boldsymbol{\Lambda}\boldsymbol{\xi}_i$. The likelihood in this form will be important for the construction of data-augmented conditional distributions for MCMC sampling. Marginalizing over $\boldsymbol{\xi}_i$ we obtain the likelihood of the observed data:

$$L(\boldsymbol{\mu}, \boldsymbol{\Lambda}, \boldsymbol{\Psi}; \mathbf{X}) = \prod_{i=1}^{n} \int f(\mathbf{x}_i | \boldsymbol{\mu}, \boldsymbol{\Lambda}, \boldsymbol{\xi}_i, \boldsymbol{\Psi}) g(\boldsymbol{\xi}_i; \mathbf{0}, \mathbf{I}_m)\, \partial \boldsymbol{\xi}_i$$
$$= \prod_{i=1}^{n} (2\pi)^{-\frac{p}{2}} |\boldsymbol{\Sigma}|^{-\frac{1}{2}} \exp\left\{ -\frac{1}{2} (\mathbf{x}_i - \boldsymbol{\mu})^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} (\mathbf{x}_i - \boldsymbol{\mu}) \right\}, \quad (2.6)$$

giving that the factor decomposition constrains the covariance structure of the $\mathbf{x}_i$ to $\boldsymbol{\Sigma} = \boldsymbol{\Lambda}\boldsymbol{\Lambda}^{\mathrm{T}} + \boldsymbol{\Psi}$. Then, for existence (vi), generally $m \leqslant [2p+1-(8p+1)^{1/2}]/2 \equiv \phi(p)$ (Ledermann, 1937), simply stating that the number of nonredundant elements in the sample covariance matrix $\mathbf{S}$ must be greater than or equal to the number of freely estimable parameters in $\boldsymbol{\Sigma}$, which places in fact an upper bound on the number of latent factors to be extracted.

2.2.2 Identifiability and Indeterminacies

It is well known that for given $\boldsymbol{\Lambda}$ and $\boldsymbol{\Psi}$, the former is defined uniquely only up to orthogonal rotation (a topic to which we will return below and throughout the paper). Correspondingly the FA literature has focussed mainly on identification of $\boldsymbol{\Psi}$, the main result of which is related to assumption (vi):

Theorem 2.1 (e.g., Bekker, Merckens, & Wansbeek, 1994). *If $m \leqslant \phi(p)$, then $\boldsymbol{\Psi}$ is locally identified for almost every $\boldsymbol{\Lambda}$ and $\boldsymbol{\Psi}$. For $m > \phi(p)$, $\boldsymbol{\Psi}$ is locally, and hence globally, not identified for almost every $\boldsymbol{\Lambda}$ and $\boldsymbol{\Psi}$.*

It is lesser known (Hayashi et al., 2007 being a notable exception) that this result is contingent upon the rank of $\boldsymbol{\Lambda}$. The implications of a failure to abide model assumption (ii) were explored by Anderson and Rubin (1956) and Geweke and Singleton (1980). Suppose that $\mathrm{rank}(\boldsymbol{\Lambda}) = r < m$. Then there exists a matrix $\mathbf{Q} \in \mathbb{R}^{m \times (m-r)}$ for which $\boldsymbol{\Lambda}\mathbf{Q} = \mathbf{0}$ and $\mathbf{Q}^{\mathrm{T}}\mathbf{Q} = \mathbf{I}_{m-r}$, such that for any $\mathbf{M} \in \mathbb{R}^{p \times (m-r)}$ with mutually orthogonal rows

$$\boldsymbol{\Sigma} = \boldsymbol{\Lambda}\boldsymbol{\Lambda}^{\mathrm{T}} + \boldsymbol{\Psi} = (\boldsymbol{\Lambda} + \mathbf{M}\mathbf{Q}^{\mathrm{T}})(\boldsymbol{\Lambda} + \mathbf{M}\mathbf{Q}^{\mathrm{T}})^{\mathrm{T}} + (\boldsymbol{\Psi} - \mathbf{M}\mathbf{M}^{\mathrm{T}}). \quad (2.7)$$

Equation (2.7) implies that no consistent estimator of $\boldsymbol{\Psi}$ exists if $\boldsymbol{\Lambda}$ fails to be of full column rank. This may induce corresponding multimodalities in the densities of $\boldsymbol{\Psi}$ and $\boldsymbol{\Lambda}$ (Lopes & West, 2004), and may lead to the Bayesian analogue of the Heywood case when using naive priors. These issues are related to the choice of factor dimensionality and the possibility of retaining too many factors and we will return to them in Section 2.4.

The FA model also copes with two inherent indeterminacies of the parameters, being: an undefined metric for the latent factors, and rotational indeterminacy of the factor solution. In our model the metric for the latent variables is set by assumption (iv). Regarding rotational indeterminacy, assume that $\mathbf{H} \in \mathbb{R}^{m \times m}$ is an arbitrary orthogonal matrix. Returning to the implied covariance structure of the observed data, we then have

$$\boldsymbol{\Sigma} = \boldsymbol{\Lambda}\boldsymbol{\Lambda}^{\mathrm{T}} + \boldsymbol{\Psi} = (\boldsymbol{\Lambda}\mathbf{H})(\boldsymbol{\Lambda}\mathbf{H})^{\mathrm{T}} + \boldsymbol{\Psi}. \quad (2.8)$$

Equation (2.8) implies that, given $\boldsymbol{\Psi}$, there is an infinite number of alternative matrices $\boldsymbol{\Lambda}\mathbf{H}$ that generate the same covariance structure as $\boldsymbol{\Lambda}$. Thus, in any solution, $\boldsymbol{\Lambda}$ can be made to satisfy $m(m-1)/2$ additional conditions, which is naturally equivalent to the number of independent elements of $\mathbf{H}$. Hence the structure of the existence condition (vi).

Note that assumption (iv) does not imply a loss of generality. Suppose that $\boldsymbol{\xi} \sim \mathcal{N}_m(\mathbf{0}, \boldsymbol{\Phi})$, with $\boldsymbol{\Phi}$ positive definite, implying the covariance structure of the $\mathbf{x}_i$ to be $\boldsymbol{\Sigma} = \boldsymbol{\Lambda}\boldsymbol{\Phi}\boldsymbol{\Lambda}^{\mathrm{T}} + \boldsymbol{\Psi}$. We may then always find $\mathbf{V} \in \mathbb{R}^{m\times m}$ such that $\boldsymbol{\Phi} = \mathbf{V}\mathbf{V}^{\mathrm{T}}$, and

$$\boldsymbol{\Sigma} = \boldsymbol{\Lambda}\boldsymbol{\Phi}\boldsymbol{\Lambda}^{\mathrm{T}}+\boldsymbol{\Psi} = (\boldsymbol{\Lambda}\mathbf{V})[\mathbf{V}^{-1}\boldsymbol{\Phi}(\mathbf{V}^{-1})^{\mathrm{T}}](\boldsymbol{\Lambda}\mathbf{V})^{\mathrm{T}}+\boldsymbol{\Psi} = (\boldsymbol{\Lambda}\mathbf{V}\mathbf{H})(\boldsymbol{\Lambda}\mathbf{V}\mathbf{H})^{\mathrm{T}}+\boldsymbol{\Psi}. \quad (2.9)$$

Now, (2.9) implies that any oblique representation has equivalent orthogonal representations. For ease of modeling and model interpretation we thus take the common factors to be independent.

2.2.3 Approaches to Uniqueness in $\boldsymbol{\Lambda}$

From the above it is clear that any method of estimation requires at a minimum $m(m-1)/2$ restrictions on $\boldsymbol{\Lambda}$ for it to be unique up to polarity reversals in its columns. The EFA tradition usually achieves this by requiring $\boldsymbol{\Lambda}^{\mathrm{T}}\boldsymbol{\Psi}^{-1}\boldsymbol{\Lambda}$ be diagonal accompanied by an order condition on the diagonal elements. Whence estimation is settled EFA traditionally endeavors on applying a rotation that satisfies certain criteria. Such criteria are usually based on the representation of a sparse factor structure. The confirmatory FA (CFA) tradition seeks to incorporate theory into the factor model while unique solutions are sought after by imposing constraints such that rotation is deemed unnecessary. These constraints emulate sparse structure and are of the fixed-value equality kind, with zero and unity being the most common values. An often used minimal alternative consists of assuring that it is possible to permute the rows of $\boldsymbol{\Lambda}$ such that $\boldsymbol{\Lambda} \equiv (\lambda_{jk})$ for $j \geqslant k$, and 0 for $j < k$. While Bayesian identifiability may be achieved in models that are underidentified from a frequentist perspective by (strong) informative prior information, the Bayesian FA (BFA) tradition usually opts for the formal notion of Bayesian identifiability (Dawid, 1979) which entails an equivalence to identifiability in the likelihood. The BFA tradition then adds the additional condition to the minimal CFA solution that at least one element in each column of the block lower triangular $\boldsymbol{\Lambda}$ should be truncated to take only positive or negative values, often, $\lambda_{jk} > 0$, for $j = k$, in order to avoid simulation draws spending time in multiple modes (cf. Geweke & Zhou, 1996; Aguilar & West, 2000; Lopes & West, 2004; Carvalho et al., 2008).

The imposition of structural constraints in $\boldsymbol{\Lambda}$ brings into question the ordering of the observed variables. Theoretically, this ordering is irrelevant when m is known as it is easily verified that $\mathbf{P}\mathbf{x}_i$, the reordering of $\mathbf{x}_i$ by any permutation matrix $\mathbf{P}$, does not alter the implied covariance structure of the observed data except for permutations in the rows and columns, that is: $\mathbf{P}\boldsymbol{\Sigma}\mathbf{P}^{\mathrm{T}}$. However, the block lower triangular form of $\boldsymbol{\Lambda}$ does imply a preferred ordering by giving weight to the first m

variables, thus determining both model interpretation and marginal computation strategies for the choice of m.

Here we set out to develop a default marginal likelihood computation strategy that is invariant to $\mathbf{Px}_i$ by employing restrictions that do not affect $\boldsymbol{\Lambda}$, i.e., from the infinite set of rotational mappings we select one that is representative. In doing so, we develop a Bayesian setting that lets go of likelihood identifiability in the strict sense and forms what can be termed a Bayesian perspective on EFA.

2.3 Marginal Likelihood Estimation

We consider a setting as in Chib (1995). Assume that $\boldsymbol{\Theta}$ can be partitioned into a finite number of parameter blocks $h = 1, \ldots, H$ and let $\boldsymbol{\vartheta}$ denote latent data. Now suppose that we have a set of $H + 1$ complete conditional densities $\{\pi(\boldsymbol{\Theta}_h|\mathbf{X}, \boldsymbol{\Theta}_s(s \neq h), \boldsymbol{\vartheta})\}_{h=1}^{H}$, $\pi(\boldsymbol{\vartheta}|\mathbf{X}, \boldsymbol{\Theta})$, amenable to a Gibbs sampling scheme; the output of which can be used to compute the marginal likelihood. The objective is to formulate a general candidate estimator method for marginal likelihood computation that gives determinate (pseudo) Bayes factors when incorporating improper noninformative prior information, and that is not detrimented by multimodality due to permutative invariance in the parameter indices. The provisions provided to this end are the use of data-augmented posterior priors and a conceptual look at the marginal likelihood for models that deal with permutative invariance of the likelihood over the parameter indices.

2.3.1 Data-Augmented Posterior Priors

Let $g(\boldsymbol{\vartheta}|\boldsymbol{\Theta})$ denote the density of $\boldsymbol{\vartheta}$ given $\boldsymbol{\Theta}$ and assume that the complete data likelihood consists of $L(\boldsymbol{\Theta}, \boldsymbol{\vartheta}; \mathbf{X})g(\boldsymbol{\vartheta}|\boldsymbol{\Theta})$. Now suppose that for the unknown model parameters $\boldsymbol{\Theta}$ only the noninformative improper prior $\pi^N(\boldsymbol{\Theta}) = \zeta(\boldsymbol{\Theta})/\kappa$ is available, where $\kappa = \int \zeta(\boldsymbol{\Theta})\, \partial\boldsymbol{\Theta}$ does not converge (strictly; does not exist). The corresponding marginal likelihood

$$m^N(\mathbf{X}) = \int L(\boldsymbol{\Theta}, \boldsymbol{\vartheta}; \mathbf{X})\pi^N(\boldsymbol{\Theta})g(\boldsymbol{\vartheta}|\boldsymbol{\Theta})\ \partial(\boldsymbol{\Theta}, \boldsymbol{\vartheta}) \tag{2.10}$$

is then defined only up to the arbitrary constant κ. Hence, the Bayes factor in such a situation would be defined only up to a ratio of arbitrary κ's when the respective model dimensions differ. Central to several approaches to this problem is the use of training samples to convert noninformative improper priors into proper 'posterior prior' distributions needed for model selection (cf. Spiegelhalter & Smith, 1982; O'Hagan, 1995; Berger & Pericchi, 1996, 2004; Pérez & Berger, 2002). For our purposes we state the definition of a standard empirical type training sample (Berger & Pericchi, 1996, 2004) in a slightly altered fashion:

Definition 2.1. *A training sample, indexed by ℓ, is a subset of the data, $\mathbf{X}(\ell)$. It is called proper if*

$$0 < \int L(\boldsymbol{\Theta}, \boldsymbol{\vartheta}(\ell); \mathbf{X}(\ell))\zeta(\boldsymbol{\Theta})g(\boldsymbol{\vartheta}(\ell)|\boldsymbol{\Theta})\ \partial(\boldsymbol{\Theta}, \boldsymbol{\vartheta}(\ell)) < \infty \tag{2.11}$$

for all models under consideration. A training sample is minimal if it is proper and no subset is proper.

Let $\mathbf{X}(\ell)^{\mathrm{T}} \equiv [\mathbf{x}(\ell)_1, \ldots, \mathbf{x}(\ell)_q]$ define p-variate observation vectors on $f = 1, \ldots, q$ subjects, where $\mathbf{x}(\ell)_f^{\mathrm{T}} \equiv [\mathrm{x}_{f1}, \ldots, \mathrm{x}_{fp}] \in \mathbb{R}^p$ and with $q < n$. Also, let $\boldsymbol{\vartheta}(\ell)$ denote corresponding latent data. Further assume that

$$L(\boldsymbol{\Theta}; \mathbf{X}(\ell)) = \int L(\boldsymbol{\Theta}, \boldsymbol{\vartheta}(\ell); \mathbf{X}(\ell))g(\boldsymbol{\vartheta}(\ell)|\boldsymbol{\Theta})\ \partial\boldsymbol{\vartheta}(\ell)$$

is either analytically tractable or available through techniques such as Gauss-Hermite integration or importance sampling. We now write the integrated likelihood posterior prior as

$$\begin{aligned}\pi^N(\boldsymbol{\Theta}|\mathbf{X}(\ell)) &= \int \pi^N(\boldsymbol{\Theta}, \boldsymbol{\vartheta}(\ell)|\mathbf{X}(\ell))\ \partial\boldsymbol{\vartheta}(\ell) \\ &= \int \pi^N(\boldsymbol{\Theta}|\boldsymbol{\vartheta}(\ell), \mathbf{X}(\ell))p(\boldsymbol{\vartheta}(\ell)|\mathbf{X}(\ell))\ \partial\boldsymbol{\vartheta}(\ell)\end{aligned}$$

where $p(\boldsymbol{\vartheta}(\ell)|\mathbf{X}(\ell))$ denotes the predictive density of latent data $\boldsymbol{\vartheta}(\ell)$ given $\mathbf{X}(\ell)$ (Tanner & Wong, 1987) and where

$$\pi^N(\boldsymbol{\Theta}, \boldsymbol{\vartheta}(\ell)|\mathbf{X}(\ell)) = \frac{L(\boldsymbol{\Theta}, \boldsymbol{\vartheta}(\ell); \mathbf{X}(\ell))\pi^N(\boldsymbol{\Theta})g(\boldsymbol{\vartheta}(\ell)|\boldsymbol{\Theta})}{\int L(\boldsymbol{\Theta}, \boldsymbol{\vartheta}(\ell); \mathbf{X}(\ell))\pi^N(\boldsymbol{\Theta})g(\boldsymbol{\vartheta}(\ell)|\boldsymbol{\Theta})\ \partial(\boldsymbol{\Theta}, \boldsymbol{\vartheta}(\ell))} \tag{2.12}$$

gives the conversion of an improper $\pi^N(\boldsymbol{\Theta})$ to a proper posterior distribution given the denominator abides condition (2.11).

It may be clear that when posteriors of these sort are used as priors to formulate marginal likelihoods and subsequently Bayes factors for the remaining data, any scaling constants cancel out. Denote the remaining observed and latent data with $\mathbf{X}(-\ell)$ and $\boldsymbol{\vartheta}(-\ell)$, respectively. We now state a first result (following directly from Definition 2.1 and (2.12)) that will help motivate our main proposition regarding an automated candidate estimator in Section 2.3.3.

Lemma 2.1. *Assuming that $L(\boldsymbol{\Theta}; \mathbf{X})$ and $L(\boldsymbol{\Theta}; \mathbf{X}(\ell))$ are either analytically or computationally tractable likelihoods, the marginal density utilizing the data-augmented posterior prior*

$$\int L(\boldsymbol{\Theta}, \boldsymbol{\vartheta}(-\ell), \mathbf{X}(\ell); \mathbf{X}(-\ell))\pi^N(\boldsymbol{\Theta}, \boldsymbol{\vartheta}(\ell)|\mathbf{X}(\ell))g(\boldsymbol{\vartheta}(-\ell)|\boldsymbol{\Theta})\ \partial(\boldsymbol{\Theta}, \boldsymbol{\vartheta}),$$

may be reformulated as:

$$\frac{\int L(\boldsymbol{\Theta}; \mathbf{X})\zeta(\boldsymbol{\Theta})\ \partial\boldsymbol{\Theta}}{\int L(\boldsymbol{\Theta}; \mathbf{X}(\ell))\zeta(\boldsymbol{\Theta})\ \partial\boldsymbol{\Theta}} = m^N(\mathbf{X}(-\ell)|\mathbf{X}(\ell)). \tag{2.13}$$

2.3.2 Marginal Likelihood Defined Over A Symmetric Group

Many latent variable models (such as mixture and latent class models) have likelihoods that are invariant under permutations of the parameter indices. If the prior information is subsequently exchangeable over these permutations then the posterior distribution is similarly symmetric. In this situation we may view the unconstrained marginal likelihood as a union of subspaces that form a symmetric group. Let us formulate the following definition:

Definition 2.2. *Let $L(\boldsymbol{\Theta}, \boldsymbol{\vartheta}; \mathbf{X})$ be invariant under (signed) permutations in the column indices of $(\boldsymbol{\Theta}, \boldsymbol{\vartheta})$, and let there be η permutations. Subsequently, let $\pi^N(\boldsymbol{\Theta})g(\boldsymbol{\vartheta}|\boldsymbol{\Theta})$ be exchangeable over these permutations. The marginal likelihood as a union of η subspaces $\mathcal{S}_d, d = 1, \ldots, \eta$, is then defined over a symmetric group if*

$$\begin{aligned} m^N(\mathbf{X}) &= \sum_{d=1}^{\eta} \int_{\mathcal{S}_d} L(\boldsymbol{\Theta}, \boldsymbol{\vartheta}; \mathbf{X}) \pi^N(\boldsymbol{\Theta}) g(\boldsymbol{\vartheta}|\boldsymbol{\Theta}) \ \partial(\boldsymbol{\Theta}, \boldsymbol{\vartheta}) \\ &= \eta \int_{\mathcal{S}_0} L(\boldsymbol{\Theta}, \boldsymbol{\vartheta}; \mathbf{X}) \pi^N(\boldsymbol{\Theta}) g(\boldsymbol{\vartheta}|\boldsymbol{\Theta}) \ \partial(\boldsymbol{\Theta}, \boldsymbol{\vartheta}), \end{aligned}$$

where $\mathcal{S}_0$ is an arbitrary subspace and $\eta < \infty$.

Neal (1999) was the first to point out that multimodal problems may strain the candidate estimator method. All subspaces of the posterior must be visited by the Gibbs sampler in order to estimate the marginal likelihood. While this will occur in theory, in practice the sampler will likely remain concentrated around a subset of the η modes of the posterior distribution. Several solutions to this problem have been put forward. One could impose direct restrictions on the parameter space. This will however truncate the posterior which may have unforeseen effects on the marginal distribution. Moreover, it will not guarantee unimodality (e.g., Celeux, Hurn, & Robert, 2000). Another approach would be to formulate a prior, say $\pi_\varphi(\boldsymbol{\Theta}, \boldsymbol{\vartheta})$, which enforces a constraint which holds on subspace φ, that is not exchangeable over permutations of the indices in $(\boldsymbol{\Theta}, \boldsymbol{\vartheta})$. The marginal obtained through this route is then (theoretically) equal to the unconstrained marginal if the constrained prior has the property $\pi_\varphi(\boldsymbol{\Theta}, \boldsymbol{\vartheta}) = \eta \cdot \pi(\boldsymbol{\Theta}, \boldsymbol{\vartheta})$, where $\pi(\boldsymbol{\Theta}, \boldsymbol{\vartheta})$ denotes the unconstrained prior (Neal, 1999). Such an approach, however, demands proper prior formulations. Also, its reliability depends on the choice of constraint on the prior (Frühwirth-Schnatter, 2004).

Our preference is to permute draws to a unique subspace (in terms of a single chosen permutation of the parameter indices) as it has a natural connection to the following result that follows directly from Lemma 2.1 and Definition 2.2.

Lemma 2.2. *The marginal likelihood $m^N(\mathbf{X}(-\ell)|\mathbf{X}(\ell))$ over the unconstrained parameter space $\mathcal{O}$, where $\mathcal{O}$ is the union of a finite number η of subspaces $\mathcal{S}_d, d = 1, \ldots, \eta$, that form a symmetric group, is equivalent to*

$$m^N(\mathbf{X}(-\ell)|\mathbf{X}(\ell)) = \frac{\displaystyle\int_{\mathcal{S}_0} L(\mathbf{\Theta};\mathbf{X})\zeta(\mathbf{\Theta})\ \partial\mathbf{\Theta}}{\displaystyle\int_{\mathcal{S}_{0'}} L(\mathbf{\Theta};\mathbf{X}(\ell))\zeta(\mathbf{\Theta})\ \partial\mathbf{\Theta}}, \tag{2.14}$$

where $\mathcal{S}_0$ *and* $\mathcal{S}_{0'}$ *denote arbitrary subspaces in* $\mathcal{O}$.

Lemma 2.2 implies that $m^N(\mathbf{X}(-\ell)|\mathbf{X}(\ell))$ takes advantage of the symmetries as the correction η over the union of subspaces is implicit and automatic given that (numerical) integration over a unique permutation subspace is possible. The remainder of this section assumes that we may permute draws from a Gibbs sampler to such a unique mode. In Section 2.4.2 we will show how this may be done for the FA setting.

2.3.3 Automated Candidate Estimator

We now develop a candidate estimator for $m^N(\mathbf{X}(-\ell)|\mathbf{X}(\ell))$ that is automated in the sense that usage of noninformative improper priors and the occurrence of multimodality due to symmetric invariance over permutations of the parameter indices do not hamper its determinability. To avoid arbitrariness the value of the noninformative prior at a candidate point must not influence the resulting estimate of the marginal. We formulate the following principle.

Principle 2.1 *Let* $\mathbf{\Theta}^*$ *denote a generic estimate of* $\mathbf{\Theta}$, *ideally at a point of high posterior density. Then, the evaluation of* $m^N(\mathbf{X}(-\ell)|\mathbf{X}(\ell))$ *at* $\mathbf{\Theta}^*$, *should not be influenced by the noninformative prior* $\pi^N(\mathbf{\Theta})$ *at* $\mathbf{\Theta}^*$.

As $m^N(\mathbf{X}(-\ell)|\mathbf{X}(\ell)) = m^N(\mathbf{X})/m^N(\mathbf{X}(\ell))$ it is possible to estimate $m^N(\mathbf{X})$ at $\mathbf{\Theta}^*_n$ and $m^N(\mathbf{X}(\ell))$ at $\mathbf{\Theta}^*_\ell$. There are then two cases in which we may abide Principle 2.1: (i) $\zeta(\mathbf{\Theta}) = 1$; (ii) $\mathbf{\Theta}^*_n = \mathbf{\Theta}^*_\ell = \mathbf{\Theta}^*$. These cases guide our main result.

Proposition 2.1. *Assume that*

A1. the unconstrained parameter space $\mathcal{O}$ *on* $\mathbf{\Theta}$ *and* $\boldsymbol{\vartheta}$ *can be expressed as the union of a finite number* η *of subspaces* $\mathcal{S}_d, d = 1, \ldots, \eta$, *that form a symmetric group;*
A2. $\{\pi(\mathbf{\Theta}_h|\mathbf{X}, \mathbf{\Theta}_s(s \neq h), \boldsymbol{\vartheta})\}_{h=1}^H$ *and* $\pi(\boldsymbol{\vartheta}|\mathbf{X}, \mathbf{\Theta})$, *and their training sample-based counterparts are strictly positive in* $\mathcal{S}_0$; *and*
A3. $\int_{\mathcal{S}_0} L(\mathbf{\Theta};\mathbf{X})\zeta(\mathbf{\Theta})\ \partial\mathbf{\Theta}$ *and* $\int_{\mathcal{S}_0} L(\mathbf{\Theta};\mathbf{X}(\ell))\zeta(\mathbf{\Theta})\ \partial\mathbf{\Theta}$ *exist.*

Let $\{\mathbf{\Theta}^*_n\}_{\mathcal{S}_0}$ *and* $\{\mathbf{\Theta}^*_\ell\}_{\mathcal{S}_0}$ *denote candidate* $\mathbf{\Theta}$ *under full data* n *and training sample* ℓ, *respectively, whose representation is in an arbitrary permutation subspace* $\mathcal{S}_0$. *Then, a simulation consistent candidate estimate of* $m^N(\mathbf{X}(-\ell)|\mathbf{X}(\ell))$ *takes the following form:*

Case (i) When abiding Principle 2.1 by $\zeta(\mathbf{\Theta}) = 1$,

$$\log_e \hat{m}^N(\mathbf{X}(-\ell)|\mathbf{X}(\ell))$$
$$= \log_e L(\{\mathbf{\Theta}_n^*\}_{\mathcal{S}_0}; \mathbf{X}) + \sum_{h=1}^{H} \log_e \hat{\pi}^N(\{\mathbf{\Theta}_{\ell h}^*\}_{\mathcal{S}_{0'}} |\mathbf{X}(\ell), \{\mathbf{\Theta}_{\ell b}^*(b < h)\}_{\mathcal{S}_{0'}})$$
$$- \log_e L(\{\mathbf{\Theta}_\ell^*\}_{\mathcal{S}_{0'}}; \mathbf{X}(\ell)) - \sum_{h=1}^{H} \log_e \hat{\pi}(\{\mathbf{\Theta}_{nh}^*\}_{\mathcal{S}_0} |\mathbf{X}, \{\mathbf{\Theta}_{nb}^*(b < h)\}_{\mathcal{S}_0}); \tag{2.15}$$

Case (ii) When abiding Principle 2.1 by $\{\mathbf{\Theta}_n^*\}_{\mathcal{S}_0} = \{\mathbf{\Theta}_\ell^*\}_{\mathcal{S}_0} = \{\mathbf{\Theta}^*\}_{\mathcal{S}_0}$,

$$\log_e \hat{m}^N(\mathbf{X}(-\ell)|\mathbf{X}(\ell))$$
$$= \log_e L(\{\mathbf{\Theta}^*\}_{\mathcal{S}_0}, \mathbf{X}(\ell); \mathbf{X}(-\ell)) + \sum_{h=1}^{H} \log_e \hat{\pi}^N(\{\mathbf{\Theta}_h^*\}_{\mathcal{S}_0} |\mathbf{X}(\ell), \{\mathbf{\Theta}_b^*(b < h)\}_{\mathcal{S}_0})$$
$$- \sum_{h=1}^{H} \log_e \hat{\pi}(\{\mathbf{\Theta}_h^*\}_{\mathcal{S}_0} |\mathbf{X}, \{\mathbf{\Theta}_b^*(b < h)\}_{\mathcal{S}_0}), \tag{2.16}$$

where $\hat{\pi}(\cdot|\cdot)$ *and* $\hat{\pi}^N(\cdot|\cdot)$ *denote Rao-Blackwellized posterior and posterior prior conditionals evaluated on a Gibbs chain over the subspace corresponding to the representation of the candidate estimate.*

Proof. We start with case (i). From (2.13) and using the basic marginal identity implied by Bayes' theorem (Besag, 1989; Chib, 1995), we may write

$$m^N(\mathbf{X}(-\ell)|\mathbf{X}(\ell)) = \frac{L(\mathbf{\Theta}; \mathbf{X})\zeta(\mathbf{\Theta})/\pi(\mathbf{\Theta}|\mathbf{X})}{L(\mathbf{\Theta}; \mathbf{X}(\ell))\zeta(\mathbf{\Theta})/\pi^N(\mathbf{\Theta}|\mathbf{X}(\ell))}. \tag{2.17}$$

From (2.14) and taking into account the principle we may estimate (2.17) as:

$$\hat{m}^N(\mathbf{X}(-\ell)|\mathbf{X}(\ell)) = \frac{L(\{\mathbf{\Theta}_n^*\}_{\mathcal{S}_0}; \mathbf{X})\hat{\pi}^N(\{\mathbf{\Theta}_\ell^*\}_{\mathcal{S}_{0'}} |\mathbf{X}(\ell))}{L(\{\mathbf{\Theta}_\ell^*\}_{\mathcal{S}_{0'}}; \mathbf{X}(\ell))\hat{\pi}(\{\mathbf{\Theta}_n^*\}_{\mathcal{S}_0} |\mathbf{X})}, \tag{2.18}$$

from which $L(\{\mathbf{\Theta}_n^*\}_{\mathcal{S}_0}; \mathbf{X})$ and $L(\{\mathbf{\Theta}_\ell^*\}_{\mathcal{S}_{0'}}; \mathbf{X}(\ell))$ are readily available. We need estimates of the posterior density $\pi(\{\mathbf{\Theta}_n^*\}_{\mathcal{S}_0} |\mathbf{X})$ and posterior prior density $\pi^N(\{\mathbf{\Theta}_\ell^*\}_{\mathcal{S}_{0'}} |\mathbf{X}(\ell))$.

By the law of total probability we have

$$\pi(\{\mathbf{\Theta}_n^*\}_{\mathcal{S}_0} |\mathbf{X}) = \pi(\{\mathbf{\Theta}_{n1}^*\}_{\mathcal{S}_0} |\mathbf{X}) \times \pi(\{\mathbf{\Theta}_{n2}^*\}_{\mathcal{S}_0} |\mathbf{X}, \{\mathbf{\Theta}_{n1}^*\}_{\mathcal{S}_0}) \times \cdots$$
$$\times \pi(\{\mathbf{\Theta}_{nH}^*\}_{\mathcal{S}_0} |\mathbf{X}, \{\mathbf{\Theta}_{n1}^*\}_{\mathcal{S}_0}, \ldots, \{\mathbf{\Theta}_{nH-1}^*\}_{\mathcal{S}_0}),$$

in which we define the typical term

$$\pi(\{\mathbf{\Theta}_{nh}^*\}_{\mathcal{S}_0} |\mathbf{X}, \{\mathbf{\Theta}_{nb}^*(b < h)\}_{\mathcal{S}_0})$$
$$= \int_{\mathcal{S}_0} \pi(\{\mathbf{\Theta}_{nh}^*\}_{\mathcal{S}_0} |\mathbf{X}, \{\mathbf{\Theta}_{nb}^*(b < h)\}_{\mathcal{S}_0}, \mathbf{\Theta}_w(w > h), \boldsymbol{\vartheta})$$
$$\partial\pi(\mathbf{\Theta}_w(w > h), \boldsymbol{\vartheta}|\mathbf{X}, \{\mathbf{\Theta}_{nb}^*(b < h)\}_{\mathcal{S}_0}). \tag{2.19}$$

Chib (1995) extended the Rao-Blackwellization approach (Gelfland & Smith, 1990) to estimate reduced conditional ordinates from the Gibbs sampling output. By sampling the complete conditional densities of $\{\boldsymbol{\Theta}_h, \boldsymbol{\Theta}_{h+1}, \ldots, \boldsymbol{\Theta}_H, \boldsymbol{\vartheta}\}$ in $\mathcal{S}_0$ (either enforced or by - computationally more convenient - *a posteriori* permutation transitions) for which $\{\boldsymbol{\Theta}_1, \boldsymbol{\Theta}_2, \ldots, \boldsymbol{\Theta}_{h-1}\}$ is set at $\{\boldsymbol{\Theta}^*_{nb}(b < h)\}_{\mathcal{S}_0}$ with typical sampling run $\{\boldsymbol{\Theta}^{(c)}_h, \boldsymbol{\Theta}^{(c)}_{h+1}, \ldots, \boldsymbol{\Theta}^{(c)}_H, \boldsymbol{\vartheta}^{(c)}\}_{\mathcal{S}_0}$, an estimate of (2.19) is obtained by

$$\hat{\pi}(\{\boldsymbol{\Theta}^*_{nh}\}_{\mathcal{S}_0} | \mathbf{X}, \{\boldsymbol{\Theta}^*_{nb}(b < h)\}_{\mathcal{S}_0}) \\ = C^{-1} \sum_{c=1}^{C} \pi(\{\boldsymbol{\Theta}^*_{nh}\}_{\mathcal{S}_0} | \mathbf{X}, \{\boldsymbol{\Theta}^*_{nb}(b < h), \boldsymbol{\Theta}^{(c)}_w(w > h), \boldsymbol{\vartheta}^{(c)}\}_{\mathcal{S}_0}).$$

Note that the draws $\{\boldsymbol{\Theta}^{(c)}_w(w > h)\}_{\mathcal{S}_0}$ and $\{\boldsymbol{\vartheta}^{(c)}\}_{\mathcal{S}_0}$ are from $\pi(\{\boldsymbol{\Theta}_w(w > h)\}_{\mathcal{S}_0} | \mathbf{X}, \{\boldsymbol{\Theta}^*_{nb}(b < h)\}_{\mathcal{S}_0})$ and $p(\{\boldsymbol{\vartheta}\}_{\mathcal{S}_0} | \mathbf{X}, \{\boldsymbol{\Theta}^*_{nb}(b < h)\}_{\mathcal{S}_0})$, respectively, as implied by (2.19). The complete conditional densities of $\{\boldsymbol{\Theta}_h, \boldsymbol{\Theta}_{h+1}, \ldots, \boldsymbol{\Theta}_H, \boldsymbol{\vartheta}\}$ inherit strict positivity in $\mathcal{S}_0$ from Condition A2. The conjunction with Condition A3 then gives that the (reduced-run) Gibbs sampler is irreducible in $\mathcal{S}_0$. Theorem 2, Corollary 1 of Tierney (1994) subsequently ensures ergodicity such that $\hat{\pi}(\{\boldsymbol{\Theta}^*_{nh}\}_{\mathcal{S}_0} | \mathbf{X}, \{\boldsymbol{\Theta}^*_{nb}(b < h)\}_{\mathcal{S}_0})$ converges to (2.19) almost surely, as $C \to \infty$.

The decomposition and estimation of the posterior prior $\pi^N(\{\boldsymbol{\Theta}^*_\ell\}_{\mathcal{S}_{0'}} | \mathbf{X}(\ell))$ is completely analogous to the decomposition and estimation of the posterior. Then,

$$\hat{m}^N(\mathbf{X}(-\ell)|\mathbf{X}(\ell)) = \frac{L(\{\boldsymbol{\Theta}^*_n\}_{\mathcal{S}_0}; \mathbf{X}) \prod_{h=1}^H \hat{\pi}^N(\{\boldsymbol{\Theta}^*_{\ell h}\}_{\mathcal{S}_{0'}} | \mathbf{X}(\ell), \{\boldsymbol{\Theta}^*_{\ell b}(b < h)\}_{\mathcal{S}_{0'}})}{L(\{\boldsymbol{\Theta}^*_\ell\}_{\mathcal{S}_{0'}}; \mathbf{X}(\ell)) \prod_{h=1}^H \hat{\pi}(\{\boldsymbol{\Theta}^*_{nh}\}_{\mathcal{S}_0} | \mathbf{X}, \{\boldsymbol{\Theta}^*_{nb}(b < h)\}_{\mathcal{S}_0})}. \tag{2.20}$$

Taking the (computationally convenient) natural logarithm completes the proof for case (i).

For case (ii) we need to realize that by using a generic $\boldsymbol{\Theta}^*$ for the evaluation of both $m^N(\mathbf{X})$ and $m^N(\mathbf{X}(\ell))$ the term $L(\{\boldsymbol{\Theta}^*_\ell\}_{\mathcal{S}_0}; \mathbf{X}(\ell))$ would factor out in (2.18). We may then obtain (2.16) along the same lines as above when replacing $\{\boldsymbol{\Theta}^*_n\}_{\mathcal{S}_0}$ and $\{\boldsymbol{\Theta}^*_\ell\}_{\mathcal{S}_{0'}}$ with $\{\boldsymbol{\Theta}^*\}_{\mathcal{S}_0}$. □

Remark 2.1. The procedure in Proposition 2.1 requires sequences of reduced Gibbs MCMC runs for both full and training data to estimate the posterior ordinates in (2.20). An alternative approach to estimating the posterior ordinates that requires only a single run from the full Gibbs sampler, is to average the Gibbs transition kernel over the draws from the posterior distribution (Ritter & Tanner, 1992). However, Chib and Jeliazkov (2001) suggest this approach is likely to be less accurate than the procedure implied above when $\boldsymbol{\Theta}$ is of high dimension and in the presence of latent data.

Remark 2.2. When the observed data likelihoods $L(\boldsymbol{\Theta}; \mathbf{X})$ and $L(\boldsymbol{\Theta}; \mathbf{X}(\ell))$ are not analytically or computationally available, or when one wants to avoid integrating the complete data likelihood over the latent data, it is possible to represent the candidate estimator in terms of the complete data likelihood. The approach of which would be analogous to the situation exemplified above. This analogy also holds for the situation in which latent data are absent.

Remark 2.3. The proposition is formulated based on the appropriateness of an MCMC scheme that combines latent data and Gibbs sampling. Such schemes are particularly useful for a broad range of Gaussian models. For (more complex) non-Gaussian models the approach from Proposition 2.1 could be naturally amended with auxiliary mixture sampling (Frühwirth-Schnatter & Wagner, 2006; Frühwirth-Schnatter & Frühwirth, 2007) or Metropolis-Hastings steps (Chib & Jeliazkov, 2001).

The estimate of $m^N(\mathbf{X}(-\ell)|\mathbf{X}(\ell))$ depends on the choice of $\mathbf{X}(\ell)$. As such it is natural to average $\hat{m}^N(\mathbf{X}(-\ell)|\mathbf{X}(\ell))$ over the $\mathbf{X}(\ell)$ in some way to increase stability and minimize dependence. Consider taking the geometric average of $\hat{m}^N(\mathbf{X}(-\ell)|\mathbf{X}(\ell))$ over $\ell = 1, \ldots, L$ training samples. Some straightforward algebra will show that the geometric intrinsic Bayes factor (GIBF, Berger & Pericchi, 1996) naturally ensues when $\hat{m}^N(\mathbf{X}(-\ell)|\mathbf{X}(\ell))$ is geometrically averaged over the same training samples for all models under consideration.

Corollary 2.1. *The estimated GIBF for any two models M_a and $M_{a'}$ on $\ell = 1, \ldots, L$ random (minimal) training samples from the set of $\binom{n}{q}$ training samples, where the respective marginals are evaluated using the candidate estimator from Proposition 2.1, is given as:*

$$\hat{B}^{GI}_{aa'} = \left\{ \prod_{\ell=1}^{L} \exp\left[\log_e \hat{m}^N_a(\mathbf{X}(-\ell)|\mathbf{X}(\ell)) - \log_e \hat{m}^N_{a'}(\mathbf{X}(-\ell)|\mathbf{X}(\ell))\right] \right\}^{L^{-1}} . \quad (2.21)$$

One might thus alternatively interpret the averaging of the automated candidate estimator as an MCMC implementation of the IBF. Such an implementation will be exemplified in more detail for the FA setting in Section 2.4.4.

Before turning to the factor analytic implementation we state some attractive properties of the default candidate estimator and the resulting GIBF. The Bayes factor in (2.21) is coherent in the sense that $B^{GI}_{a'a} = 1/B^{GI}_{aa'}$, and $B^{GI}_{aa'} = B^{GI}_{aa''}B^{GI}_{a''a'}$. Moreover, it is easy to verify that $m^N(\mathbf{X}(-\ell)|\mathbf{X}(\ell))$ and the ensuing GIBF are invariant to one-to-one transformations of the data (given that the nature of the training samples employed does not change with the transformation) or of the parameters within each model. Section 2.6 contains some deeper considerations on the automated candidate estimator - IBF connection.

2.4 Factor Analytic Implementation

The default candidate estimator can be expected to perform well in situations that are unimodal and well behaved. Factor analytic dimensionality selection forms a more stringent test of the method due to the many indeterminacies that are inherent to the problem. From Proposition 2.1 we see that we need the following to make the result applicable to this situation: posterior and posterior prior conditionals, a means of permuting draws from a Gibbs sampler on the unconstrained parameter space to a single mode (a permutation transition), a choice of candidate estimates for the parameters, and a choice of training samples. We start with reviewing the conditionals. Then we show how to choose a convenient mapping for $\mathbf{\Lambda}$ in order to deal with its rotational freedom. Such a mapping may be expressed as a symmetric group and it is subsequently shown how sample draws from any of its permutations may be permuted to a single (prespecified) subspace therein. We also make a choice regarding training sample size. These elements culminate in an algorithm that computes (2.21) for our FA problem. We then show that (2.7) can lead to a violation of Condition A2 of Proposition 2.1 such that our estimate of (2.21) is not simulation consistent. A problem that seems endemic to parameter space MCMC-based approaches towards dimensionality selection in FA. Two assessment strategies are developed whose coupling with our marginal computation strategy may curb simulation inconsistency and provide for appropriate stopping rules. In the ensuing numerical exploration we will see to an efficient choice of $\{\boldsymbol{\mu}^*, \mathbf{\Lambda}^*, \mathbf{\Psi}^*\}_{\mathcal{S}_0} = (\boldsymbol{\mu}^*, \{\mathbf{\Lambda}^*\}_{\mathcal{S}_0}, \mathbf{\Psi}^*)$.

2.4.1 Conditional Distributions

Let μ_j be the jth scalar element of $\boldsymbol{\mu}$. We assume independent priors for μ_j, λ_{jk}, and ψ_{jj}. As exemplified by Martin and McDonald (1975) and Lopes and West (Lopes & West, 2004) a prior that decays to zero at the origin is needed on the ψ_{jj} elements in order to induce proper posteriors as this prevents the posterior from placing infinite mass at $\psi_{jj} = 0$ for some j. A convenient choice is $\psi_{jj} \sim \mathcal{IG}(\nu/2, \nu d/2)$, where $\mathcal{IG}(\cdot,\cdot)$ denotes the inverse gamma distribution. Shape and scale will be chosen diffuse but proper in applications (Sections 2.4.6 and 2.5). The priors for $\boldsymbol{\mu}$ and $\mathbf{\Lambda}$ are set up such that $\forall j\ \zeta(\mu_j) = 1$ over the Lebesgue measure on $(-\infty, +\infty)$, and $\forall j, k\ \zeta(\lambda_{jk}) = 1$ over the Lebesgue measure on $(-\infty, +\infty)$. Then, taking into account assumption (iv) from Section 2.2.1, we have

$$\begin{aligned}\pi^N(\mathbf{\Theta})g(\boldsymbol{\vartheta}|\mathbf{\Theta}) &= \pi^N(\boldsymbol{\mu})\pi^N(\mathbf{\Lambda})\pi(\mathbf{\Psi})\pi(\mathbf{\Xi}) \propto \zeta(\boldsymbol{\mu}, \mathbf{\Lambda}, \mathbf{\Psi})\zeta(\mathbf{\Xi}) \\ &= \prod_{j=1}^{p} \psi_{jj}^{-(\nu/2+1)} \exp\left\{-\frac{1}{2}\frac{\nu d}{\psi_{jj}}\right\} \prod_{i=1}^{n} \exp\left\{-\frac{1}{2}\boldsymbol{\xi}_i^{\mathrm{T}}\boldsymbol{\xi}_i\right\}.\end{aligned}$$

Now, define $\mathbf{x}_j$, $\mathbf{\Lambda}_j$, and $\mathbf{1}_n$ to be the jth column of $\mathbf{X}$, the jth row of $\mathbf{\Lambda}$, and an n-dimensional unit vector, respectively. The conditional posterior distributions are then given as below.

For the conditional distribution of $\boldsymbol{\mu}$, we find

$$\pi(\boldsymbol{\mu}|\mathbf{X}, \mathbf{\Lambda}, \mathbf{\Xi}, \mathbf{\Psi}) \overset{d}{=} \mathcal{N}_p(\tilde{\boldsymbol{\mu}}, \mathbf{\Sigma}_{\boldsymbol{\mu}}), \tag{2.22}$$

where $\tilde{\boldsymbol{\mu}} = \bar{\mathbf{x}} - \mathbf{\Lambda}\bar{\boldsymbol{\xi}}$, $\mathbf{\Sigma}_{\boldsymbol{\mu}} = \mathbf{\Psi}n^{-1}$, $\bar{\mathbf{x}} = n^{-1}\sum_{i=1}^{n}\mathbf{x}_i$, and $\bar{\boldsymbol{\xi}} = n^{-1}\sum_{i=1}^{n}\boldsymbol{\xi}_i$. The rows of $\mathbf{\Lambda}$ are independent such that we need to find the conditional distribution

$\pi(\boldsymbol{\Lambda}_j^{\mathrm{T}}|\mathbf{x}_j, \mu_j, \boldsymbol{\Xi}, \psi_{jj})$, $j = 1, \ldots, p$, which is given as

$$\pi(\boldsymbol{\Lambda}_j^{\mathrm{T}}|\mathbf{x}_j, \mu_j, \boldsymbol{\Xi}, \psi_{jj}) \stackrel{d}{=} \mathcal{N}_m(\tilde{\boldsymbol{\Lambda}}_j, \boldsymbol{\Sigma}_{\boldsymbol{\Lambda}_j}), \tag{2.23}$$

where

$$\begin{aligned} \tilde{\boldsymbol{\Lambda}}_j &= (\boldsymbol{\Xi}^{\mathrm{T}}\boldsymbol{\Xi})^{-1}\boldsymbol{\Xi}^{\mathrm{T}}(\mathbf{x}_j - \mathbf{1}_n\mu_j), \\ \boldsymbol{\Sigma}_{\boldsymbol{\Lambda}_j} &= \psi_{jj}(\boldsymbol{\Xi}^{\mathrm{T}}\boldsymbol{\Xi})^{-1}. \end{aligned}$$

Also, the rows of $\boldsymbol{\Xi}$ are independent such that we need to find the conditional distribution $\pi(\boldsymbol{\xi}_i|\mathbf{x}_i, \boldsymbol{\mu}, \boldsymbol{\Lambda}, \boldsymbol{\Psi})$, $i = 1, \ldots, n$, which, with the help of the Woodbury matrix identity (Woodbury, 1950), can be found as

$$\pi(\boldsymbol{\xi}_i|\mathbf{x}_i, \boldsymbol{\mu}, \boldsymbol{\Lambda}, \boldsymbol{\Psi}) \stackrel{d}{=} \mathcal{N}_m(\tilde{\boldsymbol{\xi}}, \boldsymbol{\Sigma}_\xi), \tag{2.24}$$

where

$$\begin{aligned} \tilde{\boldsymbol{\xi}} &= \boldsymbol{\Sigma}_\xi \boldsymbol{\Lambda}^{\mathrm{T}}\boldsymbol{\Psi}^{-1}(\mathbf{x}_i - \boldsymbol{\mu}), \\ \boldsymbol{\Sigma}_\xi &= (\mathbf{I}_m + \boldsymbol{\Lambda}^{\mathrm{T}}\boldsymbol{\Psi}^{-1}\boldsymbol{\Lambda})^{-1}. \end{aligned}$$

Lastly, the conditional distribution for the diagonal elements of $\boldsymbol{\Psi}$, $\pi(\psi_{jj}|\mathbf{x}_j, \mu_j, \boldsymbol{\Lambda}_j, \boldsymbol{\Xi})$, $j = 1, \ldots, p$, can be found as

$$\pi(\psi_{jj}|\mathbf{x}_j, \mu_j, \boldsymbol{\Lambda}_j, \boldsymbol{\Xi}) \stackrel{d}{=} \mathcal{IG}((n+\nu)/2, (\beta_j + \nu d)/2), \tag{2.25}$$

where

$$\beta_j = (\mathbf{x}_j - \mathbf{1}_n\mu_j - \boldsymbol{\Xi}\boldsymbol{\Lambda}_j^{\mathrm{T}})^{\mathrm{T}}(\mathbf{x}_j - \mathbf{1}_n\mu_j - \boldsymbol{\Xi}\boldsymbol{\Lambda}_j^{\mathrm{T}}).$$

With regard to the posterior prior conditional distributions we use

$$\begin{aligned} \pi^N(\boldsymbol{\Theta})g(\boldsymbol{\vartheta}(\ell)|\boldsymbol{\Theta}) &= \pi^N(\boldsymbol{\mu})\pi^N(\boldsymbol{\Lambda})\pi(\boldsymbol{\Psi})\pi(\boldsymbol{\Xi}(\ell)) \propto \zeta(\boldsymbol{\mu}, \boldsymbol{\Lambda}, \boldsymbol{\Psi})\zeta(\boldsymbol{\Xi}(\ell)) \\ &= \prod_{j=1}^{p} \psi_{jj}^{-(\nu/2+1)} \exp\left\{-\frac{1}{2}\frac{\nu d}{\psi_{jj}}\right\} \prod_{f=1}^{q} \exp\left\{-\frac{1}{2}\boldsymbol{\xi}_f^{\mathrm{T}}\boldsymbol{\xi}_f\right\}. \end{aligned}$$

The posterior prior conditionals $\pi^N(\cdot|\cdot)$ are then found by replacing in the above conditionals $\mathbf{X}$, $\boldsymbol{\Xi}$, the index i, n, and $\mathbf{1}_n$ with $\mathbf{X}(\ell)$, $\boldsymbol{\Xi}(\ell)$, the index f, q, and $\mathbf{1}_q$ respectively.

It may be clear that, due to (2.8), the marginal conditional density for $\boldsymbol{\Lambda}$ (and subsequently $\boldsymbol{\Xi}$) will not exist without some constraint on its elements. It is to this topic we turn next.

2.4.2 Label-Switching and Polarity Reversals

Our goal here is to choose, out of the infinite number of rotational mappings, a convenient representation of $\boldsymbol{\Lambda}$ as well as devising a means of permuting draws from a Gibbs sampler over the unconstrained parameter space (that will be defined over this representation) to a single (prespecified) mode therein. To aid interpretation a convenient mapping would be to employ a rotation according to some condition for a simple (Thurstone, 1947) or sparse structure. A well known orthogonal family of analytic simple structure rotation can be defined as

$$\boldsymbol{\Gamma} = \arg\max_{\mathbf{H}} \left\{ \sum_k \sum_j \frac{(\boldsymbol{\Lambda}\mathbf{H})^4_{jk}}{y^4_j} - \frac{\gamma}{p} \sum_k \left[\sum_j \frac{(\boldsymbol{\Lambda}\mathbf{H})^2_{jk}}{y^2_j} \right]^2 : \mathbf{H}^{\mathrm{T}}\mathbf{H} = \mathbf{I}_m \right\}, \quad (2.26)$$

where y^2_j is the proportion of variance of the j'th variable explained by the corresponding m factor loadings (the 'communality' in FA terms), and $\gamma \geqslant 0$. Setting $\gamma = 1$, for example, gives the often used Varimax (Kaiser, 1958) rotation to an orthogonal simple structure.

It is easy to see that the reordering of $\mathbf{x}_i$ by any permutation matrix, such that we have $\mathbf{P}\boldsymbol{\Lambda}\mathbf{H}$ in (2.26), does not alter $\boldsymbol{\Gamma}$. It may also easily be seen that the maximization of (2.26) is invariant to label-switchings and polarity reversals in the columns of $\boldsymbol{\Lambda}$, meaning that the unconstrained parameter space based on this rotational criterium is composed of a number of subspaces that form a symmetric group. The algebraic group here being the symmetry group over the hypercube or the Coxeter group of order $2^m(m!)$.

Let $\{\boldsymbol{\Lambda}^{(c)}\}_{\mathcal{S}_d} = \boldsymbol{\Lambda}^{(c)}\boldsymbol{\Gamma}$ denote an MCMC draw of $\boldsymbol{\Lambda}$ that has been subjected to a simple structure rotation defined by (2.26). This rotated draw may be from any of $d = 1, \ldots, 2^m(m!)$ representations, that differ only in the ordering and polarity of the columns of the loadings matrix. Further, suppose we have $\{\boldsymbol{\Lambda}^*\}_{\mathcal{S}_0} = \boldsymbol{\Lambda}^*\boldsymbol{\Gamma}$, a candidate estimate in an arbitrary subspace of our Coxeter group. Our problem is now to find a signed permutation matrix $^{\pm}\mathbf{P}$ such that $\{\boldsymbol{\Lambda}^{(c)}\}_{\mathcal{S}_0} = \{\boldsymbol{\Lambda}^{(c)}\}_{\mathcal{S}_d}(^{\pm}\mathbf{P})$. Below an algorithm is given for doing so.

Algorithm 2.1 (SPMA: Signed Permutation Matrix Abstraction) The input consists of $\{\boldsymbol{\Lambda}^{(c)}\}_{\mathcal{S}_d}$ and $\{\boldsymbol{\Lambda}^*\}_{\mathcal{S}_0}$. The output is a signed permutation matrix $^{\pm}\mathbf{P}$. The algorithm uses the NEXPER routine by Nijenhuis and Wilf (1978, Chapter 7) with arguments m and $\boldsymbol{\rho}$. NEXPER produces all permutations of m letters where each permutation is obtained from its predecessor by a single transposition. The first is the identity permutation. All permutations are stored in the m-valued integer vector $\boldsymbol{\rho}$. The algorithm proceeds by collapsing all signed permutation possibilities over the polarity reversals. Subsequently, NEXPER is utilized to evaluate the remaining $m!$ possibilities with respect to a distance measure.

$^{\pm}\mathbf{P}(:,:) = \mathbf{0}$
$\{\boldsymbol{\Lambda}^{(c)}\}^{-}_{\mathcal{S}_d} = (-1)\{\boldsymbol{\Lambda}^{(c)}\}_{\mathcal{S}_d}$
for $k_1 = 1$ to m **do**
  for $k_2 = 1$ to m **do**
    if $\| \{\boldsymbol{\Lambda}^*(:,k_1)\}_{\mathcal{S}_0} - \{\boldsymbol{\Lambda}^{(c)}(:,k_2)\}_{\mathcal{S}_d} \|_2 \leqslant \| \{\boldsymbol{\Lambda}^*(:,k_1)\}_{\mathcal{S}_0} - \{\boldsymbol{\Lambda}^{(c)}(:,k_2)\}^{-}_{\mathcal{S}_d} \|_2$ **then**
      $\mathbf{C}(k_2,k_1) = \| \{\boldsymbol{\Lambda}^*(:,k_1)\}_{\mathcal{S}_0} - \{\boldsymbol{\Lambda}^{(c)}(:,k_2)\}_{\mathcal{S}_d} \|_2$

```
    else
      C(k_2, k_1) = (-1)‖ {Λ*(:, k_1)}_{S_0} − {Λ^{(c)}(:, k_2)}^−_{S_d} ‖_2
    end if
  end for
end for
for b = 1 to m! do
  call NEXPER(m, ρ)
  L^1 = 0
  for k = 1 to m do
    L^1 = L^1 + |C(ρ(k), k)|
  end for
  if b = 1 then
    ω = L^1
    p = ρ
  else
    if L^1 < ω then
      ω = L^1
      p = ρ
    else
      ω = ω
      p = p
    end if
  end if
end for
for k = 1 to m do
  if C(p(k), k) < 0 then
    ^±P(p(k), k) = −1
  else
    ^±P(p(k), k) = 1
  end if
end for
```

Algorithm 2.1 does not rely on any tolerance-dependent function and works (theoretically) for all feasible values of m. The algorithm may however loose practicality when m is large. Appendix 2A contains an alternative tolerance-dependent approach to finding ${}^{\pm}\mathbf{P}$ that is much less intensive but assumes access to terminals with high machine precision.

2.4.3 Choice of $\mathbf{X}(\boldsymbol{\ell})$

The number of parameters in factor analytic models varies with the choice of latent factor dimensionality. The choice for a minimal training sample may then leave a model empirically underidentified when m gets large, implying more parameters to be estimated than observations to estimate them with. This may exacerbate violation of Condition A2 of Proposition 2.1 (next sections) for the conditional distributions over the training data. We formulate a second principle.

Principle 2.2 *The training sample based covariance matrix* $\boldsymbol{\Sigma}$ *must be empirically identified for all models under consideration.*

A necessary condition for abiding Principle 2.2 is to set the training sample size at $p(\tilde{m}+1) - \tilde{m}(\tilde{m}-1)/2 + 1$ (the number of parameters in $\boldsymbol{\Lambda}$ and $\boldsymbol{\Psi}$, minus the number of independent elements of $\mathbf{H}$, plus 1), where $\tilde{\mathrm{m}}$ denotes the maximal factor dimensionality m for which $m \leqslant [2p + 1 - (8p+1)^{1/2}]/2$ holds (condition (vi), Section 2.2.1). Thus, the training sample for the set $\mathcal{A}$ is $\mathbf{X}(\ell) \in \mathbb{R}^{q \times p}$ with $q = p(\tilde{m}+1) - \tilde{m}(\tilde{m}-1)/2 + 1$.

2.4.4 Numerical Implementation

So we have a set $\mathcal{A}$ of models M_a that differ with respect to latent factor dimensionality m, $a = 1, \ldots, A = \tilde{m}$. We now have all the elements towards computing (2.21) for use as a stopping rule in selecting the 'optimal' dimensionality m. The recipe is given in the algorithm below. Note that with our specification of $\zeta(\boldsymbol{\mu}, \boldsymbol{\Lambda}, \boldsymbol{\Psi})$ we find ourself in case (ii) of Proposition 2.1. We will thus use a computationally more convenient generic candidate for $(\boldsymbol{\mu}, \{\boldsymbol{\Lambda}\}_{\mathcal{S}_0}, \boldsymbol{\Psi})$. Also note that the setup of (2.21) allows one to calibrate each $m_a^N(\mathbf{X}(-\ell)|\mathbf{X}(\ell))$ individually, thus reducing computational bookkeeping. Both observations are incorporated in Algorithm 2.2.

Algorithm 2.2 (SRG: Sequential Reduced Run Gibbs Sampler) The input consists of the data $\mathbf{X}$, starting values $\boldsymbol{\mu}^{(0)}$, $\{\boldsymbol{\Xi}^{(0)}\}_{\mathcal{S}_0}$, and $\boldsymbol{\Psi}^{(0)}$ for each model, as well as candidate estimates $\boldsymbol{\mu}^*$, $\{\boldsymbol{\Lambda}^*\}_{\mathcal{S}_0}$, and $\boldsymbol{\Psi}^*$ for each model. The output is a calibrated estimate of $m_a^N(\mathbf{X}(-\ell)|\mathbf{X}(\ell))$ for each model under consideration. The algorithm uses sequences of reduced Gibbs samplers to estimate the reduced conditional ordinates as implied in Proposition 2.1. Each of $\boldsymbol{\mu}, \boldsymbol{\Lambda}, \boldsymbol{\Xi}$, and $\boldsymbol{\Psi}$ is treated as a single block in the Gibbs samplers. For convenience, the first posterior conditional ordinate to be estimated is the one for $\{\boldsymbol{\Lambda}^*\}_{\mathcal{S}_0}$, as its fixation in a reduced chain eliminates the need for further permutation transitions. The algorithm assumes that for each model a sequence of given seeds is invoked at appropriate points to ensure that each model under consideration receives the same (sequence of) training samples (such that the decomposition (2.21) remains valid). We implicitly assume these seeds in the following, as well as suppress the model index a where appropriate, for sake of notational simplicity.

```
for a = 1 to A do
    Set μ*, {Λ*}_{S_0}, and Ψ*
    Set μ^(0), {Ξ^(0)}_{S_0}, and Ψ^(0)
    for c = 1 to C do
        Generate Λ^(c) from ∏_{j=1}^p π(Λ_j^T | x_j, μ_j^(c−1), {Ξ^(c−1)}_{S_0}, ψ_jj^(c−1))
        Obtain {Λ^(c)}_{S_d} using (2.26)
        call SPMA({Λ^(c)}_{S_d}, {Λ*}_{S_0}, ±P)
        {Λ^(c)}_{S_0} = {Λ^(c)}_{S_d}(±P)
        Generate μ^(c) from π(μ | X, {Λ^(c)}_{S_0}, {Ξ^(c−1)}_{S_0}, Ψ^(c−1))
        Generate {Ξ^(c)}_{S_0} from ∏_{i=1}^n π(ξ_i | x_i, μ^(c), {Λ^(c)}_{S_0}, Ψ^(c−1))
        Generate Ψ^(c) from ∏_{j=1}^p π(ψ_jj | x_j, μ_j^(c), {Λ^(c)}_{S_0}, {Ξ^(c)}_{S_0})
```

12: **end for**
13: $\log_e \hat{\pi}(\{\mathbf{\Lambda}^*\}_{\mathcal{S}_0}|\mathbf{X}) = \log_e \left\{ C^{-1} \sum_{c=1}^{C} \prod_{j=1}^{p} \pi(\{\mathbf{\Lambda}_j^{*\mathrm{T}}\}_{\mathcal{S}_0}|\mathbf{x}_j, \mu_j^{(c)}, \{\mathbf{\Xi}^{(c)}\}_{\mathcal{S}_0}, \psi_{jj}^{(c)}) \right\}$
14: **for** $c = 1$ to C **do**
15: Generate $\boldsymbol{\mu}^{(c)}$ from $\pi(\boldsymbol{\mu}|\mathbf{X}, \{\mathbf{\Lambda}^*\}_{\mathcal{S}_0}, \{\mathbf{\Xi}^{(c-1)}\}_{\mathcal{S}_0}, \mathbf{\Psi}^{(c-1)})$
16: Generate $\{\mathbf{\Xi}^{(c)}\}_{\mathcal{S}_0}$ from $\prod_{i=1}^{n} \pi(\boldsymbol{\xi}_i|\mathbf{x}_i, \boldsymbol{\mu}^{(c)}, \{\mathbf{\Lambda}^*\}_{\mathcal{S}_0}, \mathbf{\Psi}^{(c-1)})$
17: Generate $\mathbf{\Psi}^{(c)}$ from $\prod_{j=1}^{p} \pi(\psi_{jj}|\mathbf{x}_j, \mu_j^{(c)}, \{\mathbf{\Lambda}^*\}_{\mathcal{S}_0}, \{\mathbf{\Xi}^{(c)}\}_{\mathcal{S}_0})$
18: **end for**
19: $\log_e \hat{\pi}(\boldsymbol{\mu}^*|\mathbf{X}, \{\mathbf{\Lambda}^*\}_{\mathcal{S}_0}) = \log_e \left\{ C^{-1} \sum_{c=1}^{C} \pi(\boldsymbol{\mu}^*|\mathbf{X}, \{\mathbf{\Lambda}^*\}_{\mathcal{S}_0}, \{\mathbf{\Xi}^{(c)}\}_{\mathcal{S}_0}, \mathbf{\Psi}^{(c)}) \right\}$
20: **for** $c = 1$ to C **do**
21: Generate $\{\mathbf{\Xi}^{(c)}\}_{\mathcal{S}_0}$ from $\prod_{i=1}^{n} \pi(\boldsymbol{\xi}_i|\mathbf{x}_i, \boldsymbol{\mu}^*, \{\mathbf{\Lambda}^*\}_{\mathcal{S}_0}, \mathbf{\Psi}^{(c-1)})$
22: Generate $\mathbf{\Psi}^{(c)}$ from $\prod_{j=1}^{p} \pi(\psi_{jj}^*|\mathbf{x}_j, \mu_j^*, \{\mathbf{\Lambda}^*\}_{\mathcal{S}_0}, \{\mathbf{\Xi}^{(c)}\}_{\mathcal{S}_0})$
23: **end for**
24: $\log_e \hat{\pi}(\mathbf{\Psi}^*|\mathbf{X}, \boldsymbol{\mu}^*, \{\mathbf{\Lambda}^*\}_{\mathcal{S}_0}) = \log_e \left\{ C^{-1} \sum_{c=1}^{C} \prod_{j=1}^{p} \pi(\psi_{jj}^*|\mathbf{x}_j, \mu_j^*, \{\mathbf{\Lambda}^*\}_{\mathcal{S}_0}, \{\mathbf{\Xi}^{(c)}\}_{\mathcal{S}_0}) \right\}$
25: $\log_e \hat{\pi}(\boldsymbol{\mu}^*, \{\mathbf{\Lambda}^*\}_{\mathcal{S}_0}, \mathbf{\Psi}^*|\mathbf{X}) = \log_e \hat{\pi}(\{\mathbf{\Lambda}^*\}_{\mathcal{S}_0}|\mathbf{X}) + \log_e \hat{\pi}(\boldsymbol{\mu}^*|\mathbf{X}, \{\mathbf{\Lambda}^*\}_{\mathcal{S}_0}) + \log_e \hat{\pi}(\mathbf{\Psi}^*|\mathbf{X}, \boldsymbol{\mu}^*, \{\mathbf{\Lambda}^*\}_{\mathcal{S}_0})$
26: **for** $\ell = 1$ to L **do**
27: Draw a random training sample $\mathbf{Z}(\ell)$ abiding Principle 2.2
28: Obtain $\{\mathbf{\Xi}(\ell)^{(0)}\}_{\mathcal{S}_0}$
29: Obtain $\log_e L(\boldsymbol{\mu}^*, \{\mathbf{\Lambda}^*\}_{\mathcal{S}_0}, \mathbf{\Psi}^*, \mathbf{X}(\ell); \mathbf{X}(-\ell))$
30: **for** $c = 1$ to C **do**
31: repeat lines 5 to 25 with $\mathbf{X}$, $\mathbf{\Xi}$, the index i, n, and $\mathbf{1}_n$ replaced with $\mathbf{X}(\ell)$, $\mathbf{\Xi}(\ell)$, the index f, q, and $\mathbf{1}_q$ respectively, to produce $\log_e \hat{\pi}(\boldsymbol{\mu}^*, \{\mathbf{\Lambda}^*\}_{\mathcal{S}_0}, \mathbf{\Psi}^*|\mathbf{X}(\ell))$
32: **end for**
33: **end for**
34: $1/L \sum_{\ell=1}^{L} \log_e \hat{m}^N(\mathbf{X}(-\ell)|\mathbf{X}(\ell)) = 1/L \sum_{\ell=1}^{L} \log_e L(\boldsymbol{\mu}^*, \{\mathbf{\Lambda}^*\}_{\mathcal{S}_0}, \mathbf{\Psi}^*, \mathbf{X}(\ell); \mathbf{X}(-\ell)) + 1/L \sum_{\ell=1}^{L} \log_e \hat{\pi}(\boldsymbol{\mu}^*, \{\mathbf{\Lambda}^*\}_{\mathcal{S}_0}, \mathbf{\Psi}^*|\mathbf{X}(\ell)) - \log_e \hat{\pi}(\boldsymbol{\mu}^*, \{\mathbf{\Lambda}^*\}_{\mathcal{S}_0}, \mathbf{\Psi}^*|\mathbf{X})$
35: **end for**

Our calibrated estimate of $m_a^N(\mathbf{X}(-\ell)|\mathbf{X}(\ell))$ will be simulation consistent by Proposition 2.1 if we abide Conditions A2 and A3 (Condition A1 is trivially met). Condition A3 will depend on rank conditions on $\mathbf{X}^{\mathrm{T}}\mathbf{X}$, $\mathbf{X}(\ell)^{\mathrm{T}}\mathbf{X}(\ell)$ and $\mathbf{\Psi}$. Given sufficient size for the full data and the training sample, $\mathbf{X}^{\mathrm{T}}\mathbf{X}$ and $\mathbf{X}(\ell)^{\mathrm{T}}\mathbf{X}(\ell)$ will not pose problems in terms of being indefinite (it is natural to consider only positive-definite sample covariance matrices for factor analytic modeling). Positive-definiteness of $\mathbf{\Psi}$ is enforced by the prior distribution on its elements. Condition A2, in contrast, may cause problems by the nature of model and problem entertained. As m increases we are at risk of exceeding the 'true' or optimal latent factor dimensionality. It may subsequently be possible that $\mathbf{\Lambda}$ is not of full-column rank. There may then occur a situation of near-reducibility that has to be accounted for in MCMC practice as will be discussed in the following section.

2.4.5 Near-Reducibility Coping Strategies

Equation (2.7) gives that, within the mapping $\mathcal{S}_0$, corresponding multimodalities may ensue in the densities of $\mathbf{\Psi}$ and $\mathbf{\Lambda}$ when the latter is rank-deficient. The

(Gibbs) sampler can then be near-reducible if the parameter subspace $\mathcal{S}_0$ has multiple separated regions of non-negligible posterior probability, such that transition-probabilities are low. This situation clearly implies conditional densities not being strictly positive in $\mathcal{S}_0$ such that Condition A2 of Proposition 2.1 is being violated.

A near-reducible Gibbs chain may spend many consecutive iterations in (a) subregion(s) of the parameter subspace $\mathcal{S}_0$, thus affecting the viability of an MCMC-based estimate of the marginal likelihood. Consider as an example the following instance of modeling the correlation matrix taken from a simulation study by Gosh and Dunson (2009): $p = 10$, $m = 3$, $\boldsymbol{\mu} = \mathbf{0}$, and

$$\boldsymbol{\Lambda}^{\mathrm{T}} = \begin{bmatrix} .89 & .00 & .25 & .00 & .80 & .00 & .50 & .00 & .00 & .00 \\ .00 & .90 & .25 & .40 & .00 & .50 & .00 & .00 & -.30 & -.30 \\ .00 & .00 & .85 & .80 & .00 & .75 & .75 & .00 & .80 & .80 \end{bmatrix},$$

$$\boldsymbol{\Psi} = \mathrm{diag}\left[.2079 \; .19 \; .1525 \; .20 \; .36 \; .1875 \; .1875 \; 1.0 \; .27 \; .27\right].$$

We generated a dataset of size $n = 100$ with these specifications according to model (2.4). For the full data and a random training sample abiding Principle 2.2 we did the following: we ran three different Gibbs samplers with differing batches of starting values (see Section 2.4.6 for specifics). The first chain used starting values that are well within the feasible parameter space. The second chain used starting values well outside the feasible parameter space such that we heightened the probability that, as the feasible space was approached from the right, the sampler would find itself stuck in alternate separated regions in $\mathcal{S}_0$ if those would exist. The last chain used starting values for ψ_{jj} close to zero (and corresponding starting values for the other parameters), to the same purpose. Figure 2.1 gives, for each of the three-, four-, and five-factor solutions, a representation of the bivariate densities on two λ_{jk}'s for full data and the random training sample, composed of the superimposed iterations (after burn-in) stemming from the three chains mentioned above.

Figures 2.1a and 2.1b give the mentioned densities for the $m = 3$ solution. The densities are seen to be unimodal. Moreover, the starting values for the sampler do not alter the space explored in $\mathcal{S}_0$. Figures 2.1c and 2.1d give analogous densities for he $m = 4$ solution which entails slight overfactorization. The densities of the full data and the training sample are multimodal, but the strict positivity condition is not violated. Again, the choice of starting values does not alter the space explored in $\mathcal{S}_0$, such that the density stemming from each separate sampler is similar to the density for the superimposed iterations. The tale is different for the $m = 5$ solution which endeavors on more serious overfactoring. Figures 2.1e and 2.1f show that there exist within $\mathcal{S}_0$ multiple separated regions of non-negligible posterior probability for which transition probabilities are low. Samplers started at different starting values tend to remain concentrated around different (subsets of) modes. As a result, the samplers tend to 'cramp up' mass under certain modes. This exemplifies that for near-reducible situations (i) no single $\{\boldsymbol{\Theta}^*\}_{\mathcal{S}_0}$ can provide a good candidate estimate for the evaluation of its posterior probability for marginal likelihood estimation, and (ii) numerically perverse estimates of the marginal likelihood may ensue as both severe under- and overestimation of the posterior and posterior prior ordinates is possible.

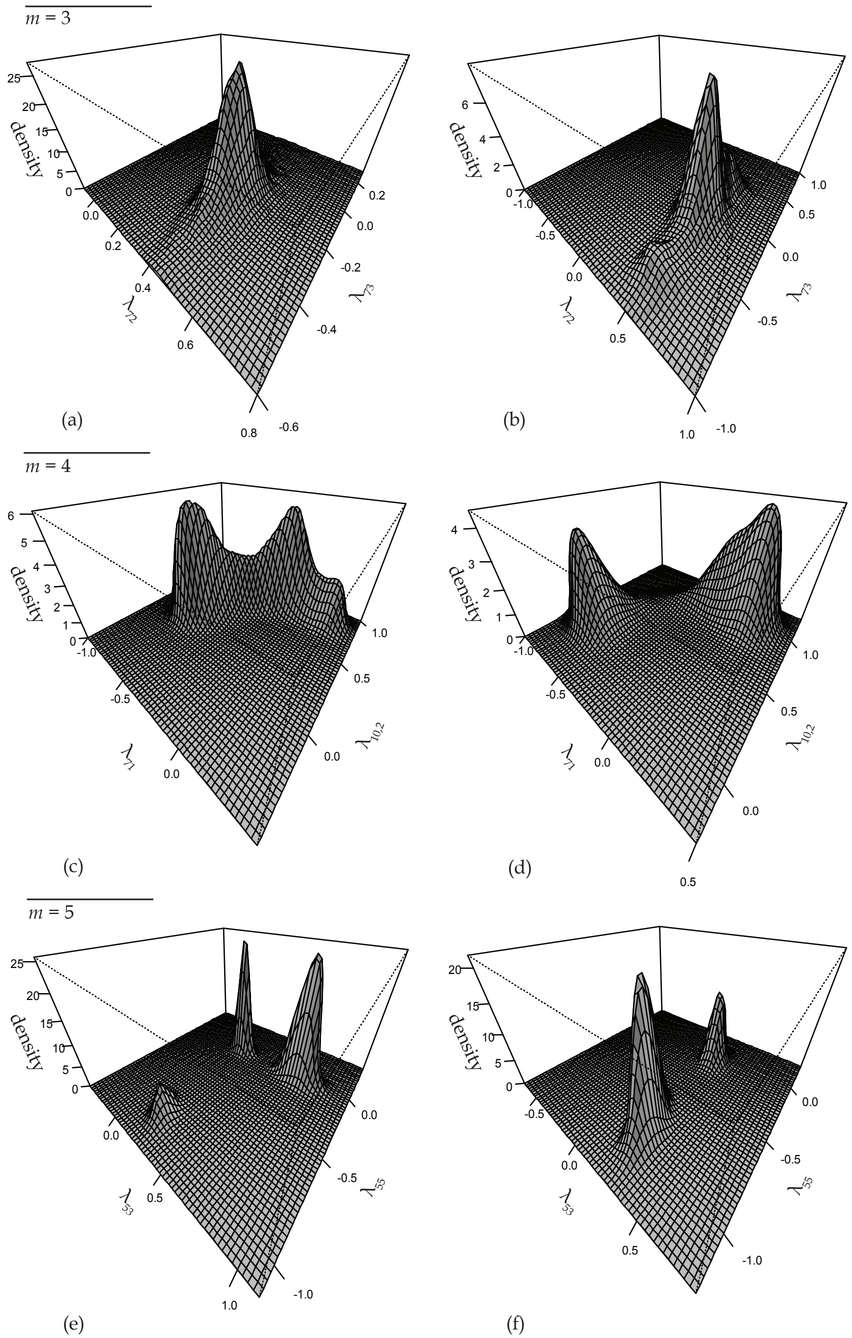

Fig. 2.1. Densities from 3 superimposed chains with varying starting values on the Gosh and Dunson (2009) data. Figures (a) and (b) give the mentioned density for the $m = 3$ solution for full data and a training sample, respectively. Figures (c) and (d) give the analogous for the $m = 4$ solution as do (e) and (f) for the $m = 5$ solution.

Gosh and Dunson (2009) only review models $m = 1, \ldots, 4$, justifying not reviewing the full list of models allowed by the existence condition by reasons of sparseness. However, for certain data it could well be possible that the model of maximum factor dimensionality is optimal. Excluding models *a priori* may thus conflict with dimensionality exploration. Note also that their method of parameter expanded path-sampling may fail if employed to the $m = 5$ and $m = 6$ solutions for reasons just exemplified. In fact, (at least) any parameter-space method of marginal likelihood or BF computation that requires the evaluation of MCMC quantities is bound to fail for the factor analytic dimensionality selection problem as m gets large and the model under consideration overfactors, because a violation of the strict positivity condition hampers proper evaluation. This may explain the apparent breakdown of many methods included in the comparison study by Lopes and West (2004). Note that usage of proper priors on the elements of $\mathbf{\Lambda}$ may exacerbate the problem as it is highly likely that informative prior information may conflict with the multimodality allowed for by the likelihood when the model overfactors. It seems that in order to avoid (heuristical) *a priori* statements on sparsity and to ensure a proper stopping rule, we need to embed the marginal likelihood computation strategy in an assessment strategy. Two possible assessment strategies are given below.

Strategy 2.1 (Semi-Automated)

i. Determine $(\boldsymbol{\mu}^*, \{\mathbf{\Lambda}^*\}_{\mathcal{S}_0}, \mathbf{\Psi}^*)$ for each model of dimensionality $m = 1, \ldots, \tilde{m}$;
ii. For each model of dimensionality $m = 2, \ldots, \tilde{m}$, run a full Gibbs sampler as in lines 4 to 12 of Algorithm 2.2 for differing batches of starting values (and possibly starting seeds) that encourage the sampler to find itself stuck in alternate separated regions in $\mathcal{S}_0$ when these exist (as the one-factor solution cannot be rank-deficient in $\mathbf{\Lambda}$, it can be excluded);
iii. For each model review the marginal posterior densities of $\{\mathbf{\Lambda}\}_{\mathcal{S}_0}$ and $\mathbf{\Psi}$ obtained in the previous step. The model of maximum dimensionality for which the strict positivity condition is not violated, say $\dot{m}$, is taken as the largest model allowed for numerical evaluation;
iv. Set $\tilde{m} = \dot{m}$ and **Call** SRG.

Strategy 2.2 (Automated)

i. Determine $(\boldsymbol{\mu}^*, \{\mathbf{\Lambda}^*\}_{\mathcal{S}_0}, \mathbf{\Psi}^*)$ for each model of dimensionality $m = 1, \ldots, \tilde{m}$;
ii. Initialize SRG for differing batches of starting values (and possibly starting seeds) that encourage the (reduced) samplers to find themselves stuck in alternate separated regions in $\mathcal{S}_0$ when these exist;
iii. Models that violate the strict positivity condition will have widely differing estimates of the marginal likelihood not attributable to sampling fluctuation. The model with the largest stable estimate of the marginal likelihood across the various initializations is to be selected 'best'.

While Strategy 2.2 has been succesfully tested in a limited setting, we will focus on employment of Strategy 2.1 in numerical explorations below. A first reason is computational friendliness of Strategy 2.1 relative to Strategy 2.2. Second, the assessment of the strict positivity condition implied by Strategy 2.1 should give a more firm ground for the exclusion of certain factor solutions relative to *a priori* statements regarding sparseness. We want to evaluate if this indeed is the case.

2.4.6 Simulation Study

Preliminaries

Algorithm 2.2 and the assessment strategies compel us to determine for each model a candidate estimate $(\boldsymbol{\mu}^*, \{\boldsymbol{\Lambda}^*\}_{\mathcal{S}_0}, \boldsymbol{\Psi}^*)$, differing batches of starting values $(\boldsymbol{\mu}^{(0)}, \{\boldsymbol{\Xi}^{(0)}\}_{\mathcal{S}_0}, \boldsymbol{\Psi}^{(0)})$ as well as the scale and shape for the prior on ψ_{jj}. We start with the latter and see to the other choices in turn.

As hyperparameters for the $\pi(\psi_{jj})$ we choose $\nu = 3$ and $\nu d = 1/2$. This setting results in a weakly informative or diffuse prior for the ψ_{jj} elements that decays to zero sufficiently rapid as ψ_{jj} tends to zero. We will not embark on assessing variations in these hyperparameter settings as previous research has established robustness of posterior probabilities against variations in ν and d (S. E. Lee & Press, 1998).

The candidate estimator method is more accurate and efficient if we choose $(\boldsymbol{\mu}^*, \{\boldsymbol{\Lambda}^*\}_{\mathcal{S}_0}, \boldsymbol{\Psi}^*)$ to be in a region of high density. In normal theory models, the posterior mean should suffice. That is why we run a preliminary Gibbs chain for each model of dimensionality m. The starting values in this preliminary Gibbs chain are based on ML common EFA, usually performed by minimizing

$$F(\boldsymbol{\Sigma}) = \mathrm{tr}(\boldsymbol{\Sigma}^{-1}\mathbf{S}) - \ln|\boldsymbol{\Sigma}^{-1}\mathbf{S}| - p. \tag{2.27}$$

The minimization of $F(\boldsymbol{\Sigma})$ is equivalent to, but computationally more convenient than, maximizing the logarithm of (2.6). The algorithm for minimizing (2.27) based on the ML normal equations $\partial F(\boldsymbol{\Sigma})/\partial\boldsymbol{\Psi} = \mathbf{0}$ and $\partial F(\boldsymbol{\Sigma})/\partial\boldsymbol{\Lambda} = \mathbf{0}$, is the Fletcher-Powell algorithm (1963) introduced into FA by Jöreskog (1967) and is widely used in many standard packages performing FA. The values $\hat{\boldsymbol{\Lambda}}$ and $\hat{\boldsymbol{\Psi}}$ that minimize (2.27) are defined as the ML estimates of $\boldsymbol{\Lambda}$ and $\boldsymbol{\Psi}$. In our setup $\hat{\boldsymbol{\Lambda}}$ will be subjected to a Varimax mapping to produce $\{\hat{\boldsymbol{\Lambda}}\}_{\mathcal{S}_0}$. The Varimax mapping will be the mapping of choice throughout the numerical explorations. The maximum likelihood estimate of $\boldsymbol{\mu}$ is simply $\hat{\boldsymbol{\mu}} = \bar{\mathbf{x}}$. Then, by regressing $\boldsymbol{\Xi}$ on $\mathbf{X}$, an estimate for the factor scores can be obtained as

$$\{\hat{\boldsymbol{\Xi}}\}_{\mathcal{S}_0} = (\mathbf{X} - \mathbf{1}_n \otimes \hat{\boldsymbol{\mu}}^{\mathrm{T}})\hat{\boldsymbol{\Psi}}^{-1}\{\hat{\boldsymbol{\Lambda}}\}_{\mathcal{S}_0}\left(\mathbf{I}_m + \{\hat{\boldsymbol{\Lambda}}\}_{\mathcal{S}_0}^{\mathrm{T}}\hat{\boldsymbol{\Psi}}^{-1}\{\hat{\boldsymbol{\Lambda}}\}_{\mathcal{S}_0}\right)^{-1}, \tag{2.28}$$

which will serve as starting value for $\boldsymbol{\Xi}$ in preliminary Gibbs sampling. We run the preliminary chain for $100,000$ iterations without burn-in and the resulting posterior means for $\boldsymbol{\mu}$, $\{\boldsymbol{\Lambda}\}_{\mathcal{S}_0}$, and $\boldsymbol{\Psi}$ are used as $(\boldsymbol{\mu}^*, \{\boldsymbol{\Lambda}^*\}_{\mathcal{S}_0}, \boldsymbol{\Psi}^*)$ in Algorithm 2.2.

For our purposes we will use three batches of starting values. A natural choice for a first batch would be to let $\boldsymbol{\mu}^*$ and $\boldsymbol{\Psi}^*$ serve as $\boldsymbol{\mu}^{(0)}$ and $\boldsymbol{\Psi}^{(0)}$. An efficient starting value $\{\boldsymbol{\Xi}^{(0)}\}_{\mathcal{S}_0}$ may then be obtained by replacing in (2.28) $\hat{\boldsymbol{\mu}}, \{\hat{\boldsymbol{\Lambda}}\}_{\mathcal{S}_0}$, and $\{\hat{\boldsymbol{\Psi}}\}$ with $\boldsymbol{\mu}^*, \{\boldsymbol{\Lambda}^*\}_{\mathcal{S}_0}$, and $\boldsymbol{\Psi}^*$ respectively. An efficient starting value $\{\boldsymbol{\Xi}(\ell)^{(0)}\}_{\mathcal{S}_0}$ may subsequently be obtained by replacing $\mathbf{X}$ and $\mathbf{1}_n$ with $\mathbf{X}(\ell)$ and $\mathbf{1}_q$, respectively. These starting values should find themselves well within the feasible space at a point of some density.

For the second batch we use $\boldsymbol{\mu}^{(0)} = 0$ and $\boldsymbol{\Psi}^{(0)} = 10\mathbf{I}_p$ for each model. Obtainment of $\boldsymbol{\Xi}^{(0)}$ and $\boldsymbol{\Xi}(\ell)^{(0)}$ in this batch is as above with λ_{jk} set at $-7.8\ \forall j,k$. These starting values are well outside the feasible parameter space as we will look at instances of modeling the correlation matrix in the numerical explorations below. The purpose is to heighten the probability that, as the feasible space is approached and entered, the (reduced) sampler(s) get(s) stuck in alternate regions in $\mathcal{S}_0$ if those are in existence for the factor solution reviewed.

For the last batch we set $\boldsymbol{\mu}^{(0)} = 0$ and $\boldsymbol{\Psi}^{(0)} = .01\mathbf{I}_p$. Obtainment of $\boldsymbol{\Xi}^{(0)}$ and $\boldsymbol{\Xi}(\ell)^{(0)}$ is then achieved with $\lambda_{jk} = \sqrt{(1-.01)/m}\ \forall j,k$ (which divides the communality implied by ψ_{jj} evenly over the λ_{jk}). These values find themselves at the boundary of the feasible space and the purpose of their usage is similar to the values in the second batch.

Naturally we could use more initializations. One could also choose to utilize each batch of starting values in conjunction with differing initialization seeds. Here we will employ five initializations: one initialization for the first batch of starting vales, and the remaining two batches are utilized in two initializations each with differing initializing seeds. In Strategy 2.1 we thus initialize a full Gibbs sampler for each model for each specified conjunction of starting values batch and initializing seed. Once the models that do not adhere the strict positivity criterion are weeded out, SRG has to be initialized once, the choice of starting values then being indifferent when using sufficient burn-in. Here, we choose the first batch of starting values for step (iv) of Strategy 2.1.

A First Numerical Exploration

As a first simulation we consider the numerical exploration employed in Lopes and West (2004). They consider a case of modeling the correlation matrix with $p = 7$, $m = 1$, $\boldsymbol{\mu} = \mathbf{0}$, and

$$\begin{aligned}\boldsymbol{\Lambda}^{\mathrm{T}} &= \left[.995\ .975\ .949\ .922\ .894\ .866\ .837\right],\\ \boldsymbol{\Psi} &= \mathrm{diag}\left[.01\ .05\ .10\ .15\ .20\ .25\ .30\right].\end{aligned}$$

We simulated 10 data sets of size $n = 100$ with the specifications above according to the model in (2.4). Step (ii) of Strategy 2.1 was performed with $10,000$ iterations after a $1,000$ iteration burn-in. SRG was invoked with $100,000$ iterations after a $1,000$ iteration burn-in. The number of training samples was taken to be 100. The maximum factor dimensionality allowed by the existence condition amounts to 3.

Step (iii) of Strategy 2.1 revealed that both the $m = 2$ and $m = 3$ solutions could display multimodalities without violating the positivity condition. SRG was thus invoked on the full list of possible models. Performance was such that the $m = 1$

solution was chosen 10/10 times, with the evidence for the chosen solution being strong to decisive (when interpreting twice the natural logarithm of the involved BFs as discussed in Kass & Raftery, 1995).

A Second Numerical Exploration

To assess if more realistic scenarios are not detrimental to performance another numerical exploration is performed. We consider a case of modeling the correlation matrix as in Ghosh and Dunson (2009), the specifications of which can be found in Section 2.4.5. Here, the solution is of higher dimension, the loadings can be relatively small and possibly negative, while the idiosyncratic variances are relatively high.

We simulated 10 data sets of size $n = 100$. Our approach was implemented along the lines specified in simulations above. Step (iii) of Strategy 2.1 revealed that for 9 data sets the strict positivity condition was violated for the $m = 5$ and $m = 6$ solutions. For 1 data set, the strict positivity condition was already violated for the $m = 4$ solution. Using the decision rules laid out in Strategy 2.1 the $m = 3$ solution was strongly favored 10/10 times. In our by no means exhaustive simulations, less optimistic factor loadings and idiosyncratic variances do not seem to detriment performance.

2.5 Pathophysiologic Factors of the Metabolic Syndrome

Continuing with the conjunction of automated candidate estimator and factor analytic dimensionality selection an illustration is given with data on anthropomorphic measures on obese and overweight children and adolescents (Weiss et al., 2004).

2.5.1 Metabolic Syndrome

The 'metabolic syndrome' (MBS) refers to the clustering of metabolic risk factors for the development of type 2 diabetes mellitus and atherosclerotic cardiovascular disease. While the aetiology of the syndrome is largely unknown, the pathophysiologic constellation of the MBS has become an important subject of recent research. Insight into the phenotypic domains that underlie the MBS is deemed important for developing efficient diagnostics and for giving pointers as to which phenotypic patterns are expressions of independent physiological processes.

Since its usage by Edwards, *et al.* (1994) FA has carved out a niche for itself in the MBS research community. For an overview of factor analytic efforts in MBS research we confine by referring to Penno, *et al.* (2006). An important question thrusting the FA efforts in MBS research is if a unifying physiology dominated by insulin resistance (resistance to insulin-stimulated glucose uptake) underlies the clustering of metabolic risk variables, or if there are multiple underlying physiologic phenotypes. Most studies employ PCA and have an inherent focus on the number of factors to be extracted. These are taken as indicative of physiologic processes that underlie the syndrome. This reveals that the factors are given a substantive interpretation and that the goal is to obtain a parsimonious explanation of the observables' covariation through a small number of explanatory factors. As such, employing common EFA would be more appropriate.

2.5.2 Data

We have data from the Yale University School of Medicine on 464 obese and overweight children and adolescents first published in Weiss, *et al.* (2004). The data contain measurements on the body mass index (BMI), blood glucose level at (fasting) baseline (GB) and two hours after (G2) oral glucose intake (both in mg/dl), fasting levels of triglycerides (trig.; mg/dl) and high-density lipoprotein (HDL) cholesterol (mg/dl), systolic and diastolic blood pressure (SBP, DBP; both in mm Hg), and insulin resistance (IR). As in Weiss, *et al.* (2004) the natural logarithm was taken of the glucose, insulin resistance and triglycerides measurements to abide the normality assumption. For more information on the data see Weiss, *et al.* (2004).

The goal is to employ the methods developed in Section 2.3 through the factor analytic implementation described in Section 2.4. We want to explore the dimensionality of the factor solution on the mentioned variables and compare the findings to the original results and other previous FA efforts regarding MBS phenotype exploration. The approach is analogous to the simulation efforts employing Strategy 2.1. The data were standardized such that the correlation structure was modeled. The Gibbs chains in Algorithm 2.2 were run for 200,000 iterations. Furthermore, 1,000 training samples were employed as was a Varimax rotation. The maximum factor dimensionality allowed by the existence condition is 4.

2.5.3 Results

Step (iii) from Strategy 2.1 revealed that the $m = 4$ solution does not abide the strict positivity condition on the conditional densities. The respective calibrated marginal density estimates (in $\log_e$) for the one-, two-, and three-factor solution are: $-16,063.865$, $-16,017.836$, and $-16,110.095$; suggesting that the two-factor model is sufficient for the data at hand. A summary of the posterior estimates of $\{\mathbf{\Lambda}\}_{\mathcal{S}_0}$ and diagonal elements of $\mathbf{\Psi}$ for this model are given in Table 2.1. The results suggest that, given these variables, there is an indication of two physiologic phenotypes that reflect impaired glucose tolerance (glucose metabolism) and dyslipidemia (lipid metabolism).

These results differ from Weiss, *et al.* (2004) and many other factor analytic efforts in which a three-factor solution was found that assigns blood pressure an independent physiologic factor. These efforts utilized heuristical stopping rules mostly, such as the scree plot and the Guttman-Kaiser rule. Moreover, the blood pressure factor in these efforts seems to be a doublet factor (see Mulaik, 2010, Section 9.5). Reviewing our three-factor solution we find very wide highest posterior density intervals. Also, the marginal posterior densities of some of the ψ_{jj} display serious

Table 2.1. Posterior Means and 95% Credible Intervals for $\{\boldsymbol{\Lambda}\}_{\mathcal{S}_0}$ and diag($\boldsymbol{\Psi}$)

	$\{\boldsymbol{\Lambda}\}_{\mathcal{S}_0}$						diag($\boldsymbol{\Psi}$)		
Item	Par.	Mean	95% CI	Par.	Mean	95% CI	Par.	Mean	95% CI
BMI	λ_{11}	.293	[.037,.432]	λ_{12}	-.096	[-.339, .067]	ψ_{11}	.902	[.783,1.036]
$\log_e$\{trig.\}	λ_{21}	.097	[-.093,.406]	λ_{22}	-.545	[-.796,-.007]	ψ_{22}	.669	[.414, .910]
HDL chol.	λ_{31}	-.049	[-.414,.161]	λ_{32}	.636	[.081, .855]	ψ_{33}	.571	[.344, .893]
$\log_e$\{IR\}	λ_{41}	.652	[.206,.841]	λ_{42}	-.203	[-.678, .181]	ψ_{44}	.470	[.320, .622]
$\log_e$\{GB\}	λ_{51}	.481	[.190,.635]	λ_{52}	.023	[-.358, .200]	ψ_{55}	.767	[.632, .914]
$\log_e$\{G2\}	λ_{61}	.366	[.038,.513]	λ_{62}	-.132	[-.429, .051]	ψ_{66}	.835	[.719, .965]
SBP	λ_{71}	.319	[.105,.516]	λ_{72}	.084	[-.131, .486]	ψ_{77}	.888	[.701,1.045]
DBP	λ_{81}	.248	[-.127,.773]	λ_{82}	.226	[.035, .842]	ψ_{88}	.861	[.308,1.071]

multimodality (although the positivity condition is not violated). These are additional indications that the three-factor solution extracts too many latent factors. The finding that blood pressure is not an independent pathophysiologic factor is consistent with epidemiologic evidence (see e.g., Liese, Mayer-Davis, & Haffner, 1998) that insulin resistance and lipid metabolism play a role in the pathogenesis of hypertension rather than hypertension being a physiologic phenotype. Moreover, the two factors connect to two main hypotheses regarding syndrome aetiology, stating that the risk-factor associations are due to abnormality of the insulin/glucose metabolism and/or abnormality of the lipid metabolism (see e.g., Liese et al., 1998).

2.6 Discussion

The purpose of this paper was twofold. First, we sought to develop an automated candidate estimator method. Second, we utilized this method as a means of exploring the dimensionality selection problem in FA. Both pillars of the paper will be discussed.

2.6.1 Comments on the Automated Candidate Estimator

An automated candidate estimator was developed by integrating the candidate estimator for marginal likelihood computation with (a) the use of training samples for converting noninformative improper priors into proper posterior distributions for model selection and (b) a conceptual marginal likelihood frame for models that deal with permutative invariance over the parameter indices. This method has several attractive properties: (i) it retains its conceptual simplicity relative to the original candidate estimator (Chib, 1995); (ii) it does not lead to indeterminate estimates of the marginal likelihood when noninformative improper priors are utilized and thus leads to determinate (pseudo) Bayes factors; (iii) as such it may be seen as an MCMC implementation of the IBF (Berger & Pericchi, 1996); (iv) which needs little new programming beyond the original candidate estimator. Regarding point (iii), utilization of the IBF can also be regarded as a means of generating specific default prior distributions for model selection: The so-called intrinsic priors (e.g., Berger & Pericchi, 1996). Obtainment of these priors requires the marginal densities defined over the training data to be analytically tractable however. One may alternatively view the automated candidate estimator as a means of generating IBF's when intrinsic priors are very difficult to obtain or analytically unavailable.

The setting of the automated candidate estimator is that of an enumerable model space with (Gaussian) models amenable to MCMC schemes based on Gibbs sampling and possible data augmentation. A broader set of models may be explored by incorporating, for example, auxiliary mixture sampling into the method.

Notwithstanding its attractiveness, the method has several drawbacks. First, there is substantial computing time involved when using standard desktop terminals. Second, the method does not seem suitable for situations in which one has little data relative to the number of parameters, as the training sample size may then approach or exceed the data size.

2.6.2 Comments on the Factor Analytic Application

The applicatory focus has been on what can be seen as a Bayesian version of common exploratory factor analysis. The FA model as given here only requires specification of a prior distribution on the idiosyncratic variances. A specification for which the model is robust. Identification of the factor loadings matrix is sought after, as opposed to fixation and truncation of certain of its elements, by a choice of rotational mapping.

With regard to the problem of selection of latent factor dimensionality we find that a crucial regularity condition for likelihood-based approaches may directly connect to a violation of a positivity condition required for ergodicity in MCMC-based approaches. This may hold considerable importance for at least parameter-space MCMC approaches towards marginal likelihood and BF computation strategies for factor analytic dimensionality selection as it can explain the apparent breakdown of certain methods employed in previous studies and as it implies that a certain computation strategy has to be embedded in an assessment strategy in order to avoid naiveness of assessment.

The findings may also hold considerable importance for those who have a preference for likelihood-based approaches towards dimensionality selection in FA. To avoid overfactoring one needs to evaluate the shape of the likelihood. Such an evaluation can be performed by some of the objective Bayesian methods given in preceding sections. Also, likelihood-based approaches may benefit from analogous assessment strategies as given above.

Next to an exploratory stopping rule, the candidate estimator method can also be applied in a more confirmatory manner, specifying only those models whose dimensions are deemed plausible. As such, Bayesian methods may bridge the EFA - CFA divide. Generally, in the CFA model all attention regarding misspecification is geared towards the pre-specified pattern of factor loadings. The evaluation of model fit in CFA is then essentially the evaluation of a diffuse hypothesis (Hoyle & Duvall, 2004) as it is unclear in the case of misspecification if the pattern of loadings or the factor dimensionality is to blame. We may give a Bayesian analogue of the 'unrestricted factor model' (Jöreskog, 1979) to separate dimensionality and pattern selection. First, embark on evaluating a series of models with differing factor dimensionality along the lines specified in previous sections. Then, whence settled on latent factor dimensionality, restrictions may straightforwardly be imposed on the factor loadings as desired. Competing factor loading patterns may again be assessed through marginal likelihood evaluation.

2A Fast Algorithm for Procrustean Matrix Smoothing

Basically, the problem of finding ${}^{\pm}\mathbf{P}$ such that $\{\boldsymbol{\Lambda}^{(c)}\}_{\mathcal{S}_0} = \{\boldsymbol{\Lambda}^{(c)}\}_{\mathcal{S}_d}({}^{\pm}\mathbf{P})$ may be approached as a question in Procrustean rotation. Computationally, Procrustean rotation depends on a numerical tolerance for rank deficiency, which is a function of machine precision. Overfactoring (and in a lesser degree underfactoring) coupled with sampling fluctuation give the possibility that a sample $\{\boldsymbol{\Lambda}^{(c)}\}$ may be drawn that does not surpass the mentioned tolerance threshold. This may especially be a problem when using terminals with relatively low precision. When high machine precision is at disposal one might proceed in finding ${}^{\pm}\mathbf{P}$ as follows. Say we want to determine a nonsingular matrix $\mathbf{T}$ that will transform, in a least squares sense, the given matrix $\{\boldsymbol{\Lambda}^{(c)}\}_{\mathcal{S}_d}$ to a prescribed factor pattern $\{\boldsymbol{\Lambda}^{*}\}_{\mathcal{S}_0}$. The orthogonal Procrustes problem may then be stated as

$$\mathbf{T} = \arg\min_{\mathbf{H}} \left\{ ||\{\boldsymbol{\Lambda}^{(c)}\}_{\mathcal{S}_d}\mathbf{H} - \{\boldsymbol{\Lambda}^{*}\}_{\mathcal{S}_0}||_F : \mathbf{H}^{\mathrm{T}}\mathbf{H} = \mathbf{I}_m \right\}. \tag{2.29}$$

Note that the Procrustean rotation may remove part of the sampling variability beyond mere permutations in the ordering and polarity of the columns of $\{\boldsymbol{\Lambda}^{(c)}\}_{\mathcal{S}_d}$. However, $\mathbf{T}$ will be 'close' to a signed permutation matrix ${}^{\pm}\mathbf{P}$ in the sense that each row and each column will have one element relatively close to plus or minus unity. We may then abstract the signed permutation matrix contained in $\mathbf{T}$ by the following algorithm.

Algorithm 2.3 (Procrustean to Signed Permutation Matrix Smoothing) The input for this algorithm is the Procrustean rotation matrix $\mathbf{T}$ from (2.29). The output is a signed permutation matrix ${}^{\pm}\mathbf{P}$. The algorithm performs what can be seen as a 'smoothing' of $\mathbf{T}$.

for $k = 1$ to m **do**
  $k^{+} = \arg\max_k \mathbf{T}(:,k)$
  $k^{-} = \arg\min_k \mathbf{T}(:,k)$
  $r^{+} = \max(\mathbf{T}(:,k))$
  $r^{-} = \min(\mathbf{T}(:,k))$
  ${}^{\pm}\mathbf{P}(:,k) = \mathbf{0}$
  if ($r^{-} < r^{+}$ and $r^{-} > 0$ and $r^{+} > 0$) **then**

    ${}^{\pm}\mathbf{P}(k^+, k) = 1$
  else if $(|r^-| < r^+$ and $r^- < 0$ and $r^+ > 0)$ **then**
    ${}^{\pm}\mathbf{P}(k^+, k) = 1$
  else if $(|r^-| > r^+$ and $r^- < 0$ and $r^+ > 0)$ **then**
    ${}^{\pm}\mathbf{P}(k^-, k) = -1$
  else
    ${}^{\pm}\mathbf{P}(k^-, k) = -1$
  end if
end for

Algorithm 2.3 cannot be used when $m = 1$ as a Procrustean rotation in such a situation is not possible. In this situation only a polarity reversal is possible, easily captured by requiring the scalar $\{\mathbf{\Lambda}^*\}^{\mathrm{T}}_{\mathcal{S}_0} \{\mathbf{\Lambda}^{(c)}\}_{\mathcal{S}_d}$ to be positive.

3

Rotational Uniqueness Conditions Under Oblique Factor Correlation Metric

Peeters, C.F.W. (2012) *Psychometrika*, *77*, 288–292

3.1 Introduction

Suppose $\mathbf{\Sigma} = \mathbf{\Lambda\Phi\Lambda}^{\mathrm{T}} + \mathbf{\Psi}$ is the usual oblique factor analysis model. Here, $\mathbf{\Lambda} \in \mathbb{R}^{p\times m}$ is a matrix of factor loadings in which each element λ_{jk} is the loading of the jth variable on the kth factor, $j = 1, \ldots, p$, $k = 1, \ldots, m$; $\mathbf{\Phi} \in \mathbb{R}^{m\times m}$ denotes the factor covariance matrix; and $\mathbf{\Psi} \in \mathbb{R}^{p\times p} \equiv \mathrm{diag}[\psi_{11}, \ldots, \psi_{pp}]$ contains the error variances. In the remainder we will assume the usual regularity assumptions: (a) $\mathrm{rank}(\mathbf{\Lambda}) = m$; (b) $\psi_{jj} > 0\ \forall j$; and (c) $(p-m)^2 - p - m \geqslant 0$, simply stating nonnegative degrees of freedom for existence.

As is well-known, the factor model is inherently underidentified, implying that $\mathbf{\Sigma}$ does not have a unique solution without imposing restrictions. Given $\mathbf{\Psi}$, two factor models defined by $\{\mathbf{\Lambda}, \mathbf{\Phi}\}$ and $\{\mathbf{\Lambda}^{\ddagger}, \mathbf{\Phi}^{\ddagger}\}$ are equivalent if there exists a mapping $\delta : \{\mathbf{\Lambda}, \mathbf{\Phi}\} \longrightarrow \{\mathbf{\Lambda}^{\ddagger}, \mathbf{\Phi}^{\ddagger}\}$ such that $\mathbf{\Sigma}(\mathbf{\Lambda}, \mathbf{\Phi}, \mathbf{\Psi}) = \mathbf{\Sigma}[\delta(\mathbf{\Lambda}, \mathbf{\Phi}), \mathbf{\Psi}]$. For the factor model we find $\mathbf{\Sigma} = (\mathbf{\Lambda R})[\mathbf{R}^{-1}\mathbf{\Phi}(\mathbf{R}^{\mathrm{T}})^{-1}](\mathbf{\Lambda R})^{\mathrm{T}} + \mathbf{\Psi}$, where $\mathbf{R} \in \mathbb{R}^{m\times m}$ is an arbitrary nonsingular matrix, implying that there is an infinite number of alternative matrices $\mathbf{\Lambda}^{\ddagger} = \mathbf{\Lambda R}$ and $\mathbf{\Phi}^{\ddagger} = \mathbf{R}^{-1}\mathbf{\Phi}(\mathbf{R}^{\mathrm{T}})^{-1}$ that generate the same covariance structure $\mathbf{\Sigma}$. The operation $\mathbf{\Lambda} \mapsto \mathbf{\Lambda R}$ is termed 'rotation', and $\mathbf{\Lambda}$ and $\mathbf{\Phi}$ are said to be globally rotationally unique *iff* $\mathbf{R} = \mathbf{I}_m$.

Finding conditions to ensure (global) rotational uniqueness has been an active area of research and debate in the factor analysis community. Building on Howe (1955), Jöreskog (1969) conjectured sufficient conditions for uniqueness for both the orthogonal ($\mathbf{\Phi} = \mathbf{I}_m$) and oblique factor model through specification of fixed elements in $\mathbf{\Lambda}$ and $\mathbf{\Phi}$. Dunn (1973) showed, especially for the orthogonal model, that Jöreskog's conditions were not sufficient. His substitute conditions based on fixed zero elements are sufficient for local rotational uniqueness only, in the sense that 2^m combinations of polarity reversals in the columns of $\mathbf{\Lambda}$ are allowed, giving that $\mathbf{R}$ is then $\mathrm{diag}[\pm 1, \ldots, \pm 1]$. Jennrich (1978) gave sufficient conditions for local rotational uniqueness under reflections for the orthogonal model when the fixed elements are arbitrary. These works inspired Jöreskog to write an addendum to his 1969 article (Jöreskog, 1979), focussing on reformulating the sufficiency conditions for the oblique factor solution with fixed zero elements. He gave the following conditions:

C1 Let $\mathbf{\Lambda}$ have at least $m-1$ fixed zeroes in each column;
C2 Let $\text{rank}(\mathbf{\Lambda}^{[k]}) = m-1$, where $\mathbf{\Lambda}^{[k]}$, $k = 1, \ldots, m$, is the submatrix of $\mathbf{\Lambda}$, consisting of the rows of $\mathbf{\Lambda}$ which have fixed zero elements in the kth column with these zeroes deleted;
C3 Let $\mathbf{\Phi}$ be a symmetric positive definite matrix with $\text{diag}(\mathbf{\Phi}) = \mathbf{I}_m$ (i.e., $\mathbf{\Phi}$ is a correlation matrix);

and conjectured that C1-C3 are sufficient for obtaining global rotational uniqueness. Moreover, he conjectured that conditions C1 and C3 are equivalent to:

C* Let $\mathbf{\Lambda}$ have at least $m-1$ fixed zeroes in each column and one fixed non-zero value in each column, the latter values being in different rows;

and subsequently proved global rotational uniqueness under pairing of conditions C2 and C*.

Many recent texts follow Jöreskog (1979) in stating that conditions C1-C3 are sufficient for (rotational) uniqueness (e.g., Hoyle & Duvall, 2004; Asparouhov & Muthén, 2009). However, it will be shown that conditions C1-C3 are *not* sufficient for global rotational uniqueness but local rotational uniqueness only and, hence, that conditions C1 and C3 are *not* equivalent to C* in terms of unicity of the solution. Although this result may be implicitly known or be considered tacit knowledge, here it is made explicit.

In the remainder, condition set C1-C3 will be amended with an additional condition. It will then be shown that the amended condition set is sufficient for obtaining global rotational uniqueness, implying that C1-C3 do not lead to global uniqueness and the non-equivalence of conditions C1 and C3 to C*. Section 3.3 concludes with a discussion.

3.2 Global Rotational Uniqueness Under $\text{diag}(\mathbf{\Phi}) = \mathbf{I}_m$

Consider the following addition to conditions C1-C3:

C4 Let in each column of $\mathbf{\Lambda}$ one parameter non-fixed by condition C1 be polarity truncated to take only positive or negative values, that is: In each column of $\mathbf{\Lambda}$ one element is to adopt either strict positivity ($\lambda_{jk} > 0$), or strict negativity ($-\lambda_{jk} > 0$).

Proposition 3.1. *Let the mapping* $\delta : \{\mathbf{\Lambda}, \mathbf{\Phi}\} \longrightarrow \{\mathbf{\Lambda}^{\ddagger}, \mathbf{\Phi}^{\ddagger}\}$ *be defined by* $\mathbf{\Lambda}^{\ddagger} = \mathbf{\Lambda}\mathbf{R}$ *and* $\mathbf{\Phi}^{\ddagger} = \mathbf{R}^{-1}\mathbf{\Phi}(\mathbf{R}^{\mathrm{T}})^{-1}$*, where* $\mathbf{R} \in \mathbb{R}^{m \times m}$ *denotes an arbitrary nonsingular matrix. If conditions* C1-C4 *hold, then* $\mathbf{R} = \mathbf{I}_m$*.*

Proof. Let conditions C1-C4 hold on $\{\mathbf{\Lambda}, \mathbf{\Phi}\}$. We start by showing that $\mathbf{R}$ is diagonal under conditions C1-C2, which can be shown with an argument analogous to Anderson (1984, pp. 576-577). Under given conditions it is always possible to find a permutation matrix $\mathbf{P}_1$ of respective dimension $p \times p$, such that $\mathbf{P}_1\mathbf{\Lambda}$ gives a block lower right triangular form on $\mathbf{\Lambda}^{[1]}$. We then have

$$\mathbf{\Lambda}' = \mathbf{P}_1\mathbf{\Lambda} = \begin{bmatrix} \mathbf{0} & \mathbf{\Lambda}^{[1]} \\ \boldsymbol{\lambda}_{(1)} & \mathbf{\Lambda}_{[1]} \end{bmatrix}, \tag{3.1}$$

where $\mathbf{0}$ is a $(m-1)$-dimensional null vector, $\boldsymbol{\lambda}_{(1)} \in \mathbb{R}^{(p-m+1)\times 1}$, $\mathbf{\Lambda}^{[1]} \in \mathbb{R}^{(m-1)\times(m-1)}$, and $\mathbf{\Lambda}_{[1]} \in \mathbb{R}^{(p-m+1)\times(m-1)}$. Now, let

$$\mathbf{R} = \begin{bmatrix} r_{11} & \mathbf{r}_{12} \\ \mathbf{r}_{21} & \mathbf{R}_{22} \end{bmatrix},$$

where r_{11} is a scalar, $\mathbf{r}_{12} \in \mathbb{R}^{1\times(m-1)}$, $\mathbf{r}_{21} \in \mathbb{R}^{(m-1)\times 1}$, and $\mathbf{R}_{22} \in \mathbb{R}^{(m-1)\times(m-1)}$. Then

$$(\mathbf{\Lambda}')^{\ddagger} = \mathbf{\Lambda}'\mathbf{R} = \begin{bmatrix} \mathbf{\Lambda}^{[1]}\mathbf{r}_{21} & \mathbf{\Lambda}^{[1]}\mathbf{R}_{22} \\ r_{11}\boldsymbol{\lambda}_{(1)} + \mathbf{\Lambda}_{[1]}\mathbf{r}_{21} & \boldsymbol{\lambda}_{(1)}\mathbf{r}_{12} + \mathbf{\Lambda}_{[1]}\mathbf{R}_{22} \end{bmatrix}.$$

As the rank of $\mathbf{\Lambda}^{[1]}$ is $m-1$, $\mathbf{r}_{21}$ should be a null vector. We may follow the same procedure for each respective remaining submatrix $\mathbf{\Lambda}^{[k]}$. That is, we can find an $(m \times m)$-dimensional permutation matrix $\mathbf{P}_2$ such that $\mathbf{P}_1'\mathbf{\Lambda}\mathbf{P}_2$ has a structure as in (3.1) with the index 1 replaced by k (we may always find a pair of permutation matrices $\mathbf{P}_1'$ and $\mathbf{P}_2$, such that $\mathbf{P}_1'\mathbf{\Lambda}\mathbf{P}_2$ gives a block lower right triangular form on $\mathbf{\Lambda}^{[k]}$). From the accompanying reorderings $\mathbf{P}_2^{\mathrm{T}}\mathbf{R}\mathbf{P}_2$ implied by the identity

$$\begin{aligned} &\mathbf{P}_1'\mathbf{\Sigma}(\mathbf{P}_1')^{\mathrm{T}} \\ &= (\mathbf{P}_1'\mathbf{\Lambda}\mathbf{R})[\mathbf{R}^{-1}\mathbf{\Phi}(\mathbf{R}^{\mathrm{T}})^{-1}](\mathbf{P}_1'\mathbf{\Lambda}\mathbf{R})^{\mathrm{T}} + \mathbf{P}_1'\mathbf{\Psi}(\mathbf{P}_1')^{\mathrm{T}} \\ &= (\mathbf{P}_1'\mathbf{\Lambda}\mathbf{P}_2\mathbf{P}_2^{\mathrm{T}}\mathbf{R}\mathbf{P}_2)[\mathbf{P}_2^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{P}_2\mathbf{P}_2^{\mathrm{T}}\mathbf{\Phi}\mathbf{P}_2\mathbf{P}_2^{\mathrm{T}}(\mathbf{R}^{\mathrm{T}})^{-1}\mathbf{P}_2](\mathbf{P}_1'\mathbf{\Lambda}\mathbf{P}_2\mathbf{P}_2^{\mathrm{T}}\mathbf{R}\mathbf{P}_2)^{\mathrm{T}} + \mathbf{P}_1'\mathbf{\Psi}(\mathbf{P}_1')^{\mathrm{T}}, \end{aligned}$$

it follows that under conditions C1-C2, $\mathbf{R} = \mathrm{diag}[r_{11}, \ldots, r_{mm}]$.

The properties of diagonal matrices are such that now $\mathbf{R}^{\mathrm{T}} = \mathbf{R}$ and $\mathbf{R}^{-1} = \mathrm{diag}[r_{11}^{-1}, \ldots, r_{mm}^{-1}]$. Superimposing condition C3 then implies

$$\phi_{kk}^{\ddagger} = r_{kk}^{-2}\phi_{kk} = r_{kk}^{-2} = 1 \quad \forall k,$$

giving that $r_{kk} = \pm 1$, and $\mathbf{R} = \mathrm{diag}[\pm 1, \ldots, \pm 1]$.

By demanding that each column of $\mathbf{\Lambda}$ has a polarity truncation, column multiplication by -1 is no longer possible as this would imply a reversal of the polarity truncation (direction inequality symbol). Superimposing condition C4 on C1-C3 thus gives that $\mathbf{R} = \mathbf{I}_m$. The proposition follows. □

Remark 3.1. The proof of Proposition 3.1 implies, contrary to previous conjectures, that conditions C1-C3 are not sufficient for global rotational uniqueness as they provide local rotational uniqueness only and, hence, that conditions C1 and C3 are not equivalent to C* as the pairing of C2-C* does provide global rotational uniqueness (Jöreskog, 1979). Moreover, the proposition indicates how inequality restrictions can aid in the attainment of global rotational uniqueness.

Remark 3.2. Instead of using strict positivity or strict negativity truncations as formulated in condition C4 we could also use strict polarity truncations by (arbitrary) constants, that is: Every column of $\mathbf{\Lambda}$ should contain either $\lambda_{jk} > c_k \in \mathbb{R}^+$ or $-\lambda_{jk} > c_k \in \mathbb{R}^+$. While this will produce global rotational uniqueness whence superimposed on C1-C3 along the same lines as C4, it may not be practical in a research setting. It may, for example, be possible to specify $\lambda_{jk} > c_k \in \mathbb{R}^+$ while the true parameter value $0 \leqslant \lambda_{jk} < c_k$ (in the positive reflection). One would then run into estimation trouble in numerical applications. A related issue lies in choosing the loading elements for column polarity fixation. For (Bayesian) estimation efficiency (see Discussion section) condition C4 should be imposed on loadings that, from prior knowledge or theory, are believed to be large.

Remark 3.3. Please note that rotational uniqueness of $\mathbf{\Lambda}$ will not guarantee identifiability of the FA model (Bollen & Jöreskog, 1985), as underidentification of $\mathbf{\Psi}$ may imply underidentification of $\mathbf{\Lambda}$. However, if the regularity assumptions stated in the introduction hold, then conditions C1-C4 will, next to rotational uniqueness, also provide identifiability. If unsure if identifiability is obtained, one could endeavor on algebraically checking (local) identification utilizing the Wald rank rule (Bekker et al., 1994).

3.3 Discussion

The following question deserves some exploration: What are reasons to prefer condition set C1-C4 above the pairing C2-C*? Before delving into possible answers it is stated why the addition of condition C4 is deemed important when working in factor correlation metric.

Not attaining global rotational uniqueness in factor correlation metric under C1-C3 does not hamper maximum likelihood estimation, as any local minimum has equivalent representations through simple polarity reflections. In obtaining factor loading standard errors or Bayesian estimates the situation is a little more intricate as the parameter space under C1-C3 is multimodal. The modes (defined over polarity reflections) should be widely separated in order for unimodal normal approximations or resampling techniques to yield valid estimates of the standard errors (see, e.g., Dolan & Molenaar, 1991). As Bayesian modeling proceeds through exploration of posterior space, posterior estimates will be flawed when transition probability between modes is non-negligible. Imposing C4 can aid when the modes are not well-separated. In Bayesian modeling, for example, imposing C4 implies a truncation of the posterior density and will restrict posterior simulation to a single mode – a fact that has been recognized (for the orthogonal factor model) by Geweke and Zhou (1996).

A first reason for preferring C1-C4 above the pairing C2-C* might be found in the topic of arbitrary units of measurement. Many (psychological) tests have units of measurement with no intrinsic meaning. Let $\mathbf{D}$ be a diagonal matrix with positive diagonal elements that indicate a change in test score units. Then $\mathbf{D\Sigma D} \equiv \mathbf{\Sigma}^{\dagger}$. If

$\mathbf{\Lambda}$ is identified by C1-C4, then $\mathbf{D\Lambda} \equiv \mathbf{\Lambda}^{\dagger}$ is similarly identified. If $\mathbf{\Lambda}$ is identified by the pairing C2-C*, then each column of $\mathbf{D\Lambda}$ has to be renormalized (Anderson, 1984, p. 557).

A second reason for preference may be that the condition set C1-C4 is less restrictive on $\mathbf{\Lambda}$. Consider the unrestricted factor model. Unrestricted solutions correspond to exploratory factor analysis (EFA) in the sense that only minimal restrictions are placed on the model to achieve at least a local rotationally unique solution for m factors. As such, an unrestricted solution for m common factors does not restrict the factor space and will yield an optimal fit for any model with m factors (Mulaik, 2010, Section 15.4). In effect, a minimal set of restrictions based on C2 and C* or C1-C4 entails a choice of rotation of the EFA model. The pairing C2-C* imposes the minimum of m^2 restrictions, equalling the number of non-redundant elements in $\mathbf{R}$, on the parameter space of $\mathbf{\Lambda}$. The condition set C1-C4 imposes only $m(m-1)$ fixed-value restrictions on $\mathbf{\Lambda}$. The parameters involved in the polarity truncations are free to be estimated in either the positive or negative range (note also that the polarity truncations do not restrict the factor space). The fewer number of restrictions on $\mathbf{\Lambda}$ under C1-C4 make it a more flexible set for unrestricted formulations of the (confirmatory) factor model.

4

Inequality Constrained Confirmatory Factor Analysis: Bayesian Specification and Model Selection

Peeters, C.F.W. (currently under revision for *J. Roy. Statist. Soc. Ser. B*)

4.1 Introduction

4.1.1 Statement of Problem

Suppose we have a matrix of standardized data $\mathbf{Z}^{\mathrm{T}} \equiv [\mathbf{z}_1, \ldots, \mathbf{z}_n]$ with $\mathbf{z}_i^{\mathrm{T}} \equiv [z_{i1}, \ldots, z_{ip}] \in \mathbb{R}^p \sim \mathcal{N}_p(\mu, \mathbf{\Sigma})$. Model evaluation for covariance structures in the classical sense then proceeds by the pairwise evaluation of either

$$H_0 : \mathbf{S} = \mathbf{\Sigma} \quad vs \quad H_1 : \mathbf{S} = \mathbf{W}, \tag{4.1}$$

where $\mathbf{\Sigma}$ denotes the correlation matrix as a function of model-dependent parameter collection $\mathbf{\Theta}$, $\mathbf{S}$ denotes the sample correlation matrix $\mathbf{Z}^{\mathrm{T}}\mathbf{Z}$, and $\mathbf{W}$ denotes an arbitrary positive definite matrix; or

$$H_0 : \mathbf{S} = \mathbf{\Sigma} \quad vs \quad H_1 : \mathbf{S} = \mathbf{\Sigma}', \tag{4.2}$$

provided $\mathbf{\Theta} \subset \mathbf{\Theta}'$. Such a setting is restrictive as substantive interest is often in selecting a 'best' model in the sense of balancing model fit and complexity from, say, the set $\mathcal{B} = \{M_1, \ldots, M_B\}$ such that we may evaluate $\forall b \neq b'$

$$M_b : \mathbf{\Sigma}_b \quad vs \quad M_{b'} : \mathbf{\Sigma}_{b'}, \tag{4.3}$$

where possibly $\mathbf{\Theta}_b \not\subset \mathbf{\Theta}_{b'}$ and/or $\mathbf{\Theta}_{b'} \not\subset \mathbf{\Theta}_b$.

The purpose of this paper is based on two pillars. First, we seek to develop Bayes factors for the setting (4.3) where

$$\boldsymbol{\theta}_b \equiv \mathrm{vec}(\mathbf{\Theta}_b) \equiv \{\boldsymbol{\theta} : \mathbf{\Omega}_b\boldsymbol{\theta} - \boldsymbol{\alpha}_b > \mathbf{0}\}, \tag{4.4}$$

in which $\mathbf{\Omega}_b$ and $\boldsymbol{\alpha}_b$ denote, respectively, a given fixed matrix and a given fixed vector, whose conjunction represents a system of linear restrictions defining model M_b. Second, we seek to utilize these Bayes factors to select among differing inequality constrained formulations of the factor analytic correlation structure. The remainder of the introduction provides the backdrop for these motivations.

4.1.2 Factor Analytic Models and Parameter Restrictions

Factor analysis (FA) assumes that a random vector $Z \in \mathbb{R}^p$ consists of correlated variables that can be grouped by their covariances into a random vector of latent factors $\Xi \in \mathbb{R}^m$, with $m < p$. In addition to being a data-reductional tool, FA as a data-analytic tool has become an appreciated aid in such diverse fields as medical research, social and behavioral science, and natural science.

An important topic in FA is the restriction of parameters. Fixed-value equality restriction of parameters is required for identification purposes, i.e., in setting the metric of the latent factors and for providing general identification in ensuring, at a minimum, existence of a solution. Equality restrictions are also used in expressing a pre-specified factor loading pattern in confirmatory factor analysis (CFA), and in testing measurement invariance across groups. Provisions for inequality restrictions on parameters in the context of FA are mainly used for the prevention of impermissible estimates, most notably in combating Heywood cases. Techniques for estimating FA models under equality and inequality restrictions have been based on the penalty function method (S. Y. Lee, 1979, 1980), the Lagrange multiplier method (S. Y. Lee, 1981), the Newton method (Jamshidian & Bentler, 1993), and the adaptive barrier method (Tsonaka & Moustaki, 2007). Techniques for inequality restricted estimation have also been based on variations on the theme of slack variables (Rindskopf, 1983, 1984; S. Y. Lee & Poon, 1985).

While pivotal to FA, restrainment of parameters in this context knows some lacunae. First, identification constraints based on, especially, exclusion restrictions often have no substantive meaning beyond being an artefact for obtaining (local) uniqueness. There is little provision to let restricted parameters express substantive theoretical ideas regarding magnitude and direction of parameter effects. Second, restricted estimation theory has received the lion's share of attention, leaving the topic of restricted-model selection out of focus. Our approach to model selection aims to come forward with respect to these gaps.

4.1.3 Motivating Example

We take interest in CFA (Jöreskog, 1969). As opposed to the abductive thrust of methods in the exploratory realm of FA (e.g., Haig, 2005), CFA seeks to determine the ability of a pre-specified factor model to fit an observed covariance matrix. Pivotal in doing so is the matrix of factor loadings $\mathbf{\Lambda}$, which indicates the correlation of the jth variable on the kth latent factor (this matrix will be more clearly defined in Section 4.2.1). The pre-specified structure is usually obtained through usage of exclusion restrictions and is subject to the lacunae mentioned above. Consider the following (running) example as an exemplification of the situation we try to avoid and the alternative we set out to create.

Example 4.1. Say we have data from the Yale University School of Medicine on 464 obese and overweight children and adolescents first published in Weiss, *et al.* (2004). The data contain measurements on the body mass index (BMI), blood glucose level at (fasting) baseline (GB) and two hours after (G2) oral glucose intake (both in mg/dl), fasting levels of triglycerides (trig.; mg/dl) and high-density lipoprotein (HDL) cholesterol (mg/dl), systolic and diastolic blood pressure (SBP, DBP; both in mm Hg), and insulin resistance (IR). As in Weiss, *et al.* (2004) the natural logarithm was taken of the glucose, insulin resistance and triglycerides measurements to abide the normality assumption.

The measurements relate to metabolic risk factors, the clustering of which is often indicated as the 'metabolic syndrome' (MBS) which is thought to be a precursor for the development of type 2 diabetes mellitus and atherosclerotic cardiovascular disease. The pathophysiologic constellation of the MBS has become an important subject of recent research and latent factors extracted from factor analytic modeling efforts are taken as indicative of physiologic processes that underlie the syndrome.

Previous research has indicated a factor solution of $m = 2$ as optimal (Peeters, invited revision a). There are at least two competing theories connected to a solution of mentioned dimensionality: (a) There are two independent pathophysiologic constellations indicating abnormalities in the glucose and lipid metabolism, respectively; and (b) The glucose and lipid metabolism are connected through insulin resistance, which is then the measure tying the syndrome together. A classical factor analyst might specify these theoretical ponderings as follows:

$$\mathbf{\Lambda}_1 = \begin{bmatrix} \lambda_{11} & 0 \\ 0 & \lambda_{22} \\ 0 & \lambda_{32} \\ \lambda_{41} & 0 \\ \lambda_{51} & 0 \\ \lambda_{61} & 0 \\ \lambda_{71} & 0 \\ 0 & \lambda_{82} \end{bmatrix} \begin{matrix} \text{BMI} \\ \log_e\{\text{trig.}\} \\ \text{HDL chol.} \\ \log_e\{\text{IR}\} \\ \log_e\{\text{GB}\} \\ \log_e\{\text{G2}\} \\ \text{SBP} \\ \text{DBP} \end{matrix}, \quad \mathbf{\Lambda}_2 = \begin{bmatrix} 0 & \lambda_{12} \\ 0 & \lambda_{22} \\ 0 & \lambda_{32} \\ \lambda_{41} & \lambda_{42} \\ \lambda_{51} & 0 \\ \lambda_{61} & 0 \\ 0 & \lambda_{72} \\ 0 & \lambda_{82} \end{bmatrix} \begin{matrix} \text{BMI} \\ \log_e\{\text{trig.}\} \\ \text{HDL chol.} \\ \log_e\{\text{IR}\} \\ \log_e\{\text{GB}\} \\ \log_e\{\text{G2}\} \\ \text{SBP} \\ \text{DBP} \end{matrix}.$$

The perceived problem with this situation is fourfold: (i) It implies a loss of information in the sense that more exclusion restrictions are applied than is usually necessary for identification of the overparameterized FA model; (ii) Exclusion restrictions may amount to errors of omission, may make the unrealistic assumption that items are factorially pure (in the population), and may induce bias in estimates of the free parameters (cf. Ferrando & Lorenzo-Seva, 2000; van Prooijen & van der Kloot, 2001). These issues are intricately connected to the well-known and widespread situation of exploratively obtained factor structures not being confirmed by CFA; (iii) Testing as in (4.2) is not possible due to non-nestedness of the model formulations, leading one to rely on heuristical devices or possibly inappropriate information criteria (see Section 4.1.4 below for further discussion); and (iv) Researchers in substantive fields usually have much stronger ideas regarding direction and magnitude of parameter effects that cannot be expressed using exclusion restrictions.

Say we want to incorporate ideas regarding factor structure using inequality restrictions (Sections 4.2.3 and 4.6 discuss in more detail the inequality constraints in which interest is taken) on the parameters in $\mathbf{\Lambda}$. A possible inequality constrained take on the two competing theories regarding factor structure might be expressed as:

$$\boldsymbol{\Lambda}_1 = \begin{bmatrix} \lambda_{11} > |\lambda_{12}| \\ |\lambda_{21}| < -\lambda_{22} \\ \lambda_{31} < \lambda_{32} \\ \lambda_{41} > |\lambda_{42}| \\ \lambda_{51} > \lambda_{52} \\ \lambda_{61} > |\lambda_{62}| \\ \lambda_{71} > |\lambda_{72}| \\ |\lambda_{81}| < -\lambda_{82} \end{bmatrix} \begin{matrix} \text{BMI} \\ \log_e\{\text{trig.}\} \\ \text{HDL chol.} \\ \log_e\{\text{IR}\} \\ \log_e\{\text{GB}\} \\ \log_e\{\text{G2}\} \\ \text{SBP} \\ \text{DBP} \end{matrix}, \quad \boldsymbol{\Lambda}_2 = \begin{bmatrix} |\lambda_{11}| & < -\lambda_{12} \\ |\lambda_{21}| & < -\lambda_{22} \\ \lambda_{31} & < \lambda_{32} \\ \lambda_{41} > .4 & \lambda_{42} < -.4 \\ \lambda_{51} & > \lambda_{52} \\ \lambda_{61} & > |\lambda_{62}| \\ \lambda_{71} & < -\lambda_{72} \\ \lambda_{81} & < -\lambda_{82} \end{bmatrix} \begin{matrix} \text{BMI} \\ \log_e\{\text{trig.}\} \\ \text{HDL chol.} \\ \log_e\{\text{IR}\} \\ \log_e\{\text{GB}\} \\ \log_e\{\text{G2}\} \\ \text{SBP} \\ \text{DBP} \end{matrix}.$$

An interesting strategy would be to incorporate these ideas regarding inequalities into a model bearing only minimal exclusion and fixed-value equality restrictions for identification purposes. We then desire to develop Bayesian strategies as they allow for the direct comparison of such models in the sense of (4.3). ■

4.1.4 Constrained-Model Selection

While not a formal model selection procedure, many applied researchers conduct a multitude of null hypothesis significance tests in order to evaluate a batch of models under consideration. In the setting of covariance structures such tests, as exemplified in (4.1) and (4.2), pair the null hypothesis with the model deemed plausible and proceed by a likelihood ratio (LR) test which, under certain regularity conditions, has a limiting χ^2 distribution. However, when the parameters are subject to functional restrictions the regularity conditions are violated when the true value of the parameter vector is on the boundary of the bounded parameter space, leading the LR test to behave as a mixture of χ^2 distributions (Chernoff, 1954). While most standard packages performing FA and covariance structure modeling in general (as for example EQS®, M*plus*®, LISREL®, IBM® SPSS® Amos) do produce constrained test statistics using techniques mentioned above, referencing these statistics to the right mixture distribution seems to be rare (Savalei & Kolenikov, 2008). Most importantly, the LR test cannot be used for testing and comparing non-nested models.

Related to the LR test is a plethora of (variants on) information criteria (IC) such as the AIC (Akaike, 1973, 1987) and BIC (Akaike, 1978; Schwarz, 1978), which are also provided for by most standard packages. One of their major advantages is that they may handle non-nestedness. However, IC have to be adapted to deal with restricted-model selection. These adaptations are connected to the appropriateness of usage of the restricted maximum likelihood (ML) estimator and finding the configuration of distinct parameters in a restricted parameter space. Anraku (1999) developed an information criterion for restricted-model selection, but it is only appropriate for simple order restrictions in ANOVA-type models.

Alternatively, a Bayesian approach to model selection may be taken. Assume that each model M_b, next to the model-specific collection of unknown parameters $\boldsymbol{\Theta}_b$ and latent data $\boldsymbol{\vartheta}_b$, is characterized by likelihood $L_b(\boldsymbol{\Theta}_b, \boldsymbol{\vartheta}_b; \mathbf{Z})$. Also assume that we have $\pi_b(\boldsymbol{\Theta}_b)g_b(\boldsymbol{\vartheta}_b|\boldsymbol{\Theta}_b)$, $b = 1,\ldots,B$, as the available prior distributions for the unknown parameters. The key quantity in Bayesian model selection is then the marginal likelihood of the data

$$m_b(\mathbf{Z}) = \int L_b(\boldsymbol{\Theta}_b, \boldsymbol{\vartheta}_b; \mathbf{Z})\pi_b(\boldsymbol{\Theta}_b)g_b(\boldsymbol{\vartheta}_b|\boldsymbol{\Theta}_b)\ \partial(\boldsymbol{\Theta}_b, \boldsymbol{\vartheta}_b), \tag{4.5}$$

which constitutes the normalizing constant for the posterior distribution $\pi_b(\boldsymbol{\Theta}_b|\mathbf{Z})$ and is the pivotal quantity in the construction of the Bayes factor (Jeffreys, 1935, 1961):

$$B_{bb'} = \frac{m_b(\mathbf{Z})}{m_{b'}(\mathbf{Z})} = \frac{\displaystyle\int L_b(\boldsymbol{\Theta}_b, \boldsymbol{\vartheta}_b; \mathbf{Z})\pi_b(\boldsymbol{\Theta}_b)g_b(\boldsymbol{\vartheta}_b|\boldsymbol{\Theta}_b)\ \partial(\boldsymbol{\Theta}_b, \boldsymbol{\vartheta}_b)}{\displaystyle\int L_{b'}(\boldsymbol{\Theta}_{b'}, \boldsymbol{\vartheta}_{b'}; \mathbf{Z})\pi_{b'}(\boldsymbol{\Theta}_{b'})g_{b'}(\boldsymbol{\vartheta}_{b'}|\boldsymbol{\Theta}_{b'})\ \partial(\boldsymbol{\Theta}_{b'}, \boldsymbol{\vartheta}_{b'})}. \tag{4.6}$$

The Bayes factor as a model selection criterion has the following advantages (cf. Kass & Raftery, 1995; S. Y. Lee, 2007, Chapter 5): (i) It provides both a measure of evidence against a competing model and a measure of support for the alternative model; (ii) It will not by default favor the alternative model in (very) large samples; (iii) The comparison of any two models does not depend on the assumption that either one is 'true'; (iv) It allows one to take model uncertainty into account, thus providing a consistent quantity for the comparison of a multitude of competing models; (v) It can handle the comparison of both nested and nonnested models.

Notwithstanding, there are several difficulties with (4.5) and thus (4.6): (i) The marginal likelihood will generally not be analytically tractable and especially when the parameter spaces are of high dimension, computation strategies can be challenging (see Kass & Raftery, 1995 and Han & Carlin, 2001 for an overview of normalizing constant computation strategies); (ii) Both the use of improper noninformative and proper but vague priors yield indeterminate answers for (4.6) when the models to be compared are of differing dimension (Jeffreys, 1961). This is undesirable as especially under default prior choices the interpretation of (4.6) as a weighted likelihood ratio is warranted (Berger & Pericchi, 2004); (iii) When not using default options for the priors $\pi_b(\boldsymbol{\Theta}_b)$, their assessment, especially in multivariate cases, may display considerable difficulty and may prove influential in the sense that differing specifications may render differing outcomes; (iv) When $\boldsymbol{\Theta}_b$ is defined as in (4.4) then under certain prior choices (4.6) looses (partially) its ability to act as a natural Ockham's razor as the implicit and automatic penalty for model complexity accounted for in (4.5) may defect for restricted parameter spaces (Mulder et al., 2010).

4.1.5 Goal and Overview

The desire to take restrainment of parameters in the context of FA beyond the purpose of identification and prevention of impermissible estimates and beyond the use of exclusion restrictions in specifying factor structure, as well as the aim to develop a Bayes factor for inequality restricted-model selection merge in the main goal of this work:

Goal. *To develop a strategy which allows one to express factor analytic structure using inequality constraints and to develop a Bayesian model selection criterion to select among competing inequality constrained (factor) structures.*

The strategy consists of choosing as a base model an unrestricted factor model, i.e., a factor model that places only minimal restrictions on the model parameters for achieving rotational uniqueness (and identification). Substantive theory will then not be represented by structural exclusions to express a pre-specified loading pattern, but by imposing inequalities on and between the free parameters in $\mathbf{\Lambda}$. It is then shown that when (i) proper but noninformative priors are chosen that are flat on the parameter space of the parameters on which inequalities are placed; and when (ii) $\mathbf{\Lambda}_b \subset \mathbf{\Lambda}_0 \; \forall b$ where $\mathbf{\Lambda}_0$ denotes a minimally restricted parameter space for purposes of identification; then (4.6) is determinate, its complexity is well-defined, and its computation is strongly simplified as under this framework model fit and complexity are explicitly connected to, respectively, the posterior and prior probability mass satisfying the constraints defining M_b. These provisions thus may curb many of the drawbacks connected to classical CFA and Bayes factors mentioned in Sections 4.1.3 and 4.1.4. Moreover, the conjunction of these provisions constitute a new take on performing a CFA.

Section 4.2 provides for the framework of the unrestricted factor model as well as a precise statement of the inequality constraints in which interest is taken. The intricacies of this model inspire the general development of a Bayes factor for inequality-constrained-model selection in Section 4.3. Section 4.4 sets out to choose priors for inequality constrained FA models that result in the Bayes factor of Section 4.3 to be well-balanced with regard to model fit and complexity in the demarcation of nested and non-nested inequality constrained factor structures. Section 4.5 discusses how standard Bayesian sampling information may be used in efficiently estimating this Bayes factor. In Section 4.6 the developed techniques are illustrated through exploring our motivating example as well as an application from the field of primatology. Section 4.7 concludes with a discussion on the developments and their implications for CFA.

4.2 The Factor Analytic Model

The model we consider here is the unrestricted FA model. After reviewing model structure and assumptions a minimal set of restrictions is given that will ensure global rotational uniqueness and in many cases will imply identifiability. Subsequently we delve into the topic of restrictions beyond identification purposes by reviewing the inequality constraints that will be considered as expressions of theory amenable to model selection. To aid notational simplicity, the model index b will be suppressed where appropriate for the remainder of the text.

4.2.1 Model Structure and Assumptions

Let $\mathbf{Z}^{\mathrm{T}} \equiv [\mathbf{z}_1, \ldots, \mathbf{z}_n]$ define standardized p-variate observation vectors on $i = 1, \ldots, n$ subjects, such that $\mathbf{z}_i^{\mathrm{T}} \equiv [z_{i1}, \ldots, z_{ip}] \in \mathbb{R}^p$ denotes a realization of the random vector $Z_i^{\mathrm{T}} \equiv [Z_{i1}, \ldots, Z_{ip}] \in \mathbb{R}^p$. Also, let $\boldsymbol{\Xi}^{\mathrm{T}} \equiv [\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_n]$ define m-variate vectors of latent factor scores on n subjects with $\boldsymbol{\xi}_i^{\mathrm{T}} \equiv [\xi_{i1}, \ldots, \xi_{im}] \in \mathbb{R}^m$. The FA model states that each random variable Z_i is a linear combination of the latent random variables, such that:

$$\underset{(p\times 1)}{\mathbf{z}_i} = \underset{(p\times 1)}{\boldsymbol{\mu}} + \underset{(p\times m)}{\boldsymbol{\Lambda}} \cdot \underset{(m\times 1)}{\boldsymbol{\xi}_i} + \underset{(p\times 1)}{\boldsymbol{\epsilon}_i} \tag{4.7}$$

with $m < p$. In (4.7) $\boldsymbol{\mu} \in \mathbb{R}^p$ denotes an overall mean vector, the $\boldsymbol{\epsilon}_i \in \mathbb{R}^p$ denote the error measurements, and $\boldsymbol{\Lambda} \in \mathbb{R}^{p\times m}$ is a matrix of factor loadings in which each element λ_{jk} is the loading of the jth variable on the kth factor, $j = 1, \ldots, p$, $k = 1, \ldots, m$.

Remark 4.1. Although we intend to model the sample correlation matrix though usage of standardized data, we do include an intercept in the model. While most studies on FA that utilize standardized or mean-centered data choose not to model $\boldsymbol{\mu}$, we treat it as a nuisance parameter.

The model maintains the following assumptions: (i) $\mathbf{z}_i \perp\!\!\!\perp \mathbf{z}_{i'}, \forall i \neq i'$; (ii) $\mathrm{rank}(\boldsymbol{\Lambda}) = m$; (iii) $\boldsymbol{\epsilon}_i \sim \mathcal{N}_p(\mathbf{0}, \boldsymbol{\Psi})$, with $\boldsymbol{\Psi} \equiv \mathrm{diag}(\psi_{11}, \ldots, \psi_{pp})$, and $\psi_{jj} > 0$; (iv) $\boldsymbol{\xi}_i \sim \mathcal{N}_m(\mathbf{0}, \boldsymbol{\Phi})$; and (v) $\boldsymbol{\xi}_i \perp\!\!\!\perp \boldsymbol{\epsilon}_{i'}, \forall i, i'$. The likelihood for the observations conditional on the realization of $\boldsymbol{\Xi}$ can then be expressed as:

$$\begin{aligned} L(\boldsymbol{\mu}, \boldsymbol{\Lambda}, \boldsymbol{\Xi}, \boldsymbol{\Psi}, \boldsymbol{\Phi}; \mathbf{Z}) &= \prod_{i=1}^{n} f(\mathbf{z}_i | \boldsymbol{\mu}, \boldsymbol{\Lambda}, \boldsymbol{\xi}_i, \boldsymbol{\Psi}, \boldsymbol{\Phi}) \\ &= \prod_{i=1}^{n} (2\pi)^{-\frac{p}{2}} |\boldsymbol{\Psi}|^{-\frac{1}{2}} \exp\left\{ -\frac{1}{2} \boldsymbol{\epsilon}_i^{\mathrm{T}} \boldsymbol{\Psi}^{-1} \boldsymbol{\epsilon}_i \right\}, \end{aligned} \tag{4.8}$$

where $\boldsymbol{\epsilon}_i = \mathbf{z}_i - \boldsymbol{\mu} - \boldsymbol{\Lambda}\boldsymbol{\xi}_i$. The likelihood in this form will be important for the construction of data-augmented conditional distributions for MCMC sampling. Marginalizing over $\boldsymbol{\xi}_i$ we obtain the likelihood of the observed data:

$$\begin{aligned} L(\boldsymbol{\mu}, \boldsymbol{\Lambda}, \boldsymbol{\Psi}, \boldsymbol{\Phi}; \mathbf{Z}) &= \prod_{i=1}^{n} \int f(\mathbf{z}_i | \boldsymbol{\mu}, \boldsymbol{\Lambda}, \boldsymbol{\xi}_i, \boldsymbol{\Psi}, \boldsymbol{\Phi}) g(\boldsymbol{\xi}_i | \boldsymbol{\Phi}) \, \partial \boldsymbol{\xi}_i \\ &= \prod_{i=1}^{n} (2\pi)^{-\frac{p}{2}} |\boldsymbol{\Sigma}|^{-\frac{1}{2}} \exp\left\{ -\frac{1}{2} (\mathbf{z}_i - \boldsymbol{\mu})^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} (\mathbf{z}_i - \boldsymbol{\mu}) \right\}, \end{aligned} \tag{4.9}$$

giving that the factor decomposition constrains the covariance structure of the $\mathbf{z}_i$ to $\boldsymbol{\Sigma} = \boldsymbol{\Lambda}\boldsymbol{\Phi}\boldsymbol{\Lambda}^{\mathrm{T}} + \boldsymbol{\Psi}$. Then, for existence (vi), generally $(p-m)^2 - p - m \geqslant 0$, simply stating that the number of nonredundant elements in the sample correlation matrix $\mathbf{S}$ must be greater than or equal to the number of freely estimable parameters in $\boldsymbol{\Sigma}$, which places an upper bound on m.

4.2.2 Uniqueness and Identifiability

It is well known that for given $\mathbf{\Lambda}$ and $\mathbf{\Psi}$, the former is defined uniquely only up to rotation. Correspondingly the FA literature has focussed mainly on identification of $\mathbf{\Psi}$, the main result of which states that $\mathbf{\Psi}$ is locally identified for almost every $\mathbf{\Lambda}$ and $\mathbf{\Psi}$ given that conditions (ii), (iv) and (vi) mentioned in Section 4.2.1 are satisfied (e.g., Bekker et al., 1994). Taking this to be the case, the FA model then copes with two inherent indeterminacies of the parameters, being: An undefined metric for the latent factors and rotational indeterminacy of the factor solution.

Regarding the latter, assume that $\mathbf{R} \in \mathbb{R}^{m \times m}$ is an arbitrary nonsingular matrix. The implied covariance structure of the data can then be stated as

$$\mathbf{\Sigma} = \mathbf{\Lambda\Phi\Lambda}^{\mathrm{T}} + \mathbf{\Psi} = (\mathbf{\Lambda R})[\mathbf{R}^{-1}\mathbf{\Phi}(\mathbf{R}^{\mathrm{T}})^{-1}](\mathbf{\Lambda R})^{\mathrm{T}} + \mathbf{\Psi}. \tag{4.10}$$

Equation (4.10) implies that, given $\mathbf{\Psi}$, there is an infinite number of alternative matrices $\mathbf{\Lambda R}$ and $\mathbf{R}^{-1}\mathbf{\Phi}(\mathbf{R}^{\mathrm{T}})^{-1}$ that generate a correlation structure equivalent to $\mathbf{\Lambda}$ and $\mathbf{\Phi}$. As $\mathbf{R}$ has m^2 independent elements, a minimum of m^2 restrictions has to be imposed on either $\mathbf{\Lambda}$ or $\mathbf{\Lambda}$ and $\mathbf{\Phi}$ to obtain a rotationally unique solution (hence the structure of existence condition (vi)).

Consider the following reformulation of and amendments to Jöreskog's conditions (1979) for metric and rotational uniqueness (Peeters, 2012):

- C1 Let $\mathbf{\Lambda}$ have at least $m-1$ fixed zeroes in each column;
- C2 Let $\mathrm{rank}(\mathbf{\Lambda}^{[k]}) = m-1$, where $\mathbf{\Lambda}^{[k]}$, $k = 1, \ldots, m$, is the submatrix of $\mathbf{\Lambda}$, consisting of the rows of $\mathbf{\Lambda}$ which have fixed zero elements in the kth column with these zeroes deleted;
- C3 Let $\mathbf{\Phi}$ be a symmetric positive definite matrix with $\mathrm{diag}(\mathbf{\Phi}) = \mathbf{I}_m$;
- C4 Let in each column of $\mathbf{\Lambda}$ one parameter non-fixed by condition C1 be polarity truncated to take only positive or negative values, that is: In each column of $\mathbf{\Lambda}$ one element is to adopt either strict positivity ($\lambda_{jk} > 0$), or strict negativity ($-\lambda_{jk} > 0$).

Proposition 4.1 (Peeters, 2012). *Let the mapping* $\delta : \{\mathbf{\Lambda}, \mathbf{\Phi}\} \longrightarrow \{\mathbf{\Lambda}^{\ddagger}, \mathbf{\Phi}^{\ddagger}\}$ *be defined by* $\mathbf{\Lambda}^{\ddagger} = \mathbf{\Lambda R}$ *and* $\mathbf{\Phi}^{\ddagger} = \mathbf{R}^{-1}\mathbf{\Phi}(\mathbf{R}^{\mathrm{T}})^{-1}$*, where* $\mathbf{R} \in \mathbb{R}^{m \times m}$ *denotes an arbitrary nonsingular matrix. If conditions* C1-C4 *hold, then* $\mathbf{R} = \mathbf{I}_m$.

Remark 4.2. Proposition 4.1 implies that the conjunction of conditions C1-C4 is sufficient to provide global rotational uniqueness of the factor solution. Global uniqueness subsequently implies unimodality of the parameter space (under given model assumptions) necessary for meaningful MCMC-sampling of the FA model. The metric for the latent factors is set by condition C3, implying that $\mathbf{\Phi}$ is a correlation matrix.

Remark 4.3. Rotational uniqueness of $\mathbf{\Lambda}$ will not guarantee identifiability of the FA model (Bollen & Jöreskog, 1985), as underidentification of $\mathbf{\Psi}$ may imply underidentification of $\mathbf{\Lambda}$. However, if regularity assumptions (ii), (iii), and (vi) from Section 4.2.1 hold, then conditions C1-C4 will, next to global rotational uniqueness, also provide identifiability. When unsure if identifiability is obtained, one could endeavor on algebraically checking (local) identification utilizing the Wald rank rule (e.g., Bekker et al., 1994).

Remark 4.4. The imposition of structural restrictions in $\mathbf{\Lambda}$ brings into question the ordering of the observed variables. Theoretically, this ordering is irrelevant when m is known as it is easily verified that $\mathbf{P}\mathbf{z}_i$, the reordering of $\mathbf{z}_i$ by any permutation matrix $\mathbf{P}$, does not alter the implied covariance structure of the observed data except for permutations in the rows and columns, that is: $\mathbf{P}\mathbf{\Sigma}\mathbf{P}^{\mathrm{T}}$. Here, we assume m to be known intrinsically or through assessment.

Remark 4.5. The conditions provide a means to design an unrestricted solution to the (Bayesian) CFA model. Unrestricted solutions correspond to exploratory factor analysis (EFA) in the sense that only minimal restrictions are placed on the model to achieve a (global) rotationally unique solution for m factors. As such, an unrestricted solution for m common factors does not restrict the factor space and will yield an optimal fit for any model with m factors (Mulaik, 2010, Section 15.4). In the EFA tradition this is usually achieved by requiring $\mathbf{\Phi} = \mathbf{I}_m$ and $\mathbf{\Lambda}^{\mathrm{T}}\mathbf{\Psi}^{-1}\mathbf{\Lambda}$ be diagonal accompanied by an order condition on the diagonal elements. Such restrictions are mere convenience for estimation purposes and whence estimation is settled traditionally rotations are employed to enhance interpretation of the solution. Through conditions C1-C4 it is possible to formulate a set of restrictions that bear meaning in terms of interpretation of the solution and as such cancel the interpretative need for post-hoc rotation.

Definition 4.1 (Unrestricted Confirmatory Factor Model). *An unrestricted confirmatory factor model (UCFM) is a FA model that places only minimal restrictions on* $\mathbf{\Lambda}$ *and* $\mathbf{\Phi}$ *for achieving global rotational uniqueness of the factor solution, with the restrictions chosen such that they convey preconceived theoretical meaning and thus render unnecessary post-hoc rotation of the solution for interpretation purposes. In the remainder parameter exclusions and inequalities for obtaining rotational uniqueness will be termed 'restrictions', while inequalities imposed on the free parameters embedded within the UCFM will be termed 'constraints' (Section 4.2.3).*

Example 4.1 (Continued). In our take on CFA substantive theory will not be represented by structural equalities to express a pre-specified factor loading pattern, but by imposing a consistent set of inequality constraints on and between the free parameters in the model. Identification is then provided for by ensuring that the set of inequality constraints is embedded within a certain minimal condition for uniqueness, i.e., within the UCFM. Abiding C1-C4, we choose the following minimal restrictions on $\mathbf{\Lambda}$ for obtaining metric and global rotational uniqueness:

$$\mathbf{\Lambda}_0 = \begin{bmatrix} \lambda_{11} & \lambda_{12} \\ \lambda_{21} & \lambda_{22} \\ \lambda_{31} = 0 & \lambda_{32} > 0 \\ \lambda_{41} & \lambda_{42} \\ \lambda_{51} > 0 & \lambda_{52} = 0 \\ \lambda_{61} & \lambda_{62} \\ \lambda_{71} & \lambda_{72} \\ \lambda_{81} & \lambda_{82} \end{bmatrix} \begin{matrix} \text{BMI} \\ \log_e\{\text{trig.}\} \\ \text{HDL chol.} \\ \log_e\{\text{IR}\} \\ \log_e\{\text{GB}\} \\ \log_e\{\text{G2}\} \\ \text{SBP} \\ \text{DBP} \end{matrix} .$$

These constraints are chosen for the following reasons. First, we believe HDL chol. to have a large loading on the second (lipid metabolism) factor while having a small loading on the first factor, and we believe $\log_e\{\text{GB}\}$ to have a large loading on the first (glucose metabolism) factor while having a small loading on the second factor. These variables then serve as an indicator of the respective factors. It is thus reasonable (also in terms of efficiency of the MCMC implementation) to specify $\{\lambda_{31}, \lambda_{52}\} = 0$ and $\{\lambda_{32}, \lambda_{51}\} > 0$. The polarity of the columns is then tied to strict positivity of λ_{32} and λ_{51}. Second, the chosen minimal restrictions comply with all competing inequality constrained formulations of factor structure to be assessed (Section 4.6.1). ■

4.2.3 Constraints Considered

Let $\mathbf{\Lambda}_0$ denote a $\mathbf{\Lambda}$ abiding (at least) the minimal conditions for uniqueness based on, say, t restrictions. The index t counts the number of restrictions under conditions C1 and C4. Now let $\boldsymbol{\lambda}_f \in \mathbb{R}^{Q=pm-t}$ be a vectorization of the free parameters in $\mathbf{\Lambda}_0$ of known factor dimensionality m. We then consider linear inequality constraints of the form:

$$\mathbf{\Omega}_b \boldsymbol{\lambda}_f - \boldsymbol{\alpha}_b > \mathbf{0} \qquad \forall b, \tag{4.11}$$

where $\mathbf{\Omega}_b \in \mathbb{R}^{L_b \times Q}$ is a constraints matrix representing a system of $l = 1, \ldots, L_b$ linear restrictions and $\boldsymbol{\alpha}_b$ denotes a real fixed-value vector of length L_b. The conjunction of $\mathbf{\Omega}_b$ and $\boldsymbol{\alpha}_b$ on $\boldsymbol{\lambda}_f$ defines model M_b. Many forms of linear inequality constraints are contained in (4.11) (e.g., Tsonaka & Moustaki, 2007). Our interest is in:

i. Fixed value constraints, where a factor loading is set to surpass a certain fixed value, i.e., $\lambda_q - \alpha_l > 0$;
ii. Positivity constraints, where a certain factor loading is set to surpass an other factor loading, i.e., $\lambda_q - \lambda_{q'} > 0$;
iii. Negativity constraints, where a factor loading is set to surpass the negative of an other factor loading, i.e, $\lambda_q + \lambda_{q'} > 0$; and, as special cases of (i), (ii), and (iii)
iv. Approximate equality constraints, where a factor loading is set to approximately equal constant e with given bound ε, or where two factor loadings are set to be approximately equal with given bound ε, i.e.,

$$|\lambda_q - e| < \varepsilon \Rightarrow \begin{cases} \lambda_q - (e-\varepsilon) > 0 \\ -\lambda_q + e + \varepsilon > 0 \end{cases}, \text{ and } |\lambda_q - \lambda_{q'}| < \varepsilon \Rightarrow \begin{cases} \lambda_q - \lambda_{q'} + \varepsilon > 0 \\ \lambda_{q'} - \lambda_q + \varepsilon > 0 \end{cases},$$

such that each row of $\mathbf{\Omega}_b$ will be a permutation of either $(1, 0, \ldots, 0)$, $(-1, 0, \ldots, 0)$, $(1, 1, 0, \ldots, 0)$, $(-1, 1, 0, \ldots, 0)$, or $(-1, -1, 0, \ldots, 0)$. Our reason for reviewing these constraints is twofold. First, as will be elaborated in Section 4.4, these constraints are adequately supported by the prior distribution. Second, these constraints prove to be most substantive from a theoretical point of view as they directly allow to express the direction and (relative) strength of factor loadings (Section 4.6).

4.3 A Bayes Factor for Constrained-Model Selection

The intricacies of the FA model in conjunction with the system of constraints in (4.11) inspire the development of a Bayes factor for the selection of inequality constrained multivariate parametric models. While notation will closely connect to the FA model, it should be clear that the development of the argument is in no way restricted to FA or Gaussian models and furthers certain previous efforts by, for example, Klugkist and Hoijtink (2007).

Pivotal in the formulation of a computationally friendly Bayes factor for model selection under inequality constraints is the nesting of competing inequality constrained models within an encompassing model (Berger & Pericchi, 1996; Klugkist & Hoijtink, 2007). The encompassing model will be given here by the model with minimal restrictions for uniqueness, i.e., the UCFM forms the base model. Inequality constraints on the parameter space enter the model through the prior distribution. The prior then taking the form of the product of a conventional (noninformative) prior distribution on the base model and an indicator function representing the inequality constraints (Geweke, 1986). For the development of our Bayes factor we need to state that in the remainder we will model the prior for $\mathbf{\Lambda}_0$ as dependent on $\mathbf{\Phi}$. The reason for doing so will become clear in Section 4.4.1, where it will also be exemplified how our approach towards Bayesian inequality-constrained-model selection differs from many previous efforts.

Proposition 4.2. *Define* $\mathbf{\Lambda}_b \equiv \{\mathbf{\Lambda}_0 : \mathbf{\Omega}_b \boldsymbol{\lambda}_f - \boldsymbol{\alpha}_b > \mathbf{0}\}$, *for which we assume that the system* $\mathbf{\Omega}_b \boldsymbol{\lambda}_f - \boldsymbol{\alpha}_b$ *is non-contradictory and does not impair or alter the conditions* C1-C4 *that define* $\mathbf{\Lambda}_0$, *giving that* $\mathbf{\Lambda}_b \subset \mathbf{\Lambda}_0$. *Let the prior under encompassing model* M_0 *for the set of parameters in the factor model be* $\pi(\boldsymbol{\mu}, \mathbf{\Psi})\pi_0(\mathbf{\Lambda}_0|\mathbf{\Phi})g(\mathbf{\Xi}|\mathbf{\Phi})\pi(\mathbf{\Phi})$, *such that the prior distribution under any constrained model* M_b *is* $\pi(\boldsymbol{\mu}, \mathbf{\Psi})\pi_b(\mathbf{\Lambda}_b|\mathbf{\Phi})g(\mathbf{\Xi}|\mathbf{\Phi})\pi(\mathbf{\Phi}) \propto \pi(\boldsymbol{\mu}, \mathbf{\Psi})\pi_0(\mathbf{\Lambda}_0|\mathbf{\Phi})\mathbb{1}_{\{\mathbf{\Omega}_b \boldsymbol{\lambda}_f - \boldsymbol{\alpha}_b > \mathbf{0}\}} g(\mathbf{\Xi}|\mathbf{\Phi})\pi(\mathbf{\Phi})$, *where* $\mathbb{1}_{\{.\}}$ *denotes an indicator function representing the system of inequality constraints defining* $\mathbf{\Lambda}_b$. *Now, assuming propriety of* $\pi_0(\boldsymbol{\mu}, \mathbf{\Lambda}_0, \mathbf{\Xi}, \mathbf{\Psi}, \mathbf{\Phi}|\mathbf{Z})$, *the Bayes factor* B_{b0} *reduces to the posterior probability mass satisfying the system of constraints that defines* $\mathbf{\Lambda}_b$ *over the prior probability mass satisfying the system of constraints that defines* $\mathbf{\Lambda}_b$.

Proof. The Bayes factor B_{b0} of constrained model M_b to encompassing model M_0, is written as

$$\frac{\int L(\boldsymbol{\mu}, \boldsymbol{\Lambda}_0, \boldsymbol{\Xi}, \boldsymbol{\Psi}, \boldsymbol{\Phi}; \mathbf{Z})\pi(\boldsymbol{\mu}, \boldsymbol{\Psi})\pi_b(\boldsymbol{\Lambda}_b|\boldsymbol{\Phi})g(\boldsymbol{\Xi}|\boldsymbol{\Phi})\pi(\boldsymbol{\Phi})\,\partial(\boldsymbol{\mu}, \boldsymbol{\Lambda}_b, \boldsymbol{\Xi}, \boldsymbol{\Psi}, \boldsymbol{\Phi})}{\int L(\boldsymbol{\mu}, \boldsymbol{\Lambda}_0, \boldsymbol{\Xi}, \boldsymbol{\Psi}, \boldsymbol{\Phi}; \mathbf{Z})\pi(\boldsymbol{\mu}, \boldsymbol{\Psi})\pi_0(\boldsymbol{\Lambda}_0|\boldsymbol{\Phi})g(\boldsymbol{\Xi}|\boldsymbol{\Phi})\pi(\boldsymbol{\Phi})\,\partial(\boldsymbol{\mu}, \boldsymbol{\Lambda}_0, \boldsymbol{\Xi}, \boldsymbol{\Psi}, \boldsymbol{\Phi})}. \quad (4.12)$$

Using the basic marginal identity (Besag, 1989; Chib, 1995) we may express (4.12) equivalently as follows:

$$\frac{L(\boldsymbol{\mu}, \boldsymbol{\Lambda}_0, \boldsymbol{\Xi}, \boldsymbol{\Psi}, \boldsymbol{\Phi}; \mathbf{Z})\pi(\boldsymbol{\mu}, \boldsymbol{\Psi})\pi_b(\boldsymbol{\Lambda}_b|\boldsymbol{\Phi})g(\boldsymbol{\Xi}|\boldsymbol{\Phi})\pi(\boldsymbol{\Phi})/\pi_b(\boldsymbol{\mu}, \boldsymbol{\Lambda}_b, \boldsymbol{\Xi}, \boldsymbol{\Psi}, \boldsymbol{\Phi}|\mathbf{Z})}{L(\boldsymbol{\mu}, \boldsymbol{\Lambda}_0, \boldsymbol{\Xi}, \boldsymbol{\Psi}, \boldsymbol{\Phi}; \mathbf{Z})\pi(\boldsymbol{\mu}, \boldsymbol{\Psi})\pi_0(\boldsymbol{\Lambda}_0|\boldsymbol{\Phi})g(\boldsymbol{\Xi}|\boldsymbol{\Phi})\pi(\boldsymbol{\Phi})/\pi_0(\boldsymbol{\mu}, \boldsymbol{\Lambda}_0, \boldsymbol{\Xi}, \boldsymbol{\Psi}, \boldsymbol{\Phi}|\mathbf{Z})}.$$

For any given value of $\{\boldsymbol{\mu}, \boldsymbol{\Lambda}_0, \boldsymbol{\Xi}, \boldsymbol{\Psi}, \boldsymbol{\Phi}\}$, say $\{\boldsymbol{\mu}^*, \boldsymbol{\Lambda}^*, \boldsymbol{\Xi}^*, \boldsymbol{\Psi}^*, \boldsymbol{\Phi}^*\}$, that is admissible under the system of constraints $\boldsymbol{\Omega}_b\boldsymbol{\lambda}_f - \boldsymbol{\alpha}_b > \mathbf{0}$, clearly $\{\boldsymbol{\mu}^*, \boldsymbol{\Lambda}^*, \boldsymbol{\Xi}^*, \boldsymbol{\Psi}^*, \boldsymbol{\Phi}^*\} \in M_b \cap M_0$. For any such value we then have:

$$\frac{L(\boldsymbol{\mu}^*, \boldsymbol{\Lambda}^*, \boldsymbol{\Xi}^*, \boldsymbol{\Psi}^*, \boldsymbol{\Phi}^*; \mathbf{Z})\pi(\boldsymbol{\mu}^*, \boldsymbol{\Psi}^*)\pi_b(\boldsymbol{\Lambda}^*|\boldsymbol{\Phi}^*)g(\boldsymbol{\Xi}^*|\boldsymbol{\Phi}^*)\pi(\boldsymbol{\Phi}^*)}{L(\boldsymbol{\mu}^*, \boldsymbol{\Lambda}^*, \boldsymbol{\Xi}^*, \boldsymbol{\Psi}^*, \boldsymbol{\Phi}^*; \mathbf{Z})\pi(\boldsymbol{\mu}^*, \boldsymbol{\Psi}^*)\pi_0(\boldsymbol{\Lambda}^*|\boldsymbol{\Phi}^*)g(\boldsymbol{\Xi}^*|\boldsymbol{\Phi}^*)\pi(\boldsymbol{\Phi}^*)} \cdot \frac{\pi_0(\boldsymbol{\mu}^*, \boldsymbol{\Lambda}^*, \boldsymbol{\Xi}^*, \boldsymbol{\Psi}^*, \boldsymbol{\Phi}^*|\mathbf{Z})}{\pi_b(\boldsymbol{\mu}^*, \boldsymbol{\Lambda}^*, \boldsymbol{\Xi}^*, \boldsymbol{\Psi}^*, \boldsymbol{\Phi}^*|\mathbf{Z})}. \quad (4.13)$$

Dividing out terms, expression (4.13) reduces to

$$\frac{\pi_0(\boldsymbol{\mu}^*, \boldsymbol{\Lambda}^*, \boldsymbol{\Xi}^*, \boldsymbol{\Psi}^*, \boldsymbol{\Phi}^*|\mathbf{Z})}{\pi_b(\boldsymbol{\mu}^*, \boldsymbol{\Lambda}^*, \boldsymbol{\Xi}^*, \boldsymbol{\Psi}^*, \boldsymbol{\Phi}^*|\mathbf{Z})} \cdot \frac{\pi_b(\boldsymbol{\Lambda}^*|\boldsymbol{\Phi}^*)}{\pi_0(\boldsymbol{\Lambda}^*|\boldsymbol{\Phi}^*)}. \quad (4.14)$$

Now, notice

$$\begin{aligned}\pi_b(\boldsymbol{\mu}^*, \boldsymbol{\Lambda}^*, \boldsymbol{\Xi}^*, \boldsymbol{\Psi}^*, \boldsymbol{\Phi}^*|\mathbf{Z}) &= \pi_0(\boldsymbol{\mu}^*, \boldsymbol{\Lambda}^*, \boldsymbol{\Xi}^*, \boldsymbol{\Psi}^*, \boldsymbol{\Phi}^*|\mathbf{Z}) \\ &\cdot \left[\int\int_{\{\boldsymbol{\Lambda}_0:\boldsymbol{\Omega}_b\boldsymbol{\lambda}_f-\boldsymbol{\alpha}_b>\mathbf{0}\}} \pi_0(\boldsymbol{\mu}, \boldsymbol{\Lambda}_0, \boldsymbol{\Xi}, \boldsymbol{\Psi}, \boldsymbol{\Phi}|\mathbf{Z})\,\partial\boldsymbol{\Lambda}_0\,\partial(\boldsymbol{\mu}, \boldsymbol{\Xi}, \boldsymbol{\Psi}, \boldsymbol{\Phi})\right]^{-1}, \quad (4.15)\end{aligned}$$

and

$$\pi_b(\boldsymbol{\Lambda}^*|\boldsymbol{\Phi}^*) = \pi_0(\boldsymbol{\Lambda}^*|\boldsymbol{\Phi}^*) \cdot \left[\int_{\{\boldsymbol{\Lambda}_0:\boldsymbol{\Omega}_b\boldsymbol{\lambda}_f-\boldsymbol{\alpha}_b>\mathbf{0}\}} \pi_0(\boldsymbol{\Lambda}_0|\boldsymbol{\Phi})\,\partial\boldsymbol{\Lambda}_0\right]^{-1}. \quad (4.16)$$

Substituting (4.15) and (4.16) in (4.14) we obtain

$$B_{b0} = \frac{\int\int_{\{\boldsymbol{\Lambda}_0:\boldsymbol{\Omega}_b\boldsymbol{\lambda}_f-\boldsymbol{\alpha}_b>\mathbf{0}\}} \pi_0(\boldsymbol{\mu}, \boldsymbol{\Lambda}_0, \boldsymbol{\Xi}, \boldsymbol{\Psi}, \boldsymbol{\Phi}|\mathbf{Z})\,\partial\boldsymbol{\Lambda}_0\,\partial(\boldsymbol{\mu}, \boldsymbol{\Xi}, \boldsymbol{\Psi}, \boldsymbol{\Phi})}{\int_{\{\boldsymbol{\Lambda}_0:\boldsymbol{\Omega}_b\boldsymbol{\lambda}_f-\boldsymbol{\alpha}_b>\mathbf{0}\}} \pi_0(\boldsymbol{\Lambda}_0|\boldsymbol{\Phi})\,\partial\boldsymbol{\Lambda}_0} \equiv \frac{f_b}{\omega_b}, \quad (4.17)$$

and the proposition follows. □

Remark 4.6. The Proposition gives that one may estimate the Bayes factor of a constrained factor model nested in a model with (minimal) uniqueness restrictions on the parameter space by a simple numerical procedure (Section 4.5): Counting the number of times an appropriate sampler for the parametric model at hand visits the permissable posterior and prior spaces defined by the system of inequality constraints.

Remark 4.7. The quantity B_{b0} has many of the properties that regular Bayes factors on non-truncated parameters have: $B_{bb} = 1$; $B_{0b} = B_{b0}^{-1}$; and $B_{bb'} = B_{b0}/B_{b'0}$. However, $0 \leqslant B_{b0} \leqslant \omega_b^{-1}$, while $B_{bb'} \geqslant 0$. The upper bound of B_{b0} should be obvious: f_b indicates the proportion of the posterior distribution satisfying the system of inequality constraints, the maximum value of which is unity, giving that the upper bound for B_{b0} is marked by the inverse of ω_b. Implications for the interpretation of B_{b0} will be discussed in Section 4.6.

The numerator in (4.17) may be seen as an explicit expression of model fit in case of constrained to base model comparison; when more constraints are supported by the data, more posterior probability mass will be located in the feasible region defined by the system of constraints and f_b will attain a larger value. The complexity of an inequality constrained model is then defined by the prior mass abiding the system of inequality constraints. Next, we will see to a choice of prior that will adequately represent complexity in an inequality constrained FA model.

4.4 Elements of Prior Specification and Complexity

It is clear that for obtainment of a well-defined Bayes factor B_{b0}, we need to explicitly connect the probability mass of the prior to a complexity criterion. We first delve into elements of prior specification. Proposition 4.2 gives that we only have to specify the prior for model M_0. The prior for any constrained model M_b directly follows from the assessment of the truncations on the prior for M_0. The specified prior connects to conditional distributions needed for the numerical machinery in Section 4.5. Our choice of prior on $\mathbf{\Lambda}_0$ accounts for reviewing correlation matrices and is subsequently shown to directly connect to a complexity criterion that, for many types of inequalities, conveys no information regarding the system (4.11), such that B_{b0} may be termed objective. For notational convenience we drop model index 0 where appropriate.

4.4.1 Elements of Prior Specification

Let μ_j be the jth scalar element of $\boldsymbol{\mu}$. We assume independency between priors for μ_j, λ_{jk}, and ψ_{jj}, such that our total prior density for model M_0 may be stated as:

$$\pi(\boldsymbol{\mu}, \mathbf{\Psi})\pi(\mathbf{\Lambda}_0|\mathbf{\Phi})g(\mathbf{\Xi}|\mathbf{\Phi})\pi(\mathbf{\Phi}) = \pi(\boldsymbol{\mu})\pi(\mathbf{\Psi})\pi(\mathbf{\Lambda}_0)\mathbb{1}_{\left\{\mathbf{I}_p - \text{diag}\left(\mathbf{\Lambda}_0\mathbf{\Phi}\mathbf{\Lambda}_0^{\text{T}}\right)>0\right\}}g(\mathbf{\Xi}|\mathbf{\Phi})\pi(\mathbf{\Phi}). \tag{4.18}$$

The prior on $\boldsymbol{\mu}$ is defined such that $\pi(\mu_j) \propto 1$ over the Lebesgue measure on $(-\infty, \infty)$. While this prior is improper it will not pose problems in terms of causing indeterminable Bayes factors, as it is a common parameter of equivalent dimension in all models under consideration.

Similarly the prior over the unrestricted elements λ_{jk} in $\boldsymbol{\Lambda}_0$ is defined such that $\pi(\lambda_{jk}) \propto 1$ over the Lebesgue measure on $(-1, 1)$ as these elements are bounded theoretically by plus and minus unity in correlation structure modeling. Let λ^p_{jk} denote loadings involved in a polarity truncation. Then we define $\pi((-)\lambda^p_{jk}) \propto 1$ over the Lebesgue measure on $(0, 1)$. The prior on $\boldsymbol{\Lambda}_0$ is dependent on $\boldsymbol{\Phi}$ through the indicator function $\mathbb{1}_{\{\cdot\}}$. The indicator function is specified to account for the identification restriction imposed by the fact that we consider $\boldsymbol{\Sigma}$ as a correlation matrix: It needs to be ensured that $\mathbf{I}_p - \operatorname{diag}\left(\boldsymbol{\Lambda}_0\boldsymbol{\Phi}\boldsymbol{\Lambda}_0^{\mathrm{T}}\right) > 0$ for $\boldsymbol{\Psi}$ to be positive definite. The prior on $\boldsymbol{\Lambda}_0$ including the indicator function, will prove important in the definition of ω_b (Section 4.4.3).

As exemplified by Martin and McDonald (1975) and Lopes and West (2004) a prior that decays to zero at the origin is needed on the ψ_{jj} elements in order to induce proper posteriors as this prevents the posterior from placing infinite mass at $\psi_{jj} = 0$ for some j. A convenient choice is $\psi_{jj} \sim \mathcal{IG}(\nu/2, \nu d/2)$, where $\mathcal{IG}(\cdot,\cdot)$ denotes the inverse gamma distribution. Shape and scale will be chosen diffuse but proper in applications (Sections 4.6.1 and 4.6.2), i.e., as hyperparameters for the $\pi(\psi_{jj})$ we choose $\nu = 3$ and $\nu d = 1/2$. This setting results in a weakly informative or diffuse prior for the ψ_{jj} elements that decays to zero sufficiently rapid as ψ_{jj} tends to zero, and places the prior mean in the middle of the interval (0,1].

The prior $\pi(\boldsymbol{\Xi}|\boldsymbol{\Phi})$ stems from assumption (iv) of Section 4.2.1, stating that $\boldsymbol{\xi}_i \sim \mathcal{N}_m(\mathbf{0}, \boldsymbol{\Phi})$. The prior on the hyperparameter $\boldsymbol{\Phi}$ will be modeled as $\pi(\boldsymbol{\Phi}) = \mathcal{IW}_m(\boldsymbol{\Upsilon}, \tau)$, where $\mathcal{IW}_m(\cdot,\cdot)$ denotes the inverse Wishart distribution. Again, we will choose shape and scale diffuse but proper in applications, i.e., $\boldsymbol{\Upsilon} = \mathbf{I}_m$ and $\tau = m + 2$. These specifications state that *a priori* we deem positive covariances between the factors to be as likely as negative covariances.

Remark 4.8. From the prior specification it is clear that we sample $\boldsymbol{\Phi}$ as a covariance matrix. However, in specifying uniqueness conditions, and in keeping with the modeling of a standardized FA covariance structure, we desire $\boldsymbol{\Phi}$ to be a correlation matrix. While there are prior specifications available for $\boldsymbol{\Phi}$ as a correlation matrix (e.g., Ansari & Jedidi, 2000), these have to be truncated by a subset of the hypercube $[-1, 1]^{m(m-1)/2}$, and as such imply less straightforward MCMC samplers. Here we opt for a more convenient road. Our prior on $\boldsymbol{\Phi}$ ensures that we sample positive definite covariance matrices. The sample covariance matrix $\boldsymbol{\Phi}$ contains all the information for a transformation to a positive definite correlation matrix, i.e., $\phi^s_{kk'} = \phi_{kk'}/\sqrt{\phi_{kk}}\cdot\sqrt{\phi_{k'k'}}$, where $\phi^s_{kk'}$ denotes the standardized covariance between the kth and k'th latent variable. This strategy allows retainment of straightforward MCMC samplers. From this point, we will use $\boldsymbol{\Phi}^s$ to denote the desired correlation version of $\boldsymbol{\Phi}$.

Remark 4.9. Our prior choices indicate that our strategy towards inequality-constrained-model selection differs from previous efforts, such as Klugkist and

Hoijtink (2007), and Mulder, Hoijtink, and Klugkist (2010). These studies revolve around inequality constraints on (multivariate) mean structures and employ diffuse proper priors or training-sample based posterior priors on the parameters of interest. However, we review regression parameters (the factor loadings are essentially regression parameters). Regression coefficients for standardized data are bound to $(-1, 1)$. A natural choice of prior for such parameters is then a uniform distribution on $(-1, 1)$, being both proper and noninformative. Complex prior specifications may then be avoided as the nuisance parameters can be treated with vague but proper prior formulations. This strategy complies (in part) with recommendations by Savage (1962).

4.4.2 Conditional Distributions

Now, define $\mathbf{z}_j$, $\mathbf{\Lambda}_{0j}$, and $\mathbf{1}_n$ to be the jth column of $\mathbf{Z}$, the jth row of $\mathbf{\Lambda}_0$, and an n-dimensional unit vector, respectively. The conditional posterior distributions, based on the priors from the previous section, are then given as below.

For the conditional distribution of $\boldsymbol{\mu}$, we find

$$\pi(\boldsymbol{\mu}|\mathbf{Z}, \mathbf{\Lambda}_0, \mathbf{\Xi}, \mathbf{\Psi}) \overset{d}{=} \mathcal{N}_p(\tilde{\boldsymbol{\mu}}, \mathbf{\Sigma}_{\boldsymbol{\mu}}), \tag{4.19}$$

where $\tilde{\boldsymbol{\mu}} = \bar{\mathbf{z}} - \mathbf{\Lambda}_0\bar{\boldsymbol{\xi}}$, $\mathbf{\Sigma}_{\boldsymbol{\mu}} = \mathbf{\Psi}n^{-1}$, $\bar{\mathbf{z}} = n^{-1}\sum_{i=1}^{n}\mathbf{z}_i$, and $\bar{\boldsymbol{\xi}} = n^{-1}\sum_{i=1}^{n}\boldsymbol{\xi}_i$. The rows of $\mathbf{\Lambda}_0$ are independent such that we need to find the conditional distribution $\pi(\mathbf{\Lambda}_{0j}^{\mathrm{T}}|\mathbf{z}_j, \mu_j, \mathbf{\Xi}, \psi_{jj}, \mathbf{\Phi}^s)$, $j = 1, \ldots, p$. This conditional is slightly different for $\mathbf{\Lambda}_{0j}^{\mathrm{T}}$ with certain restrictions relative to $\mathbf{\Lambda}_{0j}^{\mathrm{T}}$ with only free parameters. We follow Lee (2007) in developing general notation capturing both situations. Let $\mathbf{c}_j$ be a row vector of dimension m corresponding to $\mathbf{\Lambda}_{0j}$, indicating $c_{jk} = 0$ if λ_{jk} is a fixed parameter and indicating $c_{jk} = 1$ if λ_{jk} is a free parameter or a parameter involved in a polarity truncation. In addition, let $r_j = c_{j1} + \ldots + c_{jm}$. Moreover, let $\mathbf{\Lambda}_{0j}^{\star}$ and $\mathbf{\Xi}_j^{\star}$ denote the r_j-dimensional row vector containing the unknown elements in $\mathbf{\Lambda}_{0j}$ and the $(n \times r_j)$ submatrix of $\mathbf{\Xi}$ for which the columns corresponding to $c_{jk} = 0$ are deleted, respectively. We then find

$$\begin{aligned}&\pi(\mathbf{\Lambda}_{0j}^{\star\mathrm{T}}|\mathbf{z}_j, \mu_j, \mathbf{\Xi}_j^{\star}, \psi_{jj}, \mathbf{\Phi}^s)\\ &\qquad \overset{d}{=} \mathcal{N}_m(\tilde{\mathbf{\Lambda}}_{0j}^{\star}, \mathbf{\Sigma}_{\mathbf{\Lambda}_{0j}^{\star}})\mathbb{1}_{\left\{0<(-)\lambda_{jk}^{p}<1 \;\cap\; -1<\lambda_{jk'}<1 \;\forall k'\neq k \;\cap\; 1-\mathbf{\Lambda}_{0j}\mathbf{\Phi}^s\mathbf{\Lambda}_{0j}^{\mathrm{T}}>0\right\}},\end{aligned} \tag{4.20}$$

for rows which contain a loading involved in a polarity truncation, and

$$\pi(\mathbf{\Lambda}_{0j}^{\star\mathrm{T}}|\mathbf{z}_j, \mu_j, \mathbf{\Xi}_j^{\star}, \psi_{jj}, \mathbf{\Phi}^s) \overset{d}{=} \mathcal{N}_m(\tilde{\mathbf{\Lambda}}_{0j}^{\star}, \mathbf{\Sigma}_{\mathbf{\Lambda}_{0j}^{\star}})\mathbb{1}_{\left\{-1<\lambda_{jk}<1 \;\forall k \;\cap\; 1-\mathbf{\Lambda}_{0j}\mathbf{\Phi}^s\mathbf{\Lambda}_{0j}^{\mathrm{T}}>0\right\}}, \tag{4.21}$$

for rows which do not, where

$$\begin{aligned}\tilde{\mathbf{\Lambda}}_{0j}^{\star} &= (\mathbf{\Xi}_j^{\star\mathrm{T}}\mathbf{\Xi}_j^{\star})^{-1}\mathbf{\Xi}_j^{\star\mathrm{T}}(\mathbf{z}_j - \mathbf{1}_n\mu_j),\\ \mathbf{\Sigma}_{\mathbf{\Lambda}_{0j}^{\star}} &= \psi_{jj}(\mathbf{\Xi}_j^{\star\mathrm{T}}\mathbf{\Xi}_j^{\star})^{-1}.\end{aligned}$$

The rows of $\boldsymbol{\Xi}$ are also independent such that we need to find the conditional distribution $\pi(\boldsymbol{\xi}_i|\mathbf{z}_i, \boldsymbol{\mu}, \boldsymbol{\Lambda}_0, \boldsymbol{\Psi}, \boldsymbol{\Phi}^s)$, $i = 1, \ldots, n$, which, with the help of the Woodbury matrix identity (Woodbury, 1950), can be found as

$$\pi(\boldsymbol{\xi}_i|\mathbf{z}_i, \boldsymbol{\mu}, \boldsymbol{\Lambda}_0, \boldsymbol{\Psi}, \boldsymbol{\Phi}^s) \stackrel{d}{=} \mathcal{N}_m(\tilde{\boldsymbol{\xi}}, \boldsymbol{\Sigma}_\xi), \tag{4.22}$$

where

$$\begin{aligned}\tilde{\boldsymbol{\xi}} &= \boldsymbol{\Sigma}_\xi \boldsymbol{\Lambda}_0^{\mathrm{T}} \boldsymbol{\Psi}^{-1}(\mathbf{z}_i - \boldsymbol{\mu}),\\ \boldsymbol{\Sigma}_\xi &= [(\boldsymbol{\Phi}^s)^{-1} + \boldsymbol{\Lambda}_0^{\mathrm{T}} \boldsymbol{\Psi}^{-1} \boldsymbol{\Lambda}_0]^{-1}.\end{aligned}$$

The conditional distribution $\pi(\boldsymbol{\Phi}|\mathbf{Z}, \boldsymbol{\Xi})$ is easily found as

$$\pi(\boldsymbol{\Phi}|\mathbf{Z}, \boldsymbol{\Xi}) \stackrel{d}{=} \mathcal{IW}_m(\boldsymbol{\Xi}^{\mathrm{T}}\boldsymbol{\Xi} + \boldsymbol{\Upsilon}, n + \tau). \tag{4.23}$$

Lastly, the conditional distribution for the diagonal elements of $\boldsymbol{\Psi}$, $\pi(\psi_{jj}|\mathbf{z}_j, \mu_j, \boldsymbol{\Lambda}_{0j}, \boldsymbol{\Xi})$, $j = 1, \ldots, p$, can be found as

$$\pi(\psi_{jj}|\mathbf{z}_j, \mu_j, \boldsymbol{\Lambda}_{0j}, \boldsymbol{\Xi}) \stackrel{d}{=} \mathcal{IG}((n + \nu)/2, (\beta_j + \nu d)/2), \tag{4.24}$$

where

$$\beta_j = (\mathbf{z}_j - \mathbf{1}_n\mu_j - \boldsymbol{\Xi}\boldsymbol{\Lambda}_{0j}^{\mathrm{T}})^{\mathrm{T}}(\mathbf{z}_j - \mathbf{1}_n\mu_j - \boldsymbol{\Xi}\boldsymbol{\Lambda}_{0j}^{\mathrm{T}}).$$

4.4.3 Complexity

Equation (4.17) states that the complexity of an inequality constrained factor model is given by

$$\int_{\{\boldsymbol{\Lambda}_0:\boldsymbol{\Omega}_b\boldsymbol{\lambda}_f - \boldsymbol{\alpha}_b > \mathbf{0}\}} \pi_0(\boldsymbol{\Lambda}_0|\boldsymbol{\Phi})\,\partial\boldsymbol{\Lambda}_0 = \int_{\{\boldsymbol{\Lambda}_0:\boldsymbol{\Omega}_b\boldsymbol{\lambda}_f - \boldsymbol{\alpha}_b > \mathbf{0}\}} \pi_0(\boldsymbol{\Lambda}_0)\mathbb{1}_{\{\mathbf{I}_p - \mathrm{diag}(\boldsymbol{\Lambda}_0\boldsymbol{\Phi}\boldsymbol{\Lambda}_0^{\mathrm{T}})>0\}}\,\partial\boldsymbol{\Lambda}_0,$$

indicating the proportion of the domain of M_0 abiding the inequality constraints that define M_b. In this form, $\boldsymbol{\Phi}$ is unknown. Notice, however, that for any positive definite $\boldsymbol{\Phi}$ we may always find $\mathbf{V} \in \mathbb{R}^{m\times m}$ such that $\boldsymbol{\Phi} = \mathbf{V}\mathbf{V}^{\mathrm{T}}$, and $\boldsymbol{\Lambda}_0\boldsymbol{\Phi}\boldsymbol{\Lambda}_0^{\mathrm{T}} = (\boldsymbol{\Lambda}_0\mathbf{V})(\boldsymbol{\Lambda}_0\mathbf{V})^{\mathrm{T}}$. This implies that each oblique representation has equivalent orthogonal representations and suggests evaluation of complexity under orthogonality. This coincides with (integrating $\boldsymbol{\Phi}$ out of $\pi_0(\boldsymbol{\Lambda}_0, \boldsymbol{\Phi})$ and) evaluating $\boldsymbol{\Phi}$ in $\mathbb{1}_{\{\mathbf{I}_p - \mathrm{diag}(\boldsymbol{\Lambda}_0\boldsymbol{\Phi}\boldsymbol{\Lambda}_0^{\mathrm{T}})>0\}}$ at its prior expectation: $\boldsymbol{\Upsilon}(\tau - m - 1)^{-1} = \mathbf{I}_m$. Model complexity is thus formalized as follows:

Definition 4.2 (Complexity). *Let $\dot{\boldsymbol{\Lambda}}_{0j}^\star$ denote the $(m - t_j)$-dimensional row vector containing the elements in $\boldsymbol{\Lambda}_{0j}$ not involved in restrictions under conditions* C1 *and* C4 *from Section 4.2.2. Let the uniform density (conform prior specification) on $\dot{\boldsymbol{\Lambda}}_{0j}^\star$ be given by:*

$$\varrho(\dot{\boldsymbol{\Lambda}}_{0j}^\star; -1, 1) = \begin{cases} (2)^{-(m-t_j)} & \textit{if } \dot{\boldsymbol{\Lambda}}_{0j}^\star \in (-1, 1)^{m-t_j} \\ 0 & \textit{elsewhere} \end{cases}.$$

Moreover, let

$$\varrho((-)\lambda^{p}_{jk}; 0, 1) = \begin{cases} 1 \ \textit{if} \ (-)\lambda^{p}_{jk} \in (0, 1) \\ 0 \ \textit{elsewhere} \end{cases} ,$$

denote the uniform density on the factor loading elements involved in a polarity truncation as indicated by condition C4 *from Section 4.2.2 and define*

$$\Gamma_j = \begin{cases} \varrho((-)\lambda^{p}_{jk}; 0, 1) \ \textit{if} \ \lambda^{p}_{jk} \in \mathbf{\Lambda}^{\star}_{0j} \\ 1 \qquad\qquad\qquad\quad \textit{otherwise} \end{cases} .$$

In keeping with (4.17) *model complexity* ω_b *of an inequality constrained model* M_b *is then defined as the probability mass of the* j*-product of* $\varrho(\dot{\mathbf{\Lambda}}^{\star}_{0j}; -1, 1)\Gamma_j$*, bounded by the correlation restriction* $1 - \mathbf{\Lambda}_{0j}\mathbf{\Phi}\mathbf{\Lambda}^{\mathrm{T}}_{0j} > 0$ *evaluated at the prior expectation of* $\mathbf{\Phi}$*, that is located in the inequality constrained space defining* M_b*. Hence,*

$$\omega_b \equiv \int_{\{\mathbf{\Lambda}_0 : \mathbf{\Omega}_b \boldsymbol{\lambda}_f - \boldsymbol{c}_b > \mathbf{0}\}} \prod_{j=1}^{p} \varrho(\dot{\mathbf{\Lambda}}^{\star}_{0j}; -1, 1)\Gamma_j \mathbb{1}_{\{1 - \mathbf{\Lambda}_{0j}\mathbf{\Lambda}^{\mathrm{T}}_{0j} > 0\}} \, \partial \mathbf{\Lambda}_0. \qquad (4.25)$$

Remark 4.10. Definition 4.2 assumes, for sake of (notational) simplicity, that a maximum of one element per row of $\mathbf{\Lambda}_0$ is involved in a polarity truncation under Condition C4. In practice this will be the most likely case, as for estimation efficiency (see Algorithm 4.1 in Section 4.5) Condition C4 should be imposed on loadings that, from prior knowledge or theory, are believed to be large. In practice, usually one loading element per row is relatively large as an indicator of the latent factor. In obtaining global rotational uniqueness, however, it is allowed to have multiple column polarity truncations in the same row of $\mathbf{\Lambda}_0$. Definition 4.2 is easily extended to deal with this situation.

The formulation of complexity resembles the definition in Mulder, Hoijtink, and Klugkist (2010), but differs in the sense that the space represents bound proper noninformative priors on regression parameters and respects the geometry of the FA model. Regarding the latter, each row in the prior loadings matrix represents an $(m - t_j)$-dimensional unit ball. The full prior space is then given by an open $(pm - t)$-dimensional ball whose properties can be reduced to the study of the $(pm - t)$-dimensional unit ball. Model complexity is then given by the proportion of the interior of the $(pm-t)$-dimensional unit ball abiding the system of inequality constraints.

An important topic is the sensitivity of B_{b0} to the prior specification. Given the correlation bound the influence of the prior on the proportion of the posterior distribution abiding the system of inequality constraints (f_b) is negligible when data dominate the prior, as is the case given the parameterizations. Interesting is the sensitivity of ω_b to the prior specification. One could call ω_b objective when, given a system of inequality constraints, its value is independent of the parametrization of or the bound on the prior. As such, objectivity is related to the assignment of probability mass to an inequality constrained feasible region. The following definition and example formalize and exemplify objectivity of ω_b.

Definition 4.3 (Objectivity ω_b). *Let a subsystem of inequality constraints be taken to mean a connected set of parameters, in the sense that every pair of parameters in the set is connected through an (possibly incomplete) ordering or other inequality constrained structure. Let the subsystem have P_y permutations in its formulation. Then $\omega_{b|y}$, the complexity of subsystem y, is termed objective iff each permutation is assigned equal probability mass, i.e., $\omega_{b|y} = 1/P_y$ for each permutation. Model complexity ω_b is subsequently termed objective iff*

$$\omega_b = \prod_{y=1}^{Y} \omega_{b|y} = \left[\prod_{y=1}^{Y} P_y\right]^{-1} . \tag{4.26}$$

Example 4.2. Imagine we have a rotationally unique factor structure of dimension $m = 2$ on, say, 8 items, and that we specify a subsystem of constraints such as $\lambda_{11} > \lambda_{12}$. This subsystem has two permutations ($\lambda_{11} > \lambda_{12}$ and $\lambda_{12} > \lambda_{11}$) and the feasible region is defined by a semicircle. By symmetry, the area of any centered semicircle is two times the area of a circle between $(0, 0)$ and $(r, 0)$, with r denoting the radius. Let (r, o) denote polar coordinates. We may then find

$$\int_0^r \int_{-\sqrt{r^2-\lambda_{11}^2}}^{\sqrt{r^2-\lambda_{11}^2}} d\lambda_{11} d\lambda_{12} = 2r^2 \int_0^{\pi/2} \cos^2 o \, do = 2r^2 \int_0^{\pi/2} \frac{1}{2}(1+\cos^2 o) \, do$$
$$= r^2 o\Big|_0^{\pi/2} + r^2(\sin 2o)\Big|_0^{\pi/2} = \frac{\pi r^2}{2},$$

giving that, irrespective of r, $\omega_{b|y} = (\pi r^2/2)/\pi r^2 = 1/2$. For the specified prior $\omega_{b|y}$ is thus objective in the sense of Definition 4.3 with respect to inequalities of type (ii) and (iii) from Section 4.2.3.

The situation just exemplified can be generalized to transitive orderings (which are conjunctions of type (ii) or (iii) inequalities). Imagine that in the mentioned factor structure also the subsystem $\lambda_{71} > \lambda_{72} > \lambda_{82}$ is specified, which has 3! permutations. This type of ordering can be expressed through (4.25) with the following integral for which we state a useful result that may be proven by induction:

Lemma 4.1.

$$\lim_{r\to\jmath} \int_{\lambda_{j1}=-r}^{r} \int_{\lambda_{j2}=-r}^{\lambda_{j1}} \cdots \int_{\lambda_{jm}=-r}^{\lambda_{j(m-1)}} (2r)^{-m} \mathbb{1}_{\{\lambda_{j1}^2+\lambda_{j2}^2+\cdots+\lambda_{jm}^2<r^2\}} \, d\lambda_{jm}\, d\lambda_{j(m-1)} \ldots d\lambda_{j1}$$
$$= \frac{1}{m!}, \tag{4.27}$$

irrespective of the value of $\jmath$.

The complexity measure is thus also objective for complete-ordering conjunctions of type (ii) and (iii) inequalities. Objectivity may also be shown for incomplete-ordering conjunctions of type (ii) and (iii) inequalities, such as, for example, $\lambda_{41} < \lambda_{22} > \lambda_{42}$.

For the specified priors, ω_b is not objective for type (i) inequalities from Section 4.2.3. It is obvious that specifying, for example, $\lambda_{61} > \alpha_l$ which has two permutations, will not lead to a complexity of 1/2 for any α_l other than 0. Moreover, it is easy to verify that the complexity measure in this situation is also dependent on the prior bound. Similar arguments apply to inequalities of type (iv), such as $|\lambda_{62}| < .3$. Specifying such inequalities is still feasible and plausible as the space is restricted by the natural bound given the parametrization of the model. Care should be taken however, not to choose α_l too close to the boundary of the permissable space or to choose ε too small, in order to avoid analogues of Lindley-Bartlett-type paradoxes (Bartlett, 1957; Lindley, 1957). ■

Example 4.1 (Continued). Recall the factor structure with minimal restrictions for obtaining metric and rotational uniqueness:

$$\mathbf{\Lambda}_0 = \begin{bmatrix} \lambda_{11} & \lambda_{12} \\ \lambda_{21} & \lambda_{22} \\ \lambda_{31} = 0 & \lambda_{32} > 0 \\ \lambda_{41} & \lambda_{42} \\ \lambda_{51} > 0 & \lambda_{52} = 0 \\ \lambda_{61} & \lambda_{62} \\ \lambda_{71} & \lambda_{72} \\ \lambda_{81} & \lambda_{82} \end{bmatrix} \begin{matrix} \text{BMI} \\ \log_e\{\text{trig.}\} \\ \text{HDL chol.} \\ \log_e\{\text{IR}\} \\ \log_e\{\text{GB}\} \\ \log_e\{\text{G2}\} \\ \text{SBP} \\ \text{DBP} \end{matrix} .$$

The idea is to specify a theoretical factor structure by imposing inequality constraints on the free parameters in de model. Imagine one would impose $\lambda_{11} > \lambda_{12}$ and $\lambda_{71} > \lambda_{72} > \lambda_{82}$. From Example 4.2 it can be inferred that the complexity of the subsystem $\lambda_{11} > \lambda_{12}$ is 1/2 and that the complexity of the subsystem $\lambda_{71} > \lambda_{72} > \lambda_{82}$ amounts to $1/3! = 1/6$. As the subsystems are independent their conjunction amounts to a total complexity ω_b of $\omega_{b|1} \cdot \omega_{b|2} = 1/2 \cdot 1/6 = 1/12$. For elaborate inequality constrained formulations of the factor structure the complexity measure may be too tedious or cumbersome to calculate analytically. In Section 4.5 a numerical recipe for finding ω_b is given.

Notice should be given that it is possible that $\lambda_{11} > \lambda_{12}$ while λ_{12} has a larger loading, for example when λ_{11} is in the positive range while λ_{12} finds itself relatively farther in the negative range. Specifying factor structure using inequalities would naturally be geared towards expressions of (absolute) largeness of factor loadings, which is tied to the polarity of columns in $\mathbf{\Lambda}_0$. How to incorporate this will be exemplified in Section 4.6 where this example will be revisited. ■

4.5 Numerical Implementation

This section reviews the numerical nuts and bolts for making the results of Sections 4.3 and 4.4 substantively applicable to FA. First, we show how information from a Gibbs sampler may be used to assess f_b. Second, we delve into a numerical exploration of the complexity measure ω_b, as for elaborate inequality constrained formulations of the factor structure the complexity measure may be too tedious or cumbersome to calculate analytically.

Algorithm 4.1 (Gibbs Sampler for $\hat{f}_b$) The input consists of the data $\mathbf{Z}$ and starting values $\boldsymbol{\Lambda}_0^{(0)}$, $\boldsymbol{\Xi}^{(0)}$, $\boldsymbol{\Phi}^{s(0)}$, and $\boldsymbol{\Psi}^{(0)}$. The output is the estimate $\hat{f}_b$. The algorithm uses a Gibbs sampler to estimate $\hat{f}_b$, as implied possible by Proposition 4.2. Each of $\boldsymbol{\mu}, \boldsymbol{\Lambda}_0, \boldsymbol{\Phi}, \boldsymbol{\Xi}$, and $\boldsymbol{\Psi}$ is treated as a single block. From the prior specification it is clear that we sample $\boldsymbol{\Phi}$ as a covariance matrix. However, in specifying uniqueness conditions, and in keeping with the modeling of a standardized FA covariance structure, $\boldsymbol{\Phi}$ is desired to be a correlation matrix. That is why the draws $\boldsymbol{\Phi}^{(c)}$ are standardized to the correlation matrix $\boldsymbol{\Phi}^{s(c)}$. In the following, $\zeta(\cdot,\cdot)$ denotes the normal density and model indices are suppressed where appropriate, for sake of notational simplicity.

Set $\boldsymbol{\Lambda}_0^{(0)}$, $\boldsymbol{\Xi}^{(0)}$, $\boldsymbol{\Phi}^{s(0)}$, and $\boldsymbol{\Psi}^{(0)}$
for $c = 1$ to C **do**
  Generate $\boldsymbol{\mu}^{(c)}$ from $\pi(\boldsymbol{\mu}|\mathbf{Z}, \boldsymbol{\Lambda}_0^{(c-1)}, \boldsymbol{\Xi}^{(c-1)}, \boldsymbol{\Psi}^{(c-1)})$
  Generate $\boldsymbol{\Lambda}_0^{(c)}$ from $\prod_{j=1}^{p} \zeta(\tilde{\boldsymbol{\Lambda}}_{0j}^{\star}, \boldsymbol{\Sigma}_{\boldsymbol{\Lambda}_{0j}^{\star}})$
  if $\exists jk$ such that $\neg(-1 < \lambda_{jk}^{(c)} < 1)$ or $\neg(0 < (-)\lambda_{jk}^{p(c)} < 1)$ **or** $\mathbf{I}_p - \mathrm{diag}\left(\boldsymbol{\Lambda}_0^{(c)}\boldsymbol{\Phi}^{s(c-1)}\boldsymbol{\Lambda}_0^{(c)\mathrm{T}}\right) \not> 0$ **then**
    go to 4:
  else
    Generate $\boldsymbol{\Xi}^{(c)}$ from $\prod_{i=1}^{n} \pi(\boldsymbol{\xi}_i|\mathbf{z}_i, \boldsymbol{\mu}^{(c)}, \boldsymbol{\Lambda}_0^{(c)}, \boldsymbol{\Psi}^{(c-1)}, \boldsymbol{\Phi}^{s(c-1)})$
    Generate $\boldsymbol{\Phi}^{(c)}$ from $\pi(\boldsymbol{\Phi}|\mathbf{Z}, \boldsymbol{\Xi}^{(c)})$
      Set $\forall k \geqslant k'$ $\phi_{kk'}^{s(c)} = \phi_{kk'}^{(c)} / \sqrt{\phi_{kk}^{(c)}} \cdot \sqrt{\phi_{k'k'}^{(c)}}$
      Set $\forall k > k'$ $\phi_{k'k}^{s(c)} = \phi_{kk'}^{s(c)}$
    Generate $\boldsymbol{\Psi}^{(c)}$ from $\prod_{j=1}^{p} \pi(\psi_{jj}|\mathbf{z}_j, \mu_j^{(c)}, \boldsymbol{\Lambda}_{0j}^{(c)}, \boldsymbol{\Xi}^{(c)})$
  end if
end for
$\hat{f}_b = C^{-1} \sum_{c=1}^{C} \mathbb{1}_{\{\boldsymbol{\Omega}_b \boldsymbol{\lambda}_f^{(c)} - \boldsymbol{\alpha}_b > \mathbf{0}\}}$

Algorithm 4.2 (Sampler for $\hat{\omega}_b$) The input consists of the number of iterations. The output is the estimate $\hat{\omega}_b$. The algorithm proceeds by sampling $\boldsymbol{\Lambda}_0$ from its respective prior density. Subsequently, the samples are accepted or discarded according to compliance with the polarity restrictions and correlation bound, after which it is checked if the accepted draws comply with the system of inequality constraints.

Set $\varphi = 0$
for $v = 1$ to V **do**
  Generate $\boldsymbol{\Lambda}_0^{(v)}$ from $\prod_{j=1}^{p} \varrho(\dot{\boldsymbol{\Lambda}}_{0j}^{\star}; -1, 1)\Gamma_j$
  if $\mathbf{I}_p - \mathrm{diag}\left(\boldsymbol{\Lambda}_0^{(v)}\boldsymbol{\Lambda}_0^{(v)\mathrm{T}}\right) \not> 0$ **then**
    go to 3:
  else

if $\mathbb{1}_{\{\boldsymbol{\Omega}_b \boldsymbol{\lambda}_f^{(v)} - \boldsymbol{\alpha}_b > \mathbf{0}\}} = 1$ **then**
  $\varphi = \varphi + 1$
else
  $\varphi = \varphi$
end if
end if
end for
$\hat{\omega}_b = V^{-1}\varphi$

4.6 Illustrations

The techniques developed in Sections 4.2 to 4.5 will be explored numerically. Illustrations are given with data regarding anthropomorphic measures on overweight youngsters (the running example) and data on handedness in captive chimpanzees. The examples are markedly different and illustrate different uses of inequality constraints in CFA. In the running example the labeling of factors is given by previous research. The competing inequality constrained structures are then expressions of factor structure given a clearly developed idea regarding the latent variable. In the handedness example the competing inequality constrained factor structures are used to construct a labeling for the latent variables.

4.6.1 Revisiting the Running Example

Additional background

Since its usage by Edwards, *et al.* (1994) FA has carved out a niche for itself in the MBS research community. For an overview of factor analytic efforts in MBS research we confine by referring to Penno *et al.* (2006). An important question thrusting the FA efforts in MBS research is if a unifying physiology dominated by insulin resistance underlies the clustering of metabolic risk variables, or if there are multiple underlying physiologic phenotypes.

Most FA studies report on finding either three or four factors underlying the clustering of risk factors. However, Peeters (invited revision a) found a twofold phenotypic pattern dominated by abnormalities in the glucose and lipid metabolism, respectively, which resembles the initial conception of the MBS (Reaven, 1988) and is consistent with epidemiological studies (Liese et al., 1998). Here, we explore this finding by formulating constrained two-factor models that in turn emphasize different specificities for the pathophysiologic constellations.

Competing Inequality Constrained Factor Structures

Say one specifies $\lambda_{11} > \lambda_{12}$. There are three situations in which this may be correct in the sample. First, it can be that both λ_{11} and λ_{12} are positive while λ_{12} has a smaller positive loading. Second, it can be that λ_{11} is positive and λ_{12} is negative. Third, when both λ_{11} and λ_{12} are negative and λ_{12} finds itself farther in the negative range. In the last two situations the specification of $\lambda_{11} > \lambda_{12}$ may not be very informative with regard to the structure of the model. For example, λ_{11} can be larger than λ_{12} while the latter has the largest absolute loading, indicating that item or measure 1 is a better observable indicator for latent factor 2. Specifying factor structure using inequalities between loadings would naturally be geared towards expressions of (absolute) largeness of factor loadings, which is tied to the polarity of columns in $\mathbf{\Lambda}_0$. The choice of polarity truncations under Condition C4 of Section 4.2.2 fixes the polarity of each column and thus the sign of each free loading. Specifying inequalities between (or on) free loadings should take into account the anticipated sign of the loadings involved, or should anticipate lack of knowledge regarding sign. The following definition takes into account the latter situation:

Definition 4.4.

$$(-)\lambda_q > |\lambda_{q'}| \Rightarrow \begin{cases} (-)\lambda_q - \lambda_{q'} > 0 \\ (-)\lambda_q + \lambda_{q'} > 0 \end{cases} . \tag{4.28}$$

An inequality as in (4.28) allows one to specify that (the negative of) a certain factor loading is larger than an other loading, irrespective of the latter's sign. An inequality as in (4.28) is a conjunction of type (ii) and type (iii) inequalities from Section 4.2.3 and can be shown to be objective in the sense of Definition 4.3. This type of inequality is useful in formulating informative inequality constrained factor structures.

The base model was given in earlier sections. Below we formulate differing inequality constrained competing factor structures in keeping with some of the ideas explored in previous sections:

$$\mathbf{\Lambda}_1 = \begin{bmatrix} \lambda_{11} & > \; |\lambda_{12}| \\ |\lambda_{21}| & < -\lambda_{22} \\ \lambda_{31} = 0 & \lambda_{32} > 0 \\ \lambda_{41} & > \; |\lambda_{42}| \\ \lambda_{51} > 0 & \lambda_{52} = 0 \\ \lambda_{61} & > \; |\lambda_{62}| \\ \lambda_{71} & > \; |\lambda_{72}| \\ |\lambda_{81}| & < -\lambda_{82} \end{bmatrix} \begin{matrix} \text{BMI} \\ \log_e\{\text{trig.}\} \\ \text{HDL chol.} \\ \log_e\{\text{IR}\} \\ \log_e\{\text{GB}\} \\ \log_e\{\text{G2}\} \\ \text{SBP} \\ \text{DBP} \end{matrix} ,$$

$$\mathbf{\Lambda}_2 = \begin{bmatrix} |\lambda_{11}| & < -\lambda_{12} \\ |\lambda_{21}| & < -\lambda_{22} \\ \lambda_{31} = 0 & \lambda_{32} > 0 \\ \lambda_{41} > .4 & \lambda_{42} < -.4 \\ \lambda_{51} > 0 & \lambda_{52} = 0 \\ \lambda_{61} & > \; |\lambda_{62}| \\ \lambda_{71} & < -\lambda_{72} \\ \lambda_{81} & < -\lambda_{82} \end{bmatrix} \begin{matrix} \text{BMI} \\ \log_e\{\text{trig.}\} \\ \text{HDL chol.} \\ \log_e\{\text{IR}\} \\ \log_e\{\text{GB}\} \\ \log_e\{\text{G2}\} \\ \text{SBP} \\ \text{DBP} \end{matrix} ,$$

$$\boldsymbol{\Lambda}_3 = \begin{bmatrix} \lambda_{11} & > & |\lambda_{12}| \\ |\lambda_{21}| & < & -\lambda_{22} \\ \lambda_{31} = 0 & & \lambda_{32} > 0 \\ \lambda_{41} & > & |\lambda_{42}| \\ \lambda_{51} > 0 & & \lambda_{52} = 0 \\ \lambda_{61} & > & |\lambda_{62}| \\ |\lambda_{71}| < .3 & & |\lambda_{72}| < .3 \\ |\lambda_{81}| < .3 & & |\lambda_{82}| < .3 \end{bmatrix} \begin{matrix} \text{BMI} \\ \log_e\{\text{trig.}\} \\ \text{HDL chol.} \\ \log_e\{\text{IR}\} \\ \log_e\{\text{GB}\} \\ \log_e\{\text{G2}\} \\ \text{SBP} \\ \text{DBP} \end{matrix} .$$

In all models insulin resistance, blood glucose level at baseline and two hours after glucose intake are mainly related to the glucose factor, while fasting levels of triglycerides and HDL cholesterol form the base of the lipid factor. Note that for the lipid factor polarity fixation is brought about by demanding $\lambda_{32} > 0$. HDL cholesterol is generally regarded as 'good' cholesterol, meaning that the choice $\lambda_{32} > 0$ amounts to modeling a factor denoting unimpaired lipid metabolism. Hence formulations like $|\lambda_{21}| < -\lambda_{22}$, as under given polarity truncation the triglycerides item is believed to be strongly negatively related to the lipid factor.

Model 1 adds detail to the base model and the basic factors by stating that BMI and systolic blood pressure are linked to the glucose factor ($\lambda_{11} > |\lambda_{12}|$, $\lambda_{71} > |\lambda_{72}|$), while diastolic blood pressure is believed to be linked to the lipid factor ($|\lambda_{81}| < -\lambda_{82}$). Model 2 states the believe that BMI is an indicator for the lipid rather than the glucose factor ($|\lambda_{11}| < -\lambda_{12}$). Also, in this model both systolic and diastolic blood pressure are related to the lipid rather than the glucose factor, with the additional belief that both blood pressure measures will load positively on the latter ($\lambda_{71} < -\lambda_{72}$, $\lambda_{81} < -\lambda_{82}$). Moreover, the second model states that insulin resistance may be the measure tying the metabolic syndrome together, achieving a large loading on both the glucose and lipid factor ($\lambda_{41} > .4$, $\lambda_{42} < -.4$). Model 3 resembles the first model, but states that the association of systolic and diastolic blood pressure with the factors is rather loose.

Evaluation

It is efficient to initialize the sampler in Algorithm 4.1 with ML estimates. Here ML EFA estimates are used, which can be obtained from many standard packages performing FA. Let $\hat{\boldsymbol{\Lambda}}$ and $\hat{\boldsymbol{\Psi}}$ be the ML EFA estimates of $\boldsymbol{\Lambda}$ and $\boldsymbol{\Psi}$. Then $\hat{\boldsymbol{\Lambda}}$ is direct Oblimin rotated to obtain an oblique structure and an estimate $\hat{\boldsymbol{\Phi}}^s$ of $\boldsymbol{\Phi}^s$. The rotated loadings matrix then receives structural zeroes and possible column multiplications with minus unity in order to obtain $\hat{\boldsymbol{\Lambda}}_0$; a starting value for the base model loadings matrix. The ML estimate of $\boldsymbol{\mu}$ is simply $\hat{\boldsymbol{\mu}} = \bar{\mathbf{z}}$. Then, by regressing $\boldsymbol{\Xi}$ on $\mathbf{Z}$, an estimate for the factor scores can be obtained as

$$\hat{\boldsymbol{\Xi}} = (\mathbf{Z} - \mathbf{1}_n \otimes \hat{\boldsymbol{\mu}}^{\mathrm{T}})\hat{\boldsymbol{\Psi}}^{-1}\hat{\boldsymbol{\Lambda}}_0 \left[(\hat{\boldsymbol{\Phi}}^s)^{-1} + \hat{\boldsymbol{\Lambda}}_0^{\mathrm{T}}\hat{\boldsymbol{\Psi}}^{-1}\hat{\boldsymbol{\Lambda}}_0\right]^{-1},$$

which may serve as starting value for $\boldsymbol{\Xi}$. For evaluation $c = v = 3,000,000$ iterations are used. Table 4.1 contains B_{b0} for all constrained models under consideration.

Table 4.1. Estimated Bayes factors for constrained models on the metabolic syndrome data

M_1		M_2		M_3	
$\hat{f}_1 =$	1.263e-3	$\hat{f}_2 =$	.000	$\hat{f}_3 =$	.569
$\omega_1 =$	2^{-12}	$\hat{\omega}_2 =$	2.083e-4	$\hat{\omega}_3 =$	4.433e-5
$\hat{B}_{10} =$	5.175	$\hat{B}_{20} =$	.000	$\hat{B}_{30} =$	12,830.729

From Table 4.1 it can be inferred that model 3 is least complex and has the best fit, resulting in constrained model 3 deemed best in comparison to the unconstrained model. The upper bound of the Bayes factor of the constrained to the unconstrained model is the inverse of model complexity ω_b, entailing, for example, that twice the natural logarithm of the Bayes factor is no longer on the same scale as the deviance or likelihood ratio statistics. This, however, does not have to pose a problem when recognizing that the mentioned Bayes factor has its own scale relative to the complexity measure. Also note that the Bayes factor of a constrained against another constrained model (which in practice will garner most attention) does not have an upper bound and its interpretation may be referred to the usual half-units on the $\log_{10}$ scale (Jeffreys, 1961) or the suggested units on the double natural logarithm scale (Kass & Raftery, 1995). These Bayes factor can be found in Table 4.2, which represents a Bayes factor matrix.

Table 4.2. Bayes factor matrix for the metabolic syndrome data

$\hat{\text{B}}_{from\|to}$	0	1	2	3
0	1	.193	∞	7.794e-5
1	5.175	1	∞	4.033e-4
2	.000	.000	1	.000
3	12,830.729	2479.377	∞	1

Table 4.2 suggests that model 3 is unequivocally best. Given the data, there is decisive evidence in favor of model 3 against both model 1 and 2 (when interpreting twice the natural logarithm of the involved Bayes factors as discussed in Kass & Raftery, 1995). The estimates of the UCFM for the MBS data given in Table 4.3 also support model 3.

4.6.2 Handedness in Captive Chimpanzees

Background

Scientists have long thought population-level handedness to be a uniquely human trait (Warren, 1980) attributable to a specialization of brain hemispheres. This

Table 4.3. Posterior Means and 95% Credible Intervals for $\mathbf{\Lambda}_0$, diag($\mathbf{\Psi}$) and the free elements in $\mathbf{\Phi}$ on the metabolic syndrome data

	Parameter	Mean	95% CI	Parameter	Mean	95% CI	Item
$\mathbf{\Lambda}_0$	λ_{11}	.324	[.207, .440]	λ_{12}	-.068	[-.191, .055]	BMI
	λ_{21}	-.006	[-.303, .215]	λ_{22}	-.653	[-.956,-.379]	$\log_e$\{trig.\}
	λ_{31}	-	-	λ_{32}	.706	[.442, .940]	HDL chol.
	λ_{41}	.767	[.613, .921]	λ_{42}	-.179	[-.343,-.022]	$\log_e$\{IR\}
	λ_{51}	.470	[.360, .585]	λ_{52}	-	-	$\log_e$\{GB\}
	λ_{61}	.355	[.205, .492]	λ_{62}	-.124	[-.289, .036]	$\log_e$\{G2\}
	λ_{71}	.274	[.136, .416]	λ_{72}	.029	[-.118, .171]	SBP
	λ_{81}	.202	[.069, .347]	λ_{82}	.139	[-.017, .292]	DBP
$\mathbf{\Psi}, \boldsymbol{\mu}$	ψ_{11}	.881	[.769,1.008]	μ_1	3.877e-6	[-.075, .076]	BMI
	ψ_{22}	.560	[.173, .825]	μ_2	1.361e-4	[-.098, .098]	$\log_e$\{trig.\}
	ψ_{33}	.500	[.164, .813]	μ_3	-1.027e-4	[-.094, .094]	HDL chol.
	ψ_{44}	.309	[.119, .477]	μ_4	-4.299e-5	[-.109, .110]	$\log_e$\{IR\}
	ψ_{55}	.783	[.666, .908]	μ_5	-3.011e-5	[-.081, .081]	$\log_e$\{GB\}
	ψ_{66}	.833	[.721, .958]	μ_6	1.523e-5	[-.078, .079]	$\log_e$\{G2\}
	ψ_{77}	.928	[.805,1.064]	μ_7	1.575e-5	[-.072, .072]	SBP
	ψ_{88}	.954	[.826,1.094]	μ_8	-2.873e-5	[-.069, .070]	DBP
$\mathbf{\Phi}$	ϕ_{12}	-.277	[-.417,-.137]				

view, however, has been challenged by evidence that hand preferences may have evolved independently of hemisphere specialization (Hopkins & Cantalupo, 2004) and an increasing number of reports on hand preferences in various nonhuman primate species (cf. Bradshaw & Rogers, 1993; Hopkins, 1996; Wesley et al., 2002). These findings may give pointers as to whether species level handedness evolved before or after the divergence of *Pan* and *Homo* lineages.

FA is an accepted tool in human handedness research with FA efforts usually focussing on proficiency versus preference tasks (cf. Barnsley & Rabinovitch, 1970; Roszkowski, Snelbecker, & Sacks, 1981; Williams, 1986). The extracted factors are usually defined in terms of their motor or cognitive demands. Factor analytic efforts in nonhuman primate handedness research are rare and focus on assessing dimensionality and task-specificity of handedness over multiple measures of hand use (see for example Wesley et al., 2002).

Data

We have data from the Yerkes National Primate Research Center on 104 captive chimpanzees (Pan troglodytes) of both sexes first published in Wesley *et al.* (2002).

The data contain six different behavioral tasks that are used to express hand preference: (i) The coordinated bimanual tube task (T), in which the subjects receive a PVC tube with peanut butter smeared inside. The hand used to remove the treat was recorded; (ii) The fixed tube task (FT), which is similar to the former task with the exception that the PVC tube was placed above the cage floor and could only be partially pulled into the cage, thus requiring for differing postural and positional demands; (iii) The bimanual feeding task (BF), in which the chimps receive several pieces of food and typically will use one hand for feeding while holding excess food with the other. The feeding hand was recorded; (iv) The biased feeding task (F), in which the chimps receive an orange cut into several sections. The subjects typically eat the pulp before the rind, calling for manipulation of an orange slice by mouth and hand. The manipulating hand was recorded; (v) The simple reaching task (R), which records the hand used to pick up a raisin or peanut thrown into the cage; and (vi) The gesture task (G), which records the hand used for the first instance of gestural communication after the experimenter placed a piece of banana in front of the cage while directly facing the subject. Each subject was tested on each measure a multitude of times and the order of task presentation was randomized across subjects and days. For more information on the data we refer to Wesley *et al.* (2002).

The quantity for the FA is the handedness index (HI) which for each subject and measure consists of the number of left hand responses (lh) subtracted from the number of right hand responses (rh) divided by the total number of responses: HI $= (rh - lh)/(rh + lh)$. Most measures display population-level right hand preferences (Wesley et al., 2002). Interest lies in formulating constrained factor models that in turn emphasize different origins for the clustering of hand use measures.

Competing Inequality Constrained Factor Structures

First, an exploration of dimensionality of the factor solution was undertaken. The techniques developed in Peeters (invited revision a) indicated that retaining two latent factors constitutes the optimal choice (Section 4.7 discusses the interrelations between exploring dimensionality and testing confirmatory inequality constrained expressions of theory). The base model $\mathbf{\Lambda}_0$ and the competing inequality constrained models nested in $\mathbf{\Lambda}_0$ for the two-factor solution are:

$$\mathbf{\Lambda}_0 = \begin{bmatrix} \lambda_{11} > 0 & \lambda_{12} = 0 \\ \lambda_{21} = 0 & \lambda_{22} > 0 \\ \lambda_{31} & \lambda_{32} \\ \lambda_{41} & \lambda_{42} \\ \lambda_{51} & \lambda_{52} \\ \lambda_{61} & \lambda_{62} \end{bmatrix} \begin{matrix} \text{T} \\ \text{BF} \\ \text{FT} \\ \text{R} \\ \text{F} \\ \text{G} \end{matrix}, \qquad \mathbf{\Lambda}_1 = \begin{bmatrix} \lambda_{11} > 0 & \lambda_{12} = 0 \\ \lambda_{21} = 0 & \lambda_{22} > 0 \\ \lambda_{31} & > |\lambda_{32}| \\ \lambda_{41} & > |\lambda_{42}| \\ |\lambda_{51}| & < \lambda_{52} \\ |\lambda_{61}| & < \lambda_{62} \end{bmatrix} \begin{matrix} \text{T} \\ \text{BF} \\ \text{FT} \\ \text{R} \\ \text{F} \\ \text{G} \end{matrix},$$

$$\mathbf{\Lambda}_2 = \begin{bmatrix} \lambda_{11} > 0 & \lambda_{12} = 0 \\ \lambda_{21} = 0 & \lambda_{22} > 0 \\ \lambda_{31} & > |\lambda_{32}| \\ |\lambda_{41}| & < \lambda_{42} \\ |\lambda_{51}| & < \lambda_{52} \\ |\lambda_{61}| & < \lambda_{62} \end{bmatrix} \begin{matrix} \text{T} \\ \text{BF} \\ \text{FT} \\ \text{R} \\ \text{F} \\ \text{G} \end{matrix}, \quad \mathbf{\Lambda}_3 = \begin{bmatrix} \lambda_{11} > 0 & \lambda_{12} = 0 \\ \lambda_{21} = 0 & \lambda_{22} > 0 \\ \lambda_{31} & > |\lambda_{32}| \\ \lambda_{41} & > |\lambda_{42}| \\ \lambda_{51} & > |\lambda_{52}| \\ |\lambda_{61}| & < \lambda_{62} \end{bmatrix} \begin{matrix} \text{T} \\ \text{BF} \\ \text{FT} \\ \text{R} \\ \text{F} \\ \text{G} \end{matrix}.$$

The base model was chosen for its compliance with all competing inequality constrained structures. Structure $\mathbf{\Lambda}_1$ models the factors as distal versus proximal musculature. The tube, fixed tube and reach tasks imply usage of predominantly distal hand musculature, while the bias feed, feed and gesture tasks invoke predominantly proximal hand musculature. Structure $\mathbf{\Lambda}_2$ models the factors as tied to complexity of motor demands. The tube and fixed tube tasks can be seen as more involved in terms of complexity of motor demands on hand musculature. Tasks with higher motor demands are likely to invoke stronger hand preferences than those with lower motor demands. The bias feed, reach, feed, and gesture tasks may be seen as relatively less involved in terms of motor demands. A third plausible theorized explanation for the clustering of tasks may be found in unimanual versus bimanual hand use. Model 3 takes this into account by connecting the tube, fixed tube, reach, and feed tasks to a bimanual factor as these tasks imply (complementary) usage of both hands, while bias feed and gesture are connected to a unimanual factor.

Evaluation

Implementation was analogous to the MBS example. From Table 4.4 it can be seen that all models are of the same complexity. Only model 2 however, exhibits appreciable fit. Model 2 is thus the only model deemed superior to the base model.

Table 4.4. Estimated Bayes factors for constrained models on the chimpanzee handedness data

M_1	M_2	M_3
$\hat{f}_1 =$ 3.699e-2	$\hat{f}_2 =$ .792	$\hat{f}_3 =$ 3.075e-2
$\omega_1 = 2^{-8}$	$\omega_2 = 2^{-8}$	$\omega_3 = 2^{-8}$
$\hat{B}_{10} =$ 9.471	$\hat{B}_{20} =$ 202.629	$\hat{B}_{30} =$ .787

From the Bayes factor matrix given in Table 4.5 it can be seen that, given the data, there is strong evidence in favor of model 2 against both model 1 and 3. Consequently, the factors seem to be connected to motor demands. The estimates for the UCFM in Table 4.6 also comply with model 2.

Table 4.5. Bayes factor matrix for the chimpanzee handedness data

$\hat{B}_{from\|to}$	0	1	2	3
0	1	.106	4.935e-3	1.271
1	9.471	1	4.674e-2	12.034
2	202.629	21.395	1	257.47
3	.787	8.310e-2	3.884e-3	1

Table 4.6. Posterior Means and 95% Credible Intervals for $\mathbf{\Lambda}_0$, diag($\mathbf{\Psi}$) and the free elements in $\mathbf{\Phi}$ on the chimpanzee handedness data

	Parameter	Mean	95% CI	Parameter	Mean	95% CI	Item
$\mathbf{\Lambda}_0$	λ_{11}	.796	[.507, .981]	λ_{12}	-	-	T
	λ_{21}	-	-	λ_{22}	.813	[.299, .987]	BF
	λ_{31}	.701	[.443, .956]	λ_{32}	.116	[-.106, .401]	FT
	λ_{41}	.074	[-.179, .311]	λ_{42}	.423	[.142, .875]	R
	λ_{51}	.048	[-.228, .296]	λ_{52}	.338	[.070, .638]	F
	λ_{61}	-.129	[-.484, .123]	λ_{62}	.378	[.081, .896]	G
$\mathbf{\Psi}, \boldsymbol{\mu}$	ψ_{11}	.369	[.078, .752]	μ_1	-7.671e-3	[-.214, .198]	T
	ψ_{22}	.361	[.079, .940]	μ_2	-8.581e-3	[-.202, .184]	BF
	ψ_{33}	.471	[.086, .783]	μ_3	-9.683e-3	[-.213, .194]	FT
	ψ_{44}	.814	[.372,1.126]	μ_4	-1.042e-2	[-.165, .145]	R
	ψ_{55}	.886	[.646,1.187]	μ_5	-8.132e-5	[-.142, .142]	F
	ψ_{66}	.866	[.184,1.206]	μ_6	-2.245e-3	[-.155, .152]	G
$\mathbf{\Phi}$	ϕ_{12}	.182	[-.075, .437]				

4.7 Discussion

The purpose of this paper was twofold. First, it sought to develop a Bayes factor for the demarcation of competing inequality constrained models. Second, it sought to develop a strategy which allows one to express factor analytic structure using inequality constraints and to subsequently utilize the developed Bayes factor to select among differing inequality constrained formulations of the factor analytic correlation structure. Comments will be made on both endeavors.

Classical inferences for constrained problems often require specialized machinery and algorithms that are not easily generalizable across models. The framework as developed here successfully reduces the Bayes factor for inequality constrained models nested in an unconstrained model to a ratio of probability masses easily computable with an appropriate MCMC sampler. Thus providing for a flexible framework which offers all the advantages of model selection through the Bayes factor. The prior information in this Bayesian framework does not take the form of conventional priors but rather of truncations in noninformative projections of the parameter space. Information on the model in addition to the likelihood is then expressed using testable constraints on the parameter space, rather than through prior information (see also van Wesel, Hoijtink, & Klugkist, 2011, for discussion on this issue).

The strategy consists of choosing as a base model an unrestricted confirmatory factor model, i.e., a factor model that places only minimal restrictions on $\mathbf{\Lambda}$ and $\mathbf{\Phi}$ for achieving global rotational uniqueness of the factor solution, with the restrictions chosen such that they convey preconceived theoretical meaning and thus render unnecessary the post-hoc rotation of the solution for interpretation purposes. Substantive theory is then not represented by (additional) structural exclusions to express a pre-specified loading pattern, but by imposing inequalities on and between the free parameters in $\mathbf{\Lambda}$. Before delving deeper into an evaluation of such a strategy, some tempering remarks are made.

First, the applied researcher intending to use these methods must have a clear idea of the inequalities that are to be assessed, as well as the mode in which the results are to be interpreted. While this is deemed an advantage (careless use of these methods is improbable), it is fair to say that for (very) large problems the formulation of competing inequality constrained models may become computationally intensive when many cross-inequalities are specified. However, one does not have to specify inequalities between or on all free parameters and for large models inequalities may be combined with more sparse formulations of $\mathbf{\Lambda}$.

Relatedly, in this paper the exemplification of inequality constrained formulations of factor structure took the form of specified inequalities on or between loadings in rows of $\mathbf{\Lambda}$. The imposition of a system of inequality constraints is however flexible and also allows for inequalities between loadings within columns and between loadings in both different rows and columns. While such systems may be valid expressions of substantive theory, care should be taken that the feasible space defined by the system of inequality constraints does not get too small as numerical evaluation of such models may imply unfeasible numbers of iterations.

Notwithstanding the tempering remarks, the strategy coupled with the developed model selection criterion has much to offer to FA modeling. The methods as developed here may be viewed as an alternative take on performing CFA studies. Moreover, the methods above may bridge the EFA - CFA divide. Generally, in the CFA model all attention regarding misspecification is geared towards the pre-specified pattern of factor loadings. The evaluation of model fit in CFA is then essentially the evaluation of a diffuse hypothesis (Hoyle & Duvall, 2004) as it is unclear in the case of misspecification if the pattern of loadings or the factor dimensionality is to blame. A separation of dimensionality and pattern selection is proposed that may be seen as a Bayesian inequality constrained CFA analogue of the four-step approach (Mulaik & Millsap, 2000) to structural equation modeling:

1. Embark on evaluating a series of unrestricted models with differing factor dimensionality. For example along the lines specified in Peeters (invited revision a), who uses a Bayesian EFA model for selecting the optimal dimension m;

2. Whence settled on latent factor dimensionality m, specify the UCFM using conditions C1-C4 from Section 4.2.2;
3. Formulate for the UCFM competing inequality constrained factor structures making use of a system of inequality constraints on and between the free parameters in the loadings matrix;
4. Compute for each constrained model the Bayes factor from Proposition 4.2 and determine the constrained model most supported by the data.

Note that the given techniques may easily be extended to include inequality constraints on other model parameters than $\mathbf{\Lambda}$. Also, in an interpretative sense our approach offers many advantages over classical CFA: as (i) Less information is discarded in order to obtain a rotationally unique mode; (ii) Model-translated theories can be expressed using inequality constraints which connect to substantive ideas regarding direction and (relative) strength of effects rather than through (the in many ways restrictive) usage of exclusion restrictions; and (iii) It offers a more feasible way of testing exploratively obtained factor structures using confirmatory techniques.

Part II

Applications & Reflections

5

Pathophysiologic Domains Underlying the Metabolic Syndrome: An Alternative Factor Analytic Strategy

Peeters, C.F.W., Dziura, J. & van Wesel, F. (under review)

5.1 Introduction

Certain risk factors for type 2 diabetes mellitus (DM) and atherosclerotic cardiovascular disease (CVD) have long been observed to cluster together in the individual (Marañon, 1922; Kylin, 1923). This clustering has rendered renewed interest with Reaven's contention (1988, 1993) that insulin resistance forms its basis. Today, the complex of interrelated risk factors of metabolic origin is known as the 'metabolic syndrome' (MBS) (Eckel et al., 2005; Unwin, 2006). It is considered to be a major threat to current and future public health, especially as MBS might result from maladaptive human metabolism in the face of food energy abundance in combination with a sedentary lifestyle (Wilkin & Voss, 2004; Miranda et al., 2005).

Recently, the MBS concept has been hotly debated (Grundy, 2005; Kahn, Buse, Ferrannini, & Stern, 2005; Reaven, 2005a, 2005b; Blaha & Elasy, 2006; Laakso & Kovanen, 2006; Venkat Narayan, 2006; Alberti & Zimmet, 2008; Gale, 2008). A number of interrelated reasons are at the heart of the debate. The aetiology of the syndrome is largely unknown as to date it is unclear if a single pathogenetic process promotes the syndrome, or if multiple different pathogenetic processes need to concur in order for the syndrome to express itself. Notwithstanding the unknown aetiology, a number of expert groups, among which the World Health Organization (WHO, 1999) and the National Cholesterol Education Program - Adult Treatment Panel III (NCEP ATPIII, 2001), have published (slightly) different definitions of MBS intended for clinical diagnosis. However, clinical evidence on whether MBS is a better predictor of CVD and DM risk than its individual components, is equivocal (Kahn et al., 2005; Scuteri, Najjar, Morrell, & Lakatta, 2005).

The stance on the MBS concept taken here is the following. The aetiology does not have to be known for the existence of a condition to be accepted (Gale, 2008), as is the case with, for example, type 2 DM. Current knowledge is however limited, such that MBS is not considered a clinical entity, but rather as defining a state of heightened risk for DM and CVD. The syndromic approach taken is epidemiological rather than clinical. In the sense that MBS is deemed to provide a conceptual framework for the clustering of metabolic risk factors. An important step in furthering epidemiologic understanding of the syndrome is then an assessment of the pathophysiologic constellation of what are deemed to be phenotypic expressions of MBS.

Factor analysis (FA) has become an oft-used tool for evaluations of phenotypic domains underlying MBS (Meigs, 2000). FA is a multivariate technique that may reveal a pattern of reduced dimensionality among a larger set of intercorrelated variables (Mulaik, 2010). FA, however, is often poorly understood while being routinely executed (Lawlor, Ebrahim, May, & Smith, 2004), leading to diffuse findings and possibly confusing efforts in understanding MBS. This paper aims to review factor analytic efforts in the study of MBS, with emphasis on misusage of FA. An alternative factor analytic strategy is proposed that confronts weaknesses in the application of FA. A high-profile MBS data set with anthropometric measurements on overweight and obese children and adolescents is reanalyzed using the alternative strategy. The findings may give renewed cachet to both FA and its connection to MBS research.

5.2 Assessing Factor Analytic Efforts in MBS Research

5.2.1 The Factor Analytic Model

FA has come to be heavily utilized in the MBS research community since its seminal usage by Edwards, et al. (1994). An important question thrusting the FA efforts in MBS research is if a unifying physiology dominated by insulin resistance underlies the clustering of metabolic risk variables, or if there are multiple underlying physiologic phenotypes.

The common factor analytic model assumes that a random p-dimensional vector of observed variables can be grouped by their covariances or correlations into a lower-dimensional linear combination of latent variables:

$$\underset{(p\times 1)}{\mathbf{z}_i} = \underset{(p\times 1)}{\boldsymbol{\mu}} + \underset{(p\times m)}{\boldsymbol{\Lambda}} \cdot \underset{(m\times 1)}{\boldsymbol{\xi}_i} + \underset{(p\times 1)}{\boldsymbol{\epsilon}_i}. \tag{5.1}$$

In (5.1) $\mathbf{z}_i$ denotes the (possibly standardized) observed variable of dimension p for person i, $\boldsymbol{\mu}$ denotes the intercept, $\boldsymbol{\epsilon}_i$ denotes the error measurements for person i, and $\boldsymbol{\Lambda}$ is a $(p \times m)$-dimensional matrix of factor loadings in which each element λ_{jk} is the loading of the jth variable on the kth factor, $j = 1, \ldots, p$, $k = 1, \ldots, m$. Then $\boldsymbol{\xi}_i$ represents a latent variable of dimension m, with $m < p$, whose elements are referred to as common factors. In effect, FA represents a method of identifying or specifying latent factors that account for the (co)variances among observed variables by partitioning observed variance into common variance (attributable to the underlying latent common factors) and unique variance (among which are error components; Mulaik, 2010). In the standard model the random variables $\mathbf{z}_i$,

$\boldsymbol{\xi}_i$, and $\boldsymbol{\epsilon}_i$ are assumed to have Gaussian distributions. (See Appendix 5A for a more detailed overview of the FA model). The model can have both explorative and confirmative thrusts. Before assessing these thrusts some general remarks on FA are made.

A first general comment on the utilization of FA by the MBS research community is that basic model assumptions are seldom assessed. The model boasts several implicit assumptions such as a nonsingular sample covariance matrix and a reasonable proportion of variance among the observed variables being common variance. The appropriateness of these assumptions is easily (Kaiser, 1970) but rarely assessed. Moreover, the explicit distributional assumptions imply usage of observed variables of continuous metric and disqualifies the common FA model for binary and categorical observed data. Many MBS studies, however, employ standard FA on variables of non-continuous metric. In such situations extensions of the standard FA model are needed (Jöreskog, 2007; Moustaki, 2007).

A second general comment concerns interpretational overextension. The FA model cannot determine existence of MBS nor assess clinical importance of MBS as a concept (Lawlor et al., 2004). What FA *can* do is, through the latent factors, give indications of pathophysiologic domains that underlie phenotypic expressions of MBS.

5.2.2 Comments on Exploratory Efforts

In exploratory FA (EFA) both m and the meaning of latent factors are unknown. In the exploratory sense, FA is a theory-generating technique used for the identification of meaningful latent factors. Most MBS studies utilize FA in the exploratory sense (Hanley et al., 2002; Hanson, Imperatore, Bennett, & Knowler, 2002; Ford, 2003; M. Lambert et al., 2004; J.-Y. Oh, Hong, Sung, & Barrett-Connor, 2004; Ang et al., 2005; A. Ghosh, 2005; Corsetti, Rainwater, Moss, Zareba, & Sparks, 2006; Mannucci, Monami, & Rotella, 2007; Huang et al., 2008; Khader et al., 2011). Many deployments of EFA are however suboptimal.

EFA is often confused with principal components analysis (PCA). PCA is a data-reductional technique, seeking to identify components. It resembles the model in (5.1) without inclusion of error measurements, leading the components to differ conceptually and mathematically from latent factors in EFA (Widaman, 2007). Components are weighted linear combinations of observed variables seeking to efficiently explain observed variance in the data (Jolliffe, 2002), leaving the explanation of observed covariance secondary (Widaman, 2007). Many MBS studies employing FA, however, seek to obtain an explanation of the observables' *co*variation through a small number of explanatory factors. Also, a phenotype is just an expression of genotype or pathophysiology, indicating the necessity of including measurement error. As such, employing common EFA would be more appropriate.

An important decision in EFA is determining the number of factors to retain. Many methods are heuristical, relying on subjective judgments or arbitrary cut-off values. The Guttman-Kaiser rule (Guttman, 1954; Kaiser, 1960) is the most popular rule of thumb. It states that one should retain at most those factors associated with eigenvalues whose magnitude exceeds the average eigenvalue (the average eigenvalue being 1 when using standardized data). This criterion, as well as many other heuristical factor retention criteria, are prone to under- (retaining too few factors) and overfactoring (retaining too many factors) (Yeomans & Golder, 1982; Cliff, 1988).

Given the above it is disconcerting that factor analytic efforts in MBS research are usually based on what has been termed 'Little jiffy' (Kaiser, 1970): Employment of PCA, retainment of components based on the Guttmann-Kaiser rule followed by a Varimax rotation (Kaiser, 1958), and the subsequent interpretation of rotated components as if they were common factors. Such mechanical use of EFA stunts learning and interpretation (Preacher & MacCallum, 2003), and is (in part) responsible for the widely differing results obtained with EFA in MBS research.

5.2.3 Comments on Confirmatory Efforts

Confirmatory FA (CFA) is a theory-testing technique. An *a priori* factor structure is assumed, with given m, with a pre-specified loadings matrix in which exclusion constraints indicate which variables are indicators of which latent factor(s), and with possibly correlated factors and error variances. The model or models stated then remain to be tested. CFA studies are gaining interest in MBS research (Shen et al., 2003; Pladevall et al., 2006; Shah, Novak, & Stapleton, 2006; Chaoyang & Ford, 2007; Goodman et al., 2009; Boronat, Saavedra, Varillas, & Nóvoa, 2009; Solera-Martínez et al., 2011). Standard CFA however, can also be misapplied.

The evaluation of model fit in CFA is essentially the evaluation of a diffuse hypothesis as it is unclear in case of misspecification if the pattern of loadings or the factor dimensionality is to blame (Hoyle & Duvall, 2004). Moreover, specifying a pattern of factor loadings through exclusion restrictions implies a loss of information in the sense that more exclusion restrictions are applied than is usually necessary for identification of the FA model. Additionally, exclusion restrictions may amount to errors of omission, may make the unrealistic assumption that items are factorially pure (in the population), and may induce bias in estimates of the free parameters (Ferrando & Lorenzo-Seva, 2000; van Prooijen & van der Kloot, 2001). These issues are intricately connected to the well-known and widespread situation of exploratively obtained factor structures not being confirmed by CFA (Ferrando & Lorenzo-Seva, 2000).

Some recent CFA studies in MBS research claim to provide evidence that a single latent factor underlies MBS (Pladevall et al., 2006; Chaoyang & Ford, 2007; Solera-Martínez et al., 2011). These studies include four observed variables, some of which are functions of several variables usually employed as separate phenotypic items. This practice is justified by the claim that utilization of multiple measures for what is believed to be a trait will lead to a model with multiple factors, thus clouding efforts to establish a single latent factor underlying MBS. However, the inclusion of multiple sets of correlated measures does not irrevocably lead to a model with more than one latent factor, unless some measures would identify a doublet factor (see Section 'Step 1: Dimensionality Selection' below). Also, the mentioned CFA efforts may actually provide evidence for a hierarchical latent factor (Shen et al., 2003) rather than a single pathophysiological domain underlying MBS, as the usage of functions of variables implies a (partial) pre-compression of the data. Moreover, while a two-factor model can be modeled on four variables, such a model would not be meaningful given that the observables were constructed to represent a single factor; implying that the one-factor model is the only model to be meaningfully fitted on the four observed variables. The possibility to assess if there are multiple (related) pathophysiological domains underlying phenotypic expressions of MBS is then denied. A meaningful scientific method, however, allows for multiple competing theories to be tested (Platt, 1964), as single hypotheses suffer from confirmation bias.

5.3 An Alternative Factor Analytic Strategy

An alternative confirmatory factor analytic strategy is proposed that aims to confront the weaknesses in the application of FA. In a sense the strategy seeks to bridge the EFA - CFA divide so as to increase the inferential power of a factor analysis. This strategy is embedded within the Bayesian model selection approach, whose analytical advantages have been well documented (Dunson, 2001). The main Bayesian model selection criterion is the Bayes factor. This quantity incorporates model fit as well as model complexity and expresses "the evidence provided by the data in favor of one scientific theory, represented by a statistical model, as opposed to another" (Kass & Raftery, 1995). It can be used to compare any two models. Using Bayes factors, one can compute posterior model probabilities, which, assuming a uniform prior on the model space, are normalized Bayes factors. These quantities express model (un)certainty, in the sense that posterior model probabilities can be interpreted as the relative amounts of support in the data for the models under consideration.

The strategy proposed consists of the following steps:

Step 1. Determine formally the (intrinsic) number of common factors based on a weighing of model fit and model complexity. For example along the lines specified in Peeters (invited revision a), who uses a Bayesian EFA model for selecting the optimal dimension m;

Step 2. Whence settled on latent factor dimensionality m, specify an unrestricted confirmatory factor model (UCFM). A UCFM is a FA model that corresponds to EFA in the sense that only minimal restrictions are placed on the factor loadings matrix and the factor covariance matrix for achieving global rotational uniqueness of the factor solution. However, the restrictions are to be chosen such that they convey preconceived theoretical meaning and thus render unnecessary post-hoc rotation of the solution for interpretation purposes. Peeters (2012) gives minimal conditions for specifying a UCFM;

Step 3. Formulate, using the UCFM obtained in Step 2 as a base model, competing inequality constrained factor structures making use of inequality constraints on and between the free parameters in the loadings matrix. Substantive theory is then not represented by exclusion restrictions to express a pre-specified factor loading pattern, but by the imposition of inequality constraints;

Step 4. Compute the posterior model probability for each constrained model under consideration and determine the constrained model most supported by the data.

The strategy explicitly encourages the formulation of competing inequality constrained theories for (statistical) scrutiny. When used in full, the thrust of the sequence is confirmatory, with the explicit separation of dimensionality and pattern selection in order to avoid embarking on diffuse hypotheses.

5.4 Reanalysis of Data by Weiss et al. (2004)

5.4.1 Data

The data have been described elsewhere (Weiss et al., 2004). It considers a multiethnic, multiracial cohort of 464 non-diabetic obese and overweight children and adolescents. The data contain measurements on the body mass index (BMI), blood glucose level at (fasting) baseline (GB) and two hours after (G2) oral glucose intake (both in mg/dl), fasting levels of triglycerides (trig.; mg/dl) and high-density lipoprotein (HDL) cholesterol (mg/dl), systolic and diastolic blood pressure (SBP, DBP; both in mm Hg), and insulin resistance (IR). IR was measured through homeostatic model assessment (HOMA). For more information on the data, see Weiss et al. (2004).

Consideration of pediatric samples is important as current prevalence of MBS among youngsters may give indications of the future burden of DM and CVD. The measurements are in line with the American Academy of Clinical Endocrinologists (AACE) position on MBS (Einhorn et al., 2003), which emphasizes the epidemiologic pathophysiological perspective. The inclusion of IR makes it possible to test theories regarding the importance of IR in the MBS construct.

The little jiffy approach was the original factor analytic strategy for analyzing the data (Weiss et al., 2004). Here, the alternative strategy will be utilized for reanalysis. As in Weiss et al. (2004), the natural logarithm was taken of the glucose, insulin resistance and triglycerides measurements to abide the normality assumption. The data were standardized such that a case of modeling the correlation matrix is considered. The sample correlation matrix is nonsingular and the Kaiser-Meyer-Olkin test (Kaiser, 1970) indicates that a reasonable proportion of variance among the variables might be common variance.

5.4.2 Step 1: Dimensionality Selection

Posterior model probabilities are computed for each model allowed by the condition $(p-m)^2 - p - m \geqslant 0$. This inequality simply states that the number of nonredundant elements in the sample covariance matrix must be greater than or equal to the number of freely estimable parameters in the model, which places an upper bound on m. The data have eight measured variables ($p = 8$), giving that the maximum number of factors that can be subtracted equals $m = 4$. The computation strategy couples the candidate estimator method for computing Bayes factors (Chib, 1995) with the use of training samples (Berger & Pericchi, 1996). This practice allows one to obtain determinate Bayes factors and subsequent posterior model probabilities using standard diffuse conjugate or noninformative priors (S. Y. Lee, 2007). The computation strategy is embedded within a search strategy too weed out models suffering from rank deficiency in $\mathbf{\Lambda}$ as this is a direct indicator for overfactoring (see (5.4) in Appendix 5A). This search excludes the $m = 4$ model. The following posterior model probabilities are obtained when assuming that each model is equally likely *a priori*: $P(m = 1|\mathbf{Z}) = 0$; $P(m = 2|\mathbf{Z}) = 1$; and $P(m = 3|\mathbf{Z}) = 0$. The data thus support the two-factor model.

These results differ from Weiss et al. (2004) and several other factor analytic efforts in which a three-factor (or higher) solution was found. A first reason for the retention of more latent factors in these studies is the tendency of heuristic factor-selection rules to overfactor (Yeomans & Golder, 1982; Cliff, 1988). More formal selection procedures, such as the likelihood ratio test in maximum likelihood EFA and the assessment of information criteria, do not escape this tendency (Lopes & West, 2004; Hayashi et al., 2007).

Another reason for the higher factor solution in other studies might be the existence of doublet factors. Doublet factors are factors that arise as the result of common variance due to correlation between just two variables (Mulaik, 2010). Doublet factors are considered to be conceptually weak factors and the assignment of an independent latent construct to a doublet factor is contentious. SBP and DBP are variables that, within the battery of measurements on phenotypic expressions of MBS, usually correlate only with each other resulting in a doublet factor to be extracted (usually termed 'hypertension'). The strategy employed here sees past the doublet factor and indicates the more parsimonious model with two factors as optimal.

5.4.3 Step 2: Base Model Formulation

A UCFM for $m = 2$ will be formulated for confirmatory efforts. Abiding conditions given by Peeters (2012), the following minimal restrictions on $\mathbf{\Lambda}$ are chosen for global rotational uniqueness:

$$\mathbf{\Lambda}_0 = \begin{bmatrix} \lambda_{11} & \lambda_{12} \\ \lambda_{21} & \lambda_{22} \\ \lambda_{31} = 0 & \lambda_{32} > 0 \\ \lambda_{41} & \lambda_{42} \\ \lambda_{51} > 0 & \lambda_{52} = 0 \\ \lambda_{61} & \lambda_{62} \\ \lambda_{71} & \lambda_{72} \\ \lambda_{81} & \lambda_{82} \end{bmatrix} \begin{matrix} \text{BMI} \\ \log_e\{\text{trig.}\} \\ \text{HDL chol.} \\ \log_e\{\text{IR}\} \\ \log_e\{\text{GB}\} \\ \log_e\{\text{G2}\} \\ \text{SBP} \\ \text{DBP} \end{matrix} .$$

The exclusion restrictions $\{\lambda_{31}, \lambda_{52}\} = 0$ identify the model up to polarity reversals in the columns. The polarity truncations $\{\lambda_{32}, \lambda_{51}\} > 0$ then ensure global rotational uniqueness of the model. These constraints are chosen for the following reasons. First, from prior knowledge and previous analyzes a two-factor solution is deemed to consist of a glucose and a lipid factor. HDL chol. is then believed to have a large loading on a lipid (second) factor while having a small loading on the first factor, and $\log_e\{\text{GB}\}$ is believed to have a large loading on a glucose (first) factor while having a small loading on the second factor. These variables then serve as an indicator of the respective factors. It is thus reasonable to specify $\{\lambda_{31}, \lambda_{52}\} = 0$ and $\{\lambda_{32}, \lambda_{51}\} > 0$. Second, the chosen minimal restrictions comply with all competing inequality constrained formulations of factor structure to be assessed.

5.4.4 Step 3: Formulating Competing Constrained Factor Structures

Factor structure for confirmatory efforts is not represented using exclusion restrictions but by imposing inequality constraints on and between the parameters left free in the UCFM. The following inequality constrained competing factor structures are formulated:

$$\mathbf{\Lambda}_1 = \begin{bmatrix} \lambda_{11} & > & |\lambda_{12}| \\ |\lambda_{21}| & < & -\lambda_{22} \\ \lambda_{31} = 0 & & \lambda_{32} > 0 \\ \lambda_{41} & > & |\lambda_{42}| \\ \lambda_{51} > 0 & & \lambda_{52} = 0 \\ \lambda_{61} & > & |\lambda_{62}| \\ \lambda_{71} & > & |\lambda_{72}| \\ |\lambda_{81}| & < & -\lambda_{82} \end{bmatrix} \begin{matrix} \text{BMI} \\ \log_e\{\text{trig.}\} \\ \text{HDL chol.} \\ \log_e\{\text{IR}\} \\ \log_e\{\text{GB}\} \\ \log_e\{\text{G2}\} \\ \text{SBP} \\ \text{DBP} \end{matrix},$$

$$\mathbf{\Lambda}_2 = \begin{bmatrix} |\lambda_{11}| & < & -\lambda_{12} \\ |\lambda_{21}| & < & -\lambda_{22} \\ \lambda_{31} = 0 & & \lambda_{32} > 0 \\ \lambda_{41} > .4 & & \lambda_{42} < -.4 \\ \lambda_{51} > 0 & & \lambda_{52} = 0 \\ \lambda_{61} & > & |\lambda_{62}| \\ \lambda_{71} & < & -\lambda_{72} \\ \lambda_{81} & < & -\lambda_{82} \end{bmatrix} \begin{matrix} \text{BMI} \\ \log_e\{\text{trig.}\} \\ \text{HDL chol.} \\ \log_e\{\text{IR}\} \\ \log_e\{\text{GB}\} \\ \log_e\{\text{G2}\} \\ \text{SBP} \\ \text{DBP} \end{matrix},$$

$$\mathbf{\Lambda}_3 = \begin{bmatrix} \lambda_{11} & > & |\lambda_{12}| \\ |\lambda_{21}| & < & -\lambda_{22} \\ \lambda_{31} = 0 & & \lambda_{32} > 0 \\ \lambda_{41} & > & |\lambda_{42}| \\ \lambda_{51} > 0 & & \lambda_{52} = 0 \\ \lambda_{61} & > & |\lambda_{62}| \\ |\lambda_{71}| < .3 & & |\lambda_{72}| < .3 \\ |\lambda_{81}| < .3 & & |\lambda_{82}| < .3 \end{bmatrix} \begin{matrix} \text{BMI} \\ \log_e\{\text{trig.}\} \\ \text{HDL chol.} \\ \log_e\{\text{IR}\} \\ \log_e\{\text{GB}\} \\ \log_e\{\text{G2}\} \\ \text{SBP} \\ \text{DBP} \end{matrix}.$$

A formulation like $\lambda_{71} < -\lambda_{72}$ states that the negative of λ_{72} is believed to be larger than λ_{71}. Note that this is a much more informative formulation than the more usual strategy of setting $\lambda_{71} = 0$ and letting λ_{72} be free to be estimated in order to express the belief that SBP is an indicator for the second latent factor rather than the first one. In the same respect, a formulation like

$$\lambda_{61} > |\lambda_{62}| \Rightarrow \begin{cases} \lambda_{61} - \lambda_{62} > 0 \\ \lambda_{61} + \lambda_{62} > 0 \end{cases},$$

indicates the belief that λ_{61} is larger than λ_{62}, irrespective of the latter's sign. A statement like

$$|\lambda_{82}| < .3 \Rightarrow -.3 < \lambda_{82} < .3,$$

indicates the belief that λ_{82} takes a value in the interval $[-.3, .3]$.

In all models insulin resistance, blood glucose level at baseline and two hours after glucose intake are mainly related to the glucose factor, while fasting levels of triglycerides and HDL cholesterol form the base of the lipid factor. Note that for the lipid factor polarity fixation is brought about by demanding $\lambda_{32} > 0$. HDL cholesterol is generally regarded as 'good' cholesterol, meaning that the choice $\lambda_{32} > 0$ amounts to modeling a factor denoting unimpaired lipid metabolism. Hence formulations like $|\lambda_{21}| < -\lambda_{22}$, as under given polarity truncation the triglycerides item is believed to be strongly negatively related to a lipid factor.

Model 1 adds detail to the base model and the basic factors by stating that BMI and systolic blood pressure are linked to the glucose factor ($\lambda_{11} > |\lambda_{12}|$, $\lambda_{71} > |\lambda_{72}|$), while diastolic blood pressure is believed to be linked to the lipid factor ($|\lambda_{81}| < -\lambda_{82}$). Model 2 states the hypothesis that BMI is an indicator for the lipid rather than the glucose factor ($|\lambda_{11}| < -\lambda_{12}$). Also, in this model both systolic and diastolic blood pressure are related to the lipid rather than the glucose factor, with the additional belief that both blood pressure measures will load positively on the latter ($\lambda_{71} < -\lambda_{72}$, $\lambda_{81} < -\lambda_{82}$). Moreover, the second model states that insulin resistance may be the measure tying MBS together. In a multifactor model this would imply that the insulin resistance measure achieves a large or dominant loading on both the glucose and lipid factor ($\lambda_{41} > .4$, $\lambda_{42} < -.4$). Model 3 resembles the first model, but states that the association of systolic and diastolic blood pressure with the factors is rather loose.

5.4.5 Step 4: Constrained-Model Selection and Interpretation

Bayes factors for models under inequality constraints are easily computed (see, e.g., Klugkist & Hoijtink, 2007). Again, standard (diffuse) conjugate and noninformative priors are utilized. Assuming a uniform prior on the model space the following posterior model probabilities are obtained for the constrained two-factor models under consideration: $P(M_1|\mathbf{Z}) = .0004$; $P(M_2|\mathbf{Z}) = 0$; and $P(M_3|\mathbf{Z}) = .9996$. Conditioned on the data, the third model receives almost all support.

The third constrained model connects trig. and HDL chol. to a lipid metabolism factor and IR, GB, and G2 to a glucose metabolism factor. Moreover, the model states that BMI is related to the glucose metabolism factor rather than the lipid metabolism factor. The finding that blood pressure is not an independent pathophysiologic factor is consistent with epidemiologic evidence that insulin resistance and lipid metabolism play a role in the pathogenesis of hypertension rather than hypertension being a physiologic phenotype (Liese et al., 1998).

Table 5.1. Posterior Means and 95% Credible Intervals for $\mathbf{\Lambda}_0$

Factor 1			Factor 2			
Par.	Mean	95% CI	Par.	Mean	95% CI	Item
λ_{11}	.324	[.207, .440]	λ_{12}	-.068	[-.191, .055]	BMI
λ_{21}	-.006	[-.303, .215]	λ_{22}	-.653	[-.956,-.379]	$\log_e$\{trig.\}
λ_{31}	-	-	λ_{32}	.706	[.442, .940]	HDL chol.
λ_{41}	.767	[.613, .921]	λ_{42}	-.179	[-.343,-.022]	$\log_e$\{IR\}
λ_{51}	.470	[.360, .585]	λ_{52}	-	-	$\log_e$\{GB\}
λ_{61}	.355	[.205, .492]	λ_{62}	-.124	[-.289, .036]	$\log_e$\{G2\}
λ_{71}	.274	[.136, .416]	λ_{72}	.029	[-.118, .171]	SBP
λ_{81}	.202	[.069, .347]	λ_{82}	.139	[-.017, .292]	DBP

The estimates of the UCFM given in Table 5.1 also lend support to model 3. The table contains posterior means and credible intervals. A credible interval is a posterior probability interval. For example, the posterior probability that λ_{11} lies in the interval [.207, .440] is .95. (See Appendix 5B for an indication of the success of the two-factor UCFM in reproducing the sample correlation matrix).

The estimates indicate that the hypertension variables seem to be relatively weak in their association with the respective factors. The credibility intervals indicate that these variables are mostly related to the glucose metabolism factor, which would be in line with the hypothesis that hypertension is related to insulin resistance and impaired glucose metabolism (Reaven, 1988, 1993). Regarding the contention that insulin resistance is the basis for MBS: IR is related to both the glucose and lipid factors. However, the posterior mean of the loading tying IR to the lipid metabolism factor is relatively small and the upper bound of its credibility interval approaches zero. The second inequality constrained model is thus rightly not supported by the data.

The two latent factors are appreciably correlated with a posterior mean of $-.277$ and a 95% credible interval of $[-.417, -.137]$. Note that, as stated in Section 5.4.4, the second factor models unimpaired lipid metabolism. Thus, the selected model indicates that, given the data and measurements, impaired glucose metabolism and impaired lipid metabolism are positively related pathophysiologic domains. Moreover, the two factors connect to two main hypotheses regarding syndrome aetiology, stating that the risk-factor associations are due to abnormality of the insulin/glucose metabolism and/or abnormality of the lipid metabolism (Liese et al., 1998).

5.5 Discussion

Applications of FA in MBS research were evaluated. Lacunae in both EFA and CFA were discussed. It is argued that the mechanical use of EFA and misunderstandings of CFA are, at least in part, responsible for the widely differing results obtained with FA on data with phenotypic expressions of MBS.

An alternative factor analytic strategy is proposed. The strategy consists of four steps and aims to confront the weaknesses in application of the FA model, by: (i) Formally assessing optimal choice of factor dimensionality; (ii) Canceling the need for post-hoc rotation of the factor solution; (iii) Allowing to express a confirmatory factor structure through informative inequality constraints rather than through rigid exclusion restrictions; (iv) Encouraging the formulation of competing inequality constrained theory-based expressions of factor structure, in order to avoid confirmation bias.

The alternative strategy was utilized in reanalyzing a high-profile data set on which factor analyzes were previously employed. The data consider eight variables as phenotypic expressions of MBS in a cohort of non-diabetic overweight and obese children and adolescents (Weiss et al., 2004). The reanalysis based on the alternative strategy implied a more parsimonious constellation of pathophysiologic domains underlying phenotypic expressions of MBS than the original analysis (and many other analyzes). The selected two-factor solution stresses correlated factors of impaired glucose metabolism and impaired lipid metabolism. This solution does not assign hypertension a separate factor which is consistent with epidemiologic evidence that insulin resistance and lipid metabolism play a role in the pathogenesis of hypertension rather than hypertension being a physiologic phenotype (Liese et al., 1998). Moreover, there is no strong evidence of insulin resistance being dominant in both the glucose and lipid domains.

Several limitations of this study should be considered. First, the data consider a multiethnic cohort while MBS may express itself differently across ethnic groups and gender. Note, however, that the specification of a factor model through inequality constraints would also be helpful in assessing measurement (factorial) invariance across groups. Second, MBS may develop with age and with the advent of DM and CVD, implying that the data might represent a snap-shot of phenotypic expressions related to MBS. Third, the proposed factor analytic strategy is more involved than routine uses of EFA and regular CFA in the sense that it requires more computation time and puts higher cognitive demands on the researcher. Nevertheless, these drawbacks are felt to be outweighed by the advantages of the strategy and the proposed steps are deemed to form a viable analytic alternative for other studies seeking to use FA.

The findings suggest that there are two correlated pathophysiologic domains underlying the phenotypic expressions of MBS included in the analysis. These domains are characterized by impaired glucose metabolism and impaired lipid metabolism, respectively. These findings indirectly point to the possibility that several different pathogenic processes need to coincide in order to be able to identify a MBS construct. It might be timely to postulate the possible existence of a multifactorial aetiology.

5A Basics of Factor Analysis

The unrestricted factor model is considered. Let $\mathbf{Z}^{\mathrm{T}} \equiv [\mathbf{z}_1, \ldots, \mathbf{z}_n]$ define (standardized) p-variate observation vectors on $i = 1, \ldots, n$ subjects, such that $\mathbf{z}_i^{\mathrm{T}} \equiv [z_{i1}, \ldots, z_{ip}] \in \mathbb{R}^p$ denotes a realization of the random vector $Z_i^{\mathrm{T}} \equiv [Z_{i1}, \ldots, Z_{ip}] \in \mathbb{R}^p$. Also, let $\boldsymbol{\Xi}^{\mathrm{T}} \equiv [\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_n]$ define m-variate vectors of latent factor scores on n subjects with $\boldsymbol{\xi}_i^{\mathrm{T}} \equiv [\xi_{i1}, \ldots, \xi_{im}] \in \mathbb{R}^m$.

The model (5.1) maintains the following assumptions: (i) $\mathbf{z}_i \perp\!\!\!\perp \mathbf{z}_{i'}, \forall i \neq i'$; (ii) $\operatorname{rank}(\boldsymbol{\Lambda}) = m$; (iii) $\boldsymbol{\epsilon}_i \sim \mathcal{N}_p(\mathbf{0}, \boldsymbol{\Psi})$, with $\boldsymbol{\Psi} \equiv \operatorname{diag}(\psi_{11}, \ldots, \psi_{pp})$, and $\psi_{jj} > 0$; (iv) $\boldsymbol{\xi}_i \sim \mathcal{N}_m(\mathbf{0}, \boldsymbol{\Phi})$; and (v) $\boldsymbol{\xi}_i \perp\!\!\!\perp \boldsymbol{\epsilon}_{i'}, \forall i, i'$. The likelihood for the observations conditional on the realization of $\boldsymbol{\Xi}$ can then be expressed as:

$$
\begin{aligned}
L(\boldsymbol{\mu}, \boldsymbol{\Lambda}, \boldsymbol{\Xi}, \boldsymbol{\Psi}, \boldsymbol{\Phi}; \mathbf{Z}) &= \prod_{i=1}^{n} f(\mathbf{z}_i | \boldsymbol{\mu}, \boldsymbol{\Lambda}, \boldsymbol{\xi}_i, \boldsymbol{\Psi}, \boldsymbol{\Phi}) \\
&= \prod_{i=1}^{n} (2\pi)^{-\frac{p}{2}} |\boldsymbol{\Psi}|^{-\frac{1}{2}} \exp\left\{ -\frac{1}{2} \boldsymbol{\epsilon}_i^{\mathrm{T}} \boldsymbol{\Psi}^{-1} \boldsymbol{\epsilon}_i \right\},
\end{aligned}
$$

where $\boldsymbol{\epsilon}_i = \mathbf{z}_i - \boldsymbol{\mu} - \boldsymbol{\Lambda}\boldsymbol{\xi}_i$. Marginalizing over $\boldsymbol{\xi}_i$ the likelihood of the observed data can be obtained:

$$
\begin{aligned}
&L(\boldsymbol{\mu}, \boldsymbol{\Lambda}, \boldsymbol{\Psi}, \boldsymbol{\Phi}; \mathbf{Z}) \\
&\quad = \prod_{i=1}^{n} \int f(\mathbf{z}_i | \boldsymbol{\mu}, \boldsymbol{\Lambda}, \boldsymbol{\xi}_i, \boldsymbol{\Psi}, \boldsymbol{\Phi}) g(\boldsymbol{\xi}_i | \boldsymbol{\Phi}) \, \partial \boldsymbol{\xi}_i \\
&\quad = \prod_{i=1}^{n} (2\pi)^{-\frac{p}{2}} |\boldsymbol{\Lambda}\boldsymbol{\Phi}\boldsymbol{\Lambda}^{\mathrm{T}} + \boldsymbol{\Psi}|^{-\frac{1}{2}} \exp\left\{ -\frac{1}{2} (\mathbf{z}_i - \boldsymbol{\mu})^{\mathrm{T}} [\boldsymbol{\Lambda}\boldsymbol{\Phi}\boldsymbol{\Lambda}^{\mathrm{T}} + \boldsymbol{\Psi}]^{-1} (\mathbf{z}_i - \boldsymbol{\mu}) \right\},
\end{aligned}
$$

giving that the factor decomposition constrains the covariance structure of the $\mathbf{z}_i$ to

$$
\boldsymbol{\Sigma} = \boldsymbol{\Lambda}\boldsymbol{\Phi}\boldsymbol{\Lambda}^{\mathrm{T}} + \boldsymbol{\Psi}. \tag{5.2}
$$

Then, for existence (vi), generally $(p - m)^2 - p - m \geqslant 0$, simply stating that the number of nonredundant elements in the sample correlation matrix $\mathbf{S}$ must be greater than or equal to the number of freely estimable parameters in $\boldsymbol{\Sigma}$, which places an upper bound on m.

Now, $\boldsymbol{\Phi} \in \mathbb{R}^{m \times m}$ denotes the factor covariance matrix, giving that (5.2) represents an oblique model in which the latents may share covariation. Note that, for positive definite $\boldsymbol{\Phi}$, we may always find $\mathbf{V} \in \mathbb{R}^{m \times m}$ such that $\boldsymbol{\Phi} = \mathbf{V}\mathbf{V}^{\mathrm{T}}$, and

$$\mathbf{\Sigma} = \mathbf{\Lambda\Phi\Lambda}^{\mathrm{T}} + \mathbf{\Psi} = (\mathbf{\Lambda V})[\mathbf{V}^{-1}\mathbf{\Phi}(\mathbf{V}^{-1})^{\mathrm{T}}](\mathbf{\Lambda V})^{\mathrm{T}} + \mathbf{\Psi} = (\mathbf{\Lambda V})(\mathbf{\Lambda V})^{\mathrm{T}} + \mathbf{\Psi}. \quad (5.3)$$

Equation (5.3) implies that any oblique representation has equivalent orthogonal representations. The orthogonal representation makes the following statements on identification less involved.

It is well known that for given $\mathbf{\Lambda}$ and $\mathbf{\Psi}$, the former is defined uniquely only up to rotation. Correspondingly the FA literature has focussed mainly on identification of $\mathbf{\Psi}$. The main result of which is that if assumption (vi) holds, $\mathbf{\Psi}$ is almost surely identified (Bekker et al., 1994). This result is however contingent upon the rank of $\mathbf{\Lambda}$. The implications of a failure to abide model assumption (ii) were explored by Anderson and Rubin (1956) and Geweke and Singleton (1980). Suppose that $\text{rank}(\mathbf{\Lambda}) = r < m$. Then there exists a matrix $\mathbf{Q} \in \mathbb{R}^{m \times (m-r)}$ for which $(\mathbf{\Lambda V})\mathbf{Q} = \mathbf{0}$ and $\mathbf{Q}^{\mathrm{T}}\mathbf{Q} = \mathbf{I}_{m-r}$, such that for any $\mathbf{M} \in \mathbb{R}^{p \times (m-r)}$ with mutually orthogonal rows

$$\mathbf{\Sigma} = (\mathbf{\Lambda V})(\mathbf{\Lambda V})^{\mathrm{T}} + \mathbf{\Psi} = (\mathbf{\Lambda V} + \mathbf{MQ}^{\mathrm{T}})(\mathbf{\Lambda V} + \mathbf{MQ}^{\mathrm{T}})^{\mathrm{T}} + (\mathbf{\Psi} - \mathbf{MM}^{\mathrm{T}}). \quad (5.4)$$

Equation (5.4) implies that no consistent estimator of $\mathbf{\Psi}$ exists if $\mathbf{\Lambda}$ fails to be of full column rank. This may induce corresponding multimodalities in the densities of $\mathbf{\Psi}$ and $\mathbf{\Lambda}$ (Lopes & West, 2004), and is related to the choice of factor dimensionality and the possibility of retaining too many factors.

The FA model also copes with an inherent indeterminacy of the parameters, being: Rotational indeterminacy of the factor solution. Assume that $\mathbf{R} \in \mathbb{R}^{m \times m}$ is an arbitrary nonsingular matrix. Returning to the implied covariance structure of the observed data, we then have

$$\mathbf{\Sigma} = \mathbf{\Lambda\Phi\Lambda}^{\mathrm{T}} + \mathbf{\Psi} = (\mathbf{\Lambda R})[\mathbf{R}^{-1}\mathbf{\Phi}(\mathbf{R}^{\mathrm{T}})^{-1}](\mathbf{\Lambda R})^{\mathrm{T}} + \mathbf{\Psi}, \quad (5.5)$$

implying that there is an infinite number of alternative matrices $\mathbf{\Lambda}^{\ddagger} = \mathbf{\Lambda R}$ and $\mathbf{\Phi}^{\ddagger} = \mathbf{R}^{-1}\mathbf{\Phi}(\mathbf{R}^{\mathrm{T}})^{-1}$ that generate the same covariance structure $\mathbf{\Sigma}$. The operation $\mathbf{\Lambda} \mapsto \mathbf{\Lambda R}$ is termed 'rotation'. Thus, in any solution, $\mathbf{\Lambda}$ can be made to satisfy m^2 additional conditions, which is naturally equivalent to the number of independent elements of $\mathbf{R}$.

From the above it is clear that any method of estimation requires at a minimum m^2 restrictions on $\mathbf{\Lambda}$ and $\mathbf{\Phi}$. The EFA tradition usually achieves this by requiring that $\mathbf{\Phi} = \mathbf{I}_m$ and $\mathbf{\Lambda}^{\mathrm{T}}\mathbf{\Psi}^{-1}\mathbf{\Lambda}$ be diagonal accompanied by an order condition on the diagonal elements. These restrictions are arbitrary such that whence estimation is settled EFA traditionally endeavors on applying a rotation that satisfies certain criteria for interpretation purposes. Peeters (2012) has given minimal conditions for the formulation of a rotationally unique confirmatory unrestricted factor model that cancels the need for post-hoc rotation.

5B Reproduced Correlation Structure

Table 5.2 contains the sample correlations, the correlations reproduced by the factor analysis, and the residual correlations (sample correlation minus reproduced correlation).

Table 5.2. Matrix containing observed (Pearson), reproduced, and residual correlations

	1	2	3	4	5	6	7	8
1 BMI								
observed	1.000							
reproduced	1.002							
residual	-.002							
2 $\log_e$\{trig.\}								
observed	.041	1.000						
reproduced	.101	.984						
residual	-.060	.016						
3 HDL chol.								
observed	-.143	-.423	1.000					
reproduced	-.111	-.459	.998					
residual	-.032	.036	.002					
4 $\log_e$\{IR\}								
observed	.314	.259	-.254	1.000				
reproduced	.291	.250	-.276	1.006				
residual	.023	.009	.022	-.006				
5 $\log_e$\{GB\}								
observed	.071	.051	-.066	.387	1.000			
reproduced	.161	.082	-.092	.384	1.004			
residual	-.090	-.031	.026	.003	-.004			
6 $\log_e$\{G2\}								
observed	.121	.197	-.105	.336	.233	1.000		
reproduced	.141	.143	-.157	.338	.183	.999		
residual	-.020	.054	.052	-.002	.050	.001		
7 SBP								
observed	.132	.065	-.028	.190	.104	.084	1.000	
reproduced	.089	.029	-.033	.212	.125	.100	.999	
residual	.043	.036	.005	-.022	-.021	-.016	.001	
8 DBP								
observed	-.013	-.016	.084	.093	.095	-.010	.332	1.000
reproduced	.047	-.056	.059	.111	.077	.048	.047	.999
residual	-.060	.004	.025	-.018	.018	-.058	.275	.001

6

Inequality-Constrained-Model Selection for the Political Sciences

Peeters, C.F.W. & van Wesel, F. (under review)

6.1 Introduction

Parameter specification of parametric models is often rigid in the political sciences. This is meant in the sense that, most often, a parameter is either specified to be part or *not* to be part of the model. However, it would be most informative, as well as cater the need of many a substantive political scientist, when model specification would take into account direction and magnitude of parameter effects. Such provisions ask for the incorporation of inequality constraints in the model.

The incorporation of inequalities in the classical framework often implies a multi-step procedure or model-specific provisions for statistical assessment. Consider as an example an analysis of variance (ANOVA) type setting, where μ_j denotes the mean of group j, $j = 1, \ldots, J$. If one has a specific ordering of means in mind as a viable representation of the information contained in the data, say

$$M_1 : \mu_1 > \mu_2 > \cdots > \mu_J,$$

then the rejection of the null-alternative in the traditional null-hypothesis setting

$$\begin{aligned} H_0 &: \text{all } \mu_j \text{ are equal} \quad vs \\ H_a &: \text{all } \mu_j \text{ are not equal,} \end{aligned}$$

does not provide any information on M_1. One could then endeavor on post-hoc pairwise comparisons in order to retrieve information on the viability of M_1. In such a situation a multi-step procedure is followed in order to indirectly assess a model of interest. It is also possible to test H_0 against an inequality-constrained alternative such as M_1, or to pair the null-hypothesis with an inequality-constrained expression of theory (Barlow et al., 1972; Rosenthal et al., 2000; Silvapulle & Sen, 2005). However, these provisions apply to a limited set of models and focus on the evaluation of single hypotheses such that they cannot be used for the situation in which a direct comparison of a multitude of plausible inequality-constrained models is wished for. Such a situation describes a setting of model selection and is directly linked to the aim of this study: To develop and convey a more generic framework for inequality-constrained model selection in the general linear model (GLM) setting. The stance on model selection will be Bayesian and the framework will be elaborated with two extensive examples representing two different and heavily utilized linear models. The remainder of the introduction will provide the motivation and background for the stated aim.

Model selection has (at least in part) been the realm of information criteria (IC) (e.g., Akaike, 1973; Schwarz, 1978). IC offer the advantage of being able to handle model non-nestedness over many classical procedures. Anraku (1999) developed an information criterion for restricted-model selection, but it is only appropriate for simple order-restrictions in a limited set of models. In this paper, the Bayesian road is opted for. Say we consider the set of models $\mathcal{B} = \{M_1, \ldots, M_B\}$ with the aim to select the model that best balances model fit and model complexity. Assume that each model M_b, next to the model-specific collection of unknown parameters $\mathbf{\Theta}_b$, is characterized by likelihood $L_b(\mathbf{\Theta}_b; \mathbf{Z})$. Also assume that we have $\pi_b(\mathbf{\Theta}_b)$, $b = 1, \ldots, B$, as the available prior distributions for the unknown parameters. The key quantity in Bayesian model selection is then the marginal likelihood of the data

$$m_b(\mathbf{Z}) = \int L_b(\mathbf{\Theta}_b; \mathbf{Z})\pi_b(\mathbf{\Theta}_b)\ \partial\mathbf{\Theta}_b, \tag{6.1}$$

which constitutes the normalizing constant for the posterior distribution $\pi_b(\mathbf{\Theta}_b|\mathbf{Z})$ and is the pivotal quantity in the construction of the Bayes factor (Jeffreys, 1935, 1961):

$$B_{bb'} = \frac{m_b(\mathbf{Z})}{m_{b'}(\mathbf{Z})} = \frac{\int L_b(\mathbf{\Theta}_b; \mathbf{Z})\pi_b(\mathbf{\Theta}_b)\ \partial\mathbf{\Theta}_b}{\int L_{b'}(\mathbf{\Theta}_{b'}; \mathbf{Z})\pi_{b'}(\mathbf{\Theta}_{b'})\ \partial\mathbf{\Theta}_{b'}}. \tag{6.2}$$

The Bayes factor incorporates model fit as well as model complexity and expresses "the evidence provided by the data in favor of one scientific theory, represented by a statistical model, as opposed to another" (Kass & Raftery, 1995). For an overview of Bayesian statistics and methodology, see Gill (2008) or Gelman, Carlin, Stern, and Rubin (2004).

The Bayes factor as a model selection criterion has the advantages (cf. Kass & Raftery, 1995; S. Y. Lee, 2007, Chapter 5) that: (i) It provides both a measure of evidence against a competing model and a measure of support for the alternative model; (ii) The comparison of any two models does not depend on the assumption that either one is 'true'; (iii) It allows one to take model uncertainty into account, thus providing a consistent quantity for the comparison of a multitude of competing models; (v) It can handle the comparison of both nested and nonnested models. Notwithstanding, there are several difficulties with (6.1) and thus (6.2): (i) The marginal likelihood will generally not be analytically tractable and computation strategies can be challenging (see Kass & Raftery, 1995); (ii) Both the use of improper noninformative and proper but vague priors yield indeterminate answers for (6.2) when the models to be compared are of differing dimension (Jeffreys, 1961). This is undesirable as especially under default prior choices the interpretation of (6.2) as a weighted likelihood ratio is warranted (Berger & Pericchi, 2004); (iii) When not using default options for the priors $\pi_b(\boldsymbol{\Theta}_b)$, their assessment, especially in multivariate cases, may display considerable difficulty and may prove influential in the sense that differing specifications may render differing outcomes.

Early usage of the Bayesian option for inequality constraints in the linear model can be found in Chowdhury (1969) and Davis (1978). The work of Klugkist and Hoijtink (2007) formulated a computationally simple Bayes factor for normal-theory linear models bearing inequalities on the mean parameters. The methods therein, however, imply examination of the sample variance for prior specification and do not consider inequalities on regression-type parameters. Here, the work by Klugkist and Hoijtink (2007) is pursued and extended in two ways: Default type priors are incorporated in the framework as are inequalities on regression-type parameters.

Section 6.2 gives a precise statement of the inequality constraints in which interest is taken. Moreover, this section contains a general development of a computationally simple Bayes factor for the selection of inequality-constrained models. The framework given is then shown to lead to well-balanced Bayes factors under default-type priors in linear models through two extensive applications. Section 6.3 reviews inequalities on mean parameters in ANOVA modeling of data on political socialization. Section 6.4 subsequently reviews inequalities on factor loadings (which are essentially regression parameters) in a revised take on confirmatory factor analysis (CFA) modeling of data on leadership in police forces. Section 6.5 concludes with a discussion.

6.2 A Simple Bayes Factor for Constrained-Model Selection

6.2.1 Constraints Considered

Let $\boldsymbol{\theta}^{\mathfrak{c}}$ denote a vectorization of the elements of $\boldsymbol{\Theta}$ involved in a system of inequality constraints. Interest is then taken in linear inequality constraints of the form:

$$\boldsymbol{\Omega}_b\boldsymbol{\theta}^{\mathfrak{c}} - \boldsymbol{\alpha}_b > \mathbf{0}, \tag{6.3}$$

where $\boldsymbol{\Omega}_b$ is an $(L_b \times Q)$-dimensional constraints matrix representing a system of $l = 1, \ldots, L_b$ linear restrictions and $\boldsymbol{\alpha}_b$ denotes a real fixed-value vector of lenght L_b. The conjunction of $\boldsymbol{\Omega}_b$ and $\boldsymbol{\alpha}_b$ on $\boldsymbol{\theta}^{\mathfrak{c}}$ defines model M_b. Many forms of linear inequality constraints are contained in (6.3) (e.g., Tsonaka & Moustaki, 2007). Interest lies with:

i. Fixed value constraints, where a parameter is set to surpass a certain fixed value, i.e., $\theta^{\mathfrak{c}}_q - \alpha_l > 0$;
ii. Positivity constraints, where a certain parameter is set to surpass an other parameter, i.e., $\theta^{\mathfrak{c}}_q - \theta^{\mathfrak{c}}_{q'} > 0$;

iii. Negativity constraints, where a parameter is set to surpass the negative of an other parameter, i.e, $\theta_q^c + \theta_{q'}^c > 0$; and, as special cases of (i), (ii), and (iii)
iv. Approximate equality constraints, where a parameter is set to approximately equal constant e with given bound ε, or where two parameters are set to be approximately equal with given bound ε, i.e.,

$$|\theta_q^c - e| < \varepsilon \Rightarrow \begin{cases} \theta_q^c - (e - \varepsilon) > 0 \\ -\theta_q^c + e + \varepsilon > 0 \end{cases}, \text{ and } |\theta_q^c - \theta_{q'}^c| < \varepsilon \Rightarrow \begin{cases} \theta_q^c - \theta_{q'}^c + \varepsilon > 0 \\ \theta_{q'}^c - \theta_q^c + \varepsilon > 0 \end{cases},$$

such that each row of $\mathbf{\Omega}_b$ will be a permutation of either $(1, 0, \ldots, 0)$, $(-1, 0, \ldots, 0)$, $(1, 1, 0, \ldots, 0)$, $(-1, 1, 0, \ldots, 0)$, or $(-1, -1, 0, \ldots, 0)$. These constraints are adequately supported by the prior distributions, as will be elaborated in Sections 6.3 and 6.4 for mean-type and regression-type parameters, respectively.

Example 6.1. Imagine an analysis of variance (ANOVA) model is considered which hosts four group means. Suppose the group means are involved in inequality-constrained expression of theory and that one of the theories regarding the data and the model states that an incomplete ordering of the means in expected, such that $M_b : \mu_1 > \{\mu_3, \mu_4\} > \mu_2 > 2.5$. That is, the mean of group 1 is expected to be larger than both group mean 3 and group mean 4, which in turn are expected to be larger than the mean of group 2, the mean of which is assumed to be larger than 2.5. The model expectation can be expressed through the system (6.3) by:

$$\begin{bmatrix} 1 & 0 & -1 & 0 \\ 1 & 0 & 0 & -1 \\ 0 & -1 & 1 & 0 \\ 0 & -1 & 0 & 1 \\ 0 & 1 & 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \\ \mu_4 \end{bmatrix} - \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 2.5 \end{bmatrix} > \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}.$$

■

6.2.2 A Simple Bayes Factor

Pivotal in the formulation of a computationally friendly Bayes factor for model selection under inequality constraints is the nesting of competing inequality constrained models within an encompassing model (Berger & Pericchi, 1996; Klugkist & Hoijtink, 2007). In the case of inequality-constrained-model selection an encompassing model can be seen as the model bearing no inequality restrictions. Inequality constraints on the parameter space enter the model through the prior distribution. The prior then taking the form of the product of a conventional (non-informative) prior distribution on the base model and an indicator function representing the inequality constraints (Geweke, 1986).

Let $\mathbf{X}$ denote a generic vector or matrix of data and let $\boldsymbol{\vartheta}$ denote a generic vector or matrix of latent data. Moreover, let $g(\boldsymbol{\vartheta}|\mathbf{\Theta})$ denote the density of $\boldsymbol{\vartheta}$ given $\mathbf{\Theta}$ and assume that the complete data likelihood consists of $L(\mathbf{\Theta}, \boldsymbol{\vartheta}; \mathbf{X})g(\boldsymbol{\vartheta}|\mathbf{\Theta})$. Many models, such a factor analytic models or ANOVA-type models, can be expressed as (special cases of) such structure. All elements are now in place to restate more formally a result originally given by Klugkist and Hoijtink (2007):

Proposition 6.1. *Define* $\mathbf{\Theta}_b \equiv \{\mathbf{\Theta} : \mathbf{\Omega}_b \boldsymbol{\theta}^{\mathrm{c}} - \boldsymbol{\alpha}_b > \mathbf{0}\}$*, for which we assume that the system* $\mathbf{\Omega}_b \boldsymbol{\theta}^{\mathrm{c}} - \boldsymbol{\alpha}_b$ *is non-contradictory, giving that* $\mathbf{\Theta}_b \subset \mathbf{\Theta}$*. Let the prior under encompassing model* M_0 *for the set of parameters in the model be* $\pi_0(\mathbf{\Theta})g(\boldsymbol{\vartheta}|\boldsymbol{\theta}^{\mathrm{u}})$*, where* $\boldsymbol{\theta}^{\mathrm{u}}$ *denotes the elements of* $\mathbf{\Theta}$ *not involved in a system of inequality constraints. The prior distribution under any constrained model* M_b *is* $\pi_b(\mathbf{\Theta}_b)g(\boldsymbol{\vartheta}|\boldsymbol{\theta}^{\mathrm{u}}) \propto \pi_0(\mathbf{\Theta})\mathbb{1}_{\{\mathbf{\Omega}_b \boldsymbol{\theta}^{\mathrm{c}} - \boldsymbol{\alpha}_b > \mathbf{0}\}} g(\boldsymbol{\vartheta}|\boldsymbol{\theta}^{\mathrm{u}})$*, where* $\mathbb{1}_{\{\cdot\}}$ *denotes an indicator function representing the system of inequality constraints defining* $\mathbf{\Theta}_b$*. Now, assuming propriety of* $\pi_0(\mathbf{\Theta}, \boldsymbol{\vartheta}|\mathbf{X})$*, the Bayes factor* B_{b0} *reduces to the posterior probability mass satisfying the system of constraints that defines* $\mathbf{\Theta}_b$ *over the prior probability mass satisfying the system of constraints that defines* $\mathbf{\Theta}_b$*.*

Proof. The Bayes factor B_{b0} of constrained model M_b to encompassing model M_0, is written as

$$\frac{\displaystyle\int L(\mathbf{\Theta}, \boldsymbol{\vartheta}; \mathbf{X})\pi_b(\mathbf{\Theta}_b)g(\boldsymbol{\vartheta}|\boldsymbol{\theta}^{\mathrm{u}})\,\partial(\mathbf{\Theta}_b, \boldsymbol{\vartheta})}{\displaystyle\int L(\mathbf{\Theta}, \boldsymbol{\vartheta}; \mathbf{X})\pi_0(\mathbf{\Theta})g(\boldsymbol{\vartheta}|\boldsymbol{\theta}^{\mathrm{u}})\,\partial(\mathbf{\Theta}, \boldsymbol{\vartheta})}. \tag{6.4}$$

Using the basic marginal identity (Besag, 1989; Chib, 1995) we may express (6.4) equivalently as follows:

$$\frac{L(\mathbf{\Theta}, \boldsymbol{\vartheta}; \mathbf{X})\pi_b(\mathbf{\Theta}_b)g(\boldsymbol{\vartheta}|\boldsymbol{\theta}^{\mathrm{u}})/\pi_b(\mathbf{\Theta}, \boldsymbol{\vartheta}|\mathbf{X})}{L(\mathbf{\Theta}, \boldsymbol{\vartheta}; \mathbf{X})\pi_0(\mathbf{\Theta})g(\boldsymbol{\vartheta}|\boldsymbol{\theta}^{\mathrm{u}})/\pi_0(\mathbf{\Theta}, \boldsymbol{\vartheta}|\mathbf{X})}. \tag{6.5}$$

For any given value of $\{\mathbf{\Theta}, \boldsymbol{\vartheta}\}$, say $\{\mathbf{\Theta}^*, \boldsymbol{\vartheta}^*\}$, that is admissible under the system of constraints $\mathbf{\Omega}_b \boldsymbol{\theta}^{\mathrm{c}} - \boldsymbol{\alpha}_b > \mathbf{0}$, clearly $\{\mathbf{\Theta}^*, \boldsymbol{\vartheta}^*\} \in M_b, M_0$. Substituting and dividing out terms, expression (6.5) reduces to

$$\frac{\pi_0(\mathbf{\Theta}^*, \boldsymbol{\vartheta}^*|\mathbf{X})}{\pi_b(\mathbf{\Theta}^*, \boldsymbol{\vartheta}^*|\mathbf{X})} \cdot \frac{\pi_b(\mathbf{\Theta}^*)}{\pi_0(\mathbf{\Theta}^*)}. \tag{6.6}$$

Now, notice

$$\pi_b(\mathbf{\Theta}^*, \boldsymbol{\vartheta}^*|\mathbf{X}) = \pi_0(\mathbf{\Theta}^*, \boldsymbol{\vartheta}^*|\mathbf{X}) \cdot \left[\int\int_{\{\mathbf{\Theta}:\mathbf{\Omega}_b \boldsymbol{\theta}^{\mathrm{c}} - \boldsymbol{\alpha}_b > \mathbf{0}\}} \pi_0(\mathbf{\Theta}, \boldsymbol{\vartheta}|\mathbf{X})\,\partial\mathbf{\Theta}\,\partial\boldsymbol{\vartheta}\right]^{-1} \tag{6.7}$$

and

$$\pi_b(\mathbf{\Theta}^*) = \pi_0(\mathbf{\Theta}^*) \cdot \left[\int_{\{\mathbf{\Theta}:\mathbf{\Omega}_b \boldsymbol{\theta}^{\mathrm{c}} - \boldsymbol{\alpha}_b > \mathbf{0}\}} \pi_0(\mathbf{\Theta})\,\partial\mathbf{\Theta}\right]^{-1} \tag{6.8}$$

Substituting (6.7) and (6.8) in (6.6) we obtain

$$B_{b0} = \frac{\displaystyle\int\int_{\{\mathbf{\Theta}:\mathbf{\Omega}_b \boldsymbol{\theta}^{\mathrm{c}} - \boldsymbol{\alpha}_b > \mathbf{0}\}} \pi_0(\mathbf{\Theta}, \boldsymbol{\vartheta}|\mathbf{X})\,\partial\mathbf{\Theta}\,\partial\boldsymbol{\vartheta}}{\displaystyle\int_{\{\mathbf{\Theta}:\mathbf{\Omega}_b \boldsymbol{\theta}^{\mathrm{c}} - \boldsymbol{\alpha}_b > \mathbf{0}\}} \pi_0(\mathbf{\Theta})\,\partial\mathbf{\Theta}} \equiv \frac{f_b}{\omega_b}, \tag{6.9}$$

and the proposition follows. □

The Proposition gives that one may estimate the Bayes factor of a constrained model nested in an encompassing model by a very simple numerical procedure: Counting the number of times an appropriate sampler for the parametric model at hand visits the permissable posterior and prior spaces defined by the system of inequality constraints. The need for marginal computation is thus circumvented and an elegant expression for the BF is provided for in which the numerator and denominator may be interpreted as measures of model fit and model complexity, respectively. The quantity B_{b0} has many of the properties that regular Bayes factors on non-truncated parameters have: $B_{bb} = 1$; $B_{0b} = B_{b0}^{-1}$; and $B_{bb'} = B_{b0}/B_{b'0}$.

Klugkist and Hoijtink (2007) review ANOVA models. In their method the sample variance is to be examined to choose the unconstrained prior vague enough in order to avoid certain types of paradoxes. We desire Bayes factors as in (6.9) for which less arbitrary choices regarding the prior need to be made. The illustrations will be geared towards such choices: Inequalities on mean parameters with training samples (ANOVA example in Section 3), and inequalities on regression parameters with a bound uniform prior (factor analysis example in Section 4).

6.3 Illustration I: Mean Parameters in ANOVA

6.3.1 Data

People's reaction to authoritative decisions is influenced by the perceived procedural fairness behind the decision-making process. If the process is perceived to be fair, it is expected that unfavorable outcomes are more willingly accepted (cf. Lind & Tyler, 1988; Ambrose, 2002; MacCoun, 2005; Tyler, 2006). This effect might be of great practical importance when controversial governmental decisions have to be made.

A large study conducted by Esaiasson, Gilljam, and Lindholm (2007) aimed at a "further understanding of the boundary conditions for when and how procedural considerations are important for decision acceptance". In order to investigate these boundary conditions with concern to decision acceptance within a specific context - that of decisions made by governments - three decision settings were deemed of interest: Decisions made by (i) referendum, by (ii) elected representatives, and by (iii) expert bureaucrats. Furthermore, the authors used four target groups: High school students, teachers, adult citizens and political science undergraduates. All participants were subjected to a scenario-type quasi-experimental setting. The question that was asked to every participant concerned whether or not religious symbols should be banned from schools. The participants were told (1) by which decision-making process the decision was made (either referendum, representative, or expert) and (2) whether or not the outcome corresponded to the participants' own preferences, resulting in so called 'winners' (decision outcome equal to participants' preference) and 'losers' (decision outcome opposite to participants' preference). Decision acceptance was measured by an index containing the perceived fairness of the outcome, and the willingness to comply with the regulations that follow from the decision (Gilljam, Esaiasson, & Lindholm, 2010). A total of 400 high school students, 306 teachers, 521 adult citizens, and 142 political science students participated in the study.

6.3.2 The Analysis of Variance Model

The model of choice for investigation of the data is the analysis of variance model. This model can be stated as:

$$y_i = \sum_{j=1}^{J} \mu_j g_{ji} + \epsilon_i,$$

where y_i is the score on the dependent variable for $i = 1, ..., n$ persons. Group membership is denoted by g_{ji}, for $j = 1, ..., J$ groups, where g_{ji} is unity if person i is a member of group j and zero otherwise. The mean of group j is denoted by μ_j. The errors ϵ_i are assumed to be independent and normally distributed with mean zero and variance σ^2. The model likelihood is then expressed as:

$$L(\boldsymbol{\mu}, \sigma^2, \mathbf{G}; \mathbf{y}) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{ -\frac{1}{2\sigma^2} \left[y_i - \left(\sum_{j=1}^{J} \mu_j g_{ji} \right) \right]^2 \right\},$$

with $\boldsymbol{\mu}^{\mathrm{T}} = [\mu_1, \ldots, \mu_J]$, $\mathbf{G} = [\mathbf{g}_1, \ldots, \mathbf{g}_J]$, and $\mathbf{g}_j^{\mathrm{T}} = [g_{j1}, \ldots, g_{jn}]$. Inequality constraints in this type of model are imposed on the μ_j's .

6.3.3 Inequality-Constrained Theories

For reasons of simplicity the example used in this section is only one component of the complete study. As explained below, in the current illustration we focus on the group of participants receiving an outcome opposite to their preference, the so-called ‘losers’ group. Furthermore, we do not use the data gathered from the teacher target group, as the setting in which these data were gathered differed from the settings of the other three target groups.

We are interested in that part of the research that concerns the understanding of decision acceptance as a learning process. This learning process puts us in the setting of the public education system where societal norms are formed to a large extent. Expectations with concern to the level of cognitive sophistication of the voters vary. Gilljam et al. (2010) hypothesize that a high level of cognitive sophistication is needed to be able to understand the decision-making process and the fairness thereof and thus that voters who are able to understand the procedure are less influenced by the decision outcome. On the other hand, the effect of procedural fairness was found in a wide variety of settings (e.g., Price et al., 2001; Cohn, White, & Sanders, 2000), indicating that the level of cognitive sophistication does not have to be high, and thus that an easy learning process is more likely. Consequently, the authors state three possible models describing the learning process:

Table 6.1. Group means and standard deviations (in parentheses) for the ANOVA example

Should religious symbols be banned from schools?			
	Expert	Representative	Referendum
	μ_1	μ_2	μ_3
High school students	3.80 (2.65)	4.25 (3.37)	5.00 (2.81)
	$n = 69$	$n = 60$	$n = 74$
	μ_4	μ_5	μ_6
Adult citizens	4.09 (2.63)	5.53 (3.20)	6.24 (2.94)
	$n = 67$	$n = 73$	$n = 84$
	μ_7	μ_8	μ_9
Political science undergraduates	4.03 (2.71)	4.73 (3.13)	6.76 (3.25)
	$n = 29$	$n = 22$	$n = 17$

(1) A model of easy learning, i.e., all groups learn at an equal level, representing one end of the spectrum; (2) a model of tradition life-cycle learning, i.e., we learn as we get older; and (3) a model of advanced learning, i.e., only certain specialized groups learn it well, representing the other end of the spectrum.

Next to the effect of cognitive sophistication, an effect is expected with concern to the context in which the decision was made, either by means of a referendum, an elected representative, or an expert bureaucrat. It is expected that being personally involved in making a decision, as for instance in a referendum, will result in more direct acceptance of the decision outcome (Teorell, 2006). If being personally involved is not an option, decision acceptance should be best when the decision is made by a selected representative. When a representative is elected, it is assumed that the voter gave his or her consent and that the representative enjoys a certain level of trust. In addition, a representative can bring knowledge and experience to the decision-making process – knowledge and experience a voter does not necessarily have (Manin, 1997). Decisions made by expert bureaucrats may only be accepted if they are based on professional knowledge. As this type of process is more distant from the individual, decisions made this way are expected to be the least accepted.

For the current analyzes, a 3 (decision procedure) $\times$ 3 (target group) ANOVA model is appropriate. Descriptive statistics can be found in Table 6.1. Using the notation of Table 6.1, the first hypothesized model (M_1), which concerns a process of easy learning stated previously, can be formulated by usage of inequality and approximate equality constraints:

$$M_1 : \left\{ \begin{array}{ll} \mu_1 < \mu_2 < \mu_3, & \mu_1 \approx \mu_4 \approx \mu_7, \\ \mu_4 < \mu_5 < \mu_6, \,\&\, & \mu_2 \approx \mu_5 \approx \mu_8, \\ \mu_7 < \mu_8 < \mu_9, & \mu_3 \approx \mu_6 \approx \mu_9, \end{array} \right\}.$$

The left component of this model represents the expectation that the more influence a person has on the decision-making process, the more easily he or she accepts its outcome. Consequently, for each group of participants (first row high school students μ_1, μ_2, μ_3, second row adult citizens μ_4, μ_5, μ_6, and third row political science undergraduates μ_7, μ_8, μ_9) an ordering is given such that decisions made by expert bureaucrats (μ_1, μ_4, μ_7) get the lowest decision acceptance, followed by decisions made by elected representatives (μ_2, μ_5, μ_8) and decisions made by means of a referendum (μ_3, μ_6, μ_9) having the highest decision acceptance. The right component represents the model of easy learning as it states that all groups accept the decision made on approximately the same level, given an equal decision making process (represented by the three rows). The second model of life-cycle learning (M_2) differs from the first model by stating that, per type of decision-making process, high school students will display lower decision acceptance than adults and political science students, with the ordering between the latter two groups left unspecified:

$$M_2 : \left\{ \begin{array}{ll} \mu_1 < \mu_2 < \mu_3, & \mu_1 < \{\mu_4, \mu_7\}, \\ \mu_4 < \mu_5 < \mu_6, \ \& & \mu_2 < \{\mu_5, \mu_8\}, \\ \mu_7 < \mu_8 < \mu_9, & \mu_3 < \{\mu_6, \mu_9\}, \end{array} \right\}.$$

Finally the third model (M_3) represents the hypothesis of advanced learning where political science students accept the decisions made best, followed by adult citizens, followed by high school students, per decision-making process:

$$M_3 : \left\{ \begin{array}{ll} \mu_1 < \mu_2 < \mu_3, & \mu_1 < \mu_4 < \mu_7, \\ \mu_4 < \mu_5 < \mu_6, \ \& & \mu_2 < \mu_5 < \mu_8, \\ \mu_7 < \mu_8 < \mu_9, & \mu_3 < \mu_6 < \mu_9, \end{array} \right\}.$$

Note that the left component of each model-translated theory is equivalent.

6.3.4 Implementation

The prior used to calculate a default Bayes factor for inequality constrained analysis of variance is a so called ‘posterior prior’ (Berger & Pericchi, 1996; Pérez & Berger, 2002; Berger & Pericchi, 2004). A posterior prior is a reference prior (σ^{-2} for the ANOVA model; see Bernardo, 1979), updated with a small part of the data called a training sample (ℓ), such that the resulting posterior is proper. As the choice of a certain training sample is arbitrary, posterior priors are usually based on multiple training samples. The particular type of posterior prior used is the averaged restricted posterior prior (van Wesel et al., 2011), which is an approximation to the empirical expected-posterior prior (Pérez & Berger, 2002) that has good properties in representing model complexity for inequalities on mean parameters (Mulder et al., 2010; van Wesel et al., 2011). For the ANOVA model at hand this prior is denoted by:

$$\hat{\pi}_0(\boldsymbol{\mu}, \sigma^2) = \hat{\pi}_0(\boldsymbol{\mu})\hat{\pi}(\sigma^2) = \mathcal{N}_J(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\mu}}}^{-}) \cdot \mathcal{I}\chi^2(\hat{\rho}, \hat{\varpi}^2), \tag{6.10}$$

where $\mathcal{I}\chi^2(\cdot,\cdot)$ denotes a scaled inverse χ^2 distribution. The hyperparameters $\{\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}}^-_{\hat{\boldsymbol{\mu}}}, \hat{\rho}, \hat{\varpi}^2\}$ in this expression are averaged over the number of training samples $\ell = 1, ..., L$ and the number of iterations per training sample $c = 1, ..., C$. Their definitions can be found in Appendix 6A.1 along with further information on (6.10).

The posterior distribution based on the averaged constrained posterior prior can be expressed by

$$\pi_0(\boldsymbol{\mu}, \sigma^2|\mathbf{y}, \mathbf{G}) \propto L(\boldsymbol{\mu}, \sigma^2, \mathbf{G}; \mathbf{y}) \cdot \mathcal{N}_J(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}}^-_{\hat{\boldsymbol{\mu}}}) \cdot \mathcal{I}\chi^2(\hat{\rho}, \hat{\varpi}^2).$$

The posterior conditional distributions are needed for sampling. The posterior conditional distribution on $\boldsymbol{\mu}$, $\pi_0(\boldsymbol{\mu}|\mathbf{y}, \mathbf{G}, \sigma^2)$, is distributed as $\mathcal{N}_J(\tilde{\boldsymbol{\mu}}, \tilde{\boldsymbol{\Sigma}}^-_{\tilde{\boldsymbol{\mu}}})$. The posterior conditional distribution on σ^2, $\pi_0(\sigma^2|\mathbf{y}, \mathbf{G}, \boldsymbol{\mu})$, is distributed as $\mathcal{I}\chi^2(\tilde{\rho}, \tilde{\varpi}^2)$. The definitions of the parameters $\{\tilde{\boldsymbol{\mu}}, \tilde{\boldsymbol{\Sigma}}^-_{\tilde{\boldsymbol{\mu}}}, \tilde{\rho}, \tilde{\varpi}^2\}$ can be found in Appendix 6A.2. Analogous with the line of argument in Section 6.2.2, the Bayes factor for inequality-constrained-model selection in the given ANOVA setting (van Wesel et al., 2011) can be expressed by:

$$B^A_{b0} = \frac{\displaystyle\int\int_{\{\boldsymbol{\mu}:\boldsymbol{\Omega}_b\boldsymbol{\mu}-\boldsymbol{\alpha}_b>\mathbf{0}\}} \pi_0(\boldsymbol{\mu}, \sigma^2|\mathbf{y}, \mathbf{G})\, \partial\boldsymbol{\mu}\, d\sigma^2}{\displaystyle\int_{\{\boldsymbol{\mu}:\boldsymbol{\Omega}_b\boldsymbol{\mu}-\boldsymbol{\alpha}_b>\mathbf{0}\}} \hat{\pi}_0(\boldsymbol{\mu})\, \partial\boldsymbol{\mu}} \equiv \frac{f^A_b}{\omega^A_b}.$$

This Bayes factor can be estimated by obtaining the proportions, respectively, of prior and posterior samples in agreement with the inequality constraints defining the model.

6.3.5 Results

Results are generated using $L = 50,000$ training samples and $100,000$ iterations per training sample. In addition, for analyzing approximate equality constraints, we set $\varepsilon = .4$. The resulting Bayes factors for the entertained inequality-constrained models against the unconstrained model can be found in Table 6.2. As can be seen from this table, the second model, representing the expectation of traditional life-cycle learning, results in the highest Bayes factor ($\hat{B}^A_{20} = 415.17$).

Table 6.2. Estimated Bayes factors for constrained models on the decision acceptance data

M_1		M_2		M_3	
$\hat{f}^A_1 =$	2.275e-5	$\hat{f}^A_2 =$	.2491	$\hat{f}^A_3 =$	.0194
$\hat{\omega}^A_1 =$	7.7e-7	$\hat{\omega}^A_2 =$	.0006	$\hat{\omega}^A_3 =$	.0002
$\hat{B}^A_{10} =$	29.67	$\hat{B}^A_{20} =$	415.17	$\hat{B}^A_{30} =$	97.00

In order to compare the three directional models, Table 6.3 is given. What can be inferred from this table is that M_2, representing traditional life-cycle learning, has the highest Bayes factor and is preferred over both alternative models. It is therefore the most likely model within this set of models. Model M_2 is considered to be fourteen times as probable as model 1, representing a model of easy learning, and four times as probable as model 3, representing a model of advanced learning. Clearly, the model of easy learning is the least likely model within this set. This is indicated by the fact that all Bayes factors of M_3 against M_1 and M_2 are smaller than unity.

Table 6.3. Bayes factor matrix for the decision acceptance data

$\hat{\mathrm{B}}_{from\|to}$	1	2	3
1	1	.07	.31
2	13.99	1	4.28
3	3.27	.23	1

The figures tell us firstly that there is a main effect for how decisions are made: When the decision made is in contrast to what the voter desires, decision acceptance is lowest when an expert administrator makes the decision, acceptance is a little higher when a chosen representative makes the decision, and acceptance is highest when the decision is made by referendum. Also, the model of easy learning is the least probable. High school students have more difficulties accepting decisions they do not agree with than adults have. This finding indicates that high school students' level of political cognitive sophistication is lower than that of adults. When comparing the model of life-cycle learning and the model of advanced learning, we can conclude that adults, whether or not trained in decision theory, have approximately the same level of acceptance of undesired outcomes than political science undergraduates, i.e., their level of political cognitive sophistication seems to be equal.

6.4 Illustration II: Regression Parameters in CFA

6.4.1 Data

Leadership has been argued to be a key factor influencing employee behavior (e.g., Brown & Treviño, 2006). Exemplary leadership may be beneficial to any organization, but seems particularly important for police bureaucracy (Heffernan, 2002, p. 138) as acceptable leadership is needed in order for the people to accept the monopoly on societal violence a police force holds. Moreover, acceptable leadership is imperative for police officers' acceptance of the risks associated with policing (Ibidem). It has long been argued (Kohlberg, 1969) that street-level employees look to their immediate supervisors for moral guidance and authority.

The landscape of the Dutch police organization has been changing recently, in the sense that traditional moral-paternal styles of policing are being amended with more output-steering organizational styles (Mellink, 2003, p. 21; Kolthoff, 2007). The intermixing of policing traditions has led to the idea that both ethical-traditional leadership (leadership towards the morally and ethically just) and business-like organizational leadership (leadership towards organizational accomplishment) may be found within the Dutch police organization. Lasthuizen (2008) has to this end devised a questionnaire that operationalizes scales of organizational leadership styles (den Hartog, 1997) and ethical leadership styles (Craig & Gustafson, 1998; Kaptein & Wempe, 2002; Kaptein, 2003). The items constitute opinionative items on the perceived leadership characteristics of the immediate supervisor. Consider as an example the following item: 'My supervisor treats me as an individual rather than just a member of the group.' All items were measured on a six-category response scale (1. completely disagree; 2. disagree; 3. disagree more than agree; 4. agree more than disagree; 5. agree; 6. completely agree). For a complete overview of the included items see Lasthuizen (2008, Tables 6.2 and 6.4).

The questionnaire was conducted in a Dutch regional police force and received responses from 536 police officers *not* in a supervisory position. A validation of the questionnaire by Lasthuizen (2008) led to the extraction of 6 latent factors: Inspirational leadership (IL), result-oriented leadership (ROL), passive leadership (PL), role-modeling leadership (RML), integrity-focused leadership (IFL), and unethical leadership (UL). She proposes that the inspirational, result-oriented, and passive leadership styles are the *organizational* expressions of transformational (social exchange oriented), transactional (transactional exchange oriented), and laissez-faire leadership. In her view, role-modeling, integrity-focused, and unethical leadership styles are then the *ethical* expressions of transformational, transactional and laissez-faire leadership, respectively. In doing so, she theorizes, but does not assess, the existence of two higher-order latent factors (organizational and ethical leadership styles). The goal here is to assess the factor structure of the six leadership scales with the use of inequality constraints on the factor loadings, which are essentially regression parameters (Section 6.4.2). First, the factor model is reviewed before delving into inequality-constrained expressions of factor theory and its implementation.

6.4.2 The Factor Analytic Model

Quinn (2004) devised a factor model for data of mixed metric. Here, the common factor analytic model is utilized as the leadership scales abide the normality assumption. Focus is with the imposition of inequality constraints on the factor loadings. For regression-type parameters, it is natural to impose inequality constrained structure on standardized coefficients. Thus, let $\mathbf{Z}^{\mathrm{T}} \equiv [\mathbf{z}_1, \ldots, \mathbf{z}_n]$ define standardized p-variate observation vectors on $i = 1, \ldots, n$ subjects, such that $\mathbf{z}_i^{\mathrm{T}} \equiv [z_{i1}, \ldots, z_{ip}] \in \mathbb{R}^p$ denotes a realization of the random vector $Z_i^{\mathrm{T}} \equiv [Z_{i1}, \ldots, Z_{ip}] \in \mathbb{R}^p$. Also, let $\boldsymbol{\Xi}^{\mathrm{T}} \equiv [\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_n]$ define m-variate vectors of latent factor scores on n subjects with $\boldsymbol{\xi}_i^{\mathrm{T}} \equiv [\xi_{i1}, \ldots, \xi_{im}] \in \mathbb{R}^m$. The factor analysis model states that each random variable Z_i is a linear combination of the latent random variables, such that:

$$\underset{(p \times 1)}{\mathbf{z}_i} = \underset{(p \times 1)}{\boldsymbol{\mu}} + \underset{(p \times m)}{\boldsymbol{\Lambda}} \cdot \underset{(m \times 1)}{\boldsymbol{\xi}_i} + \underset{(p \times 1)}{\boldsymbol{\epsilon}_i} \tag{6.11}$$

with $m < p$. In (6.11) $\boldsymbol{\mu} \in \mathbb{R}^p$ denotes an overall mean vector, the $\boldsymbol{\epsilon}_i \in \mathbb{R}^p$ denote the error measurements, and $\boldsymbol{\Lambda} \in \mathbb{R}^{p \times m}$ is a matrix of factor loadings in which each element λ_{jk} is the loading of the jth variable on the kth factor, $j = 1, \ldots, p$, $k = 1, \ldots, m$.

The model maintains the following assumptions: (i) $\mathbf{z}_i \perp\!\!\!\perp \mathbf{z}_{i'}, \forall i \neq i'$; (ii) $\text{rank}(\boldsymbol{\Lambda}) = m$; (iii) $\boldsymbol{\epsilon}_i \sim \mathcal{N}_p(\mathbf{0}, \boldsymbol{\Psi})$, with $\boldsymbol{\Psi} \equiv \text{diag}(\psi_{11}, \ldots, \psi_{pp})$, and $\psi_{jj} > 0$; (iv) $\boldsymbol{\xi}_i \sim \mathcal{N}_m(\mathbf{0}, \boldsymbol{\Phi})$; and (v) $\boldsymbol{\xi}_i \perp\!\!\!\perp \boldsymbol{\epsilon}_{i'}, \forall i, i'$. The likelihood for the observations conditional on the realization of $\boldsymbol{\Xi}$ can then be expressed as:

$$\begin{aligned} L(\boldsymbol{\mu}, \boldsymbol{\Lambda}, \boldsymbol{\Xi}, \boldsymbol{\Psi}, \boldsymbol{\Phi}; \mathbf{Z}) &= \prod_{i=1}^{n} f(\mathbf{z}_i | \boldsymbol{\mu}, \boldsymbol{\Lambda}, \boldsymbol{\xi}_i, \boldsymbol{\Psi}, \boldsymbol{\Phi}) \\ &= \prod_{i=1}^{n} (2\pi)^{-\frac{p}{2}} |\boldsymbol{\Psi}|^{-\frac{1}{2}} \exp\left\{ -\frac{1}{2} \boldsymbol{\epsilon}_i^{\mathrm{T}} \boldsymbol{\Psi}^{-1} \boldsymbol{\epsilon}_i \right\}, \end{aligned}$$

where $\boldsymbol{\epsilon}_i = \mathbf{z}_i - \boldsymbol{\mu} - \boldsymbol{\Lambda}\boldsymbol{\xi}_i$. The likelihood in this form will be important for the construction of data-augmented conditional distributions for MCMC sampling. Marginalizing over $\boldsymbol{\xi}_i$ the observed data likelihood is obtained:

$$\begin{aligned} L(\boldsymbol{\mu}, \boldsymbol{\Lambda}, \boldsymbol{\Psi}, \boldsymbol{\Phi}; \mathbf{Z}) &= \prod_{i=1}^{n} \int f(\mathbf{z}_i | \boldsymbol{\mu}, \boldsymbol{\Lambda}, \boldsymbol{\xi}_i, \boldsymbol{\Psi}, \boldsymbol{\Phi}) \pi(\boldsymbol{\xi}_i; \mathbf{0}, \boldsymbol{\Phi}) \, \partial \boldsymbol{\xi}_i \\ &= \prod_{i=1}^{n} (2\pi)^{-\frac{p}{2}} |\boldsymbol{\Sigma}|^{-\frac{1}{2}} \exp\left\{ -\frac{1}{2} (\mathbf{z}_i - \boldsymbol{\mu})^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} (\mathbf{z}_i - \boldsymbol{\mu}) \right\}, \end{aligned}$$

giving that the factor decomposition constrains the covariance structure of the $\mathbf{z}_i$ to $\boldsymbol{\Sigma} = \boldsymbol{\Lambda}\boldsymbol{\Phi}\boldsymbol{\Lambda}^{\mathrm{T}} + \boldsymbol{\Psi}$. Then, for existence (vi), generally $(p - m)^2 - p - m \geqslant 0$, simply stating that the number of nonredundant elements in the sample correlation matrix $\mathbf{S}$ must be greater than or equal to the number of freely estimable parameters in $\boldsymbol{\Sigma}$, which places in fact an upper bound on m.

6.4.3 An Alternative Factor Analytic Strategy

A classical confirmatory factor analyst expresses theory mainly through the usage of exclusion restrictions in $\boldsymbol{\Lambda}$. In a classical CFA the ponderings by Lasthuizen (2008) given in Section 6.4.1 might then be expressed as:

$$\boldsymbol{\Lambda}_c = \begin{bmatrix} \lambda_{11} & 0 \\ \lambda_{21} & 0 \\ \lambda_{31} & 0 \\ 0 & \lambda_{42} \\ 0 & \lambda_{52} \\ 0 & \lambda_{62} \end{bmatrix} \begin{matrix} \text{IL} \\ \text{ROL} \\ \text{PL} \\ \text{RML} \\ \text{IFL} \\ \text{UL} \end{matrix} .$$

There may be several problems associated with such a (rigid) specification. First, it implies a loss of information in the sense that more exclusion restrictions are applied than is usually necessary for identification of the model. Second, exclusion restrictions may amount to errors of omission, may make the unrealistic assumption that items are factorially pure (in the population), and may induce bias in estimates of the free parameters (cf. Ferrando & Lorenzo-Seva, 2000; van Prooijen & van der Kloot, 2001). These issues are intricately connected to the well-known and widespread situation of exploratively obtained factor structures not being confirmed by CFA. Third, researchers usually have much stronger ideas regarding direction and magnitude of parameter effects that cannot be expressed using exclusion restrictions. Moreover, generally, in the CFA model all attention regarding misspecification is geared towards the pre-specified pattern of factor loadings. The evaluation of model fit in CFA is then essentially the evaluation of a diffuse hypothesis (Hoyle & Duvall, 2004) as it is unclear in the case of misspecification if the pattern of loadings or the factor dimensionality is to blame.

An alternative factor analytic strategy is proposed that separates dimensionality and pattern selection and that may be seen as a Bayesian inequality-constrained-CFA analogue of the four-step approach (Mulaik & Millsap, 2000) to structural equation modeling. The strategy is aimed at alleviating the worries stated above and consists of the following steps:

Step 1. Determine formally the (intrinsic) number of common factors based on a weighing of model fit and model complexity;

Step 2. Whence settled on latent factor dimensionality m, specify an unrestricted confirmatory factor model (UCFM). A UCFM is a FA model that places only minimal restrictions on the factor loadings matrix and the factor covariance matrix for achieving global rotational uniqueness of the factor solution, with the restrictions chosen such that they convey preconceived theoretical meaning and thus render unnecessary post-hoc rotation of the solution for interpretation purposes;

Step 3. Formulate, using the UCFM obtained in Step 2 as a base model, competing inequality constrained factor structures making use of a system of inequality constraints on and between the free parameters in the loadings matrix. Substantive theory is then not represented by exclusion restrictions to express a pre-specified factor loading pattern, but by the imposition of inequality constraints;

Step 4. Compute B_{b0} for each constrained model under consideration and determine the model most supported by the data.

Step 1 implies formal inference on the number of factors to retain, as opposed to usage of popular but heuristical procedures such as the Guttman-Kaiser rule (Guttman, 1954; Kaiser, 1960) or the scree plot (Cattell, 1966). A possible strategy is provided by Peeters (invited revision a), who uses an unconstrained Bayesian factor model for selecting the optimal dimension m. The results from this method are corroborated by the BIC, indicating $m = 2$ as optimal for the data at hand.

The UCFM in *Step 2* corresponds to exploratory factor analysis (EFA) in the sense that only minimal restrictions are placed on the model to achieve a (global) rotationally unique solution for m factors. As such, an unrestricted solution for m common factors does not restrict the factor space and will yield an optimal fit for any model with m factors (Mulaik, 2010). The UCFM, however, cancels the interpretative need for (arbitrary) post-hoc rotation. Peeters (2012) gives sufficient conditions for specifying a UCFM:

C1 Let $\mathbf{\Lambda}$ have at least $m-1$ fixed zeroes in each column;
C2 Let $\text{rank}(\mathbf{\Lambda}^{[k]}) = m-1$, where $\mathbf{\Lambda}^{[k]}$, $k = 1, \ldots, m$, is the submatrix of $\mathbf{\Lambda}$, consisting of the rows of $\mathbf{\Lambda}$ which have fixed zero elements in the kth column with these zeroes deleted;
C3 Let $\mathbf{\Phi}$ be a symmetric positive definite matrix with $\text{diag}(\mathbf{\Phi}) = \mathbf{I}_m$;
C4 Let in each column of $\mathbf{\Lambda}$ one parameter non-fixed by condition C1 be polarity truncated to take only positive or negative values, that is: In each column of $\mathbf{\Lambda}$ one element is to adopt either strict positivity ($\lambda_{jk} > 0$), or strict negativity ($-\lambda_{jk} > 0$).

Conditions C1-C4 are sufficient to provide global rotational uniqueness of the factor solution (Peeters, 2012), and lead to identification of the model when abiding the model assumptions stated in Section 6.4.2. In our take on CFA substantive theory will not be represented by structural equalities to express a pre-specified factor loading pattern, but by imposing a consistent set of inequality constraints on and between the free parameters in the model. Identification is then provided for by ensuring that the set of inequality constraints is embedded within a certain minimal condition for uniqueness, i.e. abides conditions C1-C4. Consistent with conditions C1-C4 the base model is chosen to be:

$$\mathbf{\Lambda}_0 = \begin{bmatrix} \lambda_{11} & \lambda_{12} \\ \lambda_{21} > 0 & \lambda_{22} = 0 \\ \lambda_{31} & \lambda_{32} \\ \lambda_{41} & \lambda_{42} \\ \lambda_{51} & \lambda_{52} \\ \lambda_{61} = 0 & \lambda_{62} > 0 \end{bmatrix} \begin{matrix} \text{IL} \\ \text{ROL} \\ \text{PL} \\ \text{RML} \\ \text{IFL} \\ \text{UL} \end{matrix} .$$

The chosen minimal restrictions comply with all competing inequality-constrained formulations of factor structure to be assessed (Section 6.4.4). In all constrained-model formulations the result-oriented leadership style is thought of as having a large loading on a first factor while having a small loading on a second factor, and the unethical leadership style is thought of as having a large loading on a second factor while having a small loading on a first factor. These variables then serve as an indicator of the respective factors. It is thus reasonable to specify $\{\lambda_{22}, \lambda_{61}\} = 0$ and $\{\lambda_{21}, \lambda_{62}\} > 0$. The polarity of the columns is subsequently tied to strict positivity of λ_{21} and λ_{62}.

6.4.4 Inequality-Constrained Theories

Step 3 introduces structure in the loadings matrix through inequalities rather than through exclusion restrictions. This step explicitly encourages the formulation of competing inequality constrained theories for (statistical) scrutiny. Translating Lasthuizen's (2008) idea of organizational and ethical leadership styles in an inequality-constrained factor structure gives the following loadings matrix:

$$\mathbf{\Lambda}_1 = \begin{bmatrix} \lambda_{11} & > |\lambda_{12}| \\ \lambda_{21} > 0 & \lambda_{22} = 0 \\ \lambda_{31} & > |\lambda_{32}| \\ |\lambda_{41}| & < \lambda_{42} \\ |\lambda_{51}| & < \lambda_{52} \\ \lambda_{61} = 0 & \lambda_{62} > 0 \end{bmatrix} \begin{matrix} \text{IL} \\ \text{ROL} \\ \text{PL} \\ \text{RML} \\ \text{IFL} \\ \text{UL} \end{matrix} .$$

A formulation like

$$\lambda_{11} > |\lambda_{12}| \Rightarrow \begin{cases} \lambda_{11} - \lambda_{12} > 0 \\ \lambda_{11} + \lambda_{12} > 0 \end{cases},$$

indicates the belief that (the positive of) λ_{11} is larger than λ_{12}, irrespective of the latter's sign. The structure $\mathbf{\Lambda}_1$ designates the first factor as 'organizational leadership styles', corresponding to Lasthuizen's (2008) conjecture that inspirational, result-oriented, and passive leadership styles are expressions of leadership that has an organizational thrust. Under this idea IL, ROL, and PL are then expected to positively load on the first factor and to load on the first factor mainly ($\lambda_{11} > |\lambda_{12}|, \lambda_{21} > 0, \lambda_{31} > |\lambda_{32}|$). The second factor then indicates 'ethical leadership styles' as role-modeling, integrity-focused, and unethical leadership may be seen as expressions of ethical leadership. RML, IFL, and UL are then expected to positively load on the second factor and to load on the second factor mainly ($|\lambda_{41}| < \lambda_{42}, |\lambda_{51}| < \lambda_{52}, \lambda_{62} > 0$).

Another viable but competing theory underlying the clustering of leadership styles might lie in the distinction between emphasizing either social exchange or laissez-faire (Bass, 1999). All of inspirational, result-oriented, role-modeling, and integrity-focused leadership styles imply active social exchange (Bass, 1999). Both the passive and unethical leadership styles then imply an inactive, laissez-faire stance. In such a distinction the passive leadership style can also be expected to have a strong negative loading on the active dimension. These ideas are captured in the following factor structure:

$$\mathbf{\Lambda}_2 = \begin{bmatrix} \lambda_{11} & > |\lambda_{12}| \\ \lambda_{21} > 0 & \lambda_{22} = 0 \\ \lambda_{31} < -.3 & \lambda_{32} > .3 \\ \lambda_{41} & > |\lambda_{42}| \\ \lambda_{51} & > |\lambda_{52}| \\ \lambda_{61} = 0 & \lambda_{62} > 0 \end{bmatrix} \begin{matrix} \text{IL} \\ \text{ROL} \\ \text{PL} \\ \text{RML} \\ \text{IFL} \\ \text{UL} \end{matrix} ,$$

where the first factor captures the active social exchange dimension ($\lambda_{11} > |\lambda_{12}|, \lambda_{21} > 0, \lambda_{41} > |\lambda_{42}|, \lambda_{51} > |\lambda_{52}|, \lambda_{31} < -.3$). The second factor then captures the laissez-faire dimension ($\lambda_{32} > .3, \lambda_{62} > 0$).

A third viable factor structure can be found in the juxtaposition of reinforcement style leadership with the idea of the 'full ethical range'. Theory on reinforcement leadership states that leaders mainly influence behaviors through specific rewards and punishments (Treviño, 1986). If this would be the case then the result-oriented and integrity-focused leadership styles can be expected to mainly load together on a separate factor. The idea of the 'full ethical range' implies that active transformational and transactional leadership styles are on a continuum with more laissez-faire-oriented styles. Under this idea, the inspirational, passive, role-modeling, and unethical leadership styles can then be expected to load on a separate factor. Using inequalities, these ideas are expressed as:

$$\mathbf{\Lambda}_3 = \begin{bmatrix} |\lambda_{11}| & < -\lambda_{12} \\ \lambda_{21} > 0 & \lambda_{22} = 0 \\ |\lambda_{31}| & < \lambda_{32} \\ |\lambda_{41}| & < -\lambda_{42} \\ \lambda_{51} & > -\lambda_{52} \\ \lambda_{61} = 0 & \lambda_{62} > 0 \end{bmatrix} \begin{matrix} \text{IL} \\ \text{ROL} \\ \text{PL} \\ \text{RML} \\ \text{IFL} \\ \text{UL} \end{matrix} .$$

In $\mathbf{\Lambda}_3$ the first factor expresses the reinforcement leadership dimension ($\lambda_{21} > 0, \lambda_{51} > -\lambda_{52}$). The second factor expresses the idea of the full ethical range. As polarity in the second column is connected to strict positivity of λ_{62}, then IL, RML, and IFL are expected to load on this factor in the negative range. Hence formulations like $|\lambda_{41}| < -\lambda_{42}$, stating that the negative of λ_{42} is expected to be larger than λ_{41}, irrespective of the latter's sign.

Step 4 is elaborated on in the following two subsections.

6.4.5 Implementation

The strategy regarding prior choice on regression parameters differs from the efforts around inequality constraints on (multivariate) mean structures. Regression coefficients for standardized data are bound to $(-1, 1)$. A natural choice of prior for such parameters that adequately captures model complexity is then a uniform distribution on $(-1, 1)$, being both proper and noninformative. Complex prior specifications may then be avoided as the nuisance parameters can be treated with vague but proper (conjugate) prior formulations. The prior information for the UCFM is specified such that:

$$\pi(\boldsymbol{\mu}, \mathbf{\Psi})\pi_0(\mathbf{\Lambda}_0|\mathbf{\Phi})g(\mathbf{\Xi}|\mathbf{\Phi})\pi(\mathbf{\Phi}) = \pi(\boldsymbol{\mu})\pi(\mathbf{\Psi})\pi_0(\mathbf{\Lambda}_0)\mathbb{1}_{\{\mathbf{I}_p - \text{diag}(\mathbf{\Lambda}_0\mathbf{\Phi}\mathbf{\Lambda}_0^{\text{T}})>0\}}g(\mathbf{\Xi}|\mathbf{\Phi})\pi(\mathbf{\Phi}).$$

The uniform prior on $\mathbf{\Lambda}_0$ is dependent on $\mathbf{\Phi}$ through the indicator function $\mathbb{1}_{\{.\}}$. The indicator function is specified to account for the identification restriction imposed by the fact that we consider $\mathbf{\Sigma}$ as a correlation matrix: It needs to be ensured that $\mathbf{I}_p - \text{diag}\left(\mathbf{\Lambda}_0\mathbf{\Phi}\mathbf{\Lambda}_0^{\text{T}}\right) > 0$ for $\mathbf{\Psi}$ to be positive definite. See Appendix 6B.1 for full specification of the prior elements. The Bayes factor of a factor analytic model bearing inequality constraints on the factor loadings nested within a minimally restricted model can then be expressed through (6.9) as:

$$B_{b0}^{F} = \frac{\displaystyle\int\int_{\{\boldsymbol{\Lambda}_0:\boldsymbol{\Omega}_b\boldsymbol{\lambda}_f-\boldsymbol{\alpha}_b>\mathbf{0}\}} \pi_0(\boldsymbol{\mu},\boldsymbol{\Lambda}_0,\boldsymbol{\Xi},\boldsymbol{\Psi},\boldsymbol{\Phi}|\mathbf{Z})\,\partial\boldsymbol{\Lambda}_0\partial(\boldsymbol{\mu},\boldsymbol{\Xi},\boldsymbol{\Psi},\boldsymbol{\Phi})}{\displaystyle\int_{\{\boldsymbol{\Lambda}_0:\boldsymbol{\Omega}_b\boldsymbol{\lambda}_f-\boldsymbol{\alpha}_b>\mathbf{0}\}} \pi_0(\boldsymbol{\Lambda}_0|\boldsymbol{\Phi})\,\partial\boldsymbol{\Lambda}_0} \equiv \frac{f_b^F}{\omega_b^F}.$$

Appendices 6B.2 and 6B.3 contain the resulting conditional distributions and algorithms for obtaining estimates of f_b^F and ω_b^F. It is efficient to initialize the sampler in Algorithm 6.2 (Appendix 6B.3) with ML estimates. Here ML EFA estimates are used, which can be obtained from many standard packages performing FA. Let $\hat{\boldsymbol{\Lambda}}$ and $\hat{\boldsymbol{\Psi}}$ be the ML EFA estimates of $\boldsymbol{\Lambda}$ and $\boldsymbol{\Psi}$. Then $\hat{\boldsymbol{\Lambda}}$ is direct Oblimin rotated to obtain an oblique structure and an estimate $\hat{\boldsymbol{\Phi}}^s$ of $\boldsymbol{\Phi}$ (see Appendix 6B.1). The rotated loadings matrix then receives structural zeroes and possible column multiplications with minus unity in order to obtain $\hat{\boldsymbol{\Lambda}}_0$; a starting value for the base model loadings matrix. The ML estimate of $\boldsymbol{\mu}$ is simply $\hat{\boldsymbol{\mu}} = \bar{\mathbf{z}}$. Then, by regressing $\boldsymbol{\Xi}$ on $\mathbf{Z}$, an estimate for the factor scores can be obtained as

$$\hat{\boldsymbol{\Xi}} = (\mathbf{Z} - \mathbf{1}_n \otimes \hat{\boldsymbol{\mu}}^{\mathrm{T}})\hat{\boldsymbol{\Psi}}^{-1}\hat{\boldsymbol{\Lambda}}_0\left[(\hat{\boldsymbol{\Phi}}^s)^{-1} + \hat{\boldsymbol{\Lambda}}_0^{\mathrm{T}}\hat{\boldsymbol{\Psi}}^{-1}\hat{\boldsymbol{\Lambda}}_0\right]^{-1},$$

which may serve as starting value for $\boldsymbol{\Xi}$. For evaluation 500,000 iterations are used.

6.4.6 Results

From Table 6.4 it can be inferred that model 2 is least complex and has the best fit, resulting in constrained model 2 deemed best in comparison to the unconstrained model. The Bayes factor of a constrained against another constrained model may for interpretation be referred to the usual half-units on the $\log_{10}$ scale (Jeffreys, 1961) or the suggested units on the double natural logarithm scale (Kass & Raftery, 1995). These Bayes factors can be found in Table 6.5, which represents a Bayes factor matrix.

Table 6.4. Estimated Bayes factors for constrained models on the leadership data

M_1	M_2	M_3
$\hat{f}_1^F = .000$	$\hat{f}_2^F = .533$	$\hat{f}_3^F = .000$
$\omega_1^F = 2^{-8}$	$\hat{\omega}_2^F = 1.518\text{e-}3$	$\omega_3^F = 2^{-7}$
$\hat{B}_{10}^F = .000$	$\hat{B}_{20}^F = 350.993$	$\hat{B}_{30}^F = .000$

Table 6.5 suggests that model 2 is unequivocally best. Given the data, there is decisive evidence in favor of model 2 against both model 1 and 3 (when interpreting twice the natural logarithm of the involved Bayes factors as discussed in Kass & Raftery, 1995). The estimates of the UCFM for the MBS data given in Table 6.6 also support model 2.

The results indicate that the six well-known leadership scales can be reduced to two latent factors. These latent factors represent a dimension of leadership based on active social exchange versus a dimension of laissez-faire leadership, respectively. The two latent factors are moreover appreciably correlated with a posterior mean of $-.555$ and a 95% credible interval of $[-.645, -.456]$. The results thus support leadership theories proposed by Bass (1999).

Table 6.5. Bayes factor matrix for the leadership data

$\hat{B}_{from\|to}$	0	1	2	3
0	1	∞	2.849e-3	∞
1	.000	1	.000	–
2	350.993	∞	1	∞
3	.000	–	.000	1

Table 6.6. Posterior Means and 95% Credible Intervals for $\mathbf{\Lambda}_0$, diag($\mathbf{\Psi}$) and the free elements in $\mathbf{\Phi}$ on the leadership data

	Parameter	Mean	95% CI	Parameter	Mean	95% CI	Item
$\mathbf{\Lambda}_0$	λ_{11}	.777	[.692, .860]	λ_{12}	-.215	[-.307,-.125]	IL
	λ_{21}	.687	[.605, .772]	λ_{22}	-	-	ROL
	λ_{31}	-.304	[-.401,-.202]	λ_{32}	.458	[.358, .562]	PL
	λ_{41}	.719	[.637, .803]	λ_{42}	-.204	[-.292,-.114]	RML
	λ_{51}	.627	[.542, .714]	λ_{52}	-.259	[-.349,-.168]	IFL
	λ_{61}	-	-	λ_{62}	.927	[.869, .983]	UL
$\mathbf{\Psi}, \boldsymbol{\mu}$	ψ_{11}	.153	[.115, .194]	μ_1	-3.268e-4	[-.080, .079]	IL
	ψ_{22}	.522	[.448, .603]	μ_2	-2.356e-4	[-.062, .062]	ROL
	ψ_{33}	.538	[.472, .612]	μ_3	1.679e-4	[-.063, .064]	PL
	ψ_{44}	.269	[.228, .316]	μ_4	-3.001e-4	[-.075, .074]	RML
	ψ_{55}	.351	[.303, .404]	μ_5	-2.706e-4	[-.072, .071]	IFL
	ψ_{66}	.124	[.089, .172]	μ_6	1.383e-4	[-.083, .083]	UL
$\mathbf{\Phi}$	ϕ_{12}	-.555	[-.645,-.456]				

6.5 Discussion

Many theories in political science implicitly or explicitly imply restrictions on parameters in their (statistical) model-translated counterparts. This paper presented a framework for the evaluation of general linear models that bear linear inequality constraints on mean- or regression-type parameters. The framework is inherently Bayesian and incorporates a computationally simple Bayes factor for the selection of the best model (in terms of a weighting of model fit and model complexity) out of an *a priori* set of inequality-constrained models. The framework is thus intended to support a confirmatory thrust to research. The workings of the framework were exemplified by two illustrations. The first example evaluated theories of political socialization by means of an ANOVA model with differing inequality constrained structures on the group means. Another example evaluated theories of leadership using an alternative take on confirmatory factor analysis that makes use of factor structures formulated with systems of inequality constraints on the factor loadings (regression parameters).

For elaborate systems of inequality constraints in large models the permissable parameter space may get (very) small. This may imply computational intensity in terms of the number of iterations needed in order to get an accurate estimate of B_{b0}. However, it is not necessary to impose inequalities on or between all model parameters and systems of inequalities may be combined with sparse model formulations. Another limitation of the present exposition is the focus on usage of linear inequalities in the general linear model only. Extensions to more general systems of inequalities and more general models are the subject of current study.

Notwithstanding the limitations, the framework has much to offer to the empirical political scientist seeking to use statistical methodology to evaluate theory. The framework allows for more flexible parameter specification in the sense that direction and magnitude of parameter effects may directly be accounted for in model formulation. Moreover, it allows for the direct investigation of a multitude of competing models and has the advantage that more information can be drawn from a single analysis. In short, the framework allows for more representative model-translation and evaluation of substantive political theory.

6A Implementation Details Illustration I

6A.1 Elements of Prior Specification

As stated, a constrained posterior prior is constructed for inequality-constrained-model selection in the ANOVA setting. Essentially, a posterior prior is a reference prior updated with a training sample. For the ANOVA setting the posterior prior for the unconstrained model can be stated as:

$$\begin{aligned}\pi_0(\boldsymbol{\mu}, \sigma^2|\mathbf{y}(\ell), \mathbf{G}(\ell)) &\propto L(\boldsymbol{\mu}, \sigma^2, \mathbf{G}(\ell); \mathbf{y}(\ell)) \cdot \sigma^{-2} \\ &= \prod_{j=1}^{J} \mathcal{N}(\bar{y}(\ell)_j, \sigma^2/n(\ell)_j) \cdot \mathcal{I}\chi^2(\rho(\ell), \varpi(\ell)^2), \end{aligned} \tag{6.12}$$

where $\mathbf{y}(\ell)$ denotes a training sample, $\mathbf{G}(\ell)$ denotes group membership for the elements in the training sample, $n(\ell)$ denotes training sample size whereas $n(\ell)_j$ denotes training sample size for group j, and $\bar{y}(\ell)_j$ states the mean of group j in the training sample. Moreover, the degrees of freedom for the scaled inverse χ^2 is given by $\rho(\ell) = n(\ell)-1$ while its scale is given by $\varpi(\ell)^2 = [n(\ell)-1]^{-1} \sum_{j=1}^{J} \sum_{i=1}^{n(\ell)_j} [y(\ell)_i - \bar{y}(\ell)_j]^2$, with $y(\ell)_i$ denoting the score on the dependent variable of the i-th person/element in the training sample.

For computational efficiency it is convenient to regard $\boldsymbol{\mu}$ and σ^2 as *a priori* independent so that the semi-conjugate counterpart of (6.12) will be utilized. Moreover, a restricted analogue of the semi-conjugate version of (6.12) will be used that restricts the prior information for each element μ_j to be the same. The reason for this choice is that usage of $\bar{y}(\ell)_j$ may lead, for models that are specified using inequality constraints, to a defective complexity measure implicit in the Bayes factor (see, e.g., Mulder et al., 2010; van Wesel et al., 2011). The semi-conjugate restricted posterior prior rendered is:

$$\pi_0^r(\boldsymbol{\mu}, \sigma^2 | \mathbf{y}(\ell), \mathbf{G}(\ell)) \propto \prod_{j=1}^{J} \mathcal{N}(\bar{y}(\ell), s(\ell)^2/n(\ell)) \cdot \mathcal{I}\chi^2(\rho(\ell), \varpi(\ell)^2), \qquad (6.13)$$

where $\bar{y}(\ell)$ is the mean of the training sample and where $s(\ell)^2 = n(\ell)^{-1} \sum_{i=1}^{n(\ell)} [y(\ell)_i - \bar{y}(\ell)]^2$. In this situation we let $\varpi(\ell)^2$ denote $[n(\ell)-1]^{-1} \sum_{i=1}^{n(\ell)} [y(\ell)_i - \bar{y}(\ell)]^2$. The minimal training sample (Berger & Pericchi, 1996) for obtaining a proper posterior prior of this type consists of two randomly selected observations from the total sample.

The posterior prior depends on the choice of training sample. As such it is natural to average (6.13) in some way over $\ell = 1, \ldots, L$ training samples to increase stability and minimize dependence. A natural manner of averaging is the empirical expected-posterior prior principle (Pérez & Berger, 2002), leading to:

$$\frac{1}{L} \sum_{\ell=1}^{L} \pi_0^r(\boldsymbol{\mu}, \sigma^2 | \mathbf{y}(\ell), \mathbf{G}(\ell)) \equiv \pi_0(\boldsymbol{\mu}, \sigma^2), \qquad (6.14)$$

which is essentially a mixture of posterior priors.

Posterior sampling using (6.14) is intensive (see Pérez & Berger, 2002 for algorithms). That is why an approximation to (6.14) is used in order to lessen computational burden and to speed up posterior sampling. This approximation requires mutually independent approximations of $\boldsymbol{\mu}$ and σ^2. This approximation is given by equation (6.10) stated previously:

$$\hat{\pi}_0(\boldsymbol{\mu}, \sigma^2) = \hat{\pi}_0(\boldsymbol{\mu}) \hat{\pi}(\sigma^2) = \mathcal{N}_J(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\mu}}}^{-}) \cdot \mathcal{I}\chi^2(\hat{\rho}, \hat{\varpi}^2).$$

In this expression

$$\hat{\boldsymbol{\mu}} = \mathbf{1}_J \cdot \left[(LCJ)^{-1} \sum_{\ell,c,j} \mu_j^{(\ell,c)} \right],$$

and

$$\hat{\boldsymbol{\Sigma}}^{-}_{\hat{\boldsymbol{\mu}}} = (LC)^{-1} \sum_{\ell,c} (\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}^{(\ell,c)})^{\mathrm{T}} (\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}^{(\ell,c)}),$$

need information provided by the following algorithm:

Algorithm 6.1 (Sampling μ_j)

1: **for** $\ell = 1$ to L **do**
2: Draw $\mathbf{y}(\ell)$ and corresponding $\mathbf{G}(\ell)$ randomly from $\mathbf{y}$
3: **for** $c = 1$ to C **do**
4: **for** $j = 1$ to J **do**
5: Generate μ_j from $\mathcal{N}(\bar{y}(\ell), s(\ell)^2/n(\ell))$
6: **end for**
7: **end for**
8: **end for**

Furthermore, $\hat{\rho} = n(\ell)$ and $\hat{\varpi}^2 = L^{-1} \sum_{\ell=1}^{L} \varpi(\ell)^2$ for the approximation of the mixture of L scaled inverse χ^2 distributions.

6A.2 Conditional Distributions

For the conditional distribution of $\boldsymbol{\mu}$ in the ANOVA model we find

$$\pi_0(\boldsymbol{\mu}|\mathbf{y}, \mathbf{G}, \sigma^2) \stackrel{d}{=} \mathcal{N}_J(\tilde{\boldsymbol{\mu}}, \tilde{\boldsymbol{\Sigma}}^{-}_{\tilde{\boldsymbol{\mu}}}),$$

where

$$\tilde{\boldsymbol{\mu}} = \tilde{\boldsymbol{\Sigma}}^{-}_{\tilde{\boldsymbol{\mu}}} [(\hat{\boldsymbol{\Sigma}}^{-}_{\hat{\boldsymbol{\mu}}})^{-1} \hat{\boldsymbol{\mu}} + \sigma^{-2} \mathbf{G}^{\mathrm{T}} \mathbf{G} \bar{\mathbf{y}}],$$
$$\tilde{\boldsymbol{\Sigma}}^{-}_{\tilde{\boldsymbol{\mu}}} = [(\hat{\boldsymbol{\Sigma}}^{-}_{\hat{\boldsymbol{\mu}}})^{-1} + \sigma^{-2} \mathbf{G}^{\mathrm{T}} \mathbf{G}]^{-1}.$$

For the conditional distribution of σ^2 we find

$$\pi_0(\sigma^2|\mathbf{y}, \mathbf{G}, \boldsymbol{\mu}) \stackrel{d}{=} \mathcal{I}\chi^2(\tilde{\rho}, \tilde{\varpi}^2),$$

where $\tilde{\rho} = n + n(\ell)$ and

$$\tilde{\varpi}^2 = \frac{n(\ell)\hat{\varpi}^2 + n\gamma}{n + n(\ell)},$$

with $\gamma = \frac{1}{n} \sum_{i=1}^{n} \sum_{i \in j} (y_i - \mu_j)^2$.

6B Implementation Details Illustration II

6B.1 Elements of Prior Specification

For notational convenience the model index 0 is dropped where appropriate throughout Appendix 6B. The prior on $\boldsymbol{\mu}$ is defined such that $\pi(\mu_j) \propto 1$ over the Lebesgue measure on $(-\infty, \infty)$. While this prior is improper it will not pose problems in terms of causing indeterminable Bayes factors, as it is a common parameter of equivalent dimension in all models under consideration.

Similarly the prior over the unrestricted elements λ_{jk} in $\mathbf{\Lambda}_0$ is defined such that $\pi(\lambda_{jk}) \propto 1$ over the Lebesgue measure on $(-1, 1)$ as these elements are bounded theoretically by plus and minus unity in correlation structure modeling. Let λ_{jk}^p denote loadings involved in a polarity truncation. Then we define $\pi((-)\lambda_{jk}^p) \propto 1$ over the Lebesgue measure on $(0, 1)$. The prior on $\mathbf{\Lambda}_0$ is dependent on $\mathbf{\Phi}$ through the indicator function $\mathbb{1}_{\{\cdot\}}$. The indicator function is specified to account for the identification restriction imposed by the fact that we consider $\mathbf{\Sigma}$ as a correlation matrix: It needs to be ensured that $\mathbf{I}_p - \text{diag}\left(\mathbf{\Lambda}_0\mathbf{\Phi}\mathbf{\Lambda}_0^{\mathrm{T}}\right) > 0$ for $\mathbf{\Psi}$ to be positive definite. The prior on $\mathbf{\Lambda}_0$ including the indicator function, is important in the definition of ω_b^F.

As exemplified by Martin and McDonald (1975) and Lopes and West (2004) a prior that decays to zero at the origin is needed on the ψ_{jj} elements in order to induce proper posteriors as this prevents the posterior from placing infinite mass at $\psi_{jj} = 0$ for some j. A convenient choice is $\psi_{jj} \sim \mathcal{IG}(\nu/2, \nu d/2)$, where $\mathcal{IG}(\cdot, \cdot)$ denotes the inverse gamma distribution. Shape and scale will be chosen diffuse but proper, i.e., as hyperparameters for the $\pi(\psi_{jj})$ we choose $\nu = 3$ and $\nu d = 1/2$. This setting results in a weakly informative or diffuse prior for the ψ_{jj} elements that decays to zero sufficiently rapid as ψ_{jj} tends to zero, and places the prior mean in the middle of the interval (0,1].

The prior $\pi(\mathbf{\Xi}|\mathbf{\Phi})$ stems from model assumption (iv), stating that $\boldsymbol{\xi}_i \sim \mathcal{N}_m(\mathbf{0}, \mathbf{\Phi})$. The prior on the hyperparameter $\mathbf{\Phi}$ will be modeled as $\pi(\mathbf{\Phi}) = \mathcal{IW}_m(\mathbf{\Upsilon}, \tau)$, where $\mathcal{IW}_m(\cdot, \cdot)$ denotes the inverse Wishart distribution. Again, we will choose shape and scale diffuse but proper in applications, i.e., $\mathbf{\Upsilon} = \mathbf{I}_m$ and $\tau = m + 2$.

Remark 6.1. From the prior specification it is clear that we sample $\mathbf{\Phi}$ as a covariance matrix. However, in specifying uniqueness conditions, and in keeping with the modeling of a standardized FA covariance structure, we desire $\mathbf{\Phi}$ to be a correlation matrix. While there are prior specifications available for $\mathbf{\Phi}$ as a correlation matrix (e.g., Ansari & Jedidi, 2000), these have to be truncated by a subset of the hypercube $[-1, 1]^{m(m-1)/2}$, and as such imply less straightforward MCMC samplers. Here we opt for a more convenient road. Our prior on $\mathbf{\Phi}$ ensures that we sample positive definite covariance matrices. The sample covariance matrix $\mathbf{\Phi}$ contains all the information for a transformation to a positive definite correlation matrix, i.e., $\phi_{kk'}^s = \phi_{kk'}/\sqrt{\phi_{kk}} \cdot \sqrt{\phi_{k'k'}}$, where $\phi_{kk'}^s$ denotes the standardized covariance between the kth and k'th latent variable. This strategy allows retainment of straightforward MCMC samplers. From this point, we will use $\mathbf{\Phi}^s$ to denote the desired correlation version of $\mathbf{\Phi}$.

6B.2 Conditional Distributions

Define $\mathbf{z}_j$, $\mathbf{\Lambda}_{0j}$, and $\mathbf{1}_n$ to be the jth column of $\mathbf{Z}$, the jth row of $\mathbf{\Lambda}_0$, and an n-dimensional unit vector, respectively. The conditional posterior distributions, based on the priors from the previous section, are then given as below.

For the conditional distribution of $\boldsymbol{\mu}$, we find

$$\pi(\boldsymbol{\mu}|\mathbf{Z}, \mathbf{\Lambda}_0, \mathbf{\Xi}, \mathbf{\Psi}) \stackrel{d}{=} \mathcal{N}_p(\tilde{\boldsymbol{\mu}}, \mathbf{\Sigma}_{\boldsymbol{\mu}}),$$

where $\tilde{\boldsymbol{\mu}} = \bar{\mathbf{z}} - \mathbf{\Lambda}_0\bar{\boldsymbol{\xi}}$, $\mathbf{\Sigma}_{\boldsymbol{\mu}} = \mathbf{\Psi}n^{-1}$, $\bar{\mathbf{z}} = n^{-1}\sum_{i=1}^n \mathbf{z}_i$, and $\bar{\boldsymbol{\xi}} = n^{-1}\sum_{i=1}^n \boldsymbol{\xi}_i$. The rows of $\mathbf{\Lambda}_0$ are independent such that we need to find the conditional distribution $\pi(\mathbf{\Lambda}_{0j}^{\mathrm{T}}|\mathbf{z}_j, \mu_j, \mathbf{\Xi}, \psi_{jj}, \mathbf{\Phi}^s)$, $j = 1, \ldots, p$. This conditional is slightly different for $\mathbf{\Lambda}_{0j}^{\mathrm{T}}$ with certain restrictions relative to $\mathbf{\Lambda}_{0j}^{\mathrm{T}}$ with only free parameters. We follow Lee (2007) in developing general notation capturing both situations. Let $\mathbf{c}_j$ be a row vector of dimension m corresponding to $\mathbf{\Lambda}_{0j}$, indicating $c_{jk} = 0$ if λ_{jk} is a fixed parameter and indicating $c_{jk} = 1$ if λ_{jk} is a free parameter or a parameter involved in a polarity truncation. In addition, let $r_j = c_{j1} + \ldots + c_{jm}$. Moreover, let $\mathbf{\Lambda}_{0j}^\star$ and $\mathbf{\Xi}_j^\star$ denote the r_j-dimensional row vector containing the unknown elements in $\mathbf{\Lambda}_{0j}$ and the $(n \times r_j)$ submatrix of $\mathbf{\Xi}$ for which the columns corresponding to $c_{jk} = 0$ are deleted, respectively. We then find

$$\begin{aligned}&\pi(\mathbf{\Lambda}_{0j}^{\star\mathrm{T}}|\mathbf{z}_j, \mu_j, \mathbf{\Xi}_j^\star, \psi_{jj}, \mathbf{\Phi}^s)\\ &\qquad\stackrel{d}{=} \mathcal{N}_m(\tilde{\mathbf{\Lambda}}_{0j}^\star, \mathbf{\Sigma}_{\mathbf{\Lambda}_{0j}^\star})\mathbb{1}_{\left\{0<(-)\lambda_{jk}^p<1 \;\cap\; -1<\lambda_{jk'}<1 \;\forall k'\neq k \;\cap\; 1-\mathbf{\Lambda}_{0j}\mathbf{\Phi}^s\mathbf{\Lambda}_{0j}^{\mathrm{T}}>0\right\}},\end{aligned}$$

for rows which contain a loading involved in a polarity truncation, and

$$\pi(\mathbf{\Lambda}_{0j}^{\star\mathrm{T}}|\mathbf{z}_j, \mu_j, \mathbf{\Xi}_j^\star, \psi_{jj}, \mathbf{\Phi}^s) \stackrel{d}{=} \mathcal{N}_m(\tilde{\mathbf{\Lambda}}_{0j}^\star, \mathbf{\Sigma}_{\mathbf{\Lambda}_{0j}^\star})\mathbb{1}_{\left\{-1<\lambda_{jk}<1 \;\forall k \;\cap\; 1-\mathbf{\Lambda}_{0j}\mathbf{\Phi}^s\mathbf{\Lambda}_{0j}^{\mathrm{T}}>0\right\}},$$

for rows which do not, where

$$\begin{aligned}\tilde{\mathbf{\Lambda}}_{0j}^\star &= (\mathbf{\Xi}_j^{\star\mathrm{T}}\mathbf{\Xi}_j^\star)^{-1}\mathbf{\Xi}_j^{\star\mathrm{T}}(\mathbf{z}_j - \mathbf{1}_n\mu_j),\\ \mathbf{\Sigma}_{\mathbf{\Lambda}_{0j}^\star} &= \psi_{jj}(\mathbf{\Xi}_j^{\star\mathrm{T}}\mathbf{\Xi}_j^\star)^{-1}.\end{aligned}$$

The rows of $\mathbf{\Xi}$ are also independent such that we need to find the conditional distribution $\pi(\boldsymbol{\xi}_i|\mathbf{z}_i, \boldsymbol{\mu}, \mathbf{\Lambda}_0, \mathbf{\Psi}, \mathbf{\Phi}^s)$, $i = 1, \ldots, n$, which, with the help of the Woodbury matrix identity (Woodbury, 1950), can be found as

$$\pi(\boldsymbol{\xi}_i|\mathbf{z}_i, \boldsymbol{\mu}, \mathbf{\Lambda}_0, \mathbf{\Psi}, \mathbf{\Phi}^s) \stackrel{d}{=} \mathcal{N}_m(\tilde{\boldsymbol{\xi}}, \mathbf{\Sigma}_\xi),$$

where

$$\begin{aligned}\tilde{\boldsymbol{\xi}} &= \mathbf{\Sigma}_\xi\mathbf{\Lambda}_0^{\mathrm{T}}\mathbf{\Psi}^{-1}(\mathbf{z}_i - \boldsymbol{\mu}),\\ \mathbf{\Sigma}_\xi &= [(\mathbf{\Phi}^s)^{-1} + \mathbf{\Lambda}_0^{\mathrm{T}}\mathbf{\Psi}^{-1}\mathbf{\Lambda}_0]^{-1}.\end{aligned}$$

The conditional distribution $\pi(\mathbf{\Phi}|\mathbf{Z}, \mathbf{\Xi})$ is easily found as

$$\pi(\mathbf{\Phi}|\mathbf{Z}, \mathbf{\Xi}) \stackrel{d}{=} \mathcal{IW}_m(\mathbf{\Xi}^{\mathrm{T}}\mathbf{\Xi} + \mathbf{\Upsilon}, n + \tau).$$

Lastly, the conditional distribution for the diagonal elements of $\mathbf{\Psi}$, $\pi(\psi_{jj}|\mathbf{z}_j, \mu_j, \mathbf{\Lambda}_{0j}, \mathbf{\Xi})$, $j = 1, \ldots, p$, can be found as

$$\pi(\psi_{jj}|\mathbf{z}_j, \mu_j, \mathbf{\Lambda}_{0j}, \mathbf{\Xi}) \stackrel{d}{=} \mathcal{IG}((n + \nu)/2, (\beta_j + \nu d)/2),$$

where

$$\beta_j = (\mathbf{z}_j - \mathbf{1}_n\mu_j - \mathbf{\Xi}\mathbf{\Lambda}_{0j}^{\mathrm{T}})^{\mathrm{T}}(\mathbf{z}_j - \mathbf{1}_n\mu_j - \mathbf{\Xi}\mathbf{\Lambda}_{0j}^{\mathrm{T}}).$$

6B.3 Numerical Implementation

Information from a Gibbs sampler may be used to assess f_b^F. Numerical exploration of the complexity measure ω_b^F can also be performed quite simply. As each oblique representation $\mathbf{\Lambda}\mathbf{\Phi}\mathbf{\Lambda}^{\mathrm{T}}$ has equivalent orthogonal representations it is possible to, under the given prior specifications, represent each row in the prior loadings matrix by an $(m - t_j)$-dimensional unit ball. The index t counts the number of restrictions under conditions C1 and C4, such that t_j counts the analogous for row j. The full prior space is then given by an open $(pm - t)$-dimensional ball whose properties can be reduced to the study of the $(pm - t)$-dimensional unit ball. Model complexity is then given by the proportion of the interior of the $(pm - t)$-dimensional unit ball abiding the system of inequality constraints. These observations are incorporated in the algorithms below.

Algorithm 6.2 (Gibbs Sampler for $\hat{f}_b^F$) The input consists of the data $\mathbf{Z}$ and starting values $\mathbf{\Lambda}_0^{(0)}$, $\mathbf{\Xi}^{(0)}$, $\mathbf{\Phi}^{s(0)}$, and $\mathbf{\Psi}^{(0)}$. The output is the estimate $\hat{f}_b^F$. The algorithm uses a Gibbs sampler to estimate $\hat{f}_b^F$. Each of $\boldsymbol{\mu}, \mathbf{\Lambda}_0, \mathbf{\Phi}, \mathbf{\Xi}$, and $\mathbf{\Psi}$ is treated as a single block. From the prior specification it is clear that we sample $\mathbf{\Phi}$ as a covariance matrix. However, in specifying uniqueness conditions, and in keeping with the modeling of a standardized FA covariance structure, $\mathbf{\Phi}$ is desired to be a correlation matrix. That is why the draws $\mathbf{\Phi}^{(c)}$ are standardized to the correlation matrix $\mathbf{\Phi}^{s(c)}$. In the following, $\zeta(\cdot, \cdot)$ denotes the normal density and model indices are suppressed where appropriate, for sake of notational simplicity.

1: Set $\mathbf{\Lambda}_0^{(0)}$, $\mathbf{\Xi}^{(0)}$, $\mathbf{\Phi}^{s(0)}$, and $\mathbf{\Psi}^{(0)}$
2: **for** $c = 1$ to C **do**
3:    Generate $\boldsymbol{\mu}^{(c)}$ from $\pi(\boldsymbol{\mu}|\mathbf{Z}, \mathbf{\Lambda}_0^{(c-1)}, \mathbf{\Xi}^{(c-1)}, \mathbf{\Psi}^{(c-1)})$
4:    Generate $\mathbf{\Lambda}_0^{(c)}$ from $\prod_{j=1}^{p} \zeta(\tilde{\mathbf{\Lambda}}_{0j}^{\star}, \mathbf{\Sigma}_{\mathbf{\Lambda}_{0j}^{\star}})$
5:    **if** $\exists jk$ such that $\neg(-1 < \lambda_{jk}^{(c)} < 1)$ or $\neg(0 < (-)\lambda_{jk}^{p(c)} < 1)$ **or** $\mathbf{I}_p - \mathrm{diag}\left(\mathbf{\Lambda}_0^{(c)}\mathbf{\Phi}^{s(c-1)}\mathbf{\Lambda}_0^{(c)\mathrm{T}}\right) \not> 0$ **then**
6:       go to 4:
7:    **else**
8:       Generate $\mathbf{\Xi}^{(c)}$ from $\prod_{i=1}^{n} \pi(\boldsymbol{\xi}_i|\mathbf{z}_i, \boldsymbol{\mu}^{(c)}, \mathbf{\Lambda}_0^{(c)}, \mathbf{\Psi}^{(c-1)}, \mathbf{\Phi}^{s(c-1)})$
9:       Generate $\mathbf{\Phi}^{(c)}$ from $\pi(\mathbf{\Phi}|\mathbf{Z}, \mathbf{\Xi}^{(c)})$

10: Set $\forall k \geqslant k'$ $\phi_{kk'}^{s(c)} = \phi_{kk'}^{(c)} / \sqrt{\phi_{kk}^{(c)}} \cdot \sqrt{\phi_{k'k'}^{(c)}}$
11: Set $\forall k > k'$ $\phi_{k'k}^{s(c)} = \phi_{kk'}^{s(c)}$
12: Generate $\mathbf{\Psi}^{(c)}$ from $\prod_{j=1}^{p} \pi(\psi_{jj} | \mathbf{z}_j, \mu_j^{(c)}, \mathbf{\Lambda}_{0j}^{(c)}, \mathbf{\Xi}^{(c)})$
13: **end if**
14: **end for**
15: $\hat{f}_b^F = C^{-1} \sum_{c=1}^{C} \mathbb{1}_{\{\mathbf{\Omega}_b \boldsymbol{\lambda}_f^{(c)} - \boldsymbol{\alpha}_b > \mathbf{0}\}}$

Algorithm 6.3 (Sampler for $\hat{\omega}_b^F$) The input consists of the number of iterations. The output is the estimate $\hat{\omega}_b^F$. The algorithm proceeds by sampling $\mathbf{\Lambda}_0$ from its respective prior density. Subsequently, the samples are accepted or discarded according to compliance with the polarity restrictions and correlation bound, after which it is checked if the accepted draws comply with the system of inequality constraints. In the following, $\varrho(\cdot, \cdot)$ denotes a uniform density.

1: Set $\varphi = 0$
2: **for** $v = 1$ to V **do**
3: Generate $\mathbf{\Lambda}_0^{(v)}$ from $\prod_{j=1}^{p} \varrho(\dot{\mathbf{\Lambda}}_{0j}^{\star}; -1, 1) \mathit{\Gamma}_j$
4: **if** $\mathbf{I}_p - \mathrm{diag}\left(\mathbf{\Lambda}_0^{(v)} \mathbf{\Lambda}_0^{(v)\mathrm{T}}\right) \not> 0$ **then**
5: go to 3:
6: **else**
7: **if** $\mathbb{1}_{\{\mathbf{\Omega}_b \boldsymbol{\lambda}_f^{(v)} - \boldsymbol{\alpha}_b > \mathbf{0}\}} = 1$ **then**
8: $\varphi = \varphi + 1$
9: **else**
10: $\varphi = \varphi$
11: **end if**
12: **end if**
13: **end for**
14: $\hat{\omega}_b^F = V^{-1} \varphi$

7

Reflections

Section 7.3 is partly based on Section 13.6 of: Kato, B.S. & Peeters, C.F.W. (2008) 'Inequality Constrained Multilevel Models', in: H. Hoijtink, I. Klugkist & P.A. Boelen (Eds.) *Bayesian Evaluation of Informative Hypotheses.* New York: Springer, pp. 273-295; for which C.F.W. Peeters was responsible.

This dissertation has dealt with the interplay between Bayesian statistics, constrained-model selection, and common factor analytic modeling. It has tried to provide a conceptual framework in which both the selection of the dimension of a model (Type I model selection) and the selection of truncations of the parameter space by inequalities (Type II model selection), are seen as integral parts of constrained statistical inference. Bayesian (computational) model selection procedures have been developed within this framework on the basis of which exploratory and confirmatory common factor analytic modeling were reformulated. In this concluding chapter, some reflections are provided on these efforts. First, a short summary is given. Afterwards, some limitations of underlying work are discussed. These limitations connect to Research Aims 1 and 2 mostly (see Chapter 1) and provide inroads for further research. Last, some comments are given on the perceived relation of the proposed integrative strategy (Research Aim 3) with the topic of scientific method.

7.1 Summary

The first aim for this dissertation was the construction of a conceptually and computationally simple Bayes factor for Type I constrained-model selection that is determinate under usage of improper priors and the subsequent utilization of this Bayes factor in a Bayesian EFA concerned with the selection of an optimal dimensionality for m. Chapter 2 has dealt with this aim. It connects the candidate estimator for marginal likelihood computation (Besag, 1989; Chib, 1995) with the usage of training sample priors (Berger & Pericchi, 1996) so that, pending certain conditions, a simulation consistent MCMC implementation of well-known default

Bayes factors is obtained (Berger & Pericchi, 1996). This automated candidate estimator allows for noninformative prior usage in EFA, such that the problem of interference of informative prior information with the determination of the intrinsic dimensionality m may be avoided. It was then shown that a failure to abide the regularity condition of $\mathbf{\Lambda}$ being of full rank, may result in violation of a crucial condition for simulation consistency of estimates stemming from MCMC sampling. Violation of the rank condition on $\mathbf{\Lambda}$ is however endemic when overfactoring. As m is typically unknown one is prone to overfactor for some values of m when trying to determine its optimal value by evaluating all possible values in its domain. It is this problem that has invalidated many classical approaches towards the selection of m and the Bayesian approach does not escape its gravity. However, the Bayesian approach proceeds by evaluations of the posterior space and – through such evaluations – enables the evaluation of regularity condition violation. This implies that the embedding of the automated candidate estimator in a certain assessment strategy that keeps check of the regularity condition for simulation consistency provides an appropriate stopping rule for factor analytic data compression. These results also hold importance for non-Bayesian approaches towards factor analytic dimensionality selection as they imply that for informed decisions regarding factor dimensionality, likelihood ratio and information theoretic approaches benefit from a complete exploration of the likelihood, which can be achieved by objective Bayesian methods.

The second aim of this dissertation was to construct a conceptually and computationally simple Bayes factor for Type II constrained-model selection that is geared towards inequalities on regression-type parameters and the subsequent embedding of this Bayes factor within a strategy that allows one to express factor analytic structure using inequality constraints. This aim has been given attention in Chapters 3 and 4. Chapter 3 gives a set of conditions for global rotational identification of the factor model. The condition set enables the formulation of an unrestricted confirmatory factor model (UCFM). A UCFM is a factor analysis model that places only minimal restrictions on $\mathbf{\Lambda}$ and $\mathbf{\Phi}$ for achieving global rotational uniqueness of the factor solution, with the restrictions chosen such that they convey preconceived theoretical meaning and thus render unnecessary post-hoc rotation of the solution for interpretation purposes. The UCFM is pivotal in designing a strategy for inequality-constrained CFA in Chapter 4.

In Chapter 4 a Bayes factor for Type II constrained-model selection is given. Computation of this Bayes factor is relatively simple as its expressions of model fit and model complexity are explicitly connected to, respectively, the posterior and prior probability mass satisfying the constraints defining the constrained model. This implies that one only needs to evaluate the number of times an appropriate MCMC sampler visits the permissable space. This Bayes factor is then used in rendering a take on CFA in which factor structure is expressed using inequalities. The strategy consists of choosing as a base model a UCFM and to subsequently express factor structure using inequalities on and between the free parameters in $\mathbf{\Lambda}$. Parameter restrictions in the context of CFA are then taken beyond exclusion restrictions and the prevention of impermissible estimates by allowing inequality and approximate equality constraints to express substantive theoretical ideas regarding direction and magnitude of factor loadings.

The third and final aim was to let the provisions from Aims 1 and 2 conjoin in order to develop an integrative factor analytic strategy that proposes a bridge crossing the divide between EFA and CFA and to bring this strategy to bear on substantive fields of study outside the direct realm that brought about the FA model. This aim was the focus of the second part of the dissertation. The integrative strategy, when used in full, is confirmatory in nature. It makes the selection of m through the unrestricted EFA model part of the inferential machine on (confirmatory) factor analysis in order to avoid embarking on diffuse model questions. The integrative strategy consists of the following steps: (1) Embark on evaluating a series of unrestricted (EFA) models with respect to their factor dimensionality (using, for example, the methods developed in Chapter 2); (2) Whence settled on latent factor dimensionality m, specify a UCFM; (3) Formulate, using the UCFM as a base model, competing inequality-constrained factor structures making use of a system of inequality constraints on and between the free parameters in the loadings matrix; (4) Compute for each constrained model the Type II constrained-model selection Bayes factor and determine the constrained model most supported by the data. Chapter 5 applied the integrative strategy to epidemiological data on the metabolic syndrome. Chapter 6 utilized this strategy to analyze political sciences data on decision acceptance and ethical leadership.

7.2 Limitations and Further Research

A first limitation of the present work is that, although many indeterminacies in the common factor analytic model have been reviewed, not all indeterminacies have been given full attention. Even if $\mathbf{\Lambda}$ and $\mathbf{\Psi}$ are identified, $\mathbf{\Xi}$ need not be. This is called factor (score) indeterminacy and to see how this may be the case, consider the following. The p observed variables represent a basis vector that spans a p-dimensional space. The common factor model, however, encompasses an $(m + p)$-dimensional space (the number of common and unique factors). This implies that common factors are not likely to be determined completely by the observables in the analysis, leading only part of the common factors to be estimable by the observed variables.

Factor score indeterminacy has a long history in the factor analytic literature, starting with the seminal works of Wilson (1928a, 1928b). The debate has led to strong opinions. These range from the idea that factor indeterminacy may invalidate the factor model (Guttman, 1955) to the stance that factor indeterminacy is of no practical importance as the interpretation of factors is not based on $\mathbf{\Xi}$, but on $\mathbf{\Lambda}$ and $\mathbf{\Phi}$, which remain unaltered for alternative solutions to common factor scores. In general, there are two schools of thought on the factor indeterminacy issue (see Mulaik, 2010, Chapter 13). The first is the *alternative solution position*, which emphasizes that there exist alternative solutions to estimates of common factors and that these alternative solutions may harbor different (interpretational) implications (e.g., Schönemann & Wang, 1972; Steiger, 1996). The second is the *posterior distribution position* which states that the posterior distribution, conditional on the observed data, contains the relevant information on Ξ (Bartholomew, 1981; Aitkin & Aitkin, 2005). See Mulaik (2010, Chapter 13) for an (historical) overview of the factor indeterminacy debate. It is felt that the nature of factor indeterminacy deserves more attention as does the role Bayesianism can have in the debate.

A second limitation is that attention was with the common normal-theory (exploratory and confirmatory) factor analysis model only. This dissertation sought to show that, although the common factor model has been well-studied, there is still room for learning and renewal regarding the major topics such as the selection of m and the testing of factor structure. As the common factor model is either the basis for, or a special case of, more general techniques, extensions of certain developments may give fruitful inroads for further research. An obvious extension would be the consideration of observed variables that are non-normal, or non-continuous, to accommodate research practice in which variables are often binary or categorical. Another rather obvious extension would be to position the developments regarding m selection and inequality-constrained factor structure in exploratory and confirmatory structural equation modeling. Many other considerations regarding extension are possible.

On the same note, focus has been solely with linear inequalities on the elements of $\mathbf{\Lambda}$ as alternative expressions of factor structure. The imposition and demarcation of inequalities on other parameters than $\mathbf{\Lambda}$ is straightforward and can be handled with the methods provided in Chapter 4. Still, expressions of factor structure have been considered through $\mathbf{\Lambda}$ only. In the CFA model, theory may also be represented by, for example, correlated errors. Such expressions can be accommodated within the provided framework by relaxing the UCFM such that a restricted factor solution is obtained that frees room for covariances between error terms. Similarly, nonlinear inequalities have only been used to restrict the posterior parameter space to its theoretical bounds. There may be applications where nonlinear inequalities may also be wished for as expressions of factor structure. In such situations the prior information and the definition of model complexity for inequality-constrained factor models are in need of reassessment.

A final limitation to be discussed is that of computational intensity. In general the automated candidate estimator procedure for Type I model selection will be intensive for matrix-variate parameters and a substantial number of parameter blocks. The proposed procedure for the selection of m then increases in computational intensity as m grows large. The procedure for computing the Bayes factor for inequality-constrained-model selection looses efficiency as the feasible space defined by the system of inequality constraints gets smaller and larger portions of the unconstrained posterior space are discarded. For large factor analytic problems (meaning relatively large m and/or relatively large p) computational intensity can be substantial when many cross-inequalities are specified. Specifically, the methods as developed here are suitable, when having access to desktop terminals only, for problems of moderate (data) size.

The computational intensity does not discard the methods as developed here. There was a time when simply performing a factor analysis could take weeks (and could earn one a Ph.D.; Steiger, 1996). With growing computer power the proposed methods will gain in computational feasibility for larger problems. Notwithstanding, the problem of computational intensity may also propose avenues for further research. A first avenue would be to see if the proposed methods can be properly accommodated with (adaptations of) approximation methods, such as integrated nested Laplace approximations (Rue, Martino, & Chopin, 2009). For the methods regarding Type II model selection computation time can be reduced by not specifying inequalities between or on all free parameters and for large models inequalities may be combined with more sparse formulations of $\mathbf{\Lambda}$. A second avenue for further research would then be an exploration of the question how small the feasible region may get for simulation efforts to stay reasonably efficient. In a deeper sense the accommodation of the developed methods in faster computational procedures will allow for an exploration of the extent to which the pointers and pitfalls indicated by the present study of the classic common factor model geared towards moderately-sized data translate to factor analysis for massive (high-throughput) data.

7.3 Some Comments on Method

In treating the factor model one basically treats inference for observational studies. Research Aim 3 proposed an integrative strategy for factor analytic studies that is suitable for (at least) moderately-sized problems. The developments in Chapter 2 can of course be used to perform solely an EFA. In this respect one would use EFA in a theory-generating manner with a Peircean abductive logic (Haig, 2005). It is also possible to use the developments in Chapter 4 in a purely confirmatory manner, by postulating a choice of m for which confirmatory structure is formulated (possibly using inequalities). Many flavors of integration of EFA and CFA are possible. The integrative strategy as proposed in this dissertation uses both EFA and CFA in an inferential manner, utilizing an unrestricted model (i) for the selection of m and (ii) as a base model for the formulation of factor structure through inequalities once the dimension m is chosen.

When used in full the thrust of the proposed integrative strategy is confirmatory. When used on the same data it is assumed that the EFA in the integrative strategy is used for the selection of m only, not for the assessment of rotated factor structure on the basis of which confirmatory structure is formulated. This would lead to the unwanted practice of hypothesizing after results are known (Kerr, 1998), and as such imposes physical as well as philosophical restrictions on a meaningful scientific method. In case one wants to peek at the factor structure obtained with EFA it is recommended that cross-validatory efforts are undertaken such as (i) splitting the data so that one batch may be used in Step 1 while another batch is used in Steps 2 to 4, or (ii) collecting new data for corroboration in Steps 2 to 4 after one performs Step 1.

A basic principle of scientific inference is that a good fit of a model to a set of data never proves the truth of that model. Indeed if one does find the best fitting model, it may not be theoretically plausible or represent the actual state of affairs. No (statistical) technique can prove that a model is correct, at best one can give evidence that a certain model or set of models may or may not be a plausible representation of the unobservable forces that generated the data set at hand. The crux of a meaningful scientific method is thus the exclusion of plausible alternatives. As no technique is able to prove that a model translation of theory is correct the most substantive central question of interest becomes: *Which of the theories best fits the data?* Meaningful evaluation of substantive theory then asks for a statistical framework which is able to simultaneously compare non-trivially constrained complex model alternatives. The integrative strategy on common factor analysis is believed to provide such a framework for the factor analytic setting.

Moreover, the integrative strategy and the ideas regarding constrained-model selection are seen as part of a particular philosophy of method for observational studies. Fields based on observational studies are usually devoid from the subject matter that lends itself for laboratory settings as the set of possible experiments for these fields is usually unknowable. For example, in the behavioral sciences it is impossible to characterize the set of possible experiments for human (induced) systems. It is especially for such fields that a philosophy of method aimed at the exclusion of plausible alternative theories is important. The constrained Bayesian approach to factor analyzes explicitly encourages researchers to formulate plausible competing theories for confirmatory analysis, and offers a framework in which one is able to simultaneously evaluate all possible alternative model translated theories with regard to model fit and model complexity. As such it has a strong connection with the hypothetico-deductive scientific method (Popper, 2002) and the concept of strong inference (Platt, 1964). This method of scientific advance has, coupled to Bayesian (constrained-)model selection, the following form (also see Platt, 1964): (i) Devise on the basis of previous knowledge (such as a former exploratory data analysis on preliminary data, previous results, or expert opinion) alternative theories. These alternative theories may, to express specificity, have inequality constraints among the parameters of its constituent hypotheses; (ii) Devise a crucial experiment whose possible outcomes will be able to demarcate maximally the alternative theories or (when experiments are not possible) establish which observational data one would need to exclude one or more of the theories; (iii) Perform the experiment or obtain the observational data and establish the 'best model(s)' with the inequality constrained Bayesian confirmatory data analysis framework; (iv) Repeat the cycle by refining the model(s) that remain(s).

References

Aguilar, O., & West, M. (2000). Bayesian dynamic factor models and portfolio allocation. *Journal of Business & Economic Statistics*, *18*, 338-357.

Aitkin, M., & Aitkin, I. (2005). Bayesian inference for factor scores. In A. Maydeu-Olivares & J. J. McArdle (Eds.), *Contemporary psychometrics: A festschrift for Roderick P. McDonald* (pp. 207–221). Mahwah, NJ: Lawrence Eerlbaum.

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B. N. Petrov & F. Csaki (Eds.), *Second international symposium on information theory* (pp. 267–281). Budapest: Akademiai Kaido.

Akaike, H. (1978). A Bayesian analysis of the minimum AIC procedure. *Annals of the Institute of Statistical Mathematics, Part A*, *30*, 9-14.

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

Alberti, K. G. M. M., & Zimmet, P. Z. (2008). Should we dump the metabolic syndrome?: No. *British Medical Journal*, *336*, 641.

Ambrose, M. L. (2002). Contemporary justice research: A new look at familiar questions. *Organizational Behavior and Human Decision Processes*, *89*, 803-812.

Anderson, T. W. (1984). *An introduction to multivariate statistical analysis.* New York: Wiley.

Anderson, T. W., & Rubin, H. (1956). Statistical inference in factor analysis. In J. Neyman (Ed.), *Proceedings of the third berkeley symposium on mathematical statistics and probability* (Vol. V). Berkeley: University of California Press.

Ang, L. W., Ma, S., Cutter, J., Chew, S. K., Tan, C. E., & Tai, E. S. (2005). The metabolic syndrome in Chinese, Malays, and Asian Indians: Factor analysis of data from the 1998 Singapore National Health Survey. *Diabetes Research and Clinical Practice*, *67*, 53-62.

Anraku, K. (1999). An information criterion for parameters under a simple order restriction. *Biometrika*, *86*, 141-152.

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

Asparouhov, T., & Muthén, B. (2009). Exploratory structural equation modeling. *Structural Equation Modeling*, *16*, 397-438.

Bagoly, Z., Borgonovo, L., Mészáros, A., Balázs, L. G., & Horváth, I. (2009). Factor analysis of the long gamma-ray bursts. *Astronomy & Astrophysics*, *493*, 51-54.

Barlow, R. E., Bartholomew, D. J., Bremner, J. M., & Brunk, H. D. (1972). *Statistical inference under order restrictions: Theory and application of isotonic regression.* New York: Wiley.

Barnsley, R. H., & Rabinovitch, M. S. (1970). Handedness: Proficiency versus stated preference. *Perceptual and Motor Skills*, *30*, 343-362.

Bartholomew, D. J. (1981). Posterior analysis of the factor model. *British Journal of Mathematical and Statistical Psychology*, *34*, 93-99.

Bartholomew, D. J. (1987). *Latent variable models and factor analysis.* New York: Oxford University Press.

Bartlett, M. S. (1957). A comment on D. V. Lindley's "statistical paradox". *Biometrika*, *44*, 533-534.

Basilevksy, A. (1994). *Statistical factor analysis and related methods: Theory and application.* New York: John Wiley & Sons.

Bass, B. M. (1999). Two decades of research and development in transformational leadership. *European Journal of Work and Organizational Psychology*, *8*, 9-32.

Bayes, T. (1958). An essay towards solving a problem in the doctrine of chances. *Biometrika*, *45*, 293-315. (reproduced from: The Philosophical Transactions (1763), **53**, 370–418; with a foreword by G.A. Barnard)

Bekker, P. A., Merckens, A., & Wansbeek, T. J. (1994). *Identification, equivalent models, and computer algebra.* Boston: Academic Press.

Berger, J. O., & Pericchi, L. R. (1996). The intrinsic Bayes factor for model selection and prediction. *Journal of the American Statistical Association*, *91*, 109-122.

Berger, J. O., & Pericchi, L. R. (2004). Training samples in objective Bayesian model selection. *The Annals of Statistics*, *32*, 841-869.

Bernardo, J. M. (1979). Reference posterior distributions for Bayesian inference. *Journal of the Royal Statistical Society, Series B*, *41*, 113-147.

Besag, J. (1989). A candidate's formula: A curious result in Bayesian prediction. *Biometrika*, *76*, 183.

Blaha, M., & Elasy, T. A. (2006). Clinical use of the metabolic syndrome: Why the confusion? *Clinical Diabetes*, *24*, 125-131.

Bollen, K. A. (1989). *Structural equations with latent variables.* New York: John Wiley & Sons.

Bollen, K. A., & Jöreskog, K. G. (1985). Uniqueness does not imply identification. A note on confirmatory factor analysis. *Sociological Methods & Research*, *14*, 155-163.

Boronat, M., Saavedra, P., Varillas, V. F., & Nóvoa, F. J. (2009). Use of confirmatory factor analysis for the identification of new components of the metabolic syndrome: The role of plasminogen activator inhibitor-1 and Haemoglobin A1c. *Nutrition, Metabolism & Cardiovascular Diseases*, *19*, 271-276.

Bozdogan, H., & Shigemasu, K. (1998). Bayesian factor analysis model and choosing the number of factors using a new informational complexity criterion. In A. Rizzi, M. Vichi, & H. Bock (Eds.), *Advances in data sciences and classification* (pp. 335–342). Berlin: Springer Verlag.

Bradshaw, J., & Rogers, L. J. (1993). *The evolution of lateral asymmetries, language, tool use and intellect.* San Diego: Academic Press.

Brown, M. E., & Treviño, L. K. (2006). Ethical leadership: A review and future directions. *The Leadership Quarterly*, *17*, 595-616.

Cangelosi, R., & Goriely, A. (2007). Component retention in principal component analysis with application to cDNA microarray data. *Biology Direct*, *2*, DOI: 10.1186/1745-6150-2-2.

Carvalho, C., Chang, J., Lucas, J., Nevins, J. R., Wang, Q., & West, M. (2008). High-dimensional sparse factor modeling: Applications in gene expression genomics. *Journal of the American Statistical Association*, *103*, 1438-1456.

Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research*, *1*, 245-276.

Celeux, G., Hurn, M., & Robert, C. P. (2000). Computational and inferential difficulties with mixture posterior distributions. *Journal of the American Statistical Association*, *95*, 957-970.

Chaoyang, L., & Ford, E. S. (2007). Is there a single underlying factor for the metabolic syndrome in adolescents? *Diabetes Care*, *30*, 1556-1561.

Chernoff, H. (1954). On the distribution of the likelihood ratio. *The Annals of Mathematical Statistics*, *25*, 573-578.

Chib, S. (1995). Marginal likelihood from the Gibbs output. *Journal of the American Statistical Association*, *90*, 1313-1321.

Chib, S., & Jeliazkov, I. (2001). Marginal likelihood from the Metropolis-Hastings output. *Journal of the American Statistical Association*, *96*, 270-281.

Chowdhurry, S. R. (1969). *A Bayesian approach in multiple regression analysis with inequality constraints* (Tech. Rep. No. EIT.6). Department of Econometrics, Economics Institute Tilburg.

Cliff, N. (1988). The eigenvalues-greater-than-one rule and the reliability of components. *Psychological Bulletin*, *103*, 276-279.

Cohen, A., Kemperman, J. H. B., & Sackrowitz, H. B. (2000). Properties of likelihood inference for order restricted models. *Journal of Multivariate Analysis*, *72*, 50-77.

Cohn, E. S., White, S. O., & Sanders, J. (2000). Distributive and procedural justice in seven nations. *Law and Human Behavior*, *24*, 553-580.

Corsetti, J. P., Rainwater, D. L., Moss, A. J., Zareba, W., & Sparks, C. E. (2006). High lipoprotein-associated phospholipase A_2 is a risk factor for recurrent coronary events in postinfarction patients. *Clinical Chemistry*, *52*, 1331-1338.

Craig, S. B., & Gustafson, S. B. (1998). Perceived leader integrity scale: An instrument for assessing employee perceptions of leader integrity. *The Leadership*

Quarterly, *9*, 127-145.

Davis, W. W. (1978). Bayesian analysis of the linear model subject to linear inequality constraints. *Journal of the American Statistical Association*, *73*, 573-579.

Dawid, A. P. (1979). Conditional independence in statistical theory (with discussion). *Journal of the Royal Statistical Society, Series B*, *41*, 1-31.

den Hartog, D. N. (1997). *Inspirational leadership.* Enschede: Ipskamp BV.

Dijkstra, R. L., Robertson, T., & Silvapulle, M. J. (2002). Special issue Statistical Inference under Inequality Constraints. *Journal of Statistical Planning and Inference*, *107*, 1-378.

Dolan, C. V., & Molenaar, P. C. M. (1991). A comparison of four methods of calculating standard errors of maximum-likelihood estimates in the analysis of covariance structure. *British Journal of Mathematical & Statistical Psychology*, *44*, 359-368.

Dunn, J. E. (1973). A note on a sufficiency condition for uniqueness of a restricted factor matrix. *Psychometrika*, *38*, 141-143.

Dunson, D. B. (2001). Commentary: Practical advantages of Bayesian analysis of epidemiologic data. *American Journal of Epidemiology*, *153*, 1222-1226.

Eckel, R. H., Grundy, S. M., & Zimmet, P. Z. (2005). The metabolic syndrome. *The Lancet*, *365*, 1415-1428.

Edwards, K. L., Austin, M. A., Newman, B., Mayer, E., Krauss, R. M., & Selby, J. (1994). Multivariate analysis of the insulin resistance syndrome in women. *Arteriosclerosis, Thrombosis, and Vascular Biology*, *14*, 1940-1945.

Einhorn, D., Reaven, G. M., Cobin, R. H., Ford, E. S., Ganda, O. P., Handelsman, Y., et al. (2003). American College of of Endocrinology position statement on the insulin resistance syndrome. *Endocrine Practice*, *9*, 237-252.

Esaiasson, P., Gilljam, M., & Lindholm, T. (2007). *When winning is everything. An experimental account of determinants of decision-acceptence in schools, in the general public and among students of political science.* (Paper presented at the 4th European Consortium for Political Research General Conference, Pisa)

Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults. (2001). Executive Summary of the Third Report of the National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III). *Journal of the American Medical Association*, *285*, 2486-2497.

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*, *4*, 272-299.

Ferrando, P. J., & Lorenzo-Seva, U. (2000). Unrestricted versus restricted factor analysis of multidimensional test items: Some aspects of the problem and some suggestions. *Psicológica*, *21*, 301-323.

Fletcher, R., & Powell, M. J. D. (1963). A rapidly convergent descent method for minimization. *Computer Journal*, *2*, 163-168.

Fokoué, E. (2009). Bayesian computation of the intrinsic structure of factor analytic models. *Journal of Data Science*, *7*, 285-311.

Ford, E. S. (2003). Factor analysis and defining the metabolic syndrome. *Ethnicity & Disease*, *13*, 429-437.

Frühwirth-Schnatter, S. (2004). Estimating marginal likelihoods for mixture and Markov switching models using bridge sampling techniques. *Econometrics Journal*, *7*, 143-167.

Frühwirth-Schnatter, S., & Frühwirth, R. (2007). Auxiliary mixture sampling with applications to logistic models. *Computational Statistics & Data Analysis*, *51*, 3509-3528.

Frühwirth-Schnatter, S., & Wagner, H. (2006). Auxiliary mixture sampling for parameter-driven models of time series of counts with applications to state space modelling. *Biometrika*, *93*, 827-841.

Gale, E. A. M. (2008). Should we dump the metabolic syndrome?: Yes. *British Medical Journal*, *336*, 640.

Gelfland, A. E., & Smith, A. F. E. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, *85*, 398-409.

Gelman, A., Carlin, B. P., Stern, H. S., & Rubin, D. B. (2004). *Bayesian data analysis.* Boca Raton: Chapman & Hall/CRC.

Gelman, A., & Meng, X. L. (1998). Simulating normalizing constants: From importance sampling to bridge sampling to path sampling. *Statistical Science*, *13*, 163-185.

Geweke, J. F. (1986). Exact inference in the inequality constrained normal linear regression model. *Journal of Applied Econometrics*, *1*, 127-141.

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

Geweke, J. F., & Zhou, G. (1996). Measuring the pricing error of the arbitrage pricing theory. *Review of Financial Studies*, *9*, 557-587.

Ghosh, A. (2005). Factor analysis of metabolic syndrome among the middle-aged Bengalee Hindu men of Calcutta, India. *Diabetes/Metabolism Research and Reviews*, *21*, 58-64.

Ghosh, J., & Dunson, D. B. (2009). Default prior distributions and efficient posterior computation in Bayesian factor analysis. *Journal of Computational and Graphical Statistics*, *18*, 306-320.

Gill, J. (2008). *Bayesian methods: A social and behavioral sciences approach* (2nd ed.). Boca Raton: Chapman & Hall/CRC.

Gilljam, M., Esaiasson, P., & Lindholm, T. (2010). Opening up the black box of pupil influence: An experimental comparison between decisions made by elected pupil councils, by pupils in referenda, and by the teaching staff. *Educational Assessment, Evaluation and Accountability*, *22*, 73-88.

Goodman, E., Li, C., Tu, Y.-K., Ford, E. S., Sun, S. S., & Huang, T. T.-K. (2009). Stability of the factor structure of the metabolic syndrome across pubertal development: Confirmatory factor analyses of three alternative models. *The Journal of Pediatrics*, *155*, S5.e1-S5.e8.

Gorsuch, R. L., & Nelson, J. (1981). *CNG scree test: An objective procedure for determining the number of factors.* (Paper presented at the annual meeting of the Society for Multivariate Experimental Psychology)

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

Grundy, S. M. (2005). Point: The metabolic syndrome still lives. *Clinical Chemistry*, *51*, 1352-1354.

Guttman, L. (1954). Some necessary conditions for common factor analysis. *Psychometrika*, *19*, 149-162.

Guttman, L. (1955). The determinacy of factor score matrices with implications for five other basic problems of common-factor theory. *British Journal of Statistical Psychology*, *8*, 65-81.

Haig, B. D. (2005). Exploratory factor analysis, theory generation, and scientific method. *Multivariate Behavioral Research*, *40*, 303-329.

Han, C., & Carlin, B. P. (2001). Markov chain Monte Carlo methods for computing Bayes factors: A comparative review. *Journal of the American Statistical Association*, *96*, 1122-1132.

Hanley, A. J. G., Karter, A. J., Festa, A., D'Agostino, R., Wagenknecht, L. E., Savage, P., et al. (2002). Factor analysis of metabolic syndrome using directly measured insulin sensitivity: The insulin resistance atherosclerosis study. *Diabetes*, *51*, 2642-2647.

Hanson, R. L., Imperatore, G., Bennett, P. H., & Knowler, W. C. (2002). Components of the "metabolic syndrome" and incidence of type 2 diabetes. *Diabetes*, *51*, 3120-3127.

Hayashi, K., Bentler, P. M., & Yuan, K.-H. (2007). On the likelihood ratio test for the number of factors in EFA. *Structural Equation Modeling*, *14*, 505-526.

Heffernan, W. C. (2002). Three types of leadership. In R. Adlam & P. Villiers (Eds.), *Police leadership in the twenty-first century: Philosophy, doctrine and developments* (pp. 134–145). Hampshire, UK: Waterside Press.

Heywood, H. B. (1931). On finite sequences of real numbers. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, *134*, 486-501.

Hoijtink, H., Klugkist, I., & Boelen, P. (2008). *Bayesian evaluation of informative hypotheses.* New York: Springer.

Hong, S., Mitchell, S. K., & Harshman, R. A. (2006). Bootstrap scree tests: A Monte Carlo simulation and applications to published data. *British Journal of Mathematical & Statistical Psychology*, *59*, 35-57.

Hopkins, W. D. (1996). Chimpanzee handedness revisited: 54 years since Finch (1941). *Psychonomic Bulletin and Review*, *3*, 449-457.

Hopkins, W. D., & Cantalupo, C. (2004). Handedness in chimpanzees (*Pan troglodytes*) is associated with asymmetries of the primary motor cortex but not with homologous language areas. *Behavioral Neuroscience*, *118*, 1176-1183.

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, *30*, 179-185.

Howe, W. G. (1955). *Some contributions to factor analysis* (Tech. Rep. No. ORNL-1919). Oak Ridge, Tennessee: Oak Ridge National Laboratory.

Hoyle, R. H., & Duvall, J. L. (2004). Determining the number of factors in exploratory and confirmatory factor analysis. In D. Kaplan (Ed.), *The SAGE handbook of quantitative methodology for the social sciences* (pp. 301–315). Thousand Oaks, CA: SAGE.

Huang, P., Kraja, A. T., Tang, W., Hunt, S. C., North, K. E., Lewis, C. E., et al. (2008). Factor relationships of metabolic syndrome and echocardiographic phenotypes in the HyperGEN study. *Journal of Hypertension*, *26*, 1360-1366.

Jackson, D. A. (1993). Stopping rules in principal components analysis: A comparison of heuristical and statistical approaches. *Ecology*, *74*, 2204-2214.

Jamshidian, M., & Bentler, P. M. (1993). A modified newton method for constrained estimation in covariance structure analysis. *Computational Statistics & Data Analysis*, *15*, 133-146.

Jefferys, W., & Berger, J. O. (1992). Ockham's razor and Bayesian analysis. *American Scientist*, *80*, 64-72.

Jeffreys, H. (1935). Some tests of significance, treated by the theory of probability. *Proceedings of the Cambridge Philosophy Society*, *31*, 203-222.

Jeffreys, H. (1961). *Theory of probability* (3rd ed.). Oxford: Oxford University Press.

Jennrich, R. I. (1978). Rotational equivalence of factor loading matrices with specified values. *Psychometrika*, *43*, 421-426.

Jolliffe, I. T. (2002). *Principal component analysis* (2nd ed.). New York: Springer.

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

Jöreskog, K. G. (1969). A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika*, *34*, 183-202.

Jöreskog, K. G. (1979). Author's addendum. In J. Magidson (Ed.), *Advances in factor analysis and structural equation models* (pp. 40–43). Cambridge, Massachusetts: Abt Books.

Jöreskog, K. G. (2007). Factor analysis and its extensions. In R. Cudeck & R. C. MacCallum (Eds.), *Factor analysis at 100: Historical developments and future directions* (pp. 47–77). Mahwah, NJ: Lawrence Erlbaum Associates.

Jurs, S., Zoski, K., & Mueller, R. (1993). *Using linear regression to determine the number of factors to retain in factor analysis and the number of issues to retain in Delphi studies and other surveys.* (Paper presented at the annual meeting of the American Educational Research Association)

Kahn, R., Buse, J., Ferrannini, E., & Stern, M. (2005). The metabolic syndrome: Time for a critical appraisal. Joint statement from the American Diabetes Association and the European Association for the Study of Diabetes. *Diabetes Care*, *28*, 2289-2304.

Kaiser, H. F. (1958). The Varimax criterion for analytic rotation in factor analysis. *Psychometrika*, *23*, 187-200.

Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement*, *20*, 141-151.

Kaiser, H. F. (1970). A second generation little jiffy. *Psychometrika*, *35*, 401-415.

Kano, Y. (2007). Selection of manifest variables. In S. Y. Lee (Ed.), *Handbook of latent variable and related models* (pp. 65–86). Amsterdam: Elsevier.

Kano, Y., & Harada, A. (2000). Stepwise variable selection in factor analysis. *Psychometrika*, *65*, 7-22.

Kaptein, M. (2003). The diamond of managerial integrity. *European Management Review*, *21*, 99-108.

Kaptein, M., & Wempe, J. (2002). *The balanced company: A theory of corporate integrity.* Oxford: Oxford University Press.

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

Kato, B. S., & Hoijtink, H. (2006). A Bayesian approach to inequality constrained linear mixed models: Estimation and model selection. *Statistical Modelling*, *6*, 231-249.

Kato, B. S., & Peeters, C. F. W. (2008). Inequality constrained multilevel models. In H. J. A. Hoijtink, I. G. Klugkist, & P. A. Boelen (Eds.), *Bayesian evaluation of informative hypotheses* (pp. 273–295). New York: Springer.

Kerr, N. L. (1998). HARKing: Hypothesizing After the Results are Known. *Personality and Social Psychology Review*, *2*, 196-217.

Khader, Y. S., Batieha, A., Jaddou, H., Batieha, Z., El-Khateeb, M., & Ajlouni, K. (2011). Factor analysis of cardiometabolic risk factors clustering in children and adolescents. *Metabolic Syndrome and Related Disorders*, *9*, 151-156.

Klugkist, I., & Hoijtink, H. (2007). The Bayes factor for inequality and about equality constrained models. *Computational Statistics & Data Analysis*, *51*, 6367-6379.

Kohlberg, L. (1969). State and sequence: The cognitive-development approach to socialization. In D. A. Goslin (Ed.), *Handbook of socialization theory and research* (pp. 347–480). Chicago: Rand McNally.

Kolthoff, E. (2007). *Ethics and new public management: Empirical research into the effects of businesslike government on ethics and integrity.* The Hague: BJu Legal Publishers.

Kudô, A. (1963). A multivariate analogue of the one-sided test. *Biometrika*, *50*, 403-418.

Kylin, E. (1923). Studien über das Hypertonie-Hyperglykämie-Hyperurikämie Syndrom. *Zentralblatt für Innere Medizin*, *44*, 105-112.

Laakso, M., & Kovanen, P. T. (2006). Metabolic syndrome: To be or not to be? *Annals of Medicine*, *38*, 32-33.

Lambert, M., Paradis, G., O'Loughlin, J., Delvin, E. E., Hanley, J. A., & Levy, E. (2004). Insulin resistance syndrome in a representative sample of children and adolescents from Quebec, Canada. *International Journal of Obesity*, *28*, 833-841.

Lambert, Z. V., Wildt, A. R., & Durand, R. M. (1990). Assessing sampling variation relative to numbers-of-factors criteria. *Educational and Psychological Measurement*, *50*, 33-49.

Laplace, P. S. (1986). Memoir on the probability of the causes of events. *Statistical Science*, *1*, 364-378. (translated by S.M. Stigler, and reproduced from: Mémoires de Mathématique et de Physique, Presentés à l'Académie Royale des Sciences, par divers Savans & lûs dans ses Assemblées, Tome Sixìeme (1774), 621–656)

Lasthuizen, K. (2008). *Leading to integrity: Empirical research into the effects of leadership on ethics and integrity.* Amsterdam: VU University Amsterdam.

Laudy, O., & Hoijtink, H. (2007). Bayesian methods for the analysis of inequality constrained contingency tables. *Statistical Methods in Medical Research*, *16*, 123-138.

Lawlor, D. A., Ebrahim, S., May, M., & Smith, G. D. (2004). (Mis)use of factor analysis in the study of insulin resistance syndrome. *American Journal of Epidemiology*, *159*, 1013-1018.

Ledermann, W. (1937). On the rank of the reduced correlational matrix in multiple factor analysis. *Psychometrika*, *2*, 85-93.

Lee, S. E., & Press, S. J. (1998). Robustness of Bayesian factor analysis estimates. *Communications in Statistics - Theory and Methods*, *27*, 1871-1893.

Lee, S. Y. (1979). Constrained estimation in covariance structure analysis. *Biometrika*, *66*, 539-545.

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

Lee, S. Y. (1981). The multiplier method in constrained estimation of covariance structure models. *Journal of Statistical Computation and Simulation*, *12*, 247-257.

Lee, S. Y. (2007). *Structural equation modeling: A Bayesian approach.* West Sussex: John Wiley and Sons Ltd.

Lee, S. Y., & Poon, W. Y. (1985). Further developments on constrained estimation in analysis of covariance structures. *Journal of the Royal Statistical Society, series D*, *34*, 305-316.

Lee, S. Y., & Song, X. Y. (2002). Bayesian selection on the number of factors in a factor analysis model. *Behaviormetrika*, *29*, 23-39.

Liese, A. D., Mayer-Davis, E. J., & Haffner, S. M. (1998). Development of the multiple metabolic syndrome: An epidemiologic perspective. *Epidemiologic Reviews*, *20*, 157-172.

Lind, E., & Tyler, T. (1988). *The social psychology of procedural justice.* New York: Plenum.

Lindley, D. (1957). A statistical paradox. *Biometrika*, *44*, 187-192.

Lonigan, C. J., Hooe, E. S., David, C. F., & Kistner, J. (1999). Positive and negative affectivity in children: Confirmatory factor analysis of a two-factor model and its relation to symptoms of anxiety and depression. *Journal of Consulting and Clinical Psychology*, *67*, 374-386.

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

MacCoun, R. J. (2005). Voice, control, and belonging: The double-edged sword of procedural fairness. *Annual Review of Law and Social Science*, *1*, 171-201.

Manin, B. (1997). *The principles of representative government.* New York: Cambridge University Press.

Mannucci, E., Monami, M., & Rotella, C. M. (2007). How many components for the metabolic syndrome? Results of exploratory factor analysis in the FIBAR study. *Nutrition, Metabolism & Cardiovascular Diseases*, *17*, 719-726.

Marañon, G. (1922). Über Hypertonie und Zuckerkrankheit. *Zentralblatt für Innere Medizin*, *43*, 169-176.

Martin, J. K., & McDonald, R. P. (1975). Bayesian estimation in unrestricted factor analysis: A treatment for heywood cases. *Psychometrika*, *40*, 505-517.

McCrae, R. R., Zonderman, A. B., Costa, P. T., Bond, M. H., & Paunonen, S. V. (1996). Evaluating replicability of factors in the revised NEO Personality Inventory: Confirmatory factor analysis versus Procrustes rotation. *Journal of Personality and Social Psychology*, *70*, 552-566.

McDonald, R. P. (1985). *Factor analysis and related methods.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Meigs, J. B. (2000). Invited commentary: Insulin resistance syndrome? Syndrome X? Multiple metabolic syndrome? A syndrome at all? Factor analysis reveals patterns in the fabric of correlated metabolic risk factors. *American Journal of Epidemiology*, *152*, 908-911.

Mellink, A. (2003). Politieleiderschap en het middenkader. *Het tijdschrift voor de politie*, *65*, 19-21.

Miranda, P. J., DeFronzo, R. A., Califf, R. M., & Gruyton, J. R. (2005). Metabolic syndrome: Definition, pathophysiology, and mechanisms. *American Heart Journal*, *149*, 33-45.

Moustaki, I. (2007). Factor analysis and latent structure of categorical and metric data. In R. Cudeck & R. C. MacCallum (Eds.), *Factor analysis at 100: Historical developments and future directions* (pp. 293–313). Mahwah, NJ: Lawrence Erlbaum Associates.

Mulaik, S. A. (2010). *Foundations of factor analysis* (2nd ed.). Boca Raton: Chapman & Hall/CRC.

Mulaik, S. A., & Millsap, R. E. (2000). Doing the four-step right. *Structural Equation Modeling*, *7*, 36-73.

Mulder, J., Hoijtink, H., & Klugkist, I. (2010). Equality and inequality constrained multivariate linear models: Objective model selection using constrained posterior priors. *Journal of Statistical Planning and Inference*, *140*, 887-906.

Neal, R. (1999). *Erroneous results in "Marginal likelihood from the Gibbs output".* Available from `http://www.cs.toronto.edu/$\thicksim$radford/ftp/chib-letter.pdf`

Nijenhuis, A., & Wilf, H. S. (1978). *Combinatorial algorithms for computers and calculators* (2nd ed.). New York: Academic Press.

Oh, J.-Y., Hong, Y.-S., Sung, Y.-A., & Barrett-Connor, E. (2004). Prevalence and factor analysis of metabolic syndrome in an urban Korean population. *Diabetes Care*, *27*, 2027-2032.

Oh, M.-S., & Shin, D. W. (2011). A unified Bayesian inference on treatment means with order constraints. *Computational Statistics & Data Analysis*, *55*, 924-934.

O'Hagan, A. (1995). Fractional Bayes factors for model comparison. *Journal of the Royal Statistical Society, Series B*, *57*, 99-138.

Peeters, C. F. W. (2012). Rotational uniqueness conditions under oblique factor correlation metric. *Psychometrika*, *77*, 288-292.

Peeters, C. F. W. (invited revision a). An automated candidate estimator for marginal likelihood estimation with application to factor analytic dimensionality selection.

Peeters, C. F. W. (invited revision b). Inequality constrained confirmatory factor analysis: Bayesian specification and model selection.

Peeters, C. F. W., Dziura, J., & van Wesel, F. (under review). Pathophysiologic domains underlying the metabolic syndrome: An alternative factor analytic strategy.

Peeters, C. F. W., & van Wesel, F. (under review). Inequality-constrained-model selection for the political sciences.

Penno, G., Miccoli, R., Pucci, L., & Del Prato, S. (2006). The metabolic syndrome beyond the insulin resistance syndrome. *Pharmacological Research*, *53*, 457-468.

Pérez, J., & Berger, J. O. (2002). Expected-posterior prior distributions for model selection. *Biometrika*, *89*, 491-511.

Pladevall, M., Singal, B., Williams, L. K., Brotons, C., Guyer, H., Sadurni, J., et al. (2006). A single factor underlies the metabolic syndrome: A confirmatory factor analysis. *Diabetes Care*, *29*, 113-122.

Platt, J. R. (1964). Strong inference. *Science*, *146*, 347-353.

Polasek, W. (2000). *Factor analysis and outliers: A bayesian approach* (Tech. Rep.). Institute of Statistics and Econometrics, University of Basel.

Popper, K. (2002). *The logic of scientific discovery.* New York: Routledge.

Preacher, K. J., & MacCallum, R. C. (2003). Repairing Tom Swift's electric factor analysis machine. *Understanding Statistics*, *2*, 13-43.

Press, S. J. (2003). *Subjective and objective Bayesian statistics: Principles, models, and applications* (2nd ed.). New York: Wiley.

Press, S. J., & Shigemasu, K. (1999). A note on choosing the number of factors. *Communications in Statistics - Theory and Methods*, *28*, 1653-1670.

Price, K., Hall, T., Bos, K. van den, Hunton, J., Lovett, S., & Tippett, M. (2001). Features of the value function for voice and their consistency across participants from four countries: Great Britain, Mexico, the Netherlands, and the United States. *Organizational Behavior and Human Decision Processes*, *84*, 95-121.

Quinn, K. M. (2004). Bayesian factor analysis for mixed ordinal and continuous responses. *Political Analysis*, *12*, 338-353.

Reaven, G. M. (1988). Banting lecture 1988. Role of insulin resistance in human disease. *Diabetes*, *37*, 1595-1607.

Reaven, G. M. (1993). Role of insulin resistance in human disease (syndrome X): An expanded definition. *Annual Review of Medicine*, *44*, 121-131.

Reaven, G. M. (2005a). Counterpoint: Just being alive is not good enough. *Clinical Chemistry*, *51*, 1354-1357.

Reaven, G. M. (2005b). The metabolic syndrome: Requiescat in Pace. *Clinical Chemistry*, *51*, 931-938.

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

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

Ritter, C., & Tanner, M. A. (1992). Facilitating the Gibbs sampler - The Gibbs stopper and the griddy-Gibbs sampler. *Journal of the American Statistical Association*, *87*, 861-868.

Robertson, T., Wright, F. T., & Dijkstra, R. L. (1988). *Order-restricted statistical inference.* Chichester: John Wiley and Sons.

Rosenthal, R., Rosnow, R. L., & Rubin, D. B. (2000). *Contrasts and effect sizes in behavioral science: A correlational approach.* Cambridge, UK: Cambridge University Press.

Roszkowski, M. J., Snelbecker, G. E., & Sacks, R. (1981). Children's, adolescents', and adults' reports of hand preference: Homogeneity and discriminating power of selected tasks. *Journal of Clinical Neuropsychology*, *3*, 199-213.

Rowe, D. B. (2003). *Multivariate Bayesian statistics: Models for source separation and signal unmixing.* Boca Raton: Chapman & Hall/CRC.

Rue, H., Martino, S., & Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models using integrated nested Laplace approximations (with discussion). *Journal of the Royal Statistical Society, Series B*, *71*, 319-392.

Savage, L. J. (1962). *The foundations of statistical inference.* London: Methuen & Co.

Savalei, V., & Kolenikov, S. (2008). Constrained versus unconstrained estimation in structural equation modeling. *Psychological Methods*, *13*, 150-170.

Schönemann, P. H., & Wang, M.-M. (1972). Some results on factor indeterminacy. *Psychometrika*, *37*, 61-91.

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

Scuteri, A., Najjar, S. S., Morrell, C. H., & Lakatta, E. G. (2005). The metabolic syndrome in older individuals: Prevalence and prediction of cardiovascular events. The cardiovascular health study. *Diabetes Care*, *28*, 882-887.

Self, S. G., & Liang, K.-L. (1987). Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *Journal of the American Statistical Association*, *82*, 605-610.

Shah, S., Novak, S., & Stapleton, L. M. (2006). Evaluation and comparison of models of metabolic syndrome using confirmatory factor analysis. *European Journal of Epidemiology*, *21*, 343-349.

Shapiro, A. (1985). Asymptotic distribution of test statistics in the analysis of moment structures under inequality constraints. *Biometrika*, *72*, 133-144.

Shen, B.-J., Todaro, J. F., Niaura, R., McCaffery, J. M., Zhang, J., Spiro, A., et al. (2003). Are metabolic risk factors one unified syndrome? Modeling the structure of the metabolic syndrome X. *American Journal of Epidemiology*, *157*, 701-711.

Silvapulle, M. J., & Sen, P. K. (2005). *Constrained statistical inference: Inequality, order, and shape restrictions.* Hoboken, NJ: John Wiley and Sons Ltd.

Solera-Martínez, M., López-Martínez, S., Sánchez-López, M., Moya-Martínez, P., Notario-Pacheco, B., Arias-Palencia, N., et al. (2011). Validity of a single-factor model underlying the metabolic syndrome in young adults: Confirmatory factor analysis. *Revista Española de Cardología*, *64*, 379-384.

Spearman, C. (1904). "General intelligence," objectively determined and measured. *American Journal of Psychology*, *15*, 201-293.

Spiegelhalter, D. J., & Smith, A. F. M. (1982). Bayes factor for linear and log-linear models with vague prior information. *Journal of the Royal Statistical Society, Series B*, *44*, 377-387.

Sprenger, J. (2008). Statistics between inductive logic and empirical science. *Journal of Applied Logic*, *7*, 239-250.

Steiger, J. H. (1994). Factor analysis in the 1980's and the 1990's: Some old debates and some new developments. In I. Borg & P. P. Mohler (Eds.), *Trends and perspectives in empirical social research* (pp. 201–223). Berlin: Walter de Gruyter.

Steiger, J. H. (1996). Coming full circle in the history of factor indeterminacy. *Multivariate Behavioral Research*, *31*, 617-630.

Stephens, M. (2000). Bayesian analysis of mixtures models with an unknown number of components - an alternative to reversibe jump methods. *The Annals of Statistics*, *28*, 40-74.

Stoel, R. D., Garre, F. G., Dolan, C., & van den Wittenboer, G. (2006). On the likelihood ratio test in structural equation modeling when parameters are subject to boundary constraints. *Psychological Methods*, *11*, 439-455.

Takane, Y., van der Heijden, P. G. M., & Browne, M. W. (2003). On likelihood ratio tests for dimensionality selection. In T. Higuchi, Y. Iba, & M. Ishiguro (Eds.), *Proceedings of Science of Modeling: The 30th anniversary meeting of the information criterion (AIC)* (pp. 348–349). Tokyo: The Institute of Statistical Mathematics.

Tanner, M., & Wong, W. (1987). The calculation of posterior distributions by data augmentation. *Journal of the American Statistical Association*, *82*, 528-550.

Teorell, J. (2006). Political participation and three theories of democracy: A research inventory and agenda. *European Journal of Political Research*, *45*, 787-810.

Thurstone, L. L. (1947). *Multiple factor analysis.* Chicago: University of Chicago Press.

Tierney, L. (1994). Markov chains for exploring posterior distributions. *The Annals of Statistics*, *22*, 1701-1728.

Treviño, L. K. (1986). Ethical decision making in organizations: A person-situation interactionist model. *Academy of Management Review*, *11*, 601-617.

Tsonaka, R., & Moustaki, I. (2007). Parameter constraints in generalized linear latent variable models. *Computational Statistics & Data Analysis*, *51*, 4164-4177.

Tyler, T. (2006). Psychological perspectives on legitimacy and legitimation. *Annual Review of Psychology*, *57*, 375-400.

Unwin, N. (2006). The metabolic syndrome. *Journal of the Royal Society of Medicine*, *99*, 457-462.

van Prooijen, J. W., & van der Kloot, W. A. (2001). Confirmatory analysis of exploratively obtained factor structures. *Educational and Psychological Measurement*, *61*, 777-792.

van Wesel, F., Hoijtink, H., & Klugkist, I. (2011). Choosing priors for constrained analysis of variance: Methods based on training data. *Scandinavian Journal of Statistics*, *38*, 666-690.

Velicer, W. F. (1976). Determining the number of components from the matrix of partial correlations. *Psychometrika*, *31*, 321-327.

Venkat Narayan, K. M. (2006). The metabolic syndrome: Some second thoughts? *Clinical Diabetes*, *24*, 38-39.

Warren, J. M. (1980). Handedness and laterality in humans and other animals. *Physiological Psychology*, *8*, 351-359.

Waugh, R. F., & Chapman, E. S. (2005). An analysis of dimensionality using factor analysis (true-score theory) and rasch measurement: What is the difference? Which method is better? *Journal of Applied Measurement*, *6*, 80-99.

Weiss, R., Dziura, J., Burgert, T. S., Tamborlane, W. V., Taksali, S. E., Yeckel, C. W., et al. (2004). Obesity and the metabolic syndrome in children and adolescents. *The New England Journal of Medicine*, *350*, 2362-2374.

Wesley, M. J., Fernandez-Carriba, S., Hostetter, A., Pilcher, D., Poss, S., & Hopkins, W. D. (2002). Factor analysis of multiple measures of hand use in captive chimpanzees: An alternative approach to the assessment of handedness in nonhuman primates. *International Journal of Primatology*, *23*, 1155-1168.

Widaman, K. F. (2007). Common factors versus components: Principals and principles, errors and misconceptions. In R. Cudeck & R. C. MacCallum (Eds.), *Factor analysis at 100: Historical developments and future directions* (pp. 177–203). Mahwah, NJ: Lawrence Erlbaum Associates.

Wilkin, T. J., & Voss, L. D. (2004). Metabolic syndrome: Maladaptation to a modern world. *Journal of the Royal Society of Medicine*, *97*, 511-520.

Wilks, S. S. (1938). The large sample distribution of the likelihood ratio for testing composite hypotheses. *Annals of Mathematical Statistics*, *9*, 60-62.

Williams, S. M. (1986). Factor analysis of the Edinburgh Handedness Inventory. *Cortex*, *22*, 325-326.

Wilson, E. B. (1928a). On hierarchical correlation systems. *Proceedings of the National Academy of Sciences*, *14*, 283-291.

Wilson, E. B. (1928b). Review of “The abilities of man, their nature and measurement”, by C. Spearman. *Science*, *67*, 244-248.

Woodbury, M. A. (1950). *Inverting modified matrices* (Memorandum Report No. 42). Statistical Research Group, Princeton.

World Health Organization. (1999). *Definition, diagnosis and classification of diabetes mellitus and its complications. Report of a WHO Consultation. Part 1: Diagnosis and classification of diabetes mellitus* (Tech. Rep. No. WHO/NCD/NCS/99.2). Geneva: WHO.

Yeomans, K. A., & Golder, P. A. (1982). The Guttman-Kaiser criterion as a predictor of the number of common factors. *The Statistician*, *31*, 221-229.

Zoski, K., & Jurs, S. (1993). Using multiple regression to determine the number of factors to retain in factor analysis. *Multiple Linear Regression Viewpoints*, *20*, 5-9.

Nederlandse Samenvatting
– Summary in Dutch –

Overzicht

Deze dissertatie houdt zich bezig met statistische modelselectie. Dergelijke modelselectie wordt opgevat als de praktijk van het selecteren van een statistisch model, gegeven bepaalde data. Het doel is daarbij om uit een *a priori* set van gespecificeerde modellen, het model te kiezen welke de beste balans heeft tussen model-fit (hoe goed het model past bij de geobserveerde data) en model-complexiteit (de omvangrijkheid van een model, uitgedrukt in het aantal parameters). Modelselectie is daarmee een statistische vertaling van 'Ockhams scheermes'. Het gebruik van modelselectie is in deze dissertatie gekoppeld aan twee *foci*, welke beide zijn geplaatst in het perspectief van (functionele) restricties, te weten: het vaststellen van de (optimale) dimensionaliteit van een model (hier Type I modelselectie genoemd) en de selectie van toepasselijke ongelijkheidsrestricties in de parameterruimte van het model (hier Type II modelselectie genoemd).

In de activiteiten rondom modelselectie is er aandacht voor een bepaald type multivariaat model: het factoranalytisch model. Factoranalyse (FA) is de naam voor een klasse van technieken voor het modelleren van een covariantiematrix. Het idee achter een factor analyse is dat een random p-dimensionale vector X bestaat uit gecorreleerde variabelen welke gegroepeerd kunnen worden in een lager dimensionale m-vector Ξ bestaande uit latent factoren. Latente factoren zijn constructen die niet direct meetbaar zijn. Voor de bepaling van dergelijke constructen dienen oberveerbare variabelen als indicatoren. Voorbeelden van dergelijke constructen zijn 'intelligentie' uit de psychologie of 'verzwakt glucose metabolisme' uit de epidemiologie. FA wordt breed gebruikt in wetenschappelijke velden zoals sociale- en gedragswetenschappen, medische wetenschappen en natuurwetenschappen.

Het specifieke factor analytisch model waar aandacht aan wordt geschonken is het factoranalytisch basismodel, ook wel het 'common' factoranalytisch model genoemd. Hierin is de observatievector $\mathbf{x}_i$ een lineaire combinatie van de latente factorscores in $\boldsymbol{\xi}_i$:

$$\underset{(p\times 1)}{\mathbf{x}_i} = \underset{(p\times 1)}{\boldsymbol{\mu}} + \underset{(p\times m)}{\boldsymbol{\Lambda}} \cdot \underset{(m\times 1)}{\boldsymbol{\xi}_i} + \underset{(p\times 1)}{\boldsymbol{\epsilon}_i},$$

voor $i = 1, \ldots, n$ personen of objecten. In bovenstaande formule drukt $\boldsymbol{\mu}$ een vector van intercepten uit, geeft $\boldsymbol{\epsilon}_i$ de meetfouten aan, en is $\boldsymbol{\Lambda}$ een matrix van factorladingen waarbij elk element λ_{jk} de lading weergeeft van variabele j op factor k, $j = 1, \ldots, p$, $k = 1, \ldots, m$. Het model is gerelateerd aan klassieke statistische onderwerpen zoals correlatie en regressie. De expressie bekijkend, zien we in essentie een multivariaat regressiemodel, waarin de predictordata niet-geobserveerd zijn en waarin de λ_{jk}'s dienen als regressieparameters.

FA kent twee *modi operandi*: exploratief en confirmatief. In een exploratieve factoranalyse (EFA) zijn zowel m als de betekenis van latente factoren onbekend. In de exploratieve zin is factoranalyse een data-reductie instrument en mogelijk een theorie-genererende techniek voor de identificatie van betekenisvolle latente factoren. Confirmatieve factor analyse (CFA) is een data-analytisch instrument gericht op de validatie van een gehypothetiseerde covariantiestructuur. Er wordt dan een *a priori* factorstructuur verondersteld met een gegeven dimensie m en met een geprespecificeerde ladingmatrix die aangeeft welke geobserveerde variabelen indicatoren zijn van welke latente factoren. Het factoranalytisch model heeft een natuurlijke relatie met de restrictie-gebaseerde kijk op modelselectie. Wanneer FA wordt opgevat als een data-reductie instrument dan is de sleutelvraag welke dimensionaliteit van m optimaal is in de zin van balancering tussen model-fit en model-complexiteit. Wanneer FA als een data-analytisch instrument wordt gezien dan is de sleutelvraag welke hypothetische parameter-gebaseerde covariantiestructuur het beste de geobserveerde covariantiematrix representeert.

Klassieke strategieën omtrent de modelselectie vraagstukken in EFA en CFA kennen een aantal lacunes (zie hoofdstuk 1). De dimensie m van de latente data wordt vaak op heuristische wijze gekozen. Er zijn ook formele procedures voor de selectie van m. Deze zijn echter zeer gevoelig voor de schending van bepaalde regulariteitscondities. Dit alles leidt ertoe dat er in de praktijk vaak ondergefactoreerd (de extractie van te weinig latente factoren) of overgefactoreerd (de extractie van teveel latente factoren) wordt. Om deze reden is een procedure voor modelselectie gewenst welke ingebed kan worden in een strategie omtrent de evaluatie van dergelijke regulariteitscondities. Formele toetsing rondom ongelijkheidsrestricties in het factor model is alleen bruikbaar voor non-negativiteisrestricties op bepaalde variantieparameters. Daarnaast kent de praktijk van CFA een relatie met restricties die in zichzelf restrictief te noemen is. Een geprespecificeerde factorstructuur wordt doorgaans uitgedrukt met exclusierestricties in de ladingen matrix: elementen λ_{jk} worden gelijkgesteld aan 0 om aan te geven welke geobserveerde variabelen indicatoren zijn van welke latente factoren. Deze praktijk impliceert een verlies aan informatie en staat niet toe dat op een informatieve wijze richting en sterkte van factorladingen kunnen worden geëxpliciteerd. Voor CFA is een generieke Type II modelselectieprocedure gewenst welke het toestaat de factorstructuur op een alternatieve wijze uit te drukken met behulp van een systeem van ongelijkheidsrestricties. Daarnaast is het zo dat er in klassieke CFA vaak sprake is van een diffuse modelhypothese: in het geval van misspecificatie is het onduidelijk of dit ligt aan de gekozen dimensionaliteit m of aan de formulering van de factorstructuur. Kortom, er is behoefte aan de integratie van EFA en CFA in een sequentiële strategie.

Om bovenstaande wensen te realiseren wordt een Bayesiaanse aanpak van modelselectie gepropageerd. De sleutelmaat voor Bayesiaanse modelselectie is de *Bayes factor*. Deze drukt het relatieve bewijs in de data uit ten gunste van een theorie, gerepresenteerd door een statistisch model (Kass & Raftery, 1995). Er is een voorkeur voor de Bayes factor omdat deze een conceptueel consistente maat geeft voor de vergelijking van een verscheidenheid en een veelheid aan modellen. Daarnaast incorporeert de Bayes factor zowel de fit van een model als modelcomplexiteit. Ook belangrijk is dat de Bayes factor de vergelijking van niet-geneste modellen aankan. Om de Bayes factor te ontsluiten voor bovengenoemde wensen moet wel een aantal nadelen van deze maat overwonnen worden.

Ten eerste heeft de Bayes factor in vele gevallen geen analytische expressie. Computatiestrategieën voor de Bayes factor zijn daarnaast vaak moeilijk. Ook is er het probleem van *prior* verdelingen. De Bayesiaanse aanpak vraagt om de formulering van prior verdelingen. Deze zogenoemde priors reflecteren de geformaliseerde kennis, of gebrek daaraan, over de parameters in het model vóórdat de data geobserveerd zijn. In veel gevallen is een niet-informatieve prior gewenst, omdat er mogelijk weinig *a priori* kennis is over de parameters of om te beantwoorden aan het wetenschappelijke ideaal van 'objectiviteit'. Ook kan het zijn dat in hoogdimensionale modellen de formulering van uitgebreide prior verdelingen niet haalbaar is. Echter, niet-informatieve prior verdelingen zijn vaak oneigenlijk in de zin dat de normaliserende constante divergeert. In dat geval doet zich een probleem voor betreffende het gebruik van de Bayes factor voor Type I modelselectie: wanneer de te vergelijken modellen verschillen in dimensie is de Bayes factor niet determinabel indien men oneigenlijke priors gebruikt voor parameters waarin het dimensieverschil huist.

Het gebruik van de Bayes factor voor Type II modelselectie heeft recent aandacht gekregen in de literatuur (zie bijvoorbeeld Kato & Hoijtink, 2006; Laudy & Hoijtink, 2007; M.-S. Oh & Shin, 2011). In deze studies ligt de focus op de selectie van ongelijkheden op intercept-type parameters. Ongelijkheidsgerestricteerde inferentie via de Bayes factor op regressie-type parameters is nog niet serieus bestudeerd.

In deze dissertatie zijn (i) de Bayesiaanse aanpak, (ii) de kijk op modelselectie vanuit het restricties perspectief en (iii) factoranalyse intrinsiek met elkaar verbonden. De rode draad wordt gevormd door de ontwikkeling van Bayes factors en haar vergezellende computatiemethoden voor Type I en Type II modelselectie. Deze voorzieningen worden vervolgens gebruikt in de herformulering van EFA en CFA vanuit een gerestricteerd-model perspectief. De bovenstaande overwegingen leidden dan ook tot de volgende onderzoeksdoelen die de dissertatie vorm geven:

Het construeren van een conceptueel en computationeel simpele Bayes factor voor Type I gerestricteerde modelselectie welke determinabel is onder gebruik van oneigenlijke priors en de inbedding van deze Bayes factor in een strategie voor de selectie van een optimale dimensie m binnen een Bayesiaanse formulering van EFA.

Het construeren van een conceptueel en computationeel simpele Bayes factor voor Type II gerestricteerde modelselectie welke geschikt is voor ongelijkheden op regressieparameters en de inbedding van deze Bayes factor in een strategie welke het toestaat een confirmatieve factoranalytische structuur uit te drukken met behulp van ongelijkheidsrestricties.

De formulering, met behulp van de provisies van onderzoeksdoelen 1 en 2, van een integratieve factoranalytische strategie welke een mogelijke overbrugging vormt tussen EFA en CFA en het toepassen van deze alternatieve strategie op echte data.

Deel I van dit proefschrift omvat de hoofdstukken 2 tot en met 4 en houdt zich bezig met onderzoeksdoelen 1 en 2. De focus in dit deel ligt op het statistisch en computationeel modelleren van Bayesiaanse modelselectie vraagstukken ten aanzien van Type I en Type II modelselectie in het factoranalytisch model. De illustraties van de voorzieningen ontwikkeld in dit deel dienen als opstap voor deel II van het proefschrift dat hoofdstukken 5 en 6 omvat en zich bezighoudt met onderzoeksdoel 3. De focus in het tweede deel ligt dan ook bij de toelichting van de voorzieningen uit het eerste deel aan de hand van echte data. Alvorens verder in te gaan op deel I en II van dit proefschrift is in de onderstaande leeswijzer een overzicht te vinden van deze dissertatie.

	Elementen					
	Factor analyse		Inferentie onder restricties		Bayesiaanse	
Hoofdstuk	EFA	CFA	Type I	Type II	modelselectie	Doel
2	✓		✓		✓	1
3	✓	✓				2
4		✓		✓	✓	2
5	✓	✓	✓	✓	✓	3
6	✓	✓	✓	✓	✓	3

Deel I

Hoofdstuk 2 houdt zich bezig met de berekening van de marginale dichtheid via Markov chain Monte Carlo (MCMC) procedures. De Bayes factor is opgebouwd uit een ratio van marginale dichtheden. De kandidaatschattermethode (Besag, 1989; Chib, 1995) voor berekening van de marginale dichtheid wordt daarin aangepast zodat deze goed gedefiniëerde Bayes factors geeft onder gebruik van (i) niet-informatieve oneigenlijke priors en (ii) het bestaan van symmetrieën in de *a posteriori* dichtheid. Onder milde condities geeft deze aanpassing een simulatie-consistente MCMC implementatie van uit de literatuur bekende automatische Bayes factors (Berger & Pericchi, 1996). Deze geautomatiseerde kandidaatschattermethode wordt vervolgens gebruikt voor het probleem van de bepaling van de dimensionaliteit m in EFA.

De genoemde toepassing is ten eerste van belang omdat ze een test vormt op de ontworpen methoden. Ten tweede is het belangrijk omdat het in staat stelt te leren over minder bekende indeterminismen (naast die van rotatie) in het factor model. Het hoofdstuk laat dan zien dat het falen te voldoen aan een bekende regulariteitsconditie op $\mathbf{\Lambda}$ (deze moet van volle rang zijn) kan resulteren in schending van een cruciale conditie voor simulatie-consistentie van MCMC procedures. Dit impliceert dat de Bayesiaanse aanpak te maken heeft met sommige van dezelfde regulariteitscondities welke problemen veroorzaken in de klassieke aanpak. Echter, de Bayesiaanse methode drijft op een complete exploratie van de *a posteriori* parameterruimte. Dit maakt het mogelijk de geautomatiseerde kandidaatschatter te verzinken in een evaluatiestrategie welke schendingen van de regulariteitscondities beziet.Op deze manier is dan een strategie te geven voor de selectie van een optimale m in een Bayesiaanse EFA.

Hoofdstuk 3 handelt over twee sets van condities voor rotationele identificatie van het factor model. Jöreskog (1979) formuleerde twee conditiesets voor rotationele identificatie van het oblique factor model: één onder factor correlatie en één onder factor covariantie. De claim was dat beide sets equivalent zijn en zouden leiden tot globale rotationele identificatie van het FA model. Dit hoofdstuk laat zien dat de conditieset onder factor correlatie een extra conditie nodig heeft om te leiden tot globale in plaats van lokale rotationele identificatie. De aangepaste conditieset geeft vervolgens een instrument om ongerestricteerde formuleringen te geven voor het (Bayesiaanse) CFA model. Ongerestricteerde oplossingen corresponderen met EFA in díe zin dat er slechts minimale (exclusie) restricties worden gelegd op het model om te komen tot een rotationeel unieke oplossing voor m factoren. In de EFA traditie wordt dit doorgaans bereikt door het aanhouden van een orthogonaal model en de eis dat $\mathbf{\Lambda}^{\mathrm{T}}\mathbf{\Psi}^{-1}\mathbf{\Lambda}$ diagonaal moet zijn. Dit is echter een oplossing die gemak biedt bij schattingsprocedures maar geen betekenis heeft bij de interpretatie. Om deze reden vindt er in EFA vaak een *post hoc* rotatie plaats om interpretatie te bevorderen. Een ongerestricteerd confirmatief factormodel wordt hier gedefinieerd als een factormodel met slechts minimale restricties op de parameters voor het bereiken van rotationele uniciteit, waarbij de rotatierestricties via de aangepaste conditieset zo worden verkozen dat de oplossing theoretische betekenis heeft en post-hoc rotatie als zodanig overbodig wordt. Het ongerestricteerde confirmatieve factor model is een link tussen EFA en CFA en is de sleutel tot het formuleren van een strategie voor ongelijkheidsgerestricteerde CFA in hoofdstuk 4.

In hoofdstuk 4 wordt er een Bayesiaans raamwerk voorgesteld dat parameterrestricties in CFA een nieuwe betekenis geeft. In dit raamwerk wordt het mogelijk gemaakt dat ongelijkheidsrestricties expressies zijn van theoretische ideeën over de richting en sterkte van een parametereffect. De voorgestelde aanpak wordt ontsloten voor factorladingen. Het hoofdstuk focust zich eerst op de ontwikkeling van een Bayes factor voor Type II modelselectie. Deze is computationeel simpel en weet de berekening van moeilijke marginale dichtheden te vermijden. Vervolgens wordt deze Bayes factor ingebed in een strategie om onderscheid te kunnen maken tussen een veelheid aan ongelijkheidsgerestricteerde formuleringen van de factorstructuur. De strategie bestaat uit het kiezen van een ongerestricteerd confirmatief factor model als basismodel. Theorie aangaande ideeën over de factorstructuur wordt vervolgens niet uitgedrukt met (verdere) exclusierestricties, maar met het opleggen van ongelijkheidsrestricties op en tussen de vrije parameters in $\mathbf{\Lambda}$. Dit leidt tot een meer realistische en informatieve accomodatie van theoretische ideeën omtrent factorstructuur en kan gezien worden als een alternatieve kijk op CFA. Hoofdstuk 4 eindigt met een voorstel voor een integratieve strategie ten aanzien van EFA en CFA welke het hart vormt van onderzoeksdoel 3 welke uitgediept wordt in deel II.

Deel II

In deel II van dit proefschrift worden de voorzieningen uit deel I samengebracht om een alternatieve integratieve strategie rondom factor analyse te formuleren waarbij EFA en CFA gekoppeld worden. In deze aanpak gaat EFA vooraf aan ongelijkheidsgerestricteerde CFA door het deel te laten zijn van een totale inferentieële procedure waarbij de optimale dimensionaliteit m wordt gezocht vóórdat de confirmatieve structuur op het model wordt geplaatst. De strategie bestaat uit de volgende stappen:eEvalueer (i) een serie ongerestricteerde (EFA) modellen van verschillende dimensionaliteit voor de bepaling van m. Wanneer m bepaald is wordt (ii) een ongerestricteerd CFA model opgesteld als basismodel. Op basis van het ongerestricteerde basismodel kan (iii) een veelheid aan plausibele ongelijkheidsgerestricteerde factorstructuren worden gespecificeerd waarbij gebruik wordt gemaakt van ongelijkheidsrestricties op en tussen de vrije parameters in de factorladingen matrix. Reken vervolgens (iv) voor elk ongelijkheidsgerestricteerd model de type II Bayes factor uit en stel vast welk model het meest wordt ondersteund door de data.

In hoofdstuk 5 wordt de alternatieve integratieve strategie toegepast op data aangaande het metabolisch syndroom (MBS). Het metabolisch syndroom bestaat uit een clustering van risicofactoren voor diabetes mellitus en bepaalde cardiovasculaire aandoeningen. Factor analyse is substantieel deel gaan uitmaken van het onderzoek naar MBS waarbij zowel EFA als CFA als analysestrategieën worden gebruikt. De resultaten van de vele factoranalysen aangaande MBS verschillen nogal. Deze situatie is ten dele toe te schrijven aan ondoordacht gebruik van FA. In dit hoofdstuk worden ondoordachte factoranalytische pogingen in het onderzoek naar MBS beschreven. De alternatieve factoranalytische strategie wordt voorgesteld om de zwakheden van het toepassen van FA in MBS onderzoek uit te lichten. Een belangrijke dataset aangaande MBS met antropometrische gegevens van kinderen en adolescenten met overgewicht en obesitas wordt opnieuw geanalyseerd gebruik makend van de integratieve strategie. De bevindingen van deze analyse geven mogelijk hernieuwd cachet aan zowel factor analyse als haar relatie tot MBS onderzoek.

In hoofdstuk 6 wordt het raamwerk met betrekking tot Bayesiaanse ongelijkheidsgerestricteerde statistische inferentie beschreven in het licht van het gegeneraliseerde lineaire model. De focus ligt hierbij op het model voor variantieanalyse en het factoranalyse model. Type II gerestricteerde Bayesiaanse modelselectie zoals ontwikkeld in hoofdstuk 4 wordt geïllustreerd aan de hand van twee voorbeelden uit de politieke wetenschappen. Het eerste voorbeeld omvat een variantieanalyse-context terwijl het tweede voorbeeld een factoranalyse-context omvat waarbij gebruik wordt gemaakt van de integratieve factoranalytische strategie. Beide voorbeelden zijn bedoeld om de potentie van de methoden te illustreren voor onderzoek in de politieke- en bestuurswetenschappen.

Bijdragen

De bijdragen van deze dissertatie zijn direct verbonden met de eerder genoemde onderzoeksdoelen. Met het vervullen van onderzoeksdoel 1 is er een generieke computationele procedure voor de computatie van Bayes factors voor Type I modelselectie onder gebruik van oneigenlijke niet-informatieve prior verdelingen. Daarmee is ook het gebruik van oneigenlijke priors mogelijk in een Bayesiaanse formulering van EFA, zodat de impliciete orde-preferenties in reguliere priorformuleringen vermeden kunnen worden. Doordat de computationele Bayesiaanse methode vraagt om de evaluatie van de volledige *a posteriori* parameterruimte is het beter mogelijk om de schending van regulariteitscondities na te gaan. Dit leidt tot een formele Bayesiaanse procedure omtrent de selectie van m. De hierbij gebruikte strategie heeft ook potentieel voor de klassieke inferentiemethoden daaromtrent.

Met het vervullen van onderzoeksdoel 2 wordt een bijdrage geleverd aan de relatief schaarse literatuur betreffende ongelijkheidsrestricteerde inferentie op regressie-type parameters. Daarnaast worden bekende Bayesiaanse modelselectieprocedures omtrent modelselectie voor restricties op intercept-parameters geëxtrapoleerd naar regressie parameters. Ook draagt onderzoeksdoel 2 bij aan de theorie en techniek van CFA. De formulering van ongelijkheidsgerestricteerde factorstructuur staat meer realistische specificatie van model-vertalingen van theorie toe. Ideeën omtrent de sterkte en richting van factoranalytische relaties kunnen zo expliciet worden geïncludeerd in een formele modelselectie procedure. Op deze wijze is het mogelijk meer informatie te ontrekken aan een CFA.

Onderzoeksdoel 3 geeft een mogelijke weg voor de integratie van EFA en CFA. Doel 3 geeft aan dat EFA en CFA worden gezien als complementaire technieken. Daarnaast wordt EFA expliciet gezien als een inferentiële techniek omtrent de keuze van m. De voorgestelde integratie van EFA en CFA gebruikt deze dan ook in conjunctie, om zo de evaluatie van een diffuse CFA hypothese te voorkomen. De laatste bijdrage is de communicatie van deze integratieve strategie naar velden van studie buiten het psychometrische veld waarin FA haar oorsprong vindt.

Acknowledgements

Science is supposedly a dull activity. Those who hold this view clearly have not experienced the (perhaps a tad masochistic) nature of living the endeavor. Although one might never leave the confines of the office where one performs his or her intellectual labor, science implies passion. Many people have made the intense journey possible and at times very enjoyable.

Herbert Hoijtink is acknowledged for initiating and funding my research, and for stimulating discussions in its early stages. I would also like to express my gratitude towards my promotor, Peter van der Heijden. You have been imperative in the finalization of this dissertation. I thank you for your graceful promotorship.

The department of Methodology & Statistics at Utrecht University has been the location of my ivory tower for some years. Many people there passed from colleagues into the realm of friends and provided for an enjoyable working environment. Joris 'Joop' Mulder deserves special attention. Joris, we have shared the ups and downs of (Ph.D.) life as friends and office roomies. Although things have changed, one does not forget.

Many have also provided collaboration and support in both academic and non-academic endeavors that reflected positively on either this dissertation or my person. James Dziura and Sonia Caprio of the Yale University School of Medicine graciously provided their metabolic syndrome data. William Hopkins of the Yerkes National Primate Research Center send me his wonderful data on chimpanzee handedness. Peter Esaiasson of the University of Gothenburg provided the data on political decision acceptance. These exciting data enhanced the quality of the applications throughout the dissertation. The friendships and collaborations with Gjalt de Graaf, Zeger 'de Zeeg' van der Wal, Emile Kolthoff, Karin Lasthuizen, and Leo Huberts have been warm. Leo and Karin especially are to be held responsible for spurring my academic career.

From my working environment and collaborations I move on to non-work-related friends. Rockwell 'Rocky' Clancy and Kelly 'Babycakes' Vandenbossche, although we do not see each other as much as we used to, I still think about you two a lot. Berry 'kom je buiten spelen' Remmers, you're a good friend and great hometown companion. The members of the infamous Heerencub, Garrelt 'G' de

Graaf and Pieter Paul ‘PP’ Slikker, are - although they tend to damage shoulder-joints and do not display any gift whatsoever for playing Risk - loyal friends and great partners in discussion and the occasional holiday-getaway. I am proud to have you as my paranymphs.

Lu Lu, you have been very close to me for many years. Years that spanned most of the time I worked on this dissertation. I want you to know that you have been very important to me and that I will never forget our shared time. I hope we can both celebrate the blessing of ‘what if’.

Lastly, I would like to thank my family, of which all members are dear to me, but some deserve a special mention. ‘Matje’, and ‘de Noompert’, I could not wish for a better little brother or sister(-in-law). Surely, mom, dad, words cannot express my feelings for you. You are the base upon which I have founded myself.

And then there is ‘Flo’. Although I may have a greater literary penchant than the average person who likes to talk β and π, not even Joyceian capabilities would allow me to express my true feelings. All I have to say is: Our being together is what the universe would have had to culminate to; you are the one.

Biography

Carel F.W. Peeters (1982, Nijmegen) studied statistics, political science, and philosophy. He received his M.A. (summa cum laude) in Political Science specializing in Sociometrics from VU University Amsterdam (VUA), and an advanced M.Sc. (magna cum laude) in Statistics specializing in Mathematical Statistics and Applied Mathematics for the Behavioral and Biological Sciences from the Katholieke Universiteit Leuven (KUL). After commencing on a lectureship at VUA he started his Ph.D. research in Bayesian Statistics at the Department of Methodology & Statistics at Utrecht University (UU) in 2007, which resulted in underlying thesis. While endeavoring on his Ph.D. thesis he also held a research fellowship at the Strategic Chair Integrity of Governance at VUA. During his Ph.D. candidature he published scholarly work in journals such as *Sociological Methods & Research* and *Psychometrika*. He has also been employed as a research scientist at the Psychometric Research Centre of CITO, Institute for Educational Measurement. Currently he is employed as a postdoctoral researcher in multivariate statistics for integrative bioinformatics at the Department of Epidemiology and Biostatistics, VU University medical center.

Under the Bludgeonings of chance
My head is bloody, but unbowed