# Three Essays on Bayesian Factor Models

**Dissertation**

zur Erlangung des akademischen Grades
Doktor der Wirtschaftswissenschaften (Dr. rer. pol.)
am Fachbereich Wirtschaftswissenschaften
der Universität Konstanz

vorgelegt von:
Rémi Piatek
Jacob-Burckhardt-Str. 20
78464 Konstanz

Tag der mündlichen Prüfung: 23. Juli 2010

1. Referent: Prof. Dr. Winfried Pohlmeier
2. Referent: Prof. Dr. François Laisney

*A ma famille,*
*avec toute mon affection*
*et ma profonde reconnaissance*

# Acknowledgments

The realization of this dissertation has been a journey paved with the help and the support of many persons, to whom I would like to express my sincerest gratitude with these first words.

First of all, I owe the writing of this thesis to my Ph.D. advisor Prof. Dr. Winfried Pohlmeier. I am very indebted to him for offering me the great opportunity to work at his Chair of Econometrics at the University of Konstanz. Prof. Pohlmeier has a special talent recruiting people from diverse backgrounds with complementary skills and knowledge, while creating a research environment that is particularly stimulating and inspiring. I am very grateful for his open-mindedness, his constant enthusiasm, his professional advice, and his unconditional support in all the decisions I made.

I am also very thankful to Prof. Dr. François Laisney, for following me and supporting me since my very first steps in the field of Econometrics. His advice has always been very helpful and a source of great motivation.

Meeting Prof. Dr. James J. Heckman was a turning point in my studies and gave a new orientation to my research work. I am very grateful to him for inviting me to the University of Chicago, and for offering me the great chance to work with his research group, a veritable hotbed of new ideas, which is a necessary condition for innovative and successful research. I am also very thankful to Prof. Dr. Hedibert F. Lopes, for encouraging me to persevere in Bayesian Econometrics, to Georges Yates, for his invaluable help in computer programming, and to all the members of this research group from the windy city.

Working at the University of Konstanz has been more than a fulfilling professional experience, it has been an everyday pleasure. Thank you to my colleagues and friends for contributing to this: Roxana Halbleib, Li Lidan, Sandra Nolte, Derya Uysal, Laura Wichert, Anton Flossmann, Markus Jochmann, Matthias Krapf, Fabian Krüger, Hao Liu, Frieder Mokinski and Ingmar Nolte. Special thanks to Ruben Seiberlich for his steady good humor and his support during difficult times.

I would also like to thank two colleagues from the University of Mannheim, namely Pia Pinger, for the productive and pleasant collaboration, as well as Philipp Eisenhauer, for our enlightening discussions.

Finally, the last thankful words go to my family and close friends. Their unwavering support and their love gave me the strength to get through many difficult situations.

# Contents

# List of Figures

# List of Tables

# Summary

Factor analysis represents one of the most important and useful instruments in the toolbox of the analyst who needs to extract a small set of meaningful latent factors from a larger set of observed response variables. This methodology dates back to the beginning of the last century and is often attributed to Spearman (1904) in the field of psychology, who initiated the well-known yet controversial *general intelligence factor* theory—the so-called '$g$' theory. Quite recently, factor analysis has undergone a revival and has been brought to the fore in the social sciences. Two different trends can explain this awakening: first, the borderlines between the different fields of research in the social sciences have been vanishing in the last decades, and factor analysis has appeared as an obvious solution to conduct empirical investigations at the crossroads of psychology, economics and sociology, where unobserved constructs such as cognitive abilities and personality traits have to be handled. Second, new developments in statistics have given a new impulse to factor analysis, both in frequentist and Bayesian inference. Concerning the latter, modern stochastic simulation methods make it possible to tackle persistent problems such as dimension determination of the latent structure, factor and variable selection, and more generally model uncertainty.

The present dissertation consists of three stand-alone research papers that all deal with factor models from a Bayesian perspective, both in a theoretical and an empirical setup. More precisely, the thesis is organized in a progressive way as follows: Chapter 1 briefly presents the general framework of the model used throughout these pages, and surveys the simulation methods that can be implemented to conduct Bayesian inference in this kind of model. Chapters 2 and 3 introduce some important theoretical variations extending the original setup presented in Chapter 1, and illustrate how factor structure models can be empirically applied to assess the importance of cognitive and noncognitive skills in explaining various outcomes of interest.

The choice of a Bayesian rather than a frequentist approach often lends itself to great controversy. Our goal here is not to fuel the debate about the merits of the Bayesian methods relative to their classical counterparts, nor to engage in proselytizing the benefits of the former. However, a few points are worth mentioning to explain why we are adopting this approach in this dissertation. The increasing complexity of the factor models used in empirical research makes the implementation of classical methods tricky in practice. The introduction of latent variables involves for instance high-dimensional integrals in the likelihood function of the model. In this regard, Bayesian procedures relying on simulation methods such as Markov chain Monte Carlo methods clearly have a relative advantage over classical approaches. By simulating all latent variables and parameters, they bypass the problem of the untractable likelihood function in two respects: on the one hand, there is no need to derive the full likelihood function of the model, and on the other hand, the maximization problem of this complicated likelihood is merely eluded. Moreover, Bayesian procedures also enable the researcher to address questions related to factor selection, which are raising a great deal of interest since a few years in the statistical literature. Nevertheless, the methods described and used in this thesis are not exclusively reserved for Bayesian researchers. They can also be used based on their convenience, even in a frequentist framework where the identification of the model is achieved with strictly classical arguments.

The first chapter of this thesis provides an overview of the general theoretical framework for the class of factor structure models used in this thesis, as well as a short discussion on the main identification issues. But the primary purpose, and also the main contribution of this chapter, is to provide ready-to-use guidelines to the empirical researcher who is willing to perform Bayesian inference in factor structure models. In so doing, we review the Bayesian techniques that have been proposed in the literature to handle these models. More exactly, we focus on Markov chain Monte Carlo methods, a class of simulation methods that is particularly suited to this kind of problem. Full details are provided to build up the Gibbs sampler step by step, where linear, dichotomous, censored and ordered response variables can be accommodated. Some recurrent problems affecting the Gibbs sampler are then discussed, like the slow convergence of the Markov chain to the stationary distribution, or its poor mixing, and recent technical improvements are reviewed.

Finally, the standard factor structure model is often criticized for its restrictive distributional and functional assumptions. We show how the normality assumptions can be relaxed through the use of mixtures of normals, and derive all the conditional distributions required to adapt the Gibbs sampling scheme for this type

of specification. Last, we explain how the latent factors can be specified as correlated, which also represents an improvement over the traditional methods assuming independence. Relaxing these conventional assumptions introduces more flexibility, thus establishing a parallel between factors structure models and semiparametric latent variable models.

The second chapter is an empirical application of factor structure models investigating the impact of an individual's level of locus of control, a concept commonly used in social psychology (Rotter, 1966), on educational choices and wages. In this joint work with Pia Pinger, we establish that more internal individuals, i.e., who believe that reinforcement in life comes from their own actions, instead of being determined by luck or destiny, earn higher wages. However, the positive effect of a more internal locus of control only translates into labor income via the channel of education: once schooling is controlled for, the impact of locus of control on wages vanishes. The data used for our analysis are retrieved from the German Socio-Economics Panel (GSOEP), a representative longitudinal micro-dataset providing a wide range of socio-economic information on individuals in Germany. The GSOEP is particularly well-suited for our analysis, in that it allows us to exploit information on locus of control and outcomes for various cross-sections of people of different ages.

To reach our conclusions, we had to face different technical problems. To tackle measurement error and endogeneity problems that are often overlooked in the literature, although they plague many empirical studies relying on usual least squares approaches, factor models with the structure presented in Chapter 1 are implemented. To deal with the problem of truncated life-cycle data, we combine a sample of young adults who have not yet entered the labor market with a sample of working-age individuals. Producing identification of different parts of the likelihood using different samples, we are able to correct for potential biases that arise due to reverse causality and spurious correlation, and to measure the impact of *premarket* locus of control on later outcomes.

The third chapter stems from a joint work with Gabriella Conti, James J. Heckman and Hedibert F. Lopes. It presents novel Bayesian econometric methods for reducing high-dimensional data into low-dimensional aggregates using factor models to examine the effect of early-life conditions and education on health. Assessing the dimension of the latent structure of a model is one of the new challenging topics in factor analysis. Data sets with a growing number of measurements are being made available, and open new horizons for the study of human personality traits, abilities and behaviors. However, traditional factor analysis approaches show their

limits when they are applied to such large scale data sets, in that they fail to provide clear answers regarding the dimension and the role played by the latent factors in determining some outcomes of interest. To overcome these problems, several new stochastic search schemes have been proposed in the literature. We implement and extend the parsimonious Bayesian factor analysis, a modern approach developed by Frühwirth-Schnatter and Lopes (2009) that makes it possible to address all these problems simultaneously. This methodology is applied to the 1970 British Cohort Study within a life course framework, in order to analyze the effect of childhood cognitive ability and psychosocial traits on education and adult health, in a model where individuals select into education on the basis of their expected gains.

We find that, while the structure of personality and cognition is gender-invariant, the importance of these factors for adult outcomes greatly differs between genders. Noncognitive factors developed by age 10 are more important determinants of adult outcomes for males than for females. We also find that education accounts for a large part of the observed health disparities at age 30 for males, and less for females. We then show that individuals select into education depending on both market and non-market gains, and that the average effect of education for females at different margins of the distribution of the unobservables varies in a way that rationalizes the findings reported in the literature. Finally, we provide evidence that a misspecification of the latent factor structure leads to an incorrect assessment of the importance of early-life conditions in influencing both education and later-life health.

# Zusammenfassung

Die Faktorenanalyse ist ein entscheidendes Analyseinstrument, um aus einer größeren Menge von Antwortvariablen die an sich bedeutsamen Dimensionen, auch latente Faktoren genannt, zu extrahieren. Die Methode an sich wurde bereits zu Beginn des letzten Jahrhunderts entwickelt, stammt aus der Psychologie und wird im Allgemeinen Spearman (1904), dem Begründer der weit verbreiteten aber umstrittenen „Generalfaktortheorie", zugeschrieben. Seit kurzem ist die Faktorenanalyse wieder ins Zentrum des akademischen Interesses gerückt und findet ihre Anwendung heutzutage vor allem in der sozialwissenschaftlichen Forschung. Zwei Entwicklungen können als Gründe für die Wiederentdeckung der Faktorenanalyse angeführt werden: Da die Grenzen zwischen den sozialwissenschaftlichen Fachbereichen zunehmend verschwimmen, dient die Faktorenanalyse erstens dazu, empirische Fragestellungen zu beantworten. Hierbei bietet sich die Schnittstelle von Psychologie, Ökonomie und Soziologie an, für die unbeobachtete Konstrukte wie kognitive Fähigkeiten und Persönlichkeitseigenschaften eine große Rolle spielen. Zweitens haben neuere Entwicklungen aus der Statistik, sowohl in den Bereichen der bayesschen Wahrscheinlichkeitstheorie als auch der frequentistischen Inferenz, der Fakorenanalyse neue Impulse verliehen. Insbesondere mithilfe der bayesschen Wahrscheinlichkeitstheorie ist es dank moderner Computer und Simulationsmethoden heutzutage möglich, bislang ungelöste Probleme der Dimensionsreduzierung, der Bestimmung latenter Strukturen, der Faktor- und Variablenselektion und der allgemeinen Modellunsicherheit zu lösen.

Die vorgelegte Dissertation besteht aus drei unabhängigen Aufsätzen, die sich alle unter dem Oberthema der theoretischen und empirischen Faktormodellierung aus bayesscher Perspektive zusammenfassen lassen. Der Aufbau der Dissertation ist wie folgt: In Kapitel eins wird sowohl die theoretische Struktur des in den Folgekapiteln genutzten Modells im Detail erklärt, als auch eine Einordnung und Übersicht über Simulationsmethoden gegeben, die zur bayesschen Inferenz in diesem Modellrahmen genutzt werden können. In den darauffolgenden Kapiteln zwei und drei werden

dann einige wichtige theoretische Erweiterungen des in Kapitel eins beschriebenen Grundmodells präsentiert und erläutert, sowie empirische Anwendungen vorgestellt. Hierbei steht der Nutzen von Faktorstrukturmodellen für die empirische Analyse kognitiver und nichtkognitiver Fähigkeiten und deren Bedeutung für den Arbeitsmarkterfolg im Vordergrund.

Die Wahl bayesscher anstelle frequentistischer Methoden führt oft zu grundlegenden methodischen Kontroversen. Das Ziel dieser Arbeit ist es jedoch weder zu der bestehenden Diskussion um die philosophisch „richtige“ Art des methodischen Ansatzes beizutragen, noch den bayesschen Ansatz als „den Richtigen“ hervorzuheben. Die Wahl bayesscher Methoden ist in der Praxis oft allein auf die Komplexität und die große Anzahl zu schätzender Parameter zurückzuführen, welche die Nutzungsmöglichkeiten frequentistischer Methoden oft erheblich einschränken. Das Schätzen von Modellen mit latenten Variablen beinhaltet beispielsweise die Berechnung hochdimensionaler Integrale in der Wahrscheinlichkeitsfunktion des jeweiligen empirischen Modells. In dieser Hinsicht sind bayessche Simulationsverfahren wie zum Beispiel Markov Chain Monte Carlo Methoden oft besser geeignet, da sie durch die Simulation latenter Variablen und Parameter das Problem der Herleitung und der Maximierung sehr komplexer Wahrscheinlichkeitsfunktionen umgehen. Zudem ermöglichen Simulationsmethoden seit kurzem, das seit einigen Jahren in der statistischen Literatur diskutierte Problem der Faktorenselektion anzugehen. Es ist hervorzuheben, dass die in dieser Arbeit besprochenen Methoden nicht zwingend im Rahmen einer bayesschen Analyse genutzt werden müssen, sondern dass die Identifikation der Parameter klassischen Argumenten unterliegt und somit alle Schätzungen genauso mit frequentistischen Methoden durchführbar sind.

Im ersten Kapitel dieser Arbeit wird einerseits der theoretische Modellrahmen für die in den Folgekapiteln angewandten Faktorstrukturmodelle vorgegeben und erklärt. Andererseits wird auf die zentralen Identifikationsergebnisse und -probleme eingegangen. Die zentrale Intention dieser Arbeit im allgemeinen, und insbesondere die des ersten Kapitels, ist es dem empirisch orientierten Forscher Rezepte und Richtlinien an die Hand zu geben, mit deren Hilfe er sich die Methoden der bayesschen Inferenz für die Schätzung von Faktorstrukturmodellen zunutze machen kann. Daher werden im ersten Teil dieser Dissertation die wichtigsten Techniken der bayesschen Statistik zur Schätzung dieser Art von Modellen erklärend zusammengefasst. Es wird insbesondere auf Markov Chain Monte Carlo Methoden eingegangen, da es sich hierbei um eine Simulationsmethode handelt, die für die Analyse von Faktorstrukturmodellen besonders geeignet ist. Insbesondere der methodische Aufbau des Gibbs Samplers wird im Detail erklärt, wobei sowohl auf lineare als auch of dichotome, zensierte und geordnete Antwortvariablen eingegangen wird. Einige wiederkehrende Probleme des Gibbs Samplers, wie zum Beispiel Konvergenzprobleme der Markov-Ketten hin zu einer stationären Verteilung, oder Probleme, die durch die hohe Autokorrelation der Markov-Ketten entstehen, werden im Folgenden sowohl besprochen als auch die neuesten statistischen Verbesserungen in diesem Bereich diskutiert.

Nicht zuletzt werden standardmäßige Faktorstrukturmodelle oft aufgrund ihrer restriktiven Verteilungs- und Funktionalannahmen kritisiert. Um diese Annahmen zu lockern, wird insbesondere auf den Nutzen zusammengesetzter Normalverteilungen verwiesen und alle bedingten Verteilungen hergeleitet, die benötigt werden, um den Gibbs Sampler an diese Art flexibler Spezifikation anzupassen. Eine gemischte Verteilung der Latenten Faktoren dient jedoch nicht nur dazu, die traditionellen Modelle flexibler zu machen, sondern stellt auch eine Verbindung zu semiparametrischen latenten Variablenmodellen dar. Zum Schluss des ersten Kapitels wird zudem dargelegt, wie Modelle mit korrelierten latenten Faktoren spezifiziert werden können, was wiederum eine Erweiterung der traditionellen, Unabhängigkeit voraussetzenden Modelle darstellt.

Das zweite Kapitel entstammt einem gemeinsamen Forschungsprojekt mit Pia Pinger und umfasst eine empirische Anwendung der oben beschriebenen Faktorstrukturmodelle. Das Ziel der Arbeit ist es, den Einfluss der Kontrollüberzeugung (ein der Sozialpsychologie entlehntes Konzept, das auf Rotter, 1966, zurückgeführt wird) eines Individuums auf dessen Bildungsentscheidungen und Lohn zu bestimmen. Die Autoren zeigen, dass Individuen mit einer inneren Kontrollüberzeugung, das heißt die von der Wirkung ihres Handelns überzeugt sind und sich nicht von Glück oder dem Schicksal bestimmt fühlen, bessere Arbeitsmarktergebnisse in Form von höheren Löhnen erzielen. Davon abgesehen wird jedoch gezeigt, dass das Erzielen von höheren Löhnen allein auf die besseren Bildungsergebnisse der „internen“ Individuen zurückzuführen ist: Wird für Bildung kontrolliert, tendiert der Effekt der Kontrollüberzeugung gegen Null. Die der Analyse zugrunde gelegten Daten entstammen dem deutschen Sozioökonomischen Panel (SOEP), einer repräsentativen Längsschnittstudie, die eine große Anzahl sozioökonomischer Variablen für eine zufällig gezogene Stichprobe deutscher Individuen enthält. Das SOEP ist für die oben beschriebene Analyse besonders geeignet, da es sowohl psychometrische Maße der Kontrollüberzeugung als auch Bildungs- und Arbeitsmarktergebnisse für Querschnitte unterschiedlichen Alters enthält.

Verschiedene technische Probleme werden im Zuge der Datenanalyse gelöst. So eignen sich die oben beschriebenen Faktorstrukturmodelle sehr gut, um die in der bisherigen Literatur oft vernachlässigten Messfehler- und Endogenitätsprobleme zu lösen und wurden daher von den Autoren implementiert. Zudem werden die Messungen aus einer Stichprobe Siebzehnjähriger mit den Arbeitsmarktergebnissen junger Erwachsener kombiniert, um das Problem zensierter Lebenszyklen zu beheben. Um sowohl eine potentielle umgekehrte Kausalität als auch eine Scheinkorrelation ausschließen zu können, werden unterschiedliche Teile der Wahrscheinlichkeitsfunktion anhand verschiedener Stichproben identifiziert und geschätzt. Dadurch kann der unverzerrte Einfluss der sogenannten „premarket“ Kontrollüberzeugung extrahiert und somit jener Effekt, der noch nicht durch vorangegangene Arbeitsmarkterlebnisse verzerrt ist, identifiziert werden.

Das dritte Kapitel der Dissertation entspringt einem gemeinsam mit Gabriella Conti, James J. Heckman und Hedibert F. Lopes durchgeführten Forschungsprojekts. Die Autoren nutzen neuartige bayessche ökonometrische Methoden, um hochdimensionale Daten in niedrigdimensionale Aggregate zu übertragen, und somit den Effekt von frühkindlichen Bedingungen und frühkindlicher Bildung auf die Gesundheit zu erfassen. Die Dimension der latenten Struktur eines Modells gemeinsam mit seinen Parametern zu schätzen, gehört zu den größten Herausforderungen in der Faktorenanalyse und ist ein aktuell intensiv diskutiertes Thema, was vor allem auf den vermehrten Zugang zu Datensätzen mit einer großen Anzahl an Fähigkeits-, Persönlichkeits- und Verhaltensmaßen zurückzuführen ist. Traditionelle Faktormodelle können oft nur in beschränktem Maße auf solche großen Datensätze angewendet werden, da von vornherein weder die Dimension der latenten Struktur noch die Bedeutung einzelner Maße für die jeweilige abhängige Variable bestimmt sind. Um diese Probleme zu lösen, sind in der Vergangenheit bereits verschiedene stochastische Suchmechanismen entwickelt worden. Im dritten Kapitel wird das klassische Faktormodell daher erweitert und ein von Frühwirth-Schnatter und Lopes (2009) entwickelter Ansatz implementiert, der es ermöglicht alle oben genannten Probleme gleichzeitig zu lösen. Als Anwendung wird der Effekt von kognitiven und psychosozialen Eigenschaften auf die Bildungsergebnisse und die Gesundheit im Erwachsenenalter im Rahmen eines Lebenszyklusmodells für die 1970er British Cohort Studie geschätzt, wobei speziell die ergebnisorientierte Selektion berücksichtigt wird.

Die Autoren kommen zu dem Ergebnis, dass die Persönlichkeits- und Kognitionsstruktur zwar geschlechtsneutral, die Wirkung dieser Faktoren für die Gesundheit allerdings sehr heterogen ist. Im Alter von zehn Jahren gemessene nichtkognitive Faktoren haben für Männer einen weit größeren Einfluss auf die späteren Bildungs- und Gesundheitsergebnisse als für Frauen. Außerdem hat im Alter von 30 Jahren bei Männern die Schulbildung eine größere Erklärungskraft für die gesundheitlichen Unterschiede als bei Frauen. Zudem wird gezeigt, dass die Selbstselektion der Individuen ergebnisgetrieben ist und sich sowohl an der späteren Marktsituation als auch an weiteren relevanten Resultaten orientiert. Es wird deutlich, dass der positive Bildungseffekt im Einklang mit der bisherigen Literatur, insbesondere bei Frauen, in verschiedenen Bereichen der unbeobachtbaren Heterogenitätsverteilung unterschiedlich groß ausfällt. Abschließend wird im dritten Kapitel gezeigt, dass eine Misspezifikation der latenten Variablenstruktur zu einer Fehleinschätzung der Bedeutung frühkindlicher Einflüsse für Bildungs- und Gesundheitsergebnisse im Erwachsenenalter führen kann.

# Chapter 1

# Bayesian Inference for Factor Structure Models via Gibbs Sampling

## 1.1 Introduction

Factor models have a long history in statistics and have grown very popular in the social sciences. This considerable interest stems from the fact that these models make it possible to capture latent traits that cannot be directly observed, but rather unveiled through questionnaires specially designed to measure them. They are thereby particularly useful in research fields such as psychology, economics, or marketing, where personal characteristics like cognitive abilities, personality traits and preferences, among others, can be elicited through these constructs, and introduced into statistical models to predict some outcomes of interest. By combining the advantages of simultaneous equation models and factor analytic models, factor structure models allow the researcher to tackle some pervasive problems such as measurement error and endogeneity. In recent years, they have for example been extensively used to investigate the impact of latent abilities and personality traits measured with error by some indicators on various economic and social outcomes (see for example Hansen et al., 2004; Heckman et al., 2006).

Beyond the standard normal model with continuous response variables, factor structure models have been extended to accommodate discrete variables that are of widespread use in empirical analysis, and also to relax some restrictive distributional assumptions. However, the inclusion of limited dependent variables is quite challenging from a technical point of view, since it involves high-dimensional integrals that are untractable in most cases. Various approaches have been proposed from a frequentist perspective, some of them relying on analytical approximation methods like the Gaussian quadrature, others implementing numerical approximations based on simulation. However, these methods turn out to be cumbersome when the number of discrete variables increases.

For these reasons, Bayesian approaches relying on Markov chain Monte Carlo, and more especially on the Gibbs sampler, have appeared as an attractive alternative to classical methods. They are particularly well suited to models where the number of equations and parameters can be very large, and have become a key technique for their convenience (Carneiro et al., 2003). However, the features of the Gibbs sampler have not been documented in a comprehensive way for this kind of problem. Empirical researchers who are unfamiliar with Bayesian methods may therefore become discouraged and give up using this estimation strategy, or even worse, give up using factor structure models altogether. This chapter aims at filling this gap by presenting all the technical details required to construct the Gibbs sampler step by step.

Section 1.2 presents the theoretical framework, outlining the specification of the model and providing some elements of identification. In Section 1.3, the standard Gibbs sampler is derived in detail. Section 1.4 discusses the main shortcomings of the standard algorithm, with regard to the mixing and the convergence of the Markov chain. Some useful solutions are reviewed, and further extensions are then considered. Finally, Section 1.5 explains how the usual normality assumptions may be relaxed through the introduction of mixtures of normals, allowing for more flexible functional forms that tend to nonparametric approaches. The choice of the prior parameters will not be explicitly discussed here, although it is a question of first importance, and often controversial in the literature. However, some examples and their corresponding references are presented in Appendix 1.B to help selecting these prior parameters.

## 1.2 Theoretical framework

### 1.2.1 Model specification

The data consist of a $n$-sample of observations on $m$ related response variables $\boldsymbol{y}_i = (y_{1,i}, ..., y_{m,i})'$, for $i = 1, ..., n$. This vector of observed responses can contain different types of variables, e.g., continuous, dichotomous, ordinal and censored, and is linked to the same dimensional vector of latent responses $\boldsymbol{y}_i^\star$ that are assumed to linearly depend on some covariates $\boldsymbol{x}_i$ and on a set of latent factors $\boldsymbol{f}_i$ through:

$$\begin{aligned} \boldsymbol{y}_i^\star &= \boldsymbol{\beta}\boldsymbol{x}_i + \boldsymbol{\alpha}\boldsymbol{f}_i + \boldsymbol{\varepsilon}_i, \\ \boldsymbol{y}_i &= g(\boldsymbol{y}_i^\star; \boldsymbol{\lambda}) = \left(g_1(y_{1,i}^\star; \boldsymbol{\lambda}_1), ..., g_m(y_{m,i}^\star; \boldsymbol{\lambda}_m)\right)', \end{aligned} \tag{1.1}$$

where $\boldsymbol{\beta}$ is a $(m \times p)$–dimensional matrix of slope parameters, $\boldsymbol{\alpha}$ is a $(m \times k)$–dimensional matrix of factor loadings, and $\boldsymbol{\varepsilon}_i$ is a vector of independent and identically distributed error terms independent of the observed covariates. Depending on the nature of the observed response variables, the corresponding link functions $g_s(\cdot)$, $s = 1, ..., m$, can take different forms that are summarized in Table 1.1.

There are numerous examples of data in the social sciences where the variable of interest cannot be directly measured and where the latent outcome representation is very useful. Personal preferences and beliefs are for instance impossible to quantify on a continuous scale. Instead, it is common practice to use multiple-choice

**Table 1.1:** Variable types and link functions

| **Type** | **Support** | **Link function** |
|---|---|---|
| continuous | $y \in \mathbb{R}$ | $g(y^\star) = y^\star$ |
| binary | $y \in \{0, 1\}$ | $g(y^\star) = \mathbb{1}[y^\star > 0]$ |
| ordinal | $y \in \{1, ..., L\}$ | $g(y^\star, \boldsymbol{\lambda}) = \sum_{l=1}^{L} l\, \mathbb{1}[\lambda_{l-1} \leqslant y^\star < \lambda_l]$ |
| censored from below | $y \in \mathbb{R}_{\backslash(-\infty,\lambda]}$ | $g(y^\star, \lambda) = \lambda \mathbb{1}[y^\star \leqslant \lambda] + y^\star\, \mathbb{1}[y^\star > \lambda]$ |
| censored from above | $y \in \mathbb{R}_{\backslash[\lambda,+\infty)}$ | $g(y^\star, \lambda) = \lambda \mathbb{1}[y^\star \geqslant \lambda] + y^\star\, \mathbb{1}[y^\star < \lambda]$ |

**Note:** $\mathbf{1}[\cdot]$ represents the indicator function which takes value 1 if the corresponding condition is verified, 0 otherwise.

questionnaires where the different alternatives are condensed in a small number of exclusive categories.[1]

To complete the specification of the model, the distributions of the latent factors $\boldsymbol{f}_i$ and error terms $\boldsymbol{\varepsilon}_i$ are specified as multivariate normal:

$$\boldsymbol{f}_i \sim \mathcal{N}_k(\mathbf{0};\, \boldsymbol{\Psi}_{\mathbf{F}})\,, \tag{1.2}$$

$$\boldsymbol{\varepsilon}_i \sim \mathcal{N}_m(\mathbf{0};\, \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}})\,, \qquad \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}} = \operatorname{diag}\left(\sigma_1^2, ..., \sigma_m^2\right). \tag{1.3}$$

While the error terms of the different equations are assumed to be independent, the latent factors can be correlated if their covariance matrix $\boldsymbol{\Psi}_{\mathbf{F}}$ is specified as non-diagonal. All dependencies among the observed response variables $\boldsymbol{y}_i$ are explained by the common factors.[2] Furthermore, the latent factors are usually assumed to be independent of the covariates and of the error terms for identification purposes ($\mathbf{F} \perp\!\!\!\perp \mathbf{X} \perp\!\!\!\perp \boldsymbol{\varepsilon}$). The normality assumptions of Equations (1.2) and (1.3) are conventional in traditional factor analysis (see, for instance, Lopes and West, 2004) and are relaxed in Section 1.5.

### 1.2.2 Identification issues

Although the primary goal of this chapter is not identification, a few points are worth mentioning to avoid confusion when deriving the Gibbs sampler. Identification is a typical problem in factor analysis and is now well-documented. For instance, Carneiro et al. (2003) and Hansen et al. (2004) demonstrate and discuss the identification of factor structure models.

[1] In psychology, psychometric tests with questions using Likert-scale items like 'strongly disagree', 'slightly disagree', 'slightly agree', 'strongly agree' are in common use.

[2] In the factor analysis literature, the proportion of the common variance of the response variable not explained by the factors is called *uniqueness*.

First, the latent structure of the model is not identified without imposing some restrictions on the factor loading matrix $\boldsymbol{\alpha}$. For any invertible matrix $\mathbf{P}$ of dimension $(k \times k)$, the transformation of the model assigning $\tilde{\boldsymbol{\alpha}} = \boldsymbol{\alpha}\mathbf{P}^{-1}$ and $\tilde{\boldsymbol{f}}_i = \mathbf{P}\boldsymbol{f}_i$ indeed yields the same likelihood as the initial model.[3] An infinite number of observationally equivalent models can thus be found by rotating $\boldsymbol{\alpha}$. Different strategies can be implemented to solve this problem. In the case where the latent factors are uncorrelated, the standard approach consists of fixing their variances to unity ($\boldsymbol{\Psi}_{\mathbf{F}} = \mathbf{I}_k$) and assuming a full-rank lower triangular structure of the factor loading matrix (Geweke and Zhou, 1996):

$$\boldsymbol{\alpha} = \begin{pmatrix} \alpha_{1,1} & 0 & 0 & \cdots & 0 \\ \alpha_{2,1} & \alpha_{2,2} & 0 & \cdots & 0 \\ \alpha_{3,1} & \alpha_{3,2} & \alpha_{3,3} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \alpha_{k,1} & \alpha_{k,2} & \alpha_{k,3} & \cdots & \alpha_{k,k} \\ \alpha_{k+1,1} & \alpha_{k+1,2} & \alpha_{k+1,3} & \cdots & \alpha_{k+1,k} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \alpha_{m,1} & \alpha_{m,2} & \alpha_{m,3} & \cdots & \alpha_{m,k} \end{pmatrix}. \tag{1.4}$$

With this structure, $\boldsymbol{\alpha}$ has $mk - k(k-1)/2$ free elements. However, these restrictions do not rule out a sign switch of the latent factors and of the corresponding columns of the factor loading matrix.[4] To deal with this problem, the diagonal elements can be restricted to be positive or negative.

The case with correlated factors is more tricky, insofar as constraining the upper part of the factor loading matrix is not sufficient to prevent all possible transformations. Similar to the uncorrelated case, the scale of the latent factors can be set by assuming that $\boldsymbol{\Psi}_{\mathbf{F}}$ is constrained to be a correlation matrix.[5] Consider now the Cholesky decomposition $\boldsymbol{\Psi}_{\mathbf{F}} = \boldsymbol{\Gamma}\boldsymbol{\Gamma}'$, where $\boldsymbol{\Gamma}$ is a $(k \times k)$–dimensional lower-triangular matrix, and transform the model such that $\tilde{\boldsymbol{\alpha}} = \boldsymbol{\alpha}\boldsymbol{\Gamma}$ and $\tilde{\boldsymbol{f}}_i = \boldsymbol{\Gamma}^{-1}\boldsymbol{f}_i$. Because $\boldsymbol{\Gamma}$ is lower-triangular and the upper-triangular part of $\boldsymbol{\alpha}$ is restricted to zero, the product of these two matrices produces a matrix with the same lower-

[3] When the latent factors are specified as uncorrelated, the argument has to be slightly modified to preserve the diagonality of the covariance matrix $\boldsymbol{\Psi}_{\mathbf{F}}$: for any arbitrary orthogonal matrix $\mathbf{P}$ of dimension $(k \times k)$ such that $\mathbf{P}'\mathbf{P} = \mathbf{I}_k$, the transformation of the model assigning $\tilde{\boldsymbol{\alpha}} = \boldsymbol{\alpha}\mathbf{P}'$ and $\tilde{\boldsymbol{f}}_i = \mathbf{P}\boldsymbol{f}_i$ yields the same likelihood as the initial model.

[4] To better understand this problem, consider the case where the transformation matrix only has $+1$ or $-1$ on its diagonal, i.e., $\mathbf{P} = \text{diag}\,(\pm 1, \pm 1, ..., \pm 1)$.

[5] Constraining the diagonal elements of the covariance matrix to 1 is not trivial in the MCMC sampling scheme and better alternatives are available. However, let us consider this case for the purpose of presentation here.

triangular structure as $\boldsymbol{\alpha}$. Hence, this transformation shows that a model with the factor loading matrix defined as in Equation (1.4) and with correlated factors is observationally equivalent to a model with uncorrelated factors.[6]

To prevent this kind of rotation of the matrix $\boldsymbol{\alpha}$ and the identification problem it results in, more constraints are required on $\boldsymbol{\alpha}$. The use of *dedicated* response variables, i.e., variables that load exclusively on single factors, is probably the easiest way to cope with this issue. Not only does it solve the identification problem, it also makes it possible to give a clear interpretation to the latent factors in most cases, since each factor is related to a precise set of response variables (e.g., items measuring a specific personality trait). In the applied literature, models where the latent factors are identified through dedicated measurements, and where these latent factors simultaneously have an impact on some outcome variables of interest, are very common (Hansen et al., 2004).[7]

As an alternative to the above-mentioned sign constraint on the diagonal elements used to prevent the sign switching problem, some authors prefer to fix the diagonal loadings to a pre-specified value (generally to 1), rather than fixing the variances of the latent factors to 1. This is a way of anchoring the factors into real measurements by giving them an interpretable metric (Cunha and Heckman, 2007; 2008; Cunha et al., 2010). It also makes the sampling process easier in the correlated case, since covariance matrices are usually easier to draw than correlation matrices.

The last identification issues regard model specific parameters, more precisely idiosyncratic variances and cut-points in the discrete cases. Given the fact that conditional on the latent factors, the overall model in Equation (1.1) is nothing more than a simultaneous-equation model made of linear equations, standard probit, ordered probit and tobit models, the usual restrictions can be applied. We therefore set the idiosyncratic variances to 1 when the response variables are binary and ordinal, and set the cut-points such that $\lambda_0 = -\infty < \lambda_1 = 0 < \lambda_2 < \ldots < \lambda_L = -\infty$ in the ordinal case, where the threshold $\lambda_1$ needs to be set to 0 when an intercept term is included. In the censored case, the threshold $\lambda$ is assumed to be known. For instance, Hansen et al. (2004) use this kind of censored model to control for *ceiling effects* in test score achievement. They argue that people who 'hit the ceiling' in some easy tests, i.e., achieve the highest score, may still have very different abilities. In this case, the threshold is the maximum attainable score.

[6] $\mathrm{V}\left[\tilde{\boldsymbol{f}}_i\right] = \mathrm{V}\left[\boldsymbol{\Gamma}^{-1}\boldsymbol{f}_i\right] = \boldsymbol{\Gamma}^{-1}\boldsymbol{\Psi}_{\mathbf{F}}\boldsymbol{\Gamma}^{-1\prime} = \mathbf{I}_k$ since by the Cholesky decomposition $\boldsymbol{\Psi}_{\mathbf{F}} = \boldsymbol{\Gamma}\boldsymbol{\Gamma}'$.

[7] In this chapter, we do not distinguish between *measurement sytem* and *outcome system*, since from a theoretical point of view there are no differences between these sets of variables.

### 1.2.3 Likelihood and posterior

Let $\mathbf{Y} = (\boldsymbol{y}_1, ..., \boldsymbol{y}_n)'$ be the $(n \times m)$–dimensional matrix containing all observed response variables. Observed explanatory variables are contained in the $(n \times p)$–dimensional matrix $\mathbf{X} = (\boldsymbol{x}_1, ..., \boldsymbol{x}_n)'$, and the latent factors in the $(n \times k)$–dimensional matrix $\mathbf{F} = (\boldsymbol{f}_1, ..., \boldsymbol{f}_n)'$. The set of parameters specific to response variable $s$ is denoted by $\boldsymbol{\theta}_s = (\boldsymbol{\beta}_s', \boldsymbol{\alpha}_s', \sigma_s^2, \boldsymbol{\lambda}_s')'$, where $\boldsymbol{\beta}_s$ (resp. $\boldsymbol{\alpha}_s$) is the column vector containing the $s^{\text{th}}$ row of matrix $\boldsymbol{\beta}$ (resp. $\boldsymbol{\alpha}$). Finally, let $\boldsymbol{\theta} = \left(\boldsymbol{\theta}_1', ..., \boldsymbol{\theta}_m', \text{vec}\,(\boldsymbol{\Psi}_{\mathbf{F}})'\right)'$ represent the parameters of the overall model. The likelihood can be expressed as:

$$
\begin{aligned}
L(\boldsymbol{\theta}|\mathbf{Y}, \mathbf{X}) &= \int p(\mathbf{Y}|\mathbf{F}, \mathbf{X}, \boldsymbol{\theta}) p(\mathbf{F}|\boldsymbol{\theta})\, \mathrm{d}\mathbf{F}, \\
&= \int \prod_{s=1}^{m} p(\boldsymbol{y}_s|\mathbf{F}, \mathbf{X}, \boldsymbol{\theta}) p(\mathbf{F}|\boldsymbol{\theta})\, \mathrm{d}\mathbf{F}, \qquad (1.5)
\end{aligned}
$$

where $p(\cdot)$ is invariably used in this chapter to denote a probability density function, being for a prior or a posterior distribution.

Because of the latent factors $\mathbf{F}$ common across equations, the response variables are not independent and deriving a closed-form expression for the likelihood appears to be cumbersome in most cases. The latent factors have to be integrated out of the likelihood as shown in Equation (1.5), involving higher-order integrals that are difficult to deal with when the number of factors is large.

From a Bayesian perspective, the likelihood function is combined with a prior density of the parameters $p(\boldsymbol{\theta})$ to provide the posterior distribution of the parameters of interest $p(\boldsymbol{\theta}|\mathbf{Y}, \mathbf{X})$, up to a normalizing constant, through Bayes's fundamental theorem:

$$
p(\boldsymbol{\theta}|\mathbf{Y}, \mathbf{X}) \propto p(\mathbf{Y}|\boldsymbol{\theta}, \mathbf{X}) p(\boldsymbol{\theta}).
$$

However, the problem does not appear easier up to this point, since the untractable likelihood function still has to be dealt with. To circumvent this problem, the latent factors $\mathbf{F}$ and the latent response variables $\mathbf{Y}^\star = (\boldsymbol{y}_1^\star, ..., \boldsymbol{y}_n^\star)'$ can be explicitly introduced into the posterior distribution. This so-called data augmentation procedure (Tanner and Wong, 1987; van Dyk and Meng, 2001) is motivated by the fact that the posterior distribution can be expressed as:

$$
p(\boldsymbol{\theta}|\mathbf{Y}, \mathbf{X}) = \iint p(\boldsymbol{\theta}|\mathbf{F}, \mathbf{Y}^\star, \mathbf{Y}, \mathbf{X}) p(\mathbf{Y}^\star|\mathbf{F}, \mathbf{Y}, \mathbf{X}) p(\mathbf{F}|\mathbf{Y}, \mathbf{X})\, \mathrm{d}\mathbf{Y}^\star\, \mathrm{d}\mathbf{F}. \qquad (1.6)
$$

The predictive densities of $\mathbf{Y}^\star$ and $\mathbf{F}$, namely $p(\mathbf{Y}^\star|\mathbf{F},\mathbf{Y},\mathbf{X})$ and $p(\mathbf{F}|\mathbf{Y},\mathbf{X})$, can then be related to the posterior though:

$$p(\mathbf{Y}^\star|\mathbf{F},\mathbf{Y},\mathbf{X})p(\mathbf{F}|\mathbf{Y},\mathbf{X}) = \int_{\mathbf{\Theta}} p(\mathbf{Y}^\star|\boldsymbol{\theta},\mathbf{F},\mathbf{Y},\mathbf{X})p(\mathbf{F}|\boldsymbol{\theta},\mathbf{Y},\mathbf{X})\underbrace{p(\boldsymbol{\theta}|\mathbf{Y},\mathbf{X})}_{\text{posterior}}\mathrm{d}\boldsymbol{\theta}, \quad (1.7)$$

where $\mathbf{\Theta}$ represents the parameter space of $\boldsymbol{\theta}$. Equations (1.6) and (1.7) are closely related and successive substitutions between them suggest that the parameters $\boldsymbol{\theta}$, the latent $\mathbf{F}$ and the latent response variables $\mathbf{Y}^\star$ can be iteratively simulated to approximate the target posterior distribution $p(\boldsymbol{\theta}|\mathbf{Y},\mathbf{X})$. This procedure greatly simplifies the sampling process, since once the factors have been simulated, the different response variables can be regarded as independent conditional on $\mathbf{F}$, and the matrices $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ can thus be sampled row-wise. As a consequence, the whole problem can be divided into as many distinct tasks as there are different response variables. The Gibbs sampler, which proceeds by simulating each parameter—or parameter block—from its conditional distribution, is therefore particularly appropriate for this kind of problem (Casella and George, 1992).

### 1.2.4 Set-up of the Gibbs sampler

The algorithm is initialized by choosing starting values for all parameters $\boldsymbol{\theta}$, all latent outcomes $\mathbf{Y}^\star$ and all latent factors $\mathbf{F}$. Random values can be chosen, or parameter estimates from some preliminary analysis.[8] Once the initialization has been achieved, the Gibbs sampler is implemented sequentially as follows. At each iteration $(t)$:

(A) For each response variable $s = 1, ..., m$, update the parameters $\boldsymbol{\theta}_s$ and the latent outcomes $\boldsymbol{y}_s^\star$ conditional on the latent factors of the previous iteration:

- (A1) draw the slope parameters $\boldsymbol{\beta}_s^{(t)}$ from $p(\boldsymbol{\beta}_s|\boldsymbol{y}_s^{\star(t-1)},\mathbf{X},\boldsymbol{\alpha}_s^{(t-1)},\mathbf{F}^{(t-1)},\sigma_s^{2(t-1)})$,
- (A2) draw the factor loadings $\boldsymbol{\alpha}_s^{(t)}$ from $p(\boldsymbol{\alpha}_s|\boldsymbol{y}_s^{\star(t-1)},\mathbf{X},\boldsymbol{\beta}_s^{(t)},\mathbf{F}^{(t-1)},\sigma_s^{2(t-1)})$,
- (A3) draw the idiosyncratic variance $\sigma_s^{2(t)}$ from $p(\sigma_s^2|\boldsymbol{y}_s^{\star(t-1)},\mathbf{X},\boldsymbol{\beta}_s^{(t)},\boldsymbol{\alpha}_s^{(t)},\mathbf{F}^{(t-1)})$,
- (A4) for each individual $i$, draw the latent response variable $y_{s,i}^{\star(t)}$ independently from $p(y_{s,i}^\star|\boldsymbol{x}_i,\boldsymbol{\beta}_s^{(t)},\boldsymbol{\alpha}_s^{(t)},\boldsymbol{f}_i^{(t-1)},\sigma_s^{2(t)},\boldsymbol{\lambda}_s^{(t-1)})$,
- (A5) draw the model-specific parameters $\boldsymbol{\lambda}_s^{(t)}$, if any, from $p(\boldsymbol{\lambda}_s|\boldsymbol{y}_s,\boldsymbol{y}_s^{\star(t)})$.

[8] For example, exploratory factor analysis for the starting values of the factor loadings.

(B) Update the latent factors: draw $\boldsymbol{f}_i^{(t)}$ from $p(\boldsymbol{f}_i|\boldsymbol{y}_i^{\star(t)}, \boldsymbol{x}_i, \boldsymbol{\beta}^{(t)}, \boldsymbol{\alpha}^{(t)}, \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}^{(t)}, \boldsymbol{\Psi}_{\mathbf{F}}^{(t-1)})$ for each individual $i$, where $\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}^{(t)} = \text{diag}\,(\sigma_1^{2(t)}, ..., \sigma_m^{2(t)})$.

(C) Draw the parameters of the distribution of the factors $\boldsymbol{\Psi}_{\mathbf{F}}^{(t)}$ from $p(\boldsymbol{\Psi}_{\mathbf{F}}|\mathbf{F}^{(t)})$.

The goal of the Gibbs sampler is to generate a Markov chain with a stationary distribution that coincides with the posterior distribution of the parameters of interest. To do so, the algorithm cycles through Steps (A) to (C) until practical convergence is achieved. As becomes clear, data augmentation greatly simplifies the sampling scheme, since once the latent response variables and latent factors have been simulated, Steps (A1) to (A3) are identical for all types of response variables.

When the response variables are binary or ordinal, Step (A3) is not required since the idiosyncratic variances are fixed to 1 for identification purpose. Step (A5) is only carried out when the response variable is ordinal, since in the censored case the threshold is supposed to be known.

Since the factor loading matrix and the slope parameter matrix are updated row-wise, the following conventions are adopted in the following to simplify the notation. Latent factors and covariates for all individuals are represented by the compact-form matrices $\mathbf{F}$ and $\mathbf{X}$, of respective dimensions $(n \times k)$ and $(n \times p)$, while $\boldsymbol{f}_i$ and $\boldsymbol{x}_i$ represent the column vectors containing the $i^{\text{th}}$ row of the corresponding matrix. So far, $\boldsymbol{\beta}_s$ (resp. $\boldsymbol{\alpha}_s$) represented the $s^{\text{th}}$ row of $\boldsymbol{\beta}$ (resp. $\boldsymbol{\alpha}$). Henceforth, the subscript will be dropped and $\boldsymbol{\beta}$ (resp. $\boldsymbol{\alpha}$) will be used to denote the column vector of slope parameters (resp. vector of factor loadings) corresponding to the current response variable being considered by the algorithm in Step (A). In the same way, the parameter superscripts referring to the current and previous iterations will not be mentioned anymore. Obviously, conditional distributions that depend on other parameters use the latest updated values of these parameters.

## 1.3 A simple Gibbs sampler

This section provides the technical ingredients required to derive the standard Gibbs sampler for the factor structure model.

### 1.3.1 Linear part of each submodel

#### Updating the slope parameters

The vector $\boldsymbol{\beta}$ is endowed with a prior distribution $p(\boldsymbol{\beta})$ reflecting the knowledge or beliefs the researcher has about the parameters *a priori*. The standard approach consists of assuming a conjugate normal prior centered at $\boldsymbol{\mu}_{\boldsymbol{\beta}}$ and with covariance matrix $\boldsymbol{\Psi}_{\boldsymbol{\beta}}$:

$$\boldsymbol{\beta} \sim \mathcal{N}_p(\boldsymbol{\mu}_{\boldsymbol{\beta}};\, \boldsymbol{\Psi}_{\boldsymbol{\beta}})\,.$$

In case the choice of the prior parameters cannot be driven by any prior knowledge, a flat prior (i.e., noninformative) can be selected by setting $\boldsymbol{\mu}_{\boldsymbol{\beta}} = \mathbf{0}$ and infinite variance such that $\boldsymbol{\Psi}_{\boldsymbol{\beta}}^{-1} = \mathbf{0}$.[9] Since the distribution of $\boldsymbol{\beta}$ is derived conditional on the other parameters, the auxiliary outcome $\tilde{\boldsymbol{y}} = \boldsymbol{y}^\star - \mathbf{F}\boldsymbol{\alpha}$ is introduced to simplify the exposition. The application of Bayes' theorem provides:

$$\begin{aligned}
&p(\boldsymbol{\beta}|\boldsymbol{y}^\star, \mathbf{X}, \mathbf{F}, \boldsymbol{\alpha}, \sigma^2) = p(\boldsymbol{\beta}|\tilde{\boldsymbol{y}}, \mathbf{X}, \sigma^2),\\
&\propto p(\tilde{\boldsymbol{y}}|\mathbf{X}, \boldsymbol{\beta}, \sigma^2)p(\boldsymbol{\beta}),\\
&\propto \exp\left\{-\frac{1}{2\sigma^2}(\tilde{\boldsymbol{y}} - \mathbf{X}\boldsymbol{\beta})'(\tilde{\boldsymbol{y}} - \mathbf{X}\boldsymbol{\beta})\right\} \exp\left\{-\frac{1}{2}(\boldsymbol{\beta} - \boldsymbol{\mu}_{\boldsymbol{\beta}})'\boldsymbol{\Psi}_{\boldsymbol{\beta}}^{-1}(\boldsymbol{\beta} - \boldsymbol{\mu}_{\boldsymbol{\beta}})\right\},\\
&\propto \exp\left\{-\frac{1}{2}\left(\boldsymbol{\beta}'\left[\sigma^{-2}\mathbf{X}'\mathbf{X} + \boldsymbol{\Psi}_{\boldsymbol{\beta}}^{-1}\right]\boldsymbol{\beta} - 2\boldsymbol{\beta}'\left[\sigma^{-2}\mathbf{X}'\tilde{\boldsymbol{y}} + \boldsymbol{\Psi}_{\boldsymbol{\beta}}^{-1}\boldsymbol{\mu}_{\boldsymbol{\beta}}\right]\right)\right\},\\
&\propto \exp\left\{-\frac{1}{2}\left(\boldsymbol{\beta} - \left[\sigma^{-2}\mathbf{X}'\mathbf{X} + \boldsymbol{\Psi}_{\boldsymbol{\beta}}^{-1}\right]^{-1}\left[\sigma^{-2}\mathbf{X}'\tilde{\boldsymbol{y}} + \boldsymbol{\Psi}_{\boldsymbol{\beta}}^{-1}\boldsymbol{\mu}_{\boldsymbol{\beta}}\right]\right)'\left(\sigma^{-2}\mathbf{X}'\mathbf{X} + \boldsymbol{\Psi}_{\boldsymbol{\beta}}^{-1}\right)(\bullet)\right\},
\end{aligned}$$

where $(\bullet)$ represents the first factor of the corresponding sandwich matrix. Factors not involving $\boldsymbol{\beta}$ have been omitted, and the resulting normal kernel has been produced using the completion of the square. As a consequence, $\boldsymbol{\beta}$ has the following conditional distribution:

$$\boldsymbol{\beta}\,|\,\boldsymbol{y}^\star, \mathbf{F}, \boldsymbol{\alpha}, \sigma^2 \sim \mathcal{N}_p\Big([\sigma^{-2}\mathbf{X}'\mathbf{X} + \boldsymbol{\Psi}_{\boldsymbol{\beta}}^{-1}]^{-1}[\sigma^{-2}\mathbf{X}'\tilde{\boldsymbol{y}} + \boldsymbol{\Psi}_{\boldsymbol{\beta}}^{-1}\boldsymbol{\mu}_{\boldsymbol{\beta}}];\, \left[\sigma^{-2}\mathbf{X}'\mathbf{X} + \boldsymbol{\Psi}_{\boldsymbol{\beta}}^{-1}\right]^{-1}\Big)\,.$$

When noninformative priors are assumed, this conditional distribution reduces to:

$$\boldsymbol{\beta}\,|\,\boldsymbol{y}^\star, \mathbf{F}, \boldsymbol{\alpha}, \sigma^2 \sim \mathcal{N}_p\big([\mathbf{X}'\mathbf{X}]^{-1}\mathbf{X}'\tilde{\boldsymbol{y}};\, \sigma^2[\mathbf{X}'\mathbf{X}]^{-1}\big)\,.$$

[9] $\boldsymbol{\Psi}_{\boldsymbol{\beta}}^{-1}$ is called precision matrix.

### Updating the factor loadings

Conditional on the latent factors, there are no differences between the sampling schemes of the slope parameters and of the factor loadings. The factors can indeed be regarded as simple regressors once they have been simulated. This is why some authors sometimes call this type of model *factor regression models* (West, 2003).

Applying Bayes' rule for $\boldsymbol{\alpha}$ provides:

$$p(\boldsymbol{\alpha}|\boldsymbol{y}^{\star}, \mathbf{X}, \mathbf{F}, \boldsymbol{\beta}, \sigma^2) \propto p(\boldsymbol{y}^{\star}|\mathbf{X}, \mathbf{F}, \boldsymbol{\beta}, \boldsymbol{\alpha}, \sigma^2)p(\boldsymbol{\alpha}).$$

If the factor loadings are supposed to be a priori normally distributed:

$$\boldsymbol{\alpha} \sim \mathcal{N}_k(\boldsymbol{\mu}_{\boldsymbol{\alpha}};\, \boldsymbol{\Psi}_{\boldsymbol{\alpha}})\,,$$

their conditional distribution is:

$$\boldsymbol{\alpha} \,|\, \boldsymbol{y}^{\star}, \mathbf{F}, \boldsymbol{\beta}, \sigma^2 \sim \mathcal{N}_k\big([\sigma^{-2}\mathbf{F}'\mathbf{F} + \boldsymbol{\Psi}_{\boldsymbol{\alpha}}^{-1}]^{-1}[\sigma^{-2}\mathbf{F}'\tilde{\boldsymbol{y}} + \boldsymbol{\Psi}_{\boldsymbol{\alpha}}^{-1}\boldsymbol{\mu}_{\boldsymbol{\alpha}}];\, [\sigma^{-2}\mathbf{F}'\mathbf{F} + \boldsymbol{\Psi}_{\boldsymbol{\alpha}}^{-1}]^{-1}\big)\,,$$

where $\tilde{\boldsymbol{y}} = \boldsymbol{y}^{\star} - \mathbf{X}\boldsymbol{\beta}$.

It might be desired—or required for identification—to restrict some factor loadings. If a factor loading has to be fixed to a given value, it is simply assigned this value at each step of the Gibbs sampler. If sign constraints have to be implemented, the sampling procedure is very similar to the unrestricted case, with the only difference that the prior distribution is assumed to be truncated, resulting in a truncated normal distribution for the conditional distribution.

### Updating the idiosyncratic variance

Because the idiosyncratic variance has to be positive, a conjugate inverse Gamma distribution prior with following probability density function is usually adopted:

$$\sigma^2 \sim \mathcal{G}^{-1}(g_1;\, g_2)\,, \qquad\qquad p(\sigma^2) = \frac{(g_2)^{g_1}}{\Gamma(g_1)}(\sigma^2)^{-g_1-1}\exp\left\{-\frac{g_2}{\sigma^2}\right\},$$

where $\Gamma(\cdot)$ denotes the Gamma function. Combining the kernel of this density function to the kernel of the likelihood provides, though the application of Bayes' theorem:

$$\begin{aligned}&p(\sigma^2|\boldsymbol{y}^{\star}, \mathbf{X}, \mathbf{F}, \boldsymbol{\beta}, \boldsymbol{\alpha})\\ &\propto p(\boldsymbol{y}^{\star}|\mathbf{X}, \mathbf{F}, \boldsymbol{\beta}, \boldsymbol{\alpha}, \sigma^2)p(\sigma^2),\end{aligned}$$

$$\propto (\sigma^2)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(y_i^\star - \boldsymbol{x}_i'\boldsymbol{\beta} - \boldsymbol{f}_i'\boldsymbol{\alpha}\right)^2\right\}(\sigma^2)^{-g_1-1}\exp\left\{-\frac{g_2}{\sigma^2}\right\},$$
$$\propto (\sigma^2)^{-(\frac{n}{2}+g_1)-1}\exp\left\{-\frac{1}{\sigma^2}\left(g_2 + \frac{1}{2}\sum_{i=1}^{n}\left(y_i^\star - \boldsymbol{x}_i'\boldsymbol{\beta} - \boldsymbol{f}_i'\boldsymbol{\alpha}\right)^2\right)\right\}.$$

The kernel of an inverse Gamma distribution can be recognized in this last expression. Hence, the idiosyncratic variance has the following conditional distribution:

$$\sigma^2 \,|\, \boldsymbol{y}^\star, \mathbf{F}, \boldsymbol{\beta}, \boldsymbol{\alpha} \sim \mathcal{G}^{-1}\left(g_1 + \frac{n}{2};\, g_2 + \frac{1}{2}\sum_{i=1}^{n}\left(y_i^\star - \boldsymbol{x}_i'\boldsymbol{\beta} - \boldsymbol{f}_i'\boldsymbol{\alpha}\right)^2\right).$$

## 1.3.2 Latent response variables and cut-points

The set of observed variables $(\mathbf{Y}, \mathbf{X})$ can be augmented with the latent response variables $\mathbf{Y}^\star$, so that $(\mathbf{Y}, \mathbf{Y}^\star, \mathbf{X})$ can be directly analyzed and used to derive the conditional distributions of the other parameters more easily. Practically, this implies the inclusion of an additional step in the Gibbs sampler where these unobserved variables are simulated—Step (A4). In this section, the conditional distribution of $\boldsymbol{y}^\star$ is derived for each corresponding response variable. Using Bayes' theorem, it comes:

$$p(\boldsymbol{y}^\star|\boldsymbol{y}, \mathbf{X}, \mathbf{F}, \boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\lambda}) \propto p(\boldsymbol{y}|\boldsymbol{y}^\star, \boldsymbol{\lambda})p(\boldsymbol{y}^\star|\mathbf{X}, \mathbf{F}, \boldsymbol{\beta}, \boldsymbol{\alpha}). \tag{1.8}$$

Depending on the nature of the link function and on its specific parameters $\boldsymbol{\lambda}$, different sampling strategies will be implemented.

### Binary response variable

When the function $g(\cdot)$ linking the latent response variable $\boldsymbol{y}^\star$ to the observed response $\boldsymbol{y}$ is the indicator function $\mathbb{1}[\cdot]$, no specific parameters $\boldsymbol{\lambda}$ are required and the model is the standard probit if the idiosyncratic variance is fixed to 1, conditionally on the latent factors $\mathbf{F}$. The conditional distribution of $\boldsymbol{y}^\star$ can be derived from Equation (1.8) as follows:

$$p(\boldsymbol{y}^\star|\boldsymbol{y}, \mathbf{X}, \mathbf{F}, \boldsymbol{\beta}, \boldsymbol{\alpha}) \propto p(\boldsymbol{y}|\boldsymbol{y}^\star)p(\boldsymbol{y}^\star|\mathbf{X}, \mathbf{F}, \boldsymbol{\beta}, \boldsymbol{\alpha}),$$
$$\propto \prod_{i=1}^{n}\mathbb{1}[y_i^\star > 0]^{\boldsymbol{y}_i}\mathbb{1}[y_i^\star \leqslant 0]^{1-\boldsymbol{y}_i}\prod_{i=1}^{n}\exp\left\{-\frac{1}{2}(y_i^\star - \boldsymbol{x}_i'\boldsymbol{\beta} - \boldsymbol{f}_i'\boldsymbol{\alpha})^2\right\},$$

$$\propto \prod_{i=1}^{n} \left[ \mathbb{1}[y_i^\star > 0] \exp\left\{ -\frac{1}{2}(y_i^\star - \boldsymbol{x}_i'\boldsymbol{\beta} - \boldsymbol{f}_i'\boldsymbol{\alpha})^2 \right\} \right]^{\boldsymbol{y}_i}$$
$$\times \left[ \mathbb{1}[y_i^\star \leqslant 0] \exp\left\{ -\frac{1}{2}(y_i^\star - \boldsymbol{x}_i'\boldsymbol{\beta} - \boldsymbol{f}_i'\boldsymbol{\alpha})^2 \right\} \right]^{1-\boldsymbol{y}_i}.$$

Hence, the latent response variable is sampled from a truncated normal distribution, depending on the outcome of the observed binary variable:

$$y_i^\star \mid y_i, \boldsymbol{f}_i, \boldsymbol{\beta}, \boldsymbol{\alpha} \sim \begin{cases} \mathcal{TN}_{(0,+\infty)}\left(\boldsymbol{x}_i'\boldsymbol{\beta} + \boldsymbol{f}_i'\boldsymbol{\alpha};\, 1\right) & \text{if } y_i = 1, \\ \mathcal{TN}_{(-\infty,0]}\left(\boldsymbol{x}_i'\boldsymbol{\beta} + \boldsymbol{f}_i'\boldsymbol{\alpha};\, 1\right) & \text{if } y_i = 0. \end{cases}$$

### Ordinal response variable

The ordinal case is similar to the binary case, insofar as it is also a threshold-crossing model. However, since there are more than two different categories, cut-points have to be introduced to discretize the latent response variable. Assume that there are $L > 2$ different categories. In this configuration, these categories have a natural ordering generated by the latent response variable $\boldsymbol{y}^\star$:

$$\begin{aligned} & y_i = l \quad \text{if } \lambda_{l-1} \leqslant y_i^\star < \lambda_l, && l = 1, ..., L, \\ & y_i^\star \sim \mathcal{N}(\boldsymbol{x}_i'\boldsymbol{\beta} + \boldsymbol{f}_i'\boldsymbol{\alpha};\, 1), && i = 1, ..., n, \end{aligned} \tag{1.9}$$

where $\lambda_0 = -\infty$, $\lambda_1 = 0$, $\lambda_L = +\infty$ and $\boldsymbol{\lambda} = (\lambda_0, ..., \lambda_L)'$. When the error term is standard normally distributed, this is the usual ordered probit model, conditionally on $\mathbf{F}$. Albert and Chib (1993) have developed a simple data augmentation scheme for the normal case that we introduce in the next section.

***Sampling the latent response variable.*** Conditional on the observed choices $\boldsymbol{y}$, on the factors and on all parameters, the conditional distribution of the latent outcome is:

$$\begin{aligned} p(\boldsymbol{y}^\star|\boldsymbol{y}, \mathbf{X}, \mathbf{F}, \boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\lambda}) &\propto p(\boldsymbol{y}|\boldsymbol{y}^\star, \boldsymbol{\lambda}) p(\boldsymbol{y}^\star|\mathbf{X}, \mathbf{F}, \boldsymbol{\beta}, \boldsymbol{\alpha}), \\ &\propto \prod_{i=1}^{n} \left\{ \sum_{l=1}^{L} \mathbb{1}[y_i = l] \mathbb{1}[\lambda_{l-1} \leqslant y_i^\star < \lambda_l] \right\} \mathbb{1}[\lambda_0 < ... < \lambda_L] \\ &\qquad \times \prod_{i=1}^{n} \exp\left( -\frac{1}{2}(y_i^\star - \boldsymbol{x}_i'\boldsymbol{\beta} - \boldsymbol{f}_i'\boldsymbol{\alpha})^2 \right), \\ &\propto \prod_{i=1}^{n} \left\{ \sum_{l=1}^{L} \mathbb{1}[\lambda_{l-1} \leqslant y_i^\star < \lambda_l] \exp\left( -\frac{1}{2}(y_i^\star - \boldsymbol{x}_i'\boldsymbol{\beta} - \boldsymbol{f}_i'\boldsymbol{\alpha})^2 \right) \right\}. \end{aligned}$$

As a consequence, $\boldsymbol{y}_i^\star$ is sampled from the following truncated normal distribution:

$$y_i^\star \,|\, y_i, \boldsymbol{f}_i, \boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\lambda} \sim \mathcal{TN}_{[\lambda_{l-1},\lambda_l)}\left(\boldsymbol{x}_i'\boldsymbol{\beta} + \boldsymbol{f}_i'\boldsymbol{\alpha};\, 1\right) \text{ whenever } y_i = l.$$

***Sampling the cut-points.*** Bayes' theorem can be invoked as follows to derive the conditional distribution of the set of cut-points $\boldsymbol{\lambda}$:

$$\begin{aligned} p(\boldsymbol{\lambda}|\boldsymbol{y}, \boldsymbol{y}^\star, \mathbf{X}, \mathbf{F}, \boldsymbol{\beta}, \boldsymbol{\alpha}) &\propto p(\boldsymbol{y}|\boldsymbol{y}^\star, \boldsymbol{\lambda})p(\boldsymbol{\lambda}), \\ &\propto \prod_{i=1}^{n} \left\{ \sum_{l=1}^{L} \mathbb{1}[y_i = l]\mathbb{1}[\lambda_{l-1} \leqslant y_i^\star < \lambda_l] \right\} \mathbb{1}[\lambda_0 < \ldots < \lambda_L] p(\boldsymbol{\lambda}). \end{aligned}$$

Sampling the cut-points jointly would be tricky because of the ordering condition. Albert and Chib (1993) therefore proposed to draw them sequentially. The conditional distribution of $\lambda_l$, conditional on the other cut-points $\boldsymbol{\lambda}_{-l}$, is derived as:

$$\begin{aligned} p(\lambda_l|\boldsymbol{\lambda}_{-l}, \boldsymbol{y}, \boldsymbol{y}^\star, \mathbf{X}, \mathbf{F}, \boldsymbol{\beta}, \boldsymbol{\alpha}) \propto \prod_{i:y_i=l} \mathbb{1}[\lambda_{l-1} \leqslant y_i^\star < \lambda_l] \prod_{i:y_i=l+1} \mathbb{1}[\lambda_l \leqslant y_i^\star < \lambda_{l+1}] \\ \times \mathbb{1}[\lambda_{l-1} < \lambda_l < \lambda_{l+1}] p(\lambda_l), \end{aligned} \tag{1.10}$$

For each cut-point $\lambda_l$, a uniform prior on the interval bounded by $\lambda_{l-1}$ and $\lambda_{l+1}$ is used to fulfill the ordering condition. From Equation (1.10), it can be deduced that the conditional of each cut-point is a uniform distribution:

$$\lambda_l \,|\, y_i, y_i^\star, \boldsymbol{\lambda}_{-l} \sim \mathcal{U}\left( \max\left\{ \max_{i:y_i=l}\{y_i^\star\}\,;\, \lambda_{l-1} \right\};\, \min\left\{ \min_{i:y_i=l+1}\{y_i^\star\}\,;\, \lambda_{l+1} \right\} \right).$$

Section 1.4.1 discusses the drawbacks of this basic Gibbs sampling scheme for the cut-points, and presents alternative approaches that speed up convergence and offer a better mixing of the Markov chain.

### Censored response variable

The latent response variable in the censored case is observed only below or above a given threshold $\lambda$.[10] In the censored from below case, the response variable is observed only if it is above a given threshold $\lambda$, and the model can be expressed as:

$$y_i = \lambda\mathbb{1}[y_i^\star \leqslant \lambda] + y_i^\star\mathbb{1}[y_i^\star > \lambda], \tag{1.11}$$

[10] Another variation would be a double censoring, from above and from below (Rosett and Nelson, 1975). Derivation of the conditional distribution of the latent response variable would be straightforward.

$$y_i^\star \sim \mathcal{N}\left(\boldsymbol{x}_i'\boldsymbol{\beta} + \boldsymbol{f}_i'\boldsymbol{\alpha};\, \sigma^2\right).$$

If the response variable is censored from above, Equation (1.11) has to be replaced with $y_i = \lambda \mathbb{1}[y_i^\star \geqslant \lambda] + y_i^\star \mathbb{1}[y_i^\star < \lambda]$. Conditionally on $\mathbf{F}$, this is a variation of the typical Tobit model. Bayesian inference for this type of model is detailed in Chib (1992).

Only the case where the observed response variable is censored requires elaboration. For each individual $i$, the conditional density of the latent response variable given censoring from below can be expressed as:

$$\begin{aligned} p(y_i^\star | y_i = \lambda, \boldsymbol{x}_i, \lambda, \boldsymbol{f}_i', \boldsymbol{\beta}, \boldsymbol{\alpha}, \sigma^2) &\propto p(y_i = \lambda | y_i^\star, \lambda) p(y_i^\star | \boldsymbol{x}_i, \mathbf{F}, \boldsymbol{\beta}, \boldsymbol{\alpha}, \sigma^2), \\ &\propto \mathbb{1}[y_i^\star \leqslant \lambda] \exp\left\{ -\frac{1}{2\sigma^2} (y_i^\star - \boldsymbol{x}_i'\boldsymbol{\beta} - \boldsymbol{f}_i'\boldsymbol{\alpha})^2 \right\}. \end{aligned}$$

Hence, the latent response variable is sampled from the following truncated normal distribution when censoring is from below:

$$y_i^\star \mid \boldsymbol{f}_i, \boldsymbol{\beta}, \boldsymbol{\alpha}, \lambda \sim \mathcal{TN}_{(-\infty, \lambda]} \left(\boldsymbol{x}_i'\boldsymbol{\beta} + \boldsymbol{f}_i'\boldsymbol{\alpha};\, \sigma^2\right).$$

Similarly, in the case where the censoring is from above, $y_i^\star$ is sampled from:

$$y_i^\star \mid \boldsymbol{f}_i, \boldsymbol{\beta}, \boldsymbol{\alpha}, \lambda \sim \mathcal{TN}_{[\lambda, +\infty)} \left(\boldsymbol{x}_i'\boldsymbol{\beta} + \boldsymbol{f}_i'\boldsymbol{\alpha};\, \sigma^2\right).$$

In the next steps of the Gibbs sampler, the simulated $y_i^\star$ is used for the individuals whose response variable is censored, while the observed response variable is used for the other ones.

### 1.3.3 Latent factors and their covariance matrix

The latent factors are assumed to be multivariate normal $\boldsymbol{f}_i \sim \mathcal{N}_k(\mathbf{0};\, \boldsymbol{\Psi}_{\mathbf{F}})$. Step (B) of the Gibbs sampler updates the factor scores for each individual $i$, while Step (C) updates the covariance matrix of the latent factors conditional on their scores.

#### Conditional distribution of the latent factors

Since the latent factors are independently distributed across individuals, the conditional distribution for $\mathbf{F}$ can be factorized into $n$ conditionals for $\boldsymbol{f}_1$,...,$\boldsymbol{f}_n$. To derive them, note that all the contributions of $\boldsymbol{f}_i$ originate from linear regression models conditional on the latent response variables $\boldsymbol{y}^\star$:

$$\tilde{\boldsymbol{y}}_i^\star = \boldsymbol{\alpha} \boldsymbol{f}_i + \boldsymbol{\varepsilon}_i, \qquad \text{with } \tilde{\boldsymbol{y}}_i^\star = \boldsymbol{y}_i^\star - \boldsymbol{\beta} \boldsymbol{x}_i, \tag{1.12}$$

where $\tilde{\boldsymbol{y}}_i^\star$ is of dimension $(m \times 1)$, $\boldsymbol{\alpha}$ is the $(m \times k)$–dimensional factor loading matrix and $\boldsymbol{f}_i$ is the $k$–dimensional vector containing the factors to be updated for individual $i$. To complete the specification, remember that $\boldsymbol{\varepsilon}_i \sim \mathcal{N}_m(\mathbf{0};\, \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}})$.

The conditional distribution for $\boldsymbol{f}_i$ is then:

$$\begin{aligned} p(\boldsymbol{f}_i|\tilde{\boldsymbol{y}}_i^\star, \boldsymbol{\alpha}, \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}, \boldsymbol{\Psi}_{\mathbf{F}}) &\propto p(\tilde{\boldsymbol{y}}_i^\star|\boldsymbol{f}_i, \boldsymbol{\alpha}, \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}) p(\boldsymbol{f}_i|\boldsymbol{\Psi}_{\mathbf{F}}), \\ &\propto \exp\left\{ -\frac{1}{2} (\tilde{\boldsymbol{y}}_i^\star - \boldsymbol{\alpha}\boldsymbol{f}_i)' \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}^{-1} (\tilde{\boldsymbol{y}}_i^\star - \boldsymbol{\alpha}\boldsymbol{f}_i) \right\} \exp\left\{ -\frac{1}{2} \boldsymbol{f}_i' \boldsymbol{\Psi}_{\mathbf{F}}^{-1} \boldsymbol{f}_i \right\}, \\ &\propto \exp\left\{ -\frac{1}{2} \left( \boldsymbol{f}_i' \left[ \boldsymbol{\alpha}' \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}^{-1} \boldsymbol{\alpha} + \boldsymbol{\Psi}_{\mathbf{F}}^{-1} \right] \boldsymbol{f}_i - 2 \boldsymbol{f}_i' \boldsymbol{\alpha}' \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}^{-1} \tilde{\boldsymbol{y}}_i^\star \right) \right\}. \end{aligned}$$

Like for the slope parameters of the linear model described in Section 1.3.1, this last expression provides, after rearranging, the kernel of the following normal distribution:

$$\boldsymbol{f}_i \,|\, \boldsymbol{y}_i^\star, \boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}} \sim \mathcal{N}_k\Big( \left[ \boldsymbol{\alpha}' \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}^{-1} \boldsymbol{\alpha} + \boldsymbol{\Psi}_{\mathbf{F}}^{-1} \right]^{-1} \left[ \boldsymbol{\alpha}' \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}^{-1} \tilde{\boldsymbol{y}}_i^\star \right];\, \left[ \boldsymbol{\alpha}' \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}^{-1} \boldsymbol{\alpha} + \boldsymbol{\Psi}_{\mathbf{F}}^{-1} \right]^{-1} \Big). \tag{1.13}$$

In the uncorrelated case, this posterior distribution can be split into $k$ independent posteriors and the latent factors can be sampled one at a time, conditional on the other ones. This sequential sampling of the factors might be particularly useful when some factor loadings are restricted to a fixed value, or restricted to be positive or negative. In all other cases, sampling them jointly according to Equation (1.13) may be preferable, since this kind of block-sampling usually improves the mixing of the Markov chain (Liu et al., 1994).

### Factor covariance matrix

In the uncorrelated case, $\boldsymbol{\Psi}_{\mathbf{F}} = \operatorname{diag}(\psi_1, ..., \psi_k)$ and each single variance is sampled exactly in the same way as the idiosyncratic variances in Section 1.3.1. A conjugate inverse Gamma distribution $\mathcal{G}^{-1}(g_1;\, g_2)$ is assumed, and the conditional of the factor variances can be shown to be:

$$\psi_j \,|\, \mathbf{F} \sim \mathcal{G}^{-1}\left( g_1 + \frac{n}{2};\, g_2 + \frac{1}{2} \sum_i f_{i,j}^2 \right), \qquad j = 1, ..., k.$$

In the correlated case, the covariance matrix is assumed to be a priori Wishart distributed with $v$ degrees of freedom and inverse scale matrix $\boldsymbol{\Upsilon}$:

$$\boldsymbol{\Psi}_{\mathbf{F}} \sim \mathcal{W}^{-1}(v;\, \boldsymbol{\Upsilon}),$$

and applying Bayes' rule, the conditional distribution is derived as:[11]

$$\begin{aligned}
&p(\boldsymbol{\Psi}_{\mathbf{F}}|\mathbf{F}) \propto p(\mathbf{F}|\boldsymbol{\Psi}_{\mathbf{F}})p(\boldsymbol{\Psi}_{\mathbf{F}}),\\
&\propto |\boldsymbol{\Psi}_{\mathbf{F}}|^{-\frac{n}{2}} \exp\left\{-\frac{1}{2}\sum_i \boldsymbol{f}_i'\boldsymbol{\Psi}_{\mathbf{F}}^{-1}\boldsymbol{f}_i\right\} |\boldsymbol{\Psi}_{\mathbf{F}}|^{-\frac{v+k+1}{2}} \exp\left\{-\frac{1}{2}\mathrm{tr}\big(\boldsymbol{\Upsilon}\boldsymbol{\Psi}_{\mathbf{F}}^{-1}\big)\right\},\\
&\propto |\boldsymbol{\Psi}_{\mathbf{F}}|^{-\frac{n+v+k+1}{2}} \exp\left\{-\frac{1}{2}\left[\sum_i \boldsymbol{f}_i'\boldsymbol{\Psi}_{\mathbf{F}}^{-1}\boldsymbol{f}_i + \mathrm{tr}\big(\boldsymbol{\Upsilon}\boldsymbol{\Psi}_{\mathbf{F}}^{-1}\big)\right]\right\},\\
&\propto |\boldsymbol{\Psi}_{\mathbf{F}}|^{-\frac{n+v+k+1}{2}} \exp\left\{-\frac{1}{2}\left[\mathrm{tr}\left(\left[\sum_i \boldsymbol{f}_i\boldsymbol{f}_i' + \boldsymbol{\Upsilon}\right]\boldsymbol{\Psi}_{\mathbf{F}}^{-1}\right)\right]\right\}.
\end{aligned}$$

Hence, the conditional of the covariance matrix is an inverse Wishart distribution:

$$\boldsymbol{\Psi}_{\mathbf{F}}\,|\,\mathbf{F} \sim \mathcal{W}^{-1}\left(n+v;\, \sum_i \boldsymbol{f}_i\boldsymbol{f}_i' + \boldsymbol{\Upsilon}\right).$$

## 1.4 Accelerating convergence and improving mixing of the Markov chain

The Gibbs sampler described in the previous section generates a posterior sample of the parameters of interest that can be used to compute different statistics and conduct a battery of tests, depending on the goal of the researcher. However, two problems usually affect Bayesian inference relying on Markov chain Monte Carlo methods.

The first one pertains to the convergence of the chain to its stationary distribution. To tackle this problem, it is common practice to use several chains with different starting values, and to discard the first iterations of the Gibbs sampler (the so-called 'burn-in' period). If the different chains converge to the same distribution, it is reasonable to consider it as the stationary distribution and to keep the last draws of the parameters for posterior analysis.

[11]The result is obtained using a basic property of the trace operator: $\mathrm{tr}(ABC) = \mathrm{tr}(CAB)$, where $A$, $B$ and $C$ are matrices of appropriate sizes.

The second problem concerns the mixing of the Markov chain. Because of the iterative design of the Gibbs sampler, the chains of parameters are autocorrelated. The higher the autocorrelation, the slower the algorithm manages to explore the whole space of the posterior distribution. Since in practice only a finite sample can be simulated from the posterior distribution, this might be a problem to conduct posterior analysis. For instance, standard errors of the parameters of interest might be overinflated because of this autocorrelation. In empirical analysis, it is very current to *thin* the chain, i.e., to keep every $t^{\text{th}}$ iteration of the sampler to provide a posterior distribution with an apparent low autocorrelation. This is, however, circumventing the problem rather than solving it, insofar as longer chains are required and a lot of draws are thus wasted. In large data sets and/or large models with many parameters, this strategy is not attractive given the computational costs it engenders.

Factor structure models are particularly affected by these problems, like all latent class models in general. Because of the data augmentation scheme, the latent variables are usually highly autocorrelated with the other parameters. Slow convergence and poor mixing follow, and plague many analyses. In our framework, the problem is even twofold, since the latent response variables and the latent factors are simulated. Section 1.4.1 focuses on the convergence problems arising because of the correlation between the latent response variables and the cut-points in the ordinal case, and reviews the different solutions proposed in the literature. Section 1.4.2 deals with the autocorrelation of the factor loadings, and introduces a reparameterization of the model speeding up convergence and greatly improving the mixing of the chains, namely the parameter-expanded Gibbs sampler (Liu and Wu, 1999).

### 1.4.1 The problem of the cut-points

The standard Gibbs sampler introduced by Albert and Chib (1993) in Section 1.3.2 usually exhibits low convergence rate, due to the high correlation between the latent response variables and the cut-points. This problem was first noted by Cowles (1996), who suggested to jointly sample the latent response variables and the cut-points through a Hastings-within-Gibbs step. Nevertheless, this procedure has its own limitations, and alternative approaches relying on a reparameterization of the model or on a transformation of the parameters have been proposed.

### Hastings-within-Gibbs step

Instead of drawing the cut-points and the latent response variables individually from their respective full conditionals, Cowles (1996) proposed to sample them jointly through a Hastings-within-Gibbs step. To do so, a set of candidate cut-points $\tilde{\lambda}$ is drawn from a marginal proposal distribution, and then accepted with a given probability ensuring that the Markov chain converges to its equilibrium distribution:

1. *Sample proposal cut-points.* A normal prior with user-specified variance $\sigma_\lambda^2$ is assumed for each cut-point. At iteration $(t)$ of the Gibbs sampler, each candidate cut-point $\tilde{\lambda}_l$ is drawn from $\mathcal{N}\left(\lambda_l^{(t-1)};\, \sigma_\lambda^2\right)$ truncated to the interval $\left[\tilde{\lambda}_{l-1}; \lambda_{l+1}^{(t-1)}\right]$ for $l = 1, ..., L$.

2. *Compute acceptance probability $R$.* With the set of candidate cut-points in hand, the acceptance probability is computed as $\min\{R, 1\}$, with:[12]

$$R = \prod_{l=2}^{L-1} \frac{\Phi\left(\sigma_\lambda^{-1}\left[\lambda_{l+1}^{(t-1)} - \lambda_l^{(t-1)}\right]\right) - \Phi\left(\sigma_\lambda^{-1}\left[\tilde{\lambda}_{l-1} - \lambda_l^{(t-1)}\right]\right)}{\Phi\left(\sigma_\lambda^{-1}\left[\tilde{\lambda}_{l+1} - \tilde{\lambda}_l\right]\right) - \Phi\left(\sigma_\lambda^{-1}\left[\lambda_{l-1}^{(t-1)} - \tilde{\lambda}_l\right]\right)} \times \prod_{i=1}^{n} \frac{\Phi\left(\tilde{\lambda}_{y_i} - \mu_i\right) - \Phi\left(\tilde{\lambda}_{y_i-1} - \mu_i\right)}{\Phi\left(\lambda_{y_i}^{(t-1)} - \mu_i\right) - \Phi\left(\lambda_{y_i-1}^{(t-1)} - \mu_i\right)},$$

   where $\mu_i = \boldsymbol{x}_i'\boldsymbol{\beta} + \boldsymbol{f}_i'\boldsymbol{\alpha}$ is the linear predictor of $y_i^\star$ for individual $i$, and $\Phi(\cdot)$ represents the cumulative distribution function of the standard normal distribution.

3. *Accept or reject candidate $\tilde{\lambda}$.* With probability $\min\{R, 1\}$, set $\lambda^{(t)} = \tilde{\lambda}$ and update the latent outcome $\boldsymbol{y}^\star$ accordingly. Otherwise, the cut-points as well as the latent outcome of the previous iteration are kept: $\lambda^{(t)} \equiv \lambda^{(t-1)}$ and $\boldsymbol{y}^{*(t)} \equiv \boldsymbol{y}^{*(t-1)}$.

To implement this procedure, the variance of the proposal distribution has to be chosen for the cut-points, influencing in the end the acceptance ratio. A rule of thumb consists of choosing $\sigma_\lambda^2$ such that the desirable acceptance rate lies between 25% and 50%.

This Hastings step represents a significant improvement of the standard algorithm, but selecting the tuning parameter $\sigma_\lambda^2$ is not trivial, especially in large models with multiple ordinal response variables where tuning parameters have to be chosen

[12] If the first cut-point is not restricted to zero—when there is no intercept term for instance—, the first product in the expression of $R$ starts at $l = 1$.

for each of them. Moreover, the truncated normal might not represent an appropriate proposal distribution for $\sigma^2_\lambda$, because it is not spread out enough (Nandram and Chen, 1996). Due to these potential drawbacks, alternative methods have been proposed and will now be considered.

### A useful reparameterization

In the threshold-crossing model used for the ordinal response variables, the scale of the latent response variable $\boldsymbol{y}^\star$ and of the cut-points $\boldsymbol{\lambda}$ does not matter, as long as the ordering is preserved and each individual falls into the right interval corresponding to the category she belongs to. This is actually one of the identification problems that is usually tackled by fixing the variance of the error term to 1, so as to prevent any scale transformation. Nandram and Chen (1996) exploit this feature of the ordinal model, and suggest a simple reparameterization that can potentially speed up the convergence of the algorithm in a spectacular way. Chen and Dey (2000) further elaborate this transformation for the case of correlated ordinal data models.

This reparameterization consists of rescaling the overall model by the largest finite cut-point $\lambda_{L-1}$ as follows:

$$\delta = 1/\lambda_{L-1}, \qquad \tilde{\boldsymbol{\lambda}} = \delta\boldsymbol{\lambda}, \qquad \tilde{\boldsymbol{y}}^\star = \delta\boldsymbol{y}^\star, \qquad \tilde{\boldsymbol{\beta}} = \delta\boldsymbol{\beta}, \qquad \tilde{\boldsymbol{\alpha}} = \delta\boldsymbol{\alpha}, \qquad \tilde{\boldsymbol{\varepsilon}} = \delta\boldsymbol{\varepsilon}. \tag{1.14}$$

With these transformed parameters and latent variables in hand, the original model shown in Equation (1.9) becomes:

$$\begin{aligned} y_i &= l \quad \text{if } \tilde{\lambda}_{l-1} \leqslant \tilde{y}^\star_i < \tilde{\lambda}_l, && l = 1, ..., L, \\ \tilde{y}^\star_i &\sim \mathcal{N}\left(\boldsymbol{x}'_i\tilde{\boldsymbol{\beta}} + \boldsymbol{f}'_i\tilde{\boldsymbol{\alpha}};\, \delta^2\right), && i = 1, ..., n, \end{aligned}$$

where the set of cut-points $\boldsymbol{\lambda}$ is such that $\tilde{\lambda}_0 = -\infty < \tilde{\lambda}_1 = 0 < \tilde{\lambda}_2 < ... < \tilde{\lambda}_{L-2} < \tilde{\lambda}_{L-1} = 1 < \tilde{\lambda}_L = +\infty$. The transformed model is *observationally equivalent* to the initial one, in that it provides the same outcomes and it does not change the likelihood. But then, what are the benefits of this reparameterization?

First, note that in the three-category case, the problem of autocorrelation between the latent response variables and the cut-points pointed out by Cowles (1996) vanishes, since the two cut-points are fixed to 0 and 1 in the transformed model. However, when the observed response variable has more than tree categories, the intermediate cut-points $\tilde{\lambda}_l$, $l = 2, ..., L-2$, have to be sampled, and the benefit of this approach is not obvious.

To implement the reparameterization, the design of the Gibbs sampler presented in Section 1.2.4 has to be adapted. Steps (A1), (A2) and (A4) remain unchanged, but the sampled parameters and latent response variables now have a different interpretation than in the original model, because of the rescaling. While Step (A3) is skipped in the original model, because the variance of the error term is set to 1, it now has to be inserted in the sampling scheme exactly as for the idiosyncratic variance in Section 1.3.1. Assuming an inverse Gamma distribution $\mathcal{G}^{-1}(g_1;\, g_2)$ for its prior, the conditional distribution of $\delta^2$ is:

$$\delta^2 \,|\, \boldsymbol{y}^\star, \boldsymbol{f}_i, \boldsymbol{\beta}, \boldsymbol{\alpha} \sim \mathcal{G}^{-1}\left( g_1 + \frac{n}{2};\, g_2 + \frac{1}{2} \sum_{i=1}^{n} \left( \tilde{y}_i^\star - \boldsymbol{x}_i' \tilde{\boldsymbol{\beta}} - \boldsymbol{f}_i' \tilde{\boldsymbol{\alpha}} \right)^2 \right).$$

As for the cut-points in the case where the response variable has more than three categories, they have to be sequentially updated using an appropriate sampling scheme. Nandram and Chen (1996) suggest to use a Hastings-within-Gibbs step similar to the one introduced by Cowles (1996), but relying on a Dirichlet distribution instead of the truncated normal distribution for the proposal distribution. If the analyst is interested in the parameters of the initial model, the sampled parameters are transformed back to their original definition using the relations in Equation (1.14).

This simple transformation of the model makes it possible to achieve faster convergence and smaller autocorrelations of the parameters. The intuition behind this result is pretty straightforward to understand. As aforementioned, the slow convergence is due to the high autocorrelation between the latent response variable and the cut-points. This reparameterization relaxes the constraint on the variance of the error term, at the cost of some new restrictions on the cut-points to ensure that the transformed model is equivalent to the original one. Hence, by shifting the constraints from the idiosyncratic variance to the cut-points that are responsible for the slow convergence, the algorithm is allowed to navigate faster through the parameter space.

As pointed out by Li and Tobias (2007), the only potential drawback of this reparameterization concerns the prior placed on the scaling parameter $\delta$.[13] It can indeed imply prior distributions for the parameters of the original model that are at odd with the prior knowledge or beliefs of the researcher. Sensitivity analysis should therefore be carried out to further investigate how the choice of the parameters of the prior distributions influences the final estimates.

[13]More precisely, the prior is placed on the square of the scaling parameter $\delta^2$, because of the sampling of the idiosyncratic variance in the transformed model.

### Group transformation

This method introduces an intermediate step in the Gibbs sampler that is far less demanding than the Hastings step and demonstrates good convergence performance. It consists of finding an appropriate group transformation of some parameters—possibly all of them—which does not change their target distribution. Hence, the Gibbs sampler remains the same, but its mixing behavior is likely to be greatly improved. Details and theoretical foundations of this methodology are provided in the original paper by Liu and Sabatti (2000), and explained in Raach (2006) and Fahrmeir and Raach (2007). The standard Gibbs sampler described in Section 1.3.2 is used to sample the cut-points, and the transformation is then applied to the parameters of interest.

For our purposes, we consider the partial scale group transformation:

$$\boldsymbol{\Gamma}_v = \{\gamma(\boldsymbol{\omega}) = (\gamma\omega_1, ..., \gamma\omega_v, \omega_{v+1}, ..., \omega_W), \gamma > 0\},$$

where only the first $v$ elements of the $W$–dimensional vector of parameters $\boldsymbol{\omega}$ are transformed, the others remaining unchanged. Applying the first theorem of Liu and Sabatti (2000), the goal is to sample a suitable parameter $\gamma$ from a density proportional to $\gamma^{v-1}p(\gamma(\boldsymbol{\omega}))$, conditional on $\boldsymbol{\omega}$, where $p(\cdot)$ represents the density of the parameters to be transformed. In our case, a transformation is applied on the group of parameters and latent response variables:

$$\boldsymbol{\omega} = \{y_1^\star, ..., y_n^\star, \beta_1, ..., \beta_p, \alpha_1, ..., \alpha_k, \lambda_1, ..., \lambda_{L-1}\},$$

which contains $n + p + k + L - 1$ elements, where $p$ is the number of covariates and $k$ the number of latent factors. The conditional distribution of this group of parameters is proportional to:

$$\prod_{i=1}^{n} \left[ p(y_i^\star|\boldsymbol{f}_i, \boldsymbol{x}_i, \boldsymbol{\beta}, \boldsymbol{\alpha}) \sum_{l=1}^{L} \mathbb{1}[y_i = l]\mathbb{1}[\lambda_{l-1} \leqslant y_i^\star < \lambda_l] \right] p(\boldsymbol{\beta}, \boldsymbol{\alpha})p(\boldsymbol{\lambda}). \tag{1.15}$$

To simplify the exposition, noninformative priors are assumed for the slope parameters $\boldsymbol{\beta}$ and the cut-points $\boldsymbol{\lambda}$, and therefore their priors are just normalizing constants that disappear from the conditional distribution of $\gamma$. The factor loadings $\boldsymbol{\alpha}$ are assumed to be a priori normally distributed with mean zero and finite covariance matrix $\boldsymbol{\Psi}_{\boldsymbol{\alpha}}$, and the kernel of their prior therefore remains in the expression.[14] The

[14]Note that Raach (2006) and Fahrmeir and Raach (2007) implicitly assume noninformative priors also for $\boldsymbol{\alpha}$, since the kernel of its prior distribution vanishes in their formula. The more general case with informative priors on $\boldsymbol{\beta}$ would be straightforward to derive.

target density of the scale parameter can thus be derived as:

$$\gamma^{v-1}p(\gamma(\boldsymbol{\omega}))|\boldsymbol{\omega} \propto \gamma^{v-1}\exp\left\{-\frac{1}{2}\sum_{i=1}^{n}(\gamma y_i^\star - \gamma \boldsymbol{x}_i'\boldsymbol{\beta} - \gamma \boldsymbol{f}_i'\boldsymbol{\alpha})^2\right\}\exp\left\{-\frac{1}{2}(\gamma\boldsymbol{\alpha})'\boldsymbol{\Psi}_{\boldsymbol{\alpha}}^{-1}(\gamma\boldsymbol{\alpha})\right\},$$
$$\propto (\gamma^2)^{\frac{v+1}{2}-1}\exp\left\{-\frac{1}{2}\gamma^2\left[\sum_{i=1}^{n}(y_i^\star - \boldsymbol{x}_i'\boldsymbol{\beta} - \boldsymbol{f}_i'\boldsymbol{\alpha})^2 + \boldsymbol{\alpha}'\boldsymbol{\Psi}_{\boldsymbol{\alpha}}^{-1}\boldsymbol{\alpha}\right]\right\}. \tag{1.16}$$

Note that the sum in the expression in brackets in Equation (1.15) can be dropped because it remains constant after the transformation:

$$\sum_{l=1}^{L}\mathbb{1}[y_i = l]\mathbb{1}[\gamma\lambda_{l-1} \leqslant \gamma y_i^\star < \gamma\lambda_l] = \sum_{l=1}^{L}\mathbb{1}[y_i = l]\mathbb{1}[\lambda_{l-1} \leqslant y_i^\star < \lambda_l].$$

From Equation (1.16), it can be seen that the square of the transformation parameter $\gamma^2$ has to be sampled from a Gamma distribution with shape parameter $a$ and inverse scale $b$:

$$a = \frac{v+1}{2} = \frac{n+p+k+L}{2} \quad \text{and} \quad b = \frac{\sum_{i=1}^{n}(y_i^\star - \boldsymbol{x}_i'\boldsymbol{\beta} - \boldsymbol{f}_i'\boldsymbol{\alpha})^2 + \boldsymbol{\alpha}'\boldsymbol{\Psi}_{\boldsymbol{\alpha}}^{-1}\boldsymbol{\alpha}}{2}.$$

The transformation of interest is then carried out by multiplying each element of the group $\boldsymbol{\omega}$ by the square root of the sampled $\gamma^2$.[15]

In the special case where at least one factor loading is restricted to a constant term, the target density cannot be factorized into Equation (1.16) as previously, and as a consequence $\gamma^2$ cannot be sampled from the same Gamma distribution. To solve this problem, the group transformation has to be changed in such a way that the restricted factor loadings are not scaled by $\gamma$. We do not consider the case where a factor loading is restricted to zero, because the corresponding factor simply vanishes from the distribution and the general case described before can be applied.

### 1.4.2 The parameter-expanded Gibbs sampler

Latent variables are responsible for some inertia in the Markov chain, because of their high correlation with the parameters they are directly related to. In the case of the latent response variables $\mathbf{Y}^\star$ discussed in the previous section, the cut-points $\boldsymbol{\lambda}$ are directly affected by this problem and show slow convergence in the standard

[15]Remember that $\gamma$ has to be positive in the chosen transformation group $\boldsymbol{\Gamma}$.

Gibbs sampler. Concerning the latent factors $\mathbf{F}$, the same issue arises for the factor loadings that turn out to be also highly autocorrelated.

A promising approach to accelerate the convergence of the factor loadings is the parameter-expanded Gibbs sampler (PX-Gibbs). This idea dates back to Liu et al. (1998) who, in a classical approach, used parameter expansion to accelerate the convergence of the EM-algorithm. From a Bayesian perspective, Liu and Wu (1999) exploit the same concept for the Gibbs sampler. More recently, Ghosh and Dunson (2009) and Frühwirth-Schnatter and Lopes (2009) have successfully applied this methodology to factor models. This section sketches its main features and explains the intuition behind the PX-Gibbs sampler. The theoretical foundations of the method can be found in Liu and Wu (1999).

### The principle

To achieve identification of the latent part of the model, the scale of the latent factors has to be set through the normalization either of the factor variances, or of some factor loadings (see Section 1.2.2). Besides this *inferential model*, which is identified and straightforward to estimate but exhibits slow convergence in most cases, the PX-Gibbs sampler introduces a *working model*, where these restrictions are relaxed: both the factor loadings and the factor variances are free, resulting in an over-parameterized model that is clearly non-identified. However, this working model is only used as a convenient tool to improve convergence, and a simple transformation of the parameters makes it possible to recover the parameters of the inferential model.

Let us first consider the basic framework with uncorrelated standard normal factors and the factor loading matrix defined in Equation (1.4) with positive diagonal elements. In the inferential model, $\mathbf{\Psi_F} = \mathbf{I}_k$ and $\alpha_{j,j} > 0$, $j = 1, ..., k$. The expanded model used by the PX-Gibbs is defined as:

$$\begin{aligned}
&\boldsymbol{y}_i = g(\boldsymbol{y}_i^\star; \boldsymbol{\lambda}), && \\
&\boldsymbol{y}_i^\star = \boldsymbol{x}_i'\boldsymbol{\beta} + \tilde{\boldsymbol{f}}_i{}'\tilde{\boldsymbol{\alpha}} + \boldsymbol{\varepsilon}_i, && \boldsymbol{\varepsilon}_i \sim \mathcal{N}_m(\mathbf{0};\, \mathbf{\Sigma}_{\boldsymbol{\varepsilon}}), \\
&\tilde{\boldsymbol{f}}_i \sim \mathcal{N}_k\Big(\mathbf{0};\, \tilde{\mathbf{\Psi}}_{\mathbf{F}}\Big), && \tilde{\mathbf{\Psi}}_{\mathbf{F}} = \operatorname{diag}\Big(\tilde{\psi}_1, ..., \tilde{\psi}_k\Big).
\end{aligned} \tag{1.17}$$

In this working model, over-parameterization appears because the diagonal elements of $\boldsymbol{\alpha}$ and the factor variances $\tilde{\psi}_j$ are mutually redundant. The working model can then be transformed back to the inferential model as follows:

$$\boldsymbol{\alpha} = \tilde{\boldsymbol{\Psi}}_{\mathbf{F}}^{1/2}\mathcal{S}(\tilde{\boldsymbol{\alpha}})\tilde{\boldsymbol{\alpha}}, \qquad \boldsymbol{f}_i = \tilde{\boldsymbol{\Psi}}_{\mathbf{F}}^{-1/2}\mathcal{S}(\tilde{\boldsymbol{\alpha}})\tilde{\boldsymbol{f}}_i, \tag{1.18}$$

where $\mathcal{S}(\tilde{\boldsymbol{\alpha}})$ denotes a matrix used to restore the sign constraints on the diagonal elements of the factor loading matrix and is defined as $\mathcal{S}(\tilde{\boldsymbol{\alpha}}) = \text{diag}\left(s(\tilde{\alpha}_{1,1}), ..., s(\tilde{\alpha}_{k,k})\right)$, with $s(\alpha) = 1$ if $\alpha \geqslant 0$ and $s(\alpha) = -1$ if $\alpha < 0$.

In practice, the Gibbs sampler outlined in Section 1.2.4 is implemented exactly in the same way as before, with the only difference that Step (C) has to be inserted to sample factor variances in $\tilde{\boldsymbol{\Psi}}_{\mathbf{F}}$. Ex post, the inferential parameters of interest are easily calculated using Equation (1.18). The Markov chain will have a poor mixing behavior in the working model—a direct consequence of the lack of identification—but when the parameters are transformed back to the inferential model, the mixing appears to be greatly improved. Thereby, the PX-Gibbs sampling scheme does not involve any complications and is straightforward to implement. The small additional computational costs it implies are clearly outweighed by the benefits it provides.

Nevertheless, the transformation implemented by the PX-Gibbs sampler is not innocuous for the induced prior of the factor loadings in the inferential model. Ghosh and Dunson (2009) emphasize that in the general case, the PX-Gibbs sampler induces a Student-$t$ prior distribution for the off-diagonal elements and a half-$t$ distribution (i.e., the absolute value of a Student-$t$ distribution centered at zero, see Gelman, 2006) for the diagonal elements of the factor loading matrix. Moreover, prior dependence between the elements within the columns of $\boldsymbol{\alpha}$ is also implied in the inferential model,[16] whereas in the working model prior independence prevails, as it is usually the case in the standard factor model. However, these induced priors are generally reasonable, inasmuch as the prior dependence within the columns of the factor loading matrix reflects the fact that factors with low variance imply inflated loadings.

### Further extensions

So far, only the standard factor model with uncorrelated standard normal factors has been considered in the literature (Ghosh and Dunson, 2009; Frühwirth-Schnatter and Lopes, 2009). When the latent factors are assumed to be correlated, some adjustments are required but the PX-Gibbs sampler fundamentally remains the same. More specifically, assume that in the initial model, identification is achieved

[16] This can be seen from Equation (1.18) where each column $\boldsymbol{\alpha}_{\cdot j}$ of the factor loading matrix is multiplied by $\tilde{\psi}_j^{1/2}$ and adjusted to be positive, for $j = 1, .., k$.

by restricting $\boldsymbol{\Psi}_{\mathbf{F}}$ to be a correlation matrix.[17] In the working model of the PX-Gibbs sampler, the diagonal elements of $\tilde{\boldsymbol{\Psi}}_{\mathbf{F}}$ are free and this covariance matrix is sampled using an inverse Wishart prior distribution as described in Section 1.3.3. The transformation of the working parameters back to the inferential parameters is then achieved through:

$$\mathbf{D} = \text{diag}\left(\tilde{\psi}_{1,1}, ..., \tilde{\psi}_{k,k}\right), \qquad \boldsymbol{\alpha} = \mathbf{D}^{1/2}\tilde{\boldsymbol{\alpha}}, \qquad \boldsymbol{f}_i = \mathbf{D}^{-1/2}\tilde{\boldsymbol{f}}_i,$$

where $\tilde{\psi}_{j,j}$ are the diagonal elements of $\tilde{\boldsymbol{\Psi}}_{\mathbf{F}}$. This procedure offers the additional advantage of restoring the conjugacy of the prior of the covariance matrix. With this transformation, the standard inverse Wishart prior can indeed be used without further complications. The initial configuration of the model would have required the sampling of a correlation matrix, which is far from being trivial because of the fixed diagonal elements and the positive-definiteness constraint. These constraints preclude the use of a simple conjugate inverse Wishart prior. In practice, sophisticated Metropolis-Hastings algorithms have been developed to cope with this problem (Liu, 2008), but the PX-Gibbs sampling scheme makes it possible to bypass this drawback.

The PX-Gibbs sampler represents a step forward in speeding up convergence and improving mixing, but in our opinion, its potential has not be fully exploited yet. For instance, the working model designed for the standard case in Equation (1.17) could be made more flexible by specifying a covariance matrix with non-zero off-diagonal elements instead of a diagonal covariance matrix. The transformation to recover the inferential parameters would then be performed using the Cholesky decomposition of $\tilde{\boldsymbol{\Psi}}_{\mathbf{F}}$, in the same fashion as in Section 1.2.2 where we showed that a model with correlated factors and a lower diagonal structure of the factor loading matrix is not identified and observationally equivalent to a model with uncorrelated factors. We expect this configuration of the PX-Gibbs sampler to provide even faster convergence than in Ghosh and Dunson (2009). However, this transformation would induce prior independence between all the elements of the factor loading matrix, which might not be harmless. Further investigations are therefore required to better understand the behavior of the PX-Gibbs sampler in these different contexts. We leave these open questions for future research.

[17] The alternative case where the covariance matrix of the factors is unrestricted and some factor loadings of dedicated response variables are fixed would be straightforward to derive.

## 1.5 Toward a semiparametric approach: relaxing normality assumptions

The normality assumptions of Equations (1.2) and (1.3) are quite restrictive and are likely to introduce some bias in the estimation of the parameters. For instance, if the distribution of the latent traits supposed to be captured by the factors is non-symmetric, imposing a standard normal prior will lead to inconsistent estimates of the factor loadings. Several empirical studies have for example shown that the distribution of personality traits can be highly skewed, reflecting uneven disparities in the population (Hansen et al., 2004; Cunha et al., 2005).

To introduce more flexibility, mixtures of normals have been proposed as an alternative to the standard normality assumption (e.g., Carneiro et al., 2003). Since the seminal work of Ferguson (1983), mixtures have been extensively used for semi-parametric density estimation (Diebolt and Robert, 1994; Escobar and West, 1995; Roeder and Wasserman, 1997). In this section, we show how the Gibbs sampler can be adapted when mixtures of normals are specified for the latent factors and the error terms. Section 1.5.1 describes the sampling process of the latent factors under the mixture prior, while Section 1.5.2 derives the different steps required to update the parameters of the mixture. Section 1.5.3 finally explains how mixtures of normals can be introduced for the error terms.

Identification issues related to the introduction of mixtures will not be addressed in this section, but a few points are worth mentioning to highlight the problems at stake. From a parametric point of view, only the first two moments of the distribution of the latent factors can be recovered from the covariance structure of the model, which is unfortunately not sufficient to identify the full normal mixture distribution. When the model includes a set of continuous response variables large enough to identify the distribution of the factors, the problem will not arise. From a nonparametric perspective, Carneiro et al. (2003) show how identification of the distribution of the latent factors and of the error terms can be achieved when the model consists of a combination of continuous and discrete outcomes. In a first stage, they nonparametrically identify the joint distribution of the observed and latent response variables, where for the latter, identification is only possible up to scale. With this distribution in hand, they invoke in a second stage a theorem proposed by Kotlarski (1967) to factor analyze this joint distribution, in order to achieve nonparametric identification of the distribution of the latent factors and of the error terms. Nonetheless, in the case of discrete response variables, the first stage is not straightforward to implement and requires very strong distributional and support assumptions that can be problematic to verify in practice. Moreover, the question of identification remains open when only discrete response variables are available, and calls for further investigation.

### 1.5.1 Mixture of normals as a flexible alternative for the distribution of the latent factors

The factors are assumed to follow a mixture of $q$ multivariate normal distributions:

$$\boldsymbol{f}_i \sim \sum_{c=1}^{q} \pi_c \, \mathcal{N}_k(\boldsymbol{\mu}_c;\ \boldsymbol{\Psi}_c)\,, \tag{1.19}$$

where $\boldsymbol{\mu}_c$, $\boldsymbol{\Psi}_c$ and $\pi_c$ represent the mean vector, the precision matrix and the weight of the $c^{\text{th}}$ mixture component, respectively. For identification purpose, their overall mean is restricted to zero as in the normal case:

$$\sum_{c=1}^{q} \pi_c \, \boldsymbol{\mu}_c = 0. \tag{1.20}$$

Alternatively, the means of the mixture components could all be restricted to zero to achieve identification. However, these constraints would prevent the possibility of capturing asymmetric and multimodal distributions, and would thus make the use of mixtures of normals far less attractive. Constraining the mean of the mixture to zero does not make the sampling scheme of the factors any more complicated, but when it comes to the update of the component means, the problem is not trivial and some refinements are required (Section 1.5.2).

The conditional distribution of the latent factors is derived in the same manner as in the normal case presented in Section 1.3.3, except that the prior is replaced with the density of the normal mixture and thus requires some more elaboration:

$$\begin{aligned}
p(\boldsymbol{f}_i|\tilde{\boldsymbol{y}}_i^{\star}, \boldsymbol{\alpha}, \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}) &\propto p(\tilde{\boldsymbol{y}}_i^{\star}|\boldsymbol{f}_i, \boldsymbol{\alpha}, \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}) p(\boldsymbol{f}_i),\\
&\propto \exp\left\{-\frac{1}{2}(\tilde{\boldsymbol{y}}_i^{\star} - \boldsymbol{\alpha}\boldsymbol{f}_i)'\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}^{-1}(\tilde{\boldsymbol{y}}_i^{\star} - \boldsymbol{\alpha}\boldsymbol{f}_i)\right\}\\
&\qquad \times \sum_{c=1}^{q} \pi_c |\boldsymbol{\Psi}_c|^{-1/2} \exp\left\{-\frac{1}{2}(\boldsymbol{f}_i - \boldsymbol{\mu}_c)'\boldsymbol{\Psi}_c^{-1}(\boldsymbol{f}_i - \boldsymbol{\mu}_c)\right\},\\
&\propto \sum_{c=1}^{q} \pi_c |\boldsymbol{\Psi}_c|^{-1/2} \exp\Big\{-\frac{1}{2}\left(\boldsymbol{f}_i'\left[\boldsymbol{\alpha}'\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}^{-1}\boldsymbol{\alpha} + \boldsymbol{\Psi}_c^{-1}\right]\boldsymbol{f}_i\right.\\
&\qquad \left.-2\boldsymbol{f}_i'\left[\boldsymbol{\alpha}'\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}^{-1}\tilde{\boldsymbol{y}}_i^{\star} + \boldsymbol{\Psi}_c^{-1}\boldsymbol{\mu}_c\right] + \boldsymbol{\mu}_c'\boldsymbol{\Psi}_c^{-1}\boldsymbol{\mu}_c\right)\Big\},
\end{aligned}$$

$$\propto \sum_{c=1}^{q} \pi_c |\boldsymbol{\Psi}_c|^{-1/2} \exp\left\{-\frac{1}{2}\left(\boldsymbol{\mu}_c'\boldsymbol{\Psi}_c^{-1}\boldsymbol{\mu}_c - \left[\boldsymbol{\alpha}'\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}^{-1}\tilde{\boldsymbol{y}}_i^{\star} + \boldsymbol{\Psi}_c^{-1}\boldsymbol{\mu}_c\right]' \left[\boldsymbol{\alpha}'\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}^{-1}\boldsymbol{\alpha} + \boldsymbol{\Psi}_c^{-1}\right]^{-1}[\bullet]\right)\right\}$$
$$\times \exp\left\{-\frac{1}{2}\left(\boldsymbol{f}_i - \left[\boldsymbol{\alpha}'\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}^{-1}\boldsymbol{\alpha} + \boldsymbol{\Psi}_c^{-1}\right]^{-1}\left[\boldsymbol{\alpha}'\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}^{-1}\tilde{\boldsymbol{y}}_i^{\star} + \boldsymbol{\Psi}_c^{-1}\boldsymbol{\mu}_c\right]\right)' \left(\boldsymbol{\alpha}'\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}^{-1}\boldsymbol{\alpha} + \boldsymbol{\Psi}_c^{-1}\right)(\bullet)\right\}, \tag{1.21}$$

where $(\bullet)$ represents the first factor of the corresponding sandwich matrix. The kernel of a mixture of multivariate normal distributions emerges from Equation (1.21). To further simplify its expression, the first exponential can be factorized with respect to $\boldsymbol{\mu}_c$, and after some algebra it can be shown that:[18]

$$\exp\left\{-\frac{1}{2}\left(\boldsymbol{\mu}_c'\boldsymbol{\Psi}_c^{-1}\boldsymbol{\mu}_c - \left[\boldsymbol{\alpha}'\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}^{-1}\tilde{\boldsymbol{y}}_i^{\star} + \boldsymbol{\Psi}_c^{-1}\boldsymbol{\mu}_c\right]' \left[\boldsymbol{\alpha}'\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}^{-1}\boldsymbol{\alpha} + \boldsymbol{\Psi}_c^{-1}\right]^{-1}[\bullet]\right)\right\} \tag{1.22}$$
$$\propto \exp\left\{-\frac{1}{2}\left(\boldsymbol{\mu}_c - \left[\boldsymbol{\alpha}'\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}^{-1}\boldsymbol{\alpha}\right]^{-1}\boldsymbol{\alpha}'\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}^{-1}\tilde{\boldsymbol{y}}_i^{\star}\right)' \left(\boldsymbol{\Psi}_c^{-1}\left[\boldsymbol{\alpha}'\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}^{-1}\boldsymbol{\alpha} + \boldsymbol{\Psi}_c^{-1}\right]^{-1}\boldsymbol{\alpha}'\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}^{-1}\boldsymbol{\alpha}\right)(\bullet)\right\}. \tag{1.23}$$

Using this result and writing $\mathbf{A} \equiv \boldsymbol{\alpha}'\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}^{-1}\boldsymbol{\alpha}$ and $\mathbf{B} \equiv \boldsymbol{\alpha}'\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}^{-1}\tilde{\boldsymbol{y}}_i^{\star}$ to make the notation clearer, Equation (1.21) can be re-expressed as:

$$\sum_{c=1}^{q} \pi_c |\boldsymbol{\Psi}_c|^{-1/2} \exp\left\{-\frac{1}{2}\left(\boldsymbol{\mu}_c - \mathbf{A}^{-1}\mathbf{B}\right)' \left(\boldsymbol{\Psi}_c^{-1}\left(\mathbf{A} + \boldsymbol{\Psi}_c^{-1}\right)^{-1}\mathbf{A}\right)(\bullet)\right\}$$
$$\times \exp\left\{-\frac{1}{2}\left(\boldsymbol{f}_i - [\mathbf{A} + \boldsymbol{\Psi}_c^{-1}]^{-1}[\mathbf{B} + \boldsymbol{\Psi}_c^{-1}\boldsymbol{\mu}_c]\right)'(\mathbf{A} + \boldsymbol{\Psi}_c^{-1})(\bullet)\right\},$$
$$\propto \sum_{c=1}^{q} \pi_c \left|\boldsymbol{\Psi}_c^{-1}(\mathbf{A} + \boldsymbol{\Psi}_c^{-1})^{-1}\mathbf{A}\right|^{1/2} \exp\left\{-\frac{1}{2}\left(\boldsymbol{\mu}_c - \mathbf{A}^{-1}\mathbf{B}\right)' \left(\boldsymbol{\Psi}_c^{-1}\left[\mathbf{A} + \boldsymbol{\Psi}_c^{-1}\right]^{-1}\mathbf{A}\right)(\bullet)\right\}$$
$$\times \left|\mathbf{A} + \boldsymbol{\Psi}_c^{-1}\right|^{1/2} \exp\left\{-\frac{1}{2}\left(\boldsymbol{f}_i - [\mathbf{A} + \boldsymbol{\Psi}_c^{-1}]^{-1}[\mathbf{B} + \boldsymbol{\Psi}_c^{-1}\boldsymbol{\mu}_c]\right)'(\mathbf{A} + \boldsymbol{\Psi}_c^{-1})(\bullet)\right\}.$$

Hence, the conditional of the factors $\boldsymbol{f}_i$ is a mixture of multivariate normal distributions with the following means, covariance matrices and weights:

$$\overline{\boldsymbol{\mu}}_c = \left[\boldsymbol{\alpha}'\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}^{-1}\boldsymbol{\alpha} + \boldsymbol{\Psi}_c^{-1}\right]^{-1}\left[\boldsymbol{\alpha}'\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}^{-1}\tilde{\boldsymbol{y}}_i^{\star} + \boldsymbol{\Psi}_c^{-1}\boldsymbol{\mu}_c\right],$$
$$\overline{\boldsymbol{\Psi}}_c = \left[\boldsymbol{\alpha}'\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}^{-1}\boldsymbol{\alpha} + \boldsymbol{\Psi}_c^{-1}\right]^{-1},$$
$$\overline{\pi}_c \propto \pi_c\,\phi\left(\boldsymbol{\mu}_c\,;\,\left[\boldsymbol{\alpha}'\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}^{-1}\boldsymbol{\alpha}\right]^{-1}\left[\boldsymbol{\alpha}'\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}^{-1}\tilde{\boldsymbol{y}}_i^{\star}\right]\,,\,\left[\boldsymbol{\Psi}_c^{-1}\left(\boldsymbol{\alpha}'\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}^{-1}\boldsymbol{\alpha} + \boldsymbol{\Psi}_c^{-1}\right)^{-1}\boldsymbol{\alpha}'\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}^{-1}\boldsymbol{\alpha}\right]^{-1}\right),$$

where $\phi(\cdot)$ is the pdf of the multivariate normal distribution.

[18] See Appendix 1.A for details.

### 1.5.2 Updating mixture parameters

Conditional on the updated latent factors, the parameters of their distribution can be sampled in a multi-stage procedure. The group-indicator approach of Diebolt and Robert (1994) is implemented to make the sampling easier. In the same spirit as data augmentation, a vector of indicators representing the group membership of the individuals is introduced and updated at each step of the Gibbs sampler. Conditional on this group-membership vector, the different sets of mixture parameters are sampled component-wise. For this purpose, let $g_{c,i}$ be the group indicator that is equal to one if individual $i$ belongs to group $c$, and to zero otherwise.

#### Updating group indicators

Since each individual can belong to only one mixture group, the probability mass of each $g_{c,i}$ is simply the corresponding weight $\pi_c$. As a consequence, the vector of indicators $\boldsymbol{g}_i = (g_{1,i}, ..., g_{q,i})'$ follows a multinomial distribution with the mixture weights as parameters. Let $\mathbf{G} = (\boldsymbol{g}_1, ..., \boldsymbol{g}_n)'$ be the $(n \times q)$–dimensional matrix containing all individual indicators. Conditional on group membership, $\boldsymbol{f}_i$ is normally distributed, and the posterior of the group indicators is, for individual $i$ and for all $c = 1, ..., q$:

$$\begin{aligned} p(g_{c,i} = 1|\boldsymbol{f}_i) &\propto p(\boldsymbol{f}_i|g_{c,i} = 1)p(g_{c,i} = 1), \\ &\propto \phi(\boldsymbol{f}_i; \boldsymbol{\mu}_c, \boldsymbol{\Psi}_c)\, \pi_c, \end{aligned}$$

where $\phi(\boldsymbol{f}_i; \boldsymbol{\mu}_c, \boldsymbol{\Psi}_c)$ denotes the probability density function of the multivariate normal distribution with mean $\boldsymbol{\mu}_c$ and covariance matrix $\boldsymbol{\Psi}_c$ evaluated at $\boldsymbol{f}_i$.

Thus, group membership probabilities are computed as follows in order to ensure normalization:

$$\Pr(g_{c,i} = 1) = \frac{\pi_c\, \phi(\boldsymbol{f}_i; \boldsymbol{\mu}_c, \boldsymbol{\Psi}_c)}{\sum_{r=1}^{q} \pi_r\, \phi(\boldsymbol{f}_i; \boldsymbol{\mu}_r, \boldsymbol{\Psi}_r)}, \tag{1.24}$$

and the conditional of the group indicators is therefore the following multinomial distribution:

$$\boldsymbol{g}_i \mid \boldsymbol{f}_i, \boldsymbol{\mu}, \boldsymbol{\Psi}, \boldsymbol{\pi} \sim \mathcal{M}\left( \frac{\pi_1\, \phi(\boldsymbol{f}_i; \boldsymbol{\mu}_1, \boldsymbol{\Psi}_1)}{\sum_{r=1}^{q} \pi_r\, \phi(\boldsymbol{f}_i; \boldsymbol{\mu}_r, \boldsymbol{\Psi}_r)}, \dots, \frac{\pi_q\, \phi(\boldsymbol{f}_i; \boldsymbol{\mu}_q, \boldsymbol{\Psi}_q)}{\sum_{r=1}^{q} \pi_r\, \phi(\boldsymbol{f}_i; \boldsymbol{\mu}_r, \boldsymbol{\Psi}_r)} \right).$$

### Updating mixture covariance matrices

Conditional on the latent factors, the covariance matrices of the mixture components are updated in the same way as in the normal case described in Section 1.3.3 within each mixture group. Assuming an inverse Wishart prior distribution with $v$ degrees of freedom and inverse scale matrix $\boldsymbol{\Upsilon}$ for mixture covariance matrix $\boldsymbol{\Psi}_c$:

$$\boldsymbol{\Psi}_c \sim \mathcal{W}^{-1}(v;\, \boldsymbol{\Upsilon})\,,$$

the conditional can be shown to be the following inverse Wishart distribution:

$$\boldsymbol{\Psi}_c \,|\, \mathbf{G}, \mathbf{F}, \boldsymbol{\mu}_c \sim \mathcal{W}^{-1}\left( v + n_c;\, \sum_{i: g_{c,i}=1} (\boldsymbol{f}_i - \boldsymbol{\mu}_c)(\boldsymbol{f}_i - \boldsymbol{\mu}_c)' + \boldsymbol{\Upsilon} \right),$$

where the sum is over the individuals belonging to group $c$, and $n_c$ stands for the number of individuals in this group. If there is only one latent factor in the model, or if the factors are uncorrelated and updated one at a time, an inverse Gamma prior is used instead of the inverse Wishart distribution, and the sampling scheme of the variance is achieved in the same manner as in Section 1.3.3.

### Updating mixture weights

Let $\boldsymbol{\pi} = (\pi_1, ..., \pi_c)'$ and assume a Dirichlet distribution for the prior of the weights:

$$\boldsymbol{\pi} \sim \mathcal{D}ir(a; ...; a)\,.$$

Conditional on $\boldsymbol{g}_i$, $\boldsymbol{f}_i$ is normally distributed in each mixture group and we can write $p(\boldsymbol{f}_i|\boldsymbol{g}_i, \boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\Psi}) = \prod_{c=1}^{q} [\pi_c \phi(\boldsymbol{f}_i; \boldsymbol{\mu}_c, \boldsymbol{\Psi}_c)]^{g_{c,i}}$. With this expression in hand, the conditional of the weights can be derived as:

$$\begin{aligned} p(\boldsymbol{\pi}|\mathbf{F}, \mathbf{G}, \boldsymbol{\mu}, \boldsymbol{\Psi}) &\propto p(\mathbf{F}|\mathbf{G}, \boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\Psi}) p(\boldsymbol{\pi}), \\ &\propto \prod_{i=1}^{n} \prod_{c=1}^{q} {\pi_c}^{g_{c,i}} \prod_{c=1}^{q} \pi_c^{a-1}, \\ &\propto \prod_{c=1}^{q} \pi_c^{n_c + a - 1}, \end{aligned}$$

providing the kernel of a Dirichlet distribution. Hence, $\boldsymbol{\pi} \,|\, \mathbf{G} \sim \mathcal{D}ir(n_1 + a; ...; n_c + a)$.

### Updating mixture means

Because of the zero mean constraint of the mixture, this stage of the Gibbs sampler requires some elaboration. Each mean vector is assumed to be a priori normally distributed:

$$\boldsymbol{\mu}_c \sim \mathcal{N}(\boldsymbol{\mu}_0; \boldsymbol{\Omega}_0)\,,$$

where $\boldsymbol{\mu}_c$ and $\boldsymbol{\mu}_0$ are of dimension $(k \times 1)$, and $\boldsymbol{\Omega}_0$ is a $(q \times q)$–dimensional matrix.

The zero mean restriction of Equation (1.20) can be expressed as:

$$\sum_{c=1}^{q} \pi_c\, \boldsymbol{\mu}_c = 0 \qquad \Leftrightarrow \qquad \boldsymbol{\mu}_q = -\sum_{c=1}^{q-1} \frac{\pi_c}{\pi_q}\, \boldsymbol{\mu}_c, \tag{1.25}$$

and a Gibbs sampling scheme based on the first $q-1$ means can be developed. The derivation of the conditional distribution is however not straightforward, because the zero constraint has to be integrated into both the likelihood and the prior of the component means. Using Bayes' rule, the vector $(\boldsymbol{\mu}_1', ..., \boldsymbol{\mu}_{q-1}')'$ is sampled, and finally the last mean vector $\boldsymbol{\mu}_q$ is computed to fulfill the zero-mean constraint. Two different approaches can be implemented for this purpose: the mean vectors can be either drawn sequentially across mixture components, or simultaneously.

***Drawing the means sequentially.*** Mean vectors are drawn component-wise, holding the means other than the one being updated constant. To do so, the conditional of the $r^{\text{th}}$ mean vector has to be derived conditional on the means of the other components. Let $\boldsymbol{\mu}_{-r}$ be the set of mixture mean vectors excluding the $r^{\text{th}}$ one. Then, replacing $\boldsymbol{\mu}_q$ with its expression given by Equation (1.25), the conditional distribution is, for $r = 1, ..., q-1$:

$$\begin{aligned}
&p(\boldsymbol{\mu}_r|\mathbf{F}, \mathbf{G}, \boldsymbol{\mu}_{-r}, \boldsymbol{\Psi}, \boldsymbol{\pi}) \propto p(\mathbf{F}|\boldsymbol{\mu}_r, \mathbf{G}, \boldsymbol{\mu}_{-r}, \boldsymbol{\Psi}, \boldsymbol{\pi}) p(\boldsymbol{\mu}_r, \boldsymbol{\mu}_{-r}),\\
&\propto \prod_{c=1}^{q-1} \exp\left\{-\frac{1}{2}\sum_{i:g_{c,i}=1} (\boldsymbol{f}_i - \boldsymbol{\mu}_c)'\boldsymbol{\Psi}_c^{-1}(\boldsymbol{f}_i - \boldsymbol{\mu}_c)\right\} \exp\left\{-\frac{1}{2}(\boldsymbol{\mu}_c - \boldsymbol{\mu}_0)'\boldsymbol{\Omega}_0^{-1}(\boldsymbol{\mu}_c - \boldsymbol{\mu}_0)\right\}\\
&\qquad \times \exp\left\{-\frac{1}{2}\sum_{i:g_{q,i}=1}\left(\boldsymbol{f}_i + \sum_{c=1}^{q-1}\frac{\pi_c}{\pi_q}\boldsymbol{\mu}_c\right)' \boldsymbol{\Psi}_q^{-1}\left(\boldsymbol{f}_i + \sum_{c=1}^{q-1}\frac{\pi_c}{\pi_q}\boldsymbol{\mu}_c\right)\right\}\\
&\qquad\quad \times \exp\left\{-\frac{1}{2}\left(-\sum_{c=1}^{q-1}\frac{\pi_c}{\pi_q}\boldsymbol{\mu}_c - \boldsymbol{\mu}_0\right)' \boldsymbol{\Omega}_0^{-1}\left(-\sum_{c=1}^{q-1}\frac{\pi_c}{\pi_q}\boldsymbol{\mu}_c - \boldsymbol{\mu}_0\right)\right\},
\end{aligned}$$

$$\propto \exp\left\{-\frac{1}{2}\sum_{i:g_{r,i}=1}(\boldsymbol{f}_i-\boldsymbol{\mu}_r)'\boldsymbol{\Psi}_r^{-1}(\boldsymbol{f}_i-\boldsymbol{\mu}_r)\right\}\exp\left\{-\frac{1}{2}(\boldsymbol{\mu}_r-\boldsymbol{\mu}_0)'\boldsymbol{\Omega}_0^{-1}(\boldsymbol{\mu}_r-\boldsymbol{\mu}_0)\right\}$$
$$\times\exp\left\{-\frac{1}{2}\left(\frac{\pi_r}{\pi_q}\right)^2\sum_{i:g_{q,i}=1}\left(\boldsymbol{\mu}_r+\left[\sum_{\substack{c=1\\c\neq r}}^{q-1}\frac{\pi_c}{\pi_r}\boldsymbol{\mu}_c+\frac{\pi_q}{\pi_r}\boldsymbol{f}_i\right]\right)'\boldsymbol{\Psi}_q^{-1}(\bullet)\right\}$$
$$\times\exp\left\{-\frac{1}{2}\left(\frac{\pi_r}{\pi_q}\right)^2\left(\boldsymbol{\mu}_r+\left[\sum_{\substack{c=1\\c\neq r}}^{q-1}\frac{\pi_c}{\pi_r}\boldsymbol{\mu}_c+\frac{\pi_q}{\pi_r}\boldsymbol{\mu}_0\right]\right)'\boldsymbol{\Omega}_0^{-1}(\bullet)\right\},$$
$$\propto\exp\left\{-\frac{1}{2}\left(\boldsymbol{\mu}_r'\left[n_r\boldsymbol{\Psi}_r^{-1}+\boldsymbol{\Omega}_0^{-1}+\left(\frac{\pi_r}{\pi_q}\right)^2\left(n_q\boldsymbol{\Psi}_q^{-1}+\boldsymbol{\Omega}_0^{-1}\right)\right]\boldsymbol{\mu}_r\right.\right.$$
$$-2\boldsymbol{\mu}_r'\left[\boldsymbol{\Psi}_r^{-1}\sum_{i:g_{r,i}=1}\boldsymbol{f}_i+\boldsymbol{\Omega}_0^{-1}\boldsymbol{\mu}_0-\left(\frac{\pi_r}{\pi_q}\right)^2\left(n_q\boldsymbol{\Psi}_q^{-1}+\boldsymbol{\Omega}_0^{-1}\right)\sum_{\substack{c=1\\c\neq r}}^{q-1}\frac{\pi_c}{\pi_r}\boldsymbol{\mu}_c\right.$$
$$\left.\left.\left.-\frac{\pi_r}{\pi_q}\left(\boldsymbol{\Psi}_q^{-1}\sum_{i:g_{q,i}=1}\boldsymbol{f}_i+\boldsymbol{\Omega}_0^{-1}\boldsymbol{\mu}_0\right)\right]\right)\right\},$$

where $n_r$ $(r=1,...,q)$ is the number of observations in mixture group $r$. This last expression can be factorized to provide the kernel of a multivariate normal distribution with the following covariance matrix and mean, for $r=1,...,q-1$:

$$\overline{\boldsymbol{\Omega}}_r=\left[n_r\boldsymbol{\Psi}_r^{-1}+\boldsymbol{\Omega}_0^{-1}+\left(\frac{\pi_r}{\pi_q}\right)^2\left(n_q\boldsymbol{\Psi}_q^{-1}+\boldsymbol{\Omega}_0^{-1}\right)\right]^{-1},$$
$$\overline{\boldsymbol{\mu}}_r=\overline{\boldsymbol{\Omega}}_r\left[\boldsymbol{\Psi}_r^{-1}\sum_{i:g_{r,i}=1}\boldsymbol{f}_i+\boldsymbol{\Omega}_0^{-1}\boldsymbol{\mu}_0-\left(\frac{\pi_r}{\pi_q}\right)^2\left(n_q\boldsymbol{\Psi}_q^{-1}+\boldsymbol{\Omega}_0^{-1}\right)\sum_{\substack{c=1\\c\neq r}}^{q-1}\frac{\pi_c}{\pi_r}\boldsymbol{\mu}_c\right.$$
$$\left.-\frac{\pi_r}{\pi_q}\left(\boldsymbol{\Psi}_q^{-1}\sum_{i:g_{q,i}=1}\boldsymbol{f}_i+\boldsymbol{\Omega}_0^{-1}\boldsymbol{\mu}_0\right)\right].$$

***Drawing the means simultaneously.*** To make the calculations easier to follow, compact form matrices will be used henceforth. For this purpose, some notational conventions have to be introduced. We define the following stacked vectors and block-matrices:

$$\boldsymbol{\pi}_{-q}=(\pi_1,...,\pi_{q-1})', \qquad \tilde{\boldsymbol{\pi}}=\boldsymbol{\pi}_{-q}\otimes\mathbf{I}_k,$$
$$\tilde{\boldsymbol{\mu}}=(\boldsymbol{\mu}_1',...,\boldsymbol{\mu}_{q-1}')', \qquad \tilde{\boldsymbol{\mu}}_0=\boldsymbol{\iota}_{q-1}\otimes\boldsymbol{\mu}_0,$$

$$\tilde{\mathbf{\Psi}} = \operatorname*{diag}_{c=1}^{q-1} \{\mathbf{\Psi}_c\}, \qquad \tilde{\mathbf{\Omega}}_0 = \mathbf{I}_{q-1} \otimes \mathbf{\Omega}_0,$$
$$\tilde{\boldsymbol{g}}_{-q,i} = (g_{1,i}, ..., g_{q-1,i})', \qquad \tilde{\boldsymbol{f}}_i = \boldsymbol{\iota}_{q-1} \otimes \boldsymbol{f}_i,$$

where $\otimes$ denotes the Kronecker product, $\mathbf{I}_{q-1}$ is the $(q-1)$–dimensional identity matrix, $\boldsymbol{\iota}_{q-1}$ is the $(q-1)$–dimensional vector of 1's, and diag $\{\cdot\}$ is the matrix operator stacking the specified elements to create a block-diagonal matrix.

Using this notation, the zero-mean restriction of Equation (1.25) can be rewritten as:

$$\boldsymbol{\mu}_q = -\frac{1}{\pi_q}\tilde{\boldsymbol{\pi}}'\tilde{\boldsymbol{\mu}}. \tag{1.26}$$

Since the conditional distribution is derived given group membership, the following selection matrix is useful for picking the parameters corresponding to the mixture group individual $i$ belongs to:

$$\tilde{\mathbf{G}}_i = (\tilde{\boldsymbol{g}}_{-q,i}\,\tilde{\boldsymbol{g}}'_{-q,i}) \otimes \mathbf{I}_k.$$

Replacing the last mean vector $\boldsymbol{\mu}_q$ with its expression from Equation (1.26) in the likelihood and in the prior, the conditional distribution of the first $q-1$ means is:

$$\begin{aligned}
&p(\tilde{\boldsymbol{\mu}}|\mathbf{F}, \mathbf{G}, \boldsymbol{\pi}, \mathbf{\Psi}) \propto p(\mathbf{F}|\tilde{\boldsymbol{\mu}}, \mathbf{G}, \boldsymbol{\pi}, \mathbf{\Psi})p(\tilde{\boldsymbol{\mu}}),\\
&\propto \exp\left\{-\frac{1}{2}\sum_{i=1}^{n}\left[(\tilde{\boldsymbol{f}}_i - \tilde{\boldsymbol{\mu}})'\tilde{\mathbf{G}}_i\tilde{\mathbf{\Psi}}^{-1}(\tilde{\boldsymbol{f}}_i - \tilde{\boldsymbol{\mu}}) \ + g_{q,i}\left(\boldsymbol{f}_i + \frac{1}{\pi_q}\tilde{\boldsymbol{\pi}}'\tilde{\boldsymbol{\mu}}\right)'\mathbf{\Psi}_q^{-1}\left(\boldsymbol{f}_i + \frac{1}{\pi_q}\tilde{\boldsymbol{\pi}}'\tilde{\boldsymbol{\mu}}\right)\right]\right\}\\
&\qquad \times \exp\left\{-\frac{1}{2}\left[(\tilde{\boldsymbol{\mu}} - \tilde{\boldsymbol{\mu}}_0)'\tilde{\mathbf{\Omega}}_0^{-1}(\tilde{\boldsymbol{\mu}} - \tilde{\boldsymbol{\mu}}_0) \ + \left(-\frac{1}{\pi_q}\tilde{\boldsymbol{\pi}}'\tilde{\boldsymbol{\mu}} - \boldsymbol{\mu}_0\right)'\mathbf{\Omega}_0^{-1}\left(-\frac{1}{\pi_q}\tilde{\boldsymbol{\pi}}'\tilde{\boldsymbol{\mu}} - \boldsymbol{\mu}_0\right)\right]\right\},\\
&\propto \exp\left\{-\frac{1}{2}\left[\tilde{\boldsymbol{\mu}}'\left(\sum_{i=1}^{n}\left[\tilde{\mathbf{G}}_i\tilde{\mathbf{\Psi}}^{-1} + \frac{g_{q,i}}{\pi_q^2}\tilde{\boldsymbol{\pi}}\,\mathbf{\Psi}_q^{-1}\tilde{\boldsymbol{\pi}}'\right] + \tilde{\mathbf{\Omega}}_0^{-1} + \frac{1}{\pi_q^2}\tilde{\boldsymbol{\pi}}\,\mathbf{\Omega}_0^{-1}\tilde{\boldsymbol{\pi}}'\right)\tilde{\boldsymbol{\mu}}\right.\right.\\
&\qquad \left.\left. -2\tilde{\boldsymbol{\mu}}'\left(\sum_{i=1}^{n}\left[\tilde{\mathbf{G}}_i\tilde{\mathbf{\Psi}}^{-1}\tilde{\boldsymbol{f}}_i - \frac{g_{q,i}}{\pi_q}\tilde{\boldsymbol{\pi}}\,\mathbf{\Psi}_q^{-1}\boldsymbol{f}_i\right] + \tilde{\mathbf{\Omega}}_0^{-1}\tilde{\boldsymbol{\mu}}_0 - \frac{1}{\pi_q}\tilde{\boldsymbol{\pi}}\,\mathbf{\Omega}_0^{-1}\boldsymbol{\mu}_0\right)\right]\right\}.
\end{aligned}$$

This last expression can be factorized into the kernel of a multivariate normal distribution with precision matrix:

$$\begin{aligned}
\overline{\mathbf{\Omega}}_r^{-1} &= \sum_{i=1}^{n}\left[\tilde{\mathbf{G}}_i\tilde{\mathbf{\Psi}}^{-1} + \frac{g_{q,i}}{\pi_q^2}\tilde{\boldsymbol{\pi}}\,\mathbf{\Psi}_q^{-1}\tilde{\boldsymbol{\pi}}'\right] + \tilde{\mathbf{\Omega}}_0^{-1} + \frac{1}{\pi_q^2}\tilde{\boldsymbol{\pi}}\,\mathbf{\Omega}_0^{-1}\tilde{\boldsymbol{\pi}}',\\
&= \operatorname*{diag}_{c=1}^{q-1}\left\{n_c\mathbf{\Psi}_c^{-1} + \mathbf{\Omega}_0^{-1}\right\} + \frac{1}{\pi_q^2}\tilde{\boldsymbol{\pi}}\,(n_q\mathbf{\Psi}_q^{-1} + \mathbf{\Omega}_0^{-1})\tilde{\boldsymbol{\pi}}',
\end{aligned}$$

where $n_c$ $(c = 1, ..., q)$ is the number of observations in mixture group $c$. Hence, the precision matrix of the conditional is a block matrix defined as follows:

$$[\overline{\mathbf{\Omega}}^{-1}]_{(c,c)} = n_c \mathbf{\Psi}_c^{-1} + \mathbf{\Omega}_0^{-1} + \left(\frac{\pi_c}{\pi_q}\right)^2 (n_q \mathbf{\Psi}_q^{-1} + \mathbf{\Omega}_0^{-1}) \quad \text{for diagonal block elements,}$$
$$[\overline{\mathbf{\Omega}}^{-1}]_{(c,r)} = \frac{\pi_c \pi_r}{\pi_q^2} (n_q \mathbf{\Psi}_q^{-1} + \mathbf{\Omega}_0^{-1}) \quad \text{for off-diagonal block elements,}$$

where $[\overline{\mathbf{\Omega}}^{-1}]_{(c,r)}$ is the block on the $c^{\text{th}}$ row and $r^{\text{th}}$ column of the inverse of the covariance matrix $\overline{\mathbf{\Omega}}$, for all $c, r = 1, ..., q-1$ and $c \neq r$.

As for the mean of the conditional distribution, it is equal to:

$$\begin{aligned}\overline{\boldsymbol{\mu}} &= \overline{\mathbf{\Omega}} \left( \sum_{i=1}^{n} \left[ \tilde{\mathbf{G}}_i \tilde{\mathbf{\Psi}}^{-1} \tilde{\boldsymbol{f}}_i - \frac{g_{q,i}}{\pi_q} \tilde{\boldsymbol{\pi}}\, \mathbf{\Psi}_q^{-1} \boldsymbol{f}_i \right] + \tilde{\mathbf{\Omega}}_0^{-1} \tilde{\boldsymbol{\mu}}_0 - \frac{1}{\pi_q} \tilde{\boldsymbol{\pi}}\, \mathbf{\Omega}_0^{-1} \boldsymbol{\mu}_0 \right), \\ &= \overline{\mathbf{\Omega}} \ \text{vec}_{c=1}^{q-1} \left\{ \mathbf{\Psi}_c^{-1} \sum_{i:g_{c,i}=1} \boldsymbol{f}_i + \mathbf{\Omega}_0^{-1} \boldsymbol{\mu}_0 - \frac{\pi_c}{\pi_q} \left[ \mathbf{\Psi}_q^{-1} \sum_{i:g_{q,i}=1} \boldsymbol{f}_i + \mathbf{\Omega}_0^{-1} \boldsymbol{\mu}_0 \right] \right\},\end{aligned}$$

where vec $\{\cdot\}$ is the operator vertically stacking the specified elements to create a vector.

To the best of our knowledge, this alternative consisting of drawing the mixture means simultaneously has not been implemented so far. We expect this sampling scheme to offer better performances with respect to the mixing of the Markov chain, compared to the sequential sampling, as it is usually the case for parameter block-sampling (Liu et al., 1994).

### 1.5.3 Mixed error terms

The normality assumption of the error term can also be relaxed through the introduction of a mixture of normals. This allows for more flexible functional forms that come closer to semiparametric approaches, and are likely to provide a better fit of the model. For example, Geweke and Keane (1999) introduce mixture of normals in the standard probit model and carry out a comparative study that concludes in favor of the mixture of normals probit over the standard model. In the factor model framework, Hansen et al. (2004) use a three-component mixture for the error term of the wage equation and find an improved fit of the model. Other examples can be found in Carneiro et al. (2003), Cunha et al. (2005).

The error term $\varepsilon_i$ is supposed to follow a mixture of $l$ univariate normals with mean $\nu_h$ and variance $\varphi_h^2$, for $h = 1, ..., l$:

$$\varepsilon_i \sim \sum_{h=1}^{l} w_h \mathcal{N}\left(\nu_h;\, \varphi_h^2\right), \qquad \mathrm{E}[\varepsilon_i] = \sum_{h=1}^{l} w_h \nu_h = 0. \tag{1.27}$$

The parameters of the mixture are updated in the same fashion as those of the latent factors described in Section 1.5.2, using the residuals $\boldsymbol{e} = \boldsymbol{y}^\star - \mathbf{X}\boldsymbol{\beta} - \mathbf{F}\boldsymbol{\alpha}$ at each step of the Gibbs sampler.

The introduction of a mixture of normals for the error term makes the initial problem slightly more complicated, inasmuch as the conditional distribution of the response variable is also a mixture of normals under this assumption. In the initial model for which we derived the Gibbs sampler, the latent response variable was normally distributed as:

$$y_i^\star | \boldsymbol{x}_i, \boldsymbol{f}_i, \boldsymbol{\theta} \sim \mathcal{N}\left(\boldsymbol{x}_i'\boldsymbol{\beta} + \boldsymbol{f}_i'\boldsymbol{\alpha};\, \sigma^2\right),$$

whereas with a mixed error term, the conditional distribution becomes the following mixture:

$$y_i^\star | \boldsymbol{x}_i, \boldsymbol{f}_i, \boldsymbol{\theta} \sim \sum_{h=1}^{l} w_h \mathcal{N}\left(\nu_h + \boldsymbol{x}_i'\boldsymbol{\beta} + \boldsymbol{f}_i'\boldsymbol{\alpha};\, \varphi_h^2\right).$$

The Gibbs sampler derived in Section 1.3 can therefore no longer be used in its original form. However, a simple transformation of the variables solves the problem. Once the mixture parameters have been updated, the observed and latent variables can be redefined as follows to take into account the heterogeneity of the error term:

$$\begin{aligned} \boldsymbol{y}_i^{\star,\text{mix}} &\leftarrow (\boldsymbol{y}_i^\star - \nu_h)/\varphi_h, \\ \boldsymbol{x}_i^{\text{mix}} &\leftarrow \boldsymbol{x}_i/\varphi_h, \\ \boldsymbol{f}_i^{\text{mix}} &\leftarrow \boldsymbol{f}_i/\varphi_h, \end{aligned}$$

where individual $i$ belongs to mixture group $h$. After this transformation, the error term appears to be standard normally distributed for all individuals. It is therefore enough to use these transformed variables instead of the original ones, and to set the variance to unity when deriving the conditional distributions of the other parameters. Thus, the mixture of normals for the error term can be introduced at negligible cost.

## 1.6 Conclusion

The purpose of this chapter is to review the state of the art in Bayesian inference for factor structure models. We provide all the technical details required to construct the Gibbs sampler step by step. A wide range of models can be easily accommodated within the proposed framework. For instance, limited dependent variables have for long been the bottleneck of factor structure models, and many empirical studies have so far considered them as continuous for the sake of convenience, ignoring the potential problems inherent to this misspecification of the model. The methods presented in this chapter show that it is now possible to deal with these variables in an appropriate way, thereby opening new horizons for the empirical research where discrete response variables are very common.

Some of the methods we review, and especially the refinements we suggest, are likely to substantially enhance the performance of the Gibbs sampler. Several extensions, like the use of mixtures of normals, have already been implemented in many studies, but to the best of our knowledge, their properties have not yet been investigated in a systematic way. Another question is related to the way in which these different enhancements can be combined. How would the Gibbs sampler be affected if a mixture of normals were specified for the latent factors in a parameter expanded model? Would the convergence and the mixing of the chain be improved, or on the contrary be deteriorated because of the growing number of parameters in the latent part of the model? Many open questions therefore remain, and simulation studies could for instance help better understand these features.

# Bibliography

Albert, J. H. and S. Chib (1993): "Bayesian Analysis of Binary and Polychotomous Response Data," *Journal of the American Statistical Association*, 88, 669–679.

Carneiro, P., K. Hansen, and J. J. Heckman (2003): "Estimating Distributions of Treatment Effects with an Application to the Returns to Schooling and Measurement of the Effects of Uncertainty on College Choice," *International Economic Review*, 44, 361–422.

Casella, G. and E. I. George (1992): "Explaining the Gibbs Sampler," *The American Statistician*, 46, 167–174.

CHEN, M. H. AND D. K. DEY (2000): "Bayesian Analysis for Correlated Ordinal Data Models," in *Generalized Linear Models: A Bayesian Perspective*, ed. by D. Dey, S. Ghosh, and B. Mallick, CRC Press, 133–157.

CHIB, S. (1992): "Bayes Inference in the Tobit Censored Regression Model," *Journal of Econometrics*, 51, 79–99.

COWLES, M. K. (1996): "Accelerating Monte Carlo Markov Chain Convergence for Cumulative-Link Generalized Linear Models," *Statistics and Computing*, 6, 101–111.

CUNHA, F. AND J. J. HECKMAN (2007): "The Technology of Skill Formation," *American Economic Review*, 97, 31–47.

——— (2008): "Formulating, Identifying and Estimating the Technology of Cognitive and Noncognitive Skill Formation," *Journal of Human Resources*, 43, 738–782.

CUNHA, F., J. J. HECKMAN, AND S. NAVARRO (2005): "Separating Uncertainty from Heterogeneity in Life Cycle Earnings," *Oxford Economic Papers*, 57, 191–261, The 2004 Hicks Lecture.

CUNHA, F., J. J. HECKMAN, AND S. M. SCHENNACH (2010): "Estimating the Technology of Cognitive and Noncognitive Skill Formation," *Econometrica*, 78, 883–931.

DIEBOLT, J. AND C. P. ROBERT (1994): "Estimation of Finite Mixture Distributions Through Bayesian Sampling," *Journal of the Royal Statistical Society Series B: Methodological*, 56, 363–375.

ESCOBAR, M. D. AND M. WEST (1995): "Bayesian Density Estimation and Inference Using Mixtures," *Journal of the American Statistical Association*, 90, 577–588.

FAHRMEIR, L. AND A. W. RAACH (2007): "A Bayesian Semiparametric Latent Variable Model for Mixed Responses," *Psychometrika*, 72, 327–346.

FERGUSON, T. S. (1983): "Bayesian Density Estimation by Mixtures of Normal Distributions," in *Recent Advances in Statistics: Papers in Honor of Herman Chernoff on his Sixtieth Birthday*, ed. by H. Chernoff, M. Rizvi, J. Rustagi, and D. Siegmund, New York: Academic Press, 287–302.

Frühwirth-Schnatter, S. and H. F. Lopes (2009): "Parsimonious Bayesian Factor Analysis when the Number of Factors is Unknown," Unpublished Technical Report.

Gelman, A. (2006): "Prior Distributions for Variance Parameters in Hierarchical Models," *Bayesian Analysis*, 1, 515–533.

Geweke, J. and G. Zhou (1996): "Measuring the Price of the Arbitrage Pricing Theory," *The Review of Financial Studies*, 9, 557–587.

Geweke, J. F. and M. P. Keane (1999): "Mixture of Normals Probit Models," in *Analysis of Panels and Limited Dependent Variables: A Volume in Honor of G. S. Maddala*, ed. by C. Hsiao, G. S. Maddala, K. Lahiri, and L. F. Lee, Cambridge University Press, 49–78.

Ghosh, J. and D. B. Dunson (2009): "Default Prior Distributions and Efficient Posterior Computation in Bayesian Factor Analysis," *Journal of Computational and Graphical Statistics*, 18, 306–320.

Hansen, K. T., J. J. Heckman, and K. J. Mullen (2004): "The Effect of Schooling and Ability on Achievement Test Scores," *Journal of Econometrics*, 121, 39–98.

Heckman, J. J., J. Stixrud, and S. Urzua (2006): "The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior," *Journal of Labor Economics*, 24, 411–482.

Kotlarski, I. (1967): "On Characterizing the Gamma and the Normal Distribution," *Pacific Journal of Mathematics*, 20, 69–76.

Li, M. and J. L. Tobias (2007): "Bayesian Analysis of Treatment Effects in an Ordered Potential Outcomes Model," *Advances in Econometrics*, 21, 57–91.

Liu, C., D. B. Rubin, and Y. N. Wu (1998): "Parameter Expansion to Accelerate EM: the PX-EM Algorithm," *Biometrika*, 85, 755–770.

Liu, J. S. and C. Sabatti (2000): "Generalised Gibbs Sampler and Multigrid Monte Carlo for Bayesian Computation," *Biometrika*, 87, 353–369.

Liu, J. S., W. H. Wong, and A. Kong (1994): "Covariance Structure of the Gibbs Sampler with Applications to the Comparisons of Estimators and Augmentation Schemes," *Biometrika*, 81, 27–40.

Liu, J. S. and Y. N. Wu (1999): "Parameter Expansion for Data Augmentation," *Journal of the American Statistical Association*, 94, 1264–1274.

Liu, X. (2008): "Parameter Expansion for Sampling a Correlation Matrix: An Efficient GPX-RPMH Algorithm," *Journal of Statistical Computation and Simulation*, 78, 1065–1076.

Lopes, H. F. and M. West (2004): "Bayesian Model Assessment in Factor Analysis," *Statistica Sinica*, 14, 41–67.

Nandram, B. and M.-H. Chen (1996): "Reparameterizing the Generalized Linear Model to Accelerate Gibbs Sampler Convergence," *Journal of Statistical Computation and Simulation*, 54, 129–144.

Raach, A. W. (2006): "A Bayesian Semiparametric Latent Variable Model for Binary, Ordinal and Continuous Response," Ph.D. thesis, Ludwig-Maximilians-Universität München.

Roeder, K. and L. Wasserman (1997): "Practical Bayesian Density Estimation Using Mixtures of Normals," *Journal of the American Statistical Association*, 92, 894–902.

Rosett, R. N. and F. D. Nelson (1975): "Estimation of the Two-limit Probit Regression Model," *Econometrica*, 43, 141–146.

Tanner, M. A. and W. H. Wong (1987): "The Calculation of Posterior Distributions by Data Augmentation," *Journal of the American Statistical Association*, 82, 528–540.

van Dyk, D. A. and X.-L. Meng (2001): "The Art of Data Augmentation," *Journal of Computational & Graphical Statistics*, 10, 1–50.

West, M. (2003): "Bayesian Factor Regression Models in the 'Large p, Small n' Paradigm," *Bayesian Statistics*, 7, 723–732.

## Appendix 1.A Matrix algebra

Let $\mathbf{E}$ and $\mathbf{F}$ be two invertible square matrices of the same dimension. The following results hold:

$$\begin{aligned}\mathbf{E}-\mathbf{E}(\mathbf{F}+\mathbf{E})^{-1}\mathbf{E} &= \mathbf{E}(\mathbf{F}+\mathbf{E})^{-1}(\mathbf{F}+\mathbf{E})-\mathbf{E}(\mathbf{F}+\mathbf{E})^{-1}\mathbf{E},\\ &= \mathbf{E}(\mathbf{F}+\mathbf{E})^{-1}(\mathbf{F}+\mathbf{E}-\mathbf{E}),\\ &= \mathbf{E}(\mathbf{F}+\mathbf{E})^{-1}\mathbf{F}.\end{aligned}$$

Given that $(\mathbf{EF})^{-1}=\mathbf{F}^{-1}\mathbf{E}^{-1}$, it can be shown that:

$$\begin{aligned}&(\mathbf{E}+\mathbf{F})^{-1}\mathbf{F}[\mathbf{F}(\mathbf{E}+\mathbf{F})^{-1}\mathbf{E}]^{-1}\mathbf{F}(\mathbf{E}+\mathbf{F})^{-1}+(\mathbf{E}+\mathbf{F})^{-1}\\ &\quad=(\mathbf{E}+\mathbf{F})^{-1}\mathbf{F}\mathbf{E}^{-1}+(\mathbf{E}+\mathbf{F})^{-1},\\ &\quad=(\mathbf{E}+\mathbf{F})^{-1}(\mathbf{F}\mathbf{E}^{-1}+\mathbf{I}),\\ &\quad=(\mathbf{E}+\mathbf{F})^{-1}(\mathbf{F}+\mathbf{E})\mathbf{E}^{-1},\\ &\quad=\mathbf{E}^{-1}.\end{aligned}$$

Using these identities, Equation (1.22) can be simplified as follows:

$$\begin{aligned}&\exp\left\{-\frac{1}{2}\left[\boldsymbol{\mu}_c'\boldsymbol{\Psi}_c^{-1}\boldsymbol{\mu}_c-\left(\mathbf{B}+\boldsymbol{\Psi}_c^{-1}\boldsymbol{\mu}_c\right)'\left(\mathbf{A}+\boldsymbol{\Psi}_c^{-1}\right)^{-1}(\bullet)\right]\right\}\\
&=\exp\left\{-\frac{1}{2}\left[\boldsymbol{\mu}_c'\left(\boldsymbol{\Psi}_c^{-1}-\boldsymbol{\Psi}_c^{-1}\left(\mathbf{A}+\boldsymbol{\Psi}_c^{-1}\right)^{-1}\boldsymbol{\Psi}_c^{-1}\right)\boldsymbol{\mu}_c\right.\right.\\
&\qquad\left.\left.-2\boldsymbol{\mu}_c'\boldsymbol{\Psi}_c^{-1}\left(\mathbf{A}+\boldsymbol{\Psi}_c^{-1}\right)^{-1}\mathbf{B}-\mathbf{B}'\left(\mathbf{A}+\boldsymbol{\Psi}_c^{-1}\right)^{-1}\mathbf{B}\right]\right\},\\
&=\exp\left\{-\frac{1}{2}\left[\boldsymbol{\mu}_c'\left(\boldsymbol{\Psi}_c^{-1}\left(\mathbf{A}+\boldsymbol{\Psi}_c^{-1}\right)^{-1}\mathbf{A}\right)\boldsymbol{\mu}_c\right.\right.\\
&\qquad\left.\left.-2\boldsymbol{\mu}_c'\boldsymbol{\Psi}_c^{-1}\left(\mathbf{A}+\boldsymbol{\Psi}_c^{-1}\right)^{-1}\mathbf{B}-\mathbf{B}'\left(\mathbf{A}+\boldsymbol{\Psi}_c^{-1}\right)^{-1}\mathbf{B}\right]\right\},\\
&=\exp\left\{-\frac{1}{2}\left(\left[\boldsymbol{\mu}_c-\left(\boldsymbol{\Psi}_c^{-1}\left(\mathbf{A}+\boldsymbol{\Psi}_c^{-1}\right)^{-1}\mathbf{A}\right)^{-1}\boldsymbol{\Psi}_c^{-1}\left(\mathbf{A}+\boldsymbol{\Psi}_c^{-1}\right)^{-1}\mathbf{B}\right]'\right.\right.\\
&\qquad\times\left[\boldsymbol{\Psi}_c^{-1}\left(\mathbf{A}+\boldsymbol{\Psi}_c^{-1}\right)^{-1}\mathbf{A}\right][\bullet]\\
&\quad-\mathbf{B}'\left[\left(\mathbf{A}+\boldsymbol{\Psi}_c^{-1}\right)^{-1}\boldsymbol{\Psi}_c^{-1}\left(\boldsymbol{\Psi}_c^{-1}\left(\mathbf{A}+\boldsymbol{\Psi}_c^{-1}\right)^{-1}\mathbf{A}\right)^{-1}\boldsymbol{\Psi}_c^{-1}\left(\mathbf{A}+\boldsymbol{\Psi}_c^{-1}\right)^{-1}\right.\\
&\qquad\left.\left.\left.+\left(\mathbf{A}+\boldsymbol{\Psi}_c^{-1}\right)^{-1}\right]\mathbf{B}\right)\right\},\\
&=\exp\left\{-\frac{1}{2}\left[\boldsymbol{\mu}_c-\mathbf{A}^{-1}\mathbf{B}\right]'\left[\boldsymbol{\Psi}_c^{-1}\left(\mathbf{A}+\boldsymbol{\Psi}_c^{-1}\right)^{-1}\mathbf{A}\right][\bullet]\right\}\exp\left\{\frac{1}{2}\mathbf{B}'\mathbf{A}^{-1}\mathbf{B}\right\},\end{aligned}$$

where $\mathbf{A} \equiv \boldsymbol{\alpha}'\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}^{-1}\boldsymbol{\alpha}$ and $\mathbf{B} \equiv \boldsymbol{\alpha}'\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}^{-1}\widetilde{\boldsymbol{y}}_i^*$. Since the second exponential does not depend on $\boldsymbol{f}_i$, nor on $c$, it can be absorbed into the factor of proportionality. This is why it vanishes in Equation (1.23).

## Appendix 1.B Prior specification

The choice of the prior parameters is not innocuous for the estimation results. In most cases, prior parameters are chosen to obtain a vague prior, i.e., noninformative, in order to remain as general as possible. Table 1.B.1 provides a summary of the prior distributions used for the different parameters of the model, as well as some examples of prior parameters that are often used in the literature. However, this table is just intended to provide a general overview of the magnitude of these prior parameters. It just reflects a few recent papers and is not exhaustive.

**Table 1.B.1:** Prior specifications and examples of prior parameters

| **Parameters** | **Prior distributions** | **Alternative prior parameters** |
|---|---|---|
| Slope parameters | $\boldsymbol{\beta}_s \sim \mathcal{N}_p(\boldsymbol{\mu}_{\boldsymbol{\beta}};\ \boldsymbol{\Psi}_{\boldsymbol{\beta}})$ | $\boldsymbol{\mu}_{\boldsymbol{\beta}} = \mathbf{0}_p$ and $\boldsymbol{\Psi}_{\boldsymbol{\beta}}^{-1} = \mathbf{0}_{pp}$ |
| Factor loadings | $\boldsymbol{\alpha}_s \sim \mathcal{N}_k(\boldsymbol{\mu}_{\boldsymbol{\alpha}};\ \boldsymbol{\Psi}_{\boldsymbol{\alpha}})$ | $\boldsymbol{\mu}_{\boldsymbol{\alpha}} = \mathbf{0}_k$ and $\boldsymbol{\Psi}_{\boldsymbol{\alpha}}^{-1} = \mathbf{0}_{kk}$ |
| | | $\boldsymbol{\mu}_{\boldsymbol{\alpha}} = \mathbf{0}_k$ and $\boldsymbol{\Psi}_{\boldsymbol{\alpha}} = \mathbf{I}_k$ (unit scale)[b,c,d] |
| Idiosyncratic | $\sigma_s^2 \sim \mathcal{G}^{-1}(g_1;\ g_2)$ | $g_1 = 2$ and $g_2 = 1$[d] |
| variances | | $g_1 = 1.1$ and $g_2 = 0.05$[c] |
| | $\sigma_s^2 \sim \mathcal{G}^{-1}\big(g_1;\ (g_1-1)/(\mathbf{S}_{\boldsymbol{y}}^{-1})_{ss}\big)$ | $g_1 = 2.5$ (to prevent a Heywood problem)[a,e] |
| Cut-points (ordinal case) | $\lambda_l \sim \mathcal{U}(\lambda_{l-1};\ \lambda_{l+1}),\ l = 1, \ldots, L-1$ | |
| Factor covariance matrix (normal case): $\boldsymbol{f}_i \sim \mathcal{N}_k(\mathbf{0};\ \boldsymbol{\Psi}_{\mathbf{F}})$ | | |
| correlated factors | $\boldsymbol{\Psi}_{\mathbf{F}} \sim \mathcal{W}^{-1}(v;\ \boldsymbol{\Upsilon})$ | $v = k+1$ and $\boldsymbol{\Upsilon} = \mathbf{I}_k$[d] |
| uncorrelated factors | $\boldsymbol{\Psi}_{\mathbf{F}} = \text{diag}\,(\psi_1, ..., \psi_k)$ | $g_1 = 0.5$ and $g_2 = 0.5$[a,b] |
| | $\psi_j \sim \mathcal{G}^{-1}(g_1;\ g_2),\ j = 1, \ldots, k$ | $g_1 = 2.5$ and $g_2 = 2.5$[a] |
| Factor distribution parameters (mixed case): $\boldsymbol{f}_i \sim \sum_{c=1}^{q} \pi_c \mathcal{N}_k(\boldsymbol{\mu}_c;\ \boldsymbol{\Psi}_c)$ | | |
| mixture weights | $\boldsymbol{\pi} \sim \mathcal{D}ir(a; ...; a)$ | $a = 2$[d] |
| mixture means | $\boldsymbol{\mu}_c \sim \mathcal{N}(\boldsymbol{\mu}_0;\ \boldsymbol{\Omega}_0)$ | $\boldsymbol{\mu}_0 = \mathbf{0}_k$ and $\boldsymbol{\Omega}_0 = 10\mathbf{I}_k$[d] |
| mixture covariances (correlated case) | $\boldsymbol{\Psi}_c \sim \mathcal{W}^{-1}(v;\ \boldsymbol{\Upsilon})$ | same as normal case for each component |
| mixture variances (uncorrelated case) | $\boldsymbol{\Psi}_{\mathbf{F}} = \text{diag}\,(\psi_1, ..., \psi_k)$ | |
| | $\psi_j \sim \mathcal{G}^{-1}(g_1;\ g_2),\ j = 1, \ldots, k$ | same as normal case for each component |

[a] Frühwirth-Schnatter and Lopes (2009) [b] Ghosh and Dunson (2009) [c] Lopes and West (2004) [d] Hansen et al. (2004), Carneiro et al. (2003)
[e] $(\mathbf{S}_{\boldsymbol{y}}^{-1})_{ss}$ represents the $s^{\text{th}}$ diagonal element of the inverse covariance matrix of the observed response variables $\boldsymbol{y}$.

# Chapter 2

## Maintaining (Locus of) Control?

### Assessing the Impact of Locus of Control on Education Decisions and Wages

## 2.1 Introduction

Does it make a difference if you think *you* can make a difference? Will it affect your decision making, or even your productivity? In response to such kinds of questions, the economic literature has recently come to acknowledge the considerable importance of so-called *noncognitive skills*, sometimes also referred to as *soft skills*, in explaining education choices, as well as a large variety of labor market outcomes. The present chapter focuses on *locus of control*, one dimension of the noncognitive skills that measures the extent to which individuals believe that what happens to them in life is related to their own actions and decisions, or on the contrary to fate and luck. We contribute to the existing literature on noncognitive skills by investigating the impact of locus of control on wages, while making a distinction between the direct—or *productive*—impact of locus of control, and the indirect—or *behavioral*—impact that works through education decisions.

We find that locus of control is an important predictor of the decision to obtain higher education. Furthermore, we find that *premarket* locus of control, defined as locus of control measured at the time of schooling—before the individual enters the labor market—does not significantly affect later wages after controlling for education decisions. In light of the existing literature, which finds mostly positive effects of contemporaneous locus of control measures on wages, this indicates that it is important to distinguish between *premarket* skills and those that are already influenced by labor market experience and age. Last, simulation of our model shows that moving individuals from the first to the last decile of the locus of control distribution significantly shifts the distribution of schooling choices, thus indirectly affecting later wages.

From a methodological point of view, there are two major econometric problems at stake in the economic literature on noncognitive skills and personality traits: measurement error and endogeneity (Bowles and Gintis, 2002; Borghans et al., 2008). First, the measurement error issue arises because certain traits or characteristics are measured by questions or tests that are imperfect proxies of the true latent ability. Yet, in general, most psychological measures are designed to capture a particular latent trait or skill, such that factor analytical approaches can be used to distinguish true latent abilities from measurement error (Borghans et al., 2008; Heckman et al., 2006; Hansen et al., 2004). Second, endogeneity arises in the study of the impact of locus of control on labor market outcomes for two reasons. On the one hand, the results may be flawed by reverse causality, as (anticipated) labor market outcomes may affect locus of control (e.g., see Trzcinski and Holst, 2010). For this reason, locus of control measures may reflect, rather than cause, the outcomes they are supposed to predict (Borghans et al., 2008). In this case, the coefficient on locus of control is overestimated, because of the positive covariance between the measures and the error term. On the other hand, both outcomes and measures may be affected by past labor market experiences, which are usually not accounted for. The consequence is, again, an overestimation of the locus of control coefficient due to spurious correlation.

Using data from the German Socioeconomic Panel (GSOEP), we address the problem of measurement error by extracting a latent factor reflecting locus of control. In addition, we account for the problem of reverse causality and truncated life-cycle data in that we combine information on both young individuals, who have not yet entered the labor market, and on older, working-age individuals. Our estimation approach follows the work by Heckman et al. (2006), Hansen et al. (2004), Carneiro et al. (2003). Furthermore, we build on a strategy developed in Cunha et al. (2005), which allows us to retrieve the distribution of locus of control from a sample of young individuals, and to estimate its impact on outcomes in a sample of older individuals.

This chapter proceeds as follows: Section 2.2 provides an overview of the existing literature on locus of control, where the emphasis is laid on the ambiguous results regarding the role of locus of control in determining economic outcomes. In Section 2.3, a simple framework is introduced to help understand the potential impact of locus of control on education decisions and labor market outcomes. Section 2.4 describes our estimation strategy relying on data set combination to identify the full likelihood. The Bayesian approach used to sample the parameters of interest is outlined, and an overview of the data is provided. Section 2.5 presents the results of our analysis. Section 2.6 concludes.

## 2.2 Locus of control, education and labor market outcomes: some prior evidence

The theory of human capital is solidly rooted in the economic literature since the seminal works of Mincer (1958) and Becker (1964). In this literature, human capital is defined as the stock of knowledge and personal abilities an individual possesses, and is perceived as a factor of production that can be improved through education, training and experience. However, this concept mainly refers to the cognitive abilities of an individual, while more recently, other facets of human capital have come to the forefront. Bowles and Gintis (1976) were among the first to point out what seems intuitively obvious: economic success is only partly determined by cognitive abilities and knowledge acquired in schools. Personality, incentive-enhancing preferences and socialization are other important components of human capital. Various authors from psychology, sociology, and more recently economics, have empirically investigated the importance of noncognitive traits.[1] Prominent examples are Bowles et al. (2001a,b), who view noncognitive skills as personality traits that lead to a reduction in contract enforcement costs of the employer. More recently, Heckman et al. (2006) investigated the direct impact of noncognitive skills on various outcomes. They found that a one dimensional component of noncognitive skills, comprised of self-esteem and locus of control measures, can explain many dimensions of social performance, including education decisions and labor market outcomes. In addition, various other studies exist, which relate multiple facets of noncognitive skills to labor market outcomes. Examples are Nyhus and Pons (2005), Mueller and Plug (2006), as well as Heineck and Anger (2010). Furthermore, Duncan and Dunifon (1998) emphasize the importance of motivation, as well as behavioral measures such as cleanliness, church attendance or newspaper reading as a predictor of earnings many years later. Somewhat differently, a vast literature in experimental economics is currently emerging, which analyzes the economic impact of risk aversion, reciprocity, self-confidence and time preference (Dohmen et al., 2007; Falk et al., 2006; Frey and Meier, 2004).

We decide to focus on locus of control, one of the measures of noncognitive traits that is well-established in the psychological literature, and which has frequently been suggested to affect labor market outcomes (Heckman et al., 2006; Judge and Bono, 2001; Andrisani, 1977; 1981; Osborne, 2000). Originally, locus of control is a psychological concept, generally attributed to Rotter (1966), that measures the attitude regarding the nature of the causal relationship between one's own behavior and its consequences. In this concept, which is related to self-efficacy, people who believe that they can control reinforcements in their lives are called *internalizers*. People who believe that fate, luck, or other people control reinforcements, are termed *externalizers*. Generally, externalizers—in this taxonomy, the low-ability types—do not have much confidence in their ability to influence their environment, and do not see themselves as responsible for their lives. Therefore, these individuals are generally less likely to trust their own abilities or to push themselves through difficult situations. Conversely, internalizers—the high-ability types—are likely to have more self-esteem and to trust their capability of changing circumstances for

[1] For an overview of the interrelationships between different psychological and economic concepts, see Borghans et al. (2008).

the better. These individuals perceive themselves as more capable of altering their situation, in particular their economic situation.

Locus of control is one of the most prominent concepts of noncognitive skills in the economic literature. Mostly on empirical grounds, many studies agree that it affects a variety of economic choices individuals make—behavioral impact—, while fewer studies find a direct impact on wages—productive impact. This is particularly true for education decisions, which most researchers find to be highly influenced by locus of control.[2] For instance, Coleman and DeLeire (2003) present a model of locus of control and education decisions where locus of control is viewed as a behavioral trait that affects education decisions, because it has an impact on personal beliefs about the effect of education on expected earnings. Using the National Education Longitudinal Study (NELS), the authors find locus of control to have a high and significant impact on schooling decisions, as well as on ex-ante expected earnings conditional on schooling. Similarly, recent evidence by Caliendo et al. (2010) on German unemployment data shows that locus of control is a behavioral trait that affects the subjective probability of finding a job, which in turn leads to an increased search effort and higher reservations wages. Contrary to this, using the National Longitudinal Survey of Youth (NLSY), Cebi (2007) concludes that locus of control has a productive impact on labor market outcomes and no effect on education choices.

Evidence on the effect of locus of control on labor market returns—productive impact—is mixed. For example, Andrisani (1977), using the National Longitudinal Study (NLS), finds a positive effect of locus of control on several measures of earnings and occupational attainment of young and middle-aged men. Yet, Duncan and Morgan (1981) find mostly non-significant effects of locus of control on the change in hourly earnings of individuals in the Panel Study of Income Dynamics (PSID). To our knowledge, an analysis of the impact of locus of control on labor market outcomes using German data has only been conducted by Heineck and Anger (2010), as well as by Flossmann et al. (2007), with both studies finding positive effects. We improve on these papers by accounting for education decisions, and by controlling for endogeneity issues caused by the use of contemporaneous measurements.

In the literature, four main strategies have been adopted to address the endogeneity issue. First, Duncan and Morgan (1981) and Duncan and Dunifon (1998) using the PSID, extract measures of personality traits as measured 15-25 years prior

[2] Already 40 years ago, the famous Coleman report (Coleman, 1968) reported that locus of control was not only an important predictor of academic performance, but even a more important determinant of educational achievement than any other factor in a student's background (Coleman and DeLeire, 2003).

to earnings. A similar strategy has been adopted by Heckman et al. (2006), who use locus of control measurements in the NLSY taken at age 14-22 to explain later outcomes. Second, Bowles et al. (2001b), using the National Longitudinal Survey of Young Women (NLSYW), employ contemporary measurements of locus of control, which they purge of past wage influences. Third, Osborne (2000) uses past skills to instrument for contemporaneous skill measures. Last, Cunha and Heckman (2008) explicitly model development and accumulation of skills as a technology of skill formation, in which investments in one period affect the productivity of investments in subsequent periods. However, their focus is mainly on early childhood development of skills, and not on the impact of labor market experiences and various life-time shocks on skill development and income.

## 2.3 Empirical Model

Consider a simple model where each individual chooses between obtaining higher education or not. Premarket locus of control, as imperfectly measured by a set of response variables, is captured by a latent factor $\theta$, which influences both schooling decisions and labor market outcomes. The concept of locus of control and its potential impact on education decisions and labor market outcomes is explained in Section 2.3.1, while the empirical setup of the model is detailed in Section 2.3.2.

### 2.3.1 How locus of control impacts education and labor market outcomes

In this section, we present a theoretical framework for how *premarket* locus of control may affect labor market returns. We assume that the role of locus of control for wages is potentially twofold. First, it may indirectly affect wages through its effect on education decisions, and secondly, it may have a direct influence on labor market returns after the education decision is controlled for.

In our study, locus of control is a latent variable, denoted by $\theta$, that is continuously distributed in the range $(-\infty, +\infty)$, where smaller values represent a more external locus and larger values a more internal locus of control. We assume that an individual's costs of education and wage are both functions of $\theta$. Hence, individuals with $\theta \to -\infty$ are likely to have higher costs of education and earn lower wages, while individuals with $\theta \to +\infty$ incur lower costs of obtaining a degree and earn more.

In a typical model of human capital investment, individuals decide about the level of education based on the expected returns to the respective choice, net of the costs associated with this choice. In this framework, locus of control may affect the (nonmonetary) costs of education, e.g., because individuals with a more external locus of control need to work harder than internalizers to feel well-prepared for the exams—behavioral impact. Furthermore, locus of control may be viewed as a skill with a direct impact on wages, for example because employers value having employees who exhibit a higher degree of self-efficacy, because they are more productive or act more responsibly—productive impact.

Assume that there are two education levels, denoted by $S = 0, 1$, and that agents maximize the latent utility associated with education to make their decision. Let $U^*$ denote this latent utility. The arguments of this function will be specified later. Hence, individuals attend higher education, $S = 1$, if:

$$U^* \geqslant 0,$$

and $S = 0$ otherwise. The latent utility from obtaining higher education is a function of discounted future earnings and of education costs. If wages $w_t^s$ in period $t$ conditional on schooling $s$, as well as the costs of education $C$, can all be modeled in an additively separable manner, we can specify:

$$\begin{aligned} w_t^0 &= X_{wt}\beta_0 + \theta\alpha_0 + \varepsilon_{0t}, \\ w_t^1 &= X_{wt}\beta_1 + \theta\alpha_1 + \varepsilon_{1t}, \\ C &= X_C\beta_C + \theta\alpha_C + \varepsilon_C, \end{aligned}$$

with $\mathrm{E}[\varepsilon_1|X_{wt}, \theta] = \mathrm{E}[\varepsilon_0|X_{wt}, \theta] = \mathrm{E}[\varepsilon_C|X_C, \theta] = 0$. Here $\alpha_s$, $\beta_s$ (with $s \in \{0, 1\}$) and $\alpha_C$, $\beta_C$ measure the impact of premarket locus of control $\theta$ and observable characteristics $(X_{wt}, X_C)$ on wages and education costs, respectively. Since locus of control is determined before the individual enters the labor market, it does not depend on time $t$ in our model. Moreover, $\varepsilon_{st}$ and $\varepsilon_C$ are random and independent idiosyncratic shocks. The total utility from education, accounting for the discounted flow of ex post earnings, is then:

$$\begin{aligned} U^*(X_w, X_C, \theta, \delta, t_1) = & \sum_{t=t_1}^{T} \delta^t \left( X_{wt}\beta_1 + \theta\alpha_1 + \varepsilon_{1t} \right) \\ & - \sum_{t=0}^{T} \delta^t \left( X_{wt}\beta_0 + \theta\alpha_0 + \varepsilon_{0t} \right) \\ & - \left( X_C\beta_C + \theta\alpha_C + \varepsilon_C \right), \end{aligned} \tag{2.1}$$

where $X_w = (X_{w1}, \ldots, X_{wT})$, $t_1$ represents the time required to achieve higher education, $T$ is the life horizon, and $\delta$ denotes the discount rate, which for simplicity is assumed to be constant over time.

By differentiating Equation (2.1) with respect to $\theta$, it appears that a ceteris paribus change in locus of control affects education decisions as follows:

$$\frac{\partial U^*(X_w, X_C, \theta, t_1)}{\partial \theta} = \alpha_1 \sum_{t=t_1}^{T} \delta^t - \alpha_0 \sum_{t=0}^{T} \delta^t - \alpha_C.$$

Given that $\alpha_1$ and $\alpha_0$ are independent of $t$, and making use of revealed education choices, our goal is to identify $\alpha_1$, $\alpha_0$ and $\alpha_C$. More precisely, we are investigating whether locus of control enters the education decision and outcomes both directly as a skill, in which case we would have $\alpha_1 > 0$ and $\alpha_0 > 0$, or only indirectly via the costs of education, in which case $\alpha_C < 0$. We cannot identify $\alpha_C$ directly, because we do not observe education costs. However, we can make inference on the overall impact of locus of control on education choices, and given the identification of $\alpha_1$ and $\alpha_0$, we can retrieve $\alpha_C$. More specifically, if we find that $\alpha_1 = \alpha_0 = 0$, we know that any impact of locus of control on education choices must work through $\alpha_C$.

The empirical model we specify in the next section is an approximation to this very simple theoretical framework. By combining different subsamples and using revealed schooling decisions, we are able to identify the impact of premarket locus of control on wages, and thus to make inferences about the productive—versus behavioral—impact of this specific personality trait.

### 2.3.2 Specification of the model

To investigate the impact of premarket locus of control on schooling decisions and later outcomes, we use a factor structure model in the spirit of Heckman et al. (2006), where a single latent factor is assumed to capture the latent trait of interest. The overall simultaneous equation model consists of different sets of equations using continuous, dichotomous and ordered response variables. The latent factor is common across all equations, and therefore represents the only source of dependence between the various outcomes, conditional on the observed covariates.

#### Education decision

Each agent is assumed to choose the level of schooling that maximizes her utility. The utility derived from higher education ($S^*$) is supposed to linearly depend on a vector of personal characteristics $X_S$ and on the latent factor $\theta$:

$$\begin{aligned} S &= \mathbb{1}[S^* > 0], \\ S^* &= X_S\beta_S + \theta\alpha_S + \varepsilon_S, \qquad\qquad \varepsilon_S \sim \mathcal{N}(0;\, 1)\,, \end{aligned} \tag{2.2}$$

where $\beta_S$ denotes the vector of parameters related to personal characteristics, $\alpha_S$ represents the factor loading associated with $\theta$, and $\varepsilon_S$ is an idiosyncratic error term assumed to be independent of the covariates and of the latent factor. The indicator function $\mathbb{1}[\cdot]$ is equal to 1 if the corresponding condition is verified, and to 0 otherwise. Conditional on $\theta$, this model is a standard probit when the distribution of the error term is assumed to be standard normal.

#### Labor market outcomes

Individuals with different levels of schooling become active on different segments of the labor market, where their personal characteristics, as well as their level of locus of control, may be valued differently. Labor market outcomes are modeled as a two-stage process: people first select into the labor market, and then a wage equation is estimated for those who are actually working. Observed characteristics and locus of control are allowed to play a role in both stages. Estimating the two equations simultaneously makes it possible to correct for potential sample selection bias that might affect the parameters if only the wage equation for working people were estimated (Heckman, 1979).

The labor market participation decision is assumed to be a threshold-crossing model for each level of education $s \in \{0, 1\}$, where the latent utility of working ($E_s^*$) linearly depends on a set of covariates $X_E$ through a vector of parameters $\beta_{E,s}$, and on the latent factor $\theta$ with its associated factor loading $\alpha_{E,s}$:

$$\begin{aligned} E_s &= \mathbb{1}[E_s^* > 0], \\ E_s^* &= X_E\beta_{E,s} + \theta\alpha_{E,s} + \varepsilon_{E,s}, \qquad\qquad \varepsilon_{E,s} \sim \mathcal{N}(0;\, 1)\,, \end{aligned} \tag{2.3}$$

The idiosyncratic error term $\varepsilon_{E,s}$ is assumed to be standard normal and independent of $X_E$ and $\theta$ for identification purposes. Nevertheless, this equation should not be regarded as a usual employment equation, but rather considered in a broader sense. People participating in the labor market ($E = 1$) are those who are actually active and declare a positive wage, while the group of non-participating people encompasses unemployed people, but also adult individuals who are not on the market. Therefore, this equation should be interpreted with care,[3] and serves more as a technical means to tackle the selection problem into the sample of people declaring a positive wage.

For wages, a log-linear specification with education group specific parameters is assumed:

$$Y_s = X_Y \beta_{Y,s} + \theta \alpha_{Y,s} + \varepsilon_{Y,s} \qquad \text{for } s = 0, 1, \tag{2.4}$$

where $Y_s$ represents the log hourly wage ($\ln w_s$), $X_Y$ is a set of observed covariates with the associated vector of returns $\beta_{Y,s}$, $\alpha_{Y,s}$ denotes the return to locus of control, and $\varepsilon_{Y,s}$ is an idiosyncratic error term such that $\varepsilon_{Y,s} \perp\!\!\!\perp (\theta, X_Y)$. For the specification of the error term, we relax the usual normality assumption by specifying a mixture of $h$ normal distributions with zero mean:

$$\varepsilon_{Y,s} \sim \sum_{j=1}^{h} \pi_{s,j} \, \mathcal{N}\left(\mu_{s,j};\, \omega_{s,j}^2\right), \qquad \mathrm{E}[\varepsilon_{Y,s}] = \sum_{j=1}^{h} \pi_{s,j} \, \mu_{s,j} = 0, \tag{2.5}$$

for $s = 0, 1$, where $\pi_{s,j}$, $\mu_{s,j}$ and $\omega_{s,j}^2$ denote, respectively, the weight, mean and variance of mixture component $j$. Mixtures of normals are widely used as a flexible semiparametric approach for density estimation (Ferguson, 1983; Escobar and West, 1995). In our empirical application, we find that a three-component mixture ($h = 3$) makes it possible to capture more unobserved heterogeneity, and therefore provides a better fit to the data.

Within this specification, premarket locus of control can affect labor market outcomes both directly and indirectly. The direct effect is measured by the factor loadings $\alpha_{E,s}$ and $\alpha_{Y,s}$, for $s = 0, 1$, while the indirect effect operates through the schooling decision. Two different models are considered. First, we estimate the employment and wage equations without conditioning on education, to capture the total effect of locus of control on wages. To achieve this, individuals from both schooling groups are pooled, and the subscript $s$ is therefore dropped from Equations (2.3) to (2.5). In a second stage, both direct and indirect effects are separately accounted for by specifying the model as stated above. Comparing the results from

[3]Especially for the people who achieved higher education, since in this subsample some individuals who do not participate in the labor market are still enrolled in the education system.

these two approaches turns out to be instructive to understand through which channels premarket locus of control affects labor market outcomes.

#### A measurement system for locus of control

In our data, as in most empirical applications, variables measuring latent locus of control come from a psychometric test using Likert scales with a small number of categories. Although techniques to deal with ordinal variables in a multivariate context have a long history in statistics and are now well-documented (see Jöreskog and Moustaki, 2001, for a survey of different approaches), a widespread approach in empirical research consists of ignoring ordinality and treating the manifest items as continuous. Yet, this inappropriate use may distort the results in several ways, especially when the number of categories is limited, and/or the distributions of the answers show high kurtosis.

In this chapter, the ordinal nature of the $K$ measurements is explicitly accounted for by specifying that each individual has a latent level of agreement $M_k^*$ with the corresponding statement $k$ of the corresponding test, for $k = 1, ..., K$. This latent level of agreement is assumed to linearly depend on some covariates $X_M$ and on the factor $\theta$, and is discretized by a set of cut-points $\{\gamma_k\}$ to produce the observed measurement, with $C$ different alternative ordered answers as follows:

$$\begin{aligned} M_k &= c \quad \text{if } \gamma_{k,c-1} \leqslant M_k^* < \gamma_{k,c}, && c = 1, ..., C, \\ M_k^* &= X_M \beta_{M,k} + \theta \alpha_{M,k} + \varepsilon_{M,k}, && \text{for } k = 1, ..., K, \end{aligned} \tag{2.6}$$

where $\beta_{M,k}$ denotes the vector of parameters associated with $X_M$, $\alpha_{M,k}$ represents the factor loading, and the idiosyncratic error term $\varepsilon_{M,k}$ is assumed to be standard normal and independent of $\theta$ and $X_M$. Assuming standard normality for the error term is the usual solution adopted to guarantee invariance of the latent response variable to scale transformation. As for the cut-points, they are such that $\gamma_{k,0} = -\infty < \gamma_{k,1} = 0 < ... < \gamma_{k,C-1} < +\infty = \gamma_{k,C}$.

#### Latent factor for locus of control

To complete the specification of the model, one last distributional assumption is required for the latent factor $\theta$. In a similar framework, Carneiro et al. (2003), Hansen et al. (2004) achieve nonparametric identification of the latent factors thanks to some independence and support assumptions. When the measurement system consists of a combination of discrete and continuous outcomes, they first nonparametrically identify the joint distribution of the observed and latent measurements, before turning to the identification of the latent factors and error terms using a theorem proposed by Kotlarski (1967). In our case, this identification strategy cannot be applied, insofar as the measurements are all discrete. Nonparametric identification of the latent factor distribution, as well as of the error term distributions, would only be possible if we first managed to nonparametrically identify the joint distribution of the latent measurements. This preliminary stage appears to be difficult when dealing with discrete variables, and requires very strong distributional and support assumptions. Another problem arises because the covariates are sparse and common across measurement equations. The lack of variability and of exclusion restrictions for each measurement makes nonparametric identification and the use of more flexible distributional assumptions such as mixtures impossible. For these reasons, and for the sake of simplicity, we specify a normal distribution and make the following independence assumption:

$$\theta \sim \mathcal{N}\big(0; \sigma_\theta^2\big)\,, \qquad\qquad \theta \perp\!\!\!\perp (X, \varepsilon),$$

where $X = (X_S, X_E, X_Y, X_M)$ and $\varepsilon = (\varepsilon_S, \{\varepsilon_{E,s}\}, \{\varepsilon_{Y,s}\}, \{\varepsilon_{M,k}\})$.

Since the variance of the latent factor is not constrained, we need to impose one restriction to set the scale of $\theta$. For this purpose of identification, we fix one of the factor loadings to a given value in the measurement system.

## 2.4 Estimation strategy

The identification strategy that relies on data set combination is presented in Section 2.4.1, while the estimation method and the data are detailed in Section 2.4.2. The parameters of interest are simulated through the implementation of Bayesian Markov chain Monte Carlo techniques.

### 2.4.1 Combining data sets to identify the model likelihood

Ideally, we would have access to a data set where individuals are observed at different periods of their life cycle. The likelihood of the model for such an hypothetical sample can be expressed as

$$L(\psi|S,E,Y,M,X) = \int_\Theta \prod_{s=0}^{1} \big[\Pr(S=s|X_S,\theta,\psi)\, f(E_s|X_E,\theta,\psi) f(Y_s|X_Y,\theta,\psi)\big]^{\mathbf{1}[S=s]}$$

$$\times \prod_{k=1}^{K} f(M_k|X_M, \theta, \psi)\ \mathrm{d}F_\theta(\theta), \tag{2.7}$$

where $\psi$ represents the vector containing all model parameters, $f(\cdot)$ invariantly denotes a density function, and $F_\theta(\cdot)$ is the cumulative distribution function (cdf) of the latent factor $\theta$ on the support $\Theta$. In our case, this would require information on people's labor market outcomes and personal background, as well as on their premarket locus of control. Estimation based on the likelihood (2.7) would be straightforward.

Unfortunately, the structure of the GSOEP only offers this opportunity for a subsample of the population, which turns out to be too small to conduct any relevant analysis. Although the GSOEP is a longitudinal study, youth are surveyed since 2000 only, and many of them still have not entered the labor market in the last available wave of the survey in 2008. We therefore have to face a major dilemma: on the one hand, we have a large data set of working-age people—adult sample—, but without any information on their locus of control at the time of schooling. On the other hand, a sample of 17-year-olds is available—youth sample—, including premarket locus of control measurements, but labor market outcomes only for a very small group of mostly low-educated individuals. The adult and the youth samples can nevertheless be combined to overcome this problem. We rely on an idea implemented in Cunha et al. (2005), which consists of identifying one part of the likelihood in each subsample, getting rid of the unobserved response variables by integrating them out of the likelihood.

To understand the mechanisms of the data set combination, consider the following sketch of proof. First, derive the contribution to the likelihood of a person with higher education. Since her future labor market participation and wage cannot be observed, they are integrated out to provide

$$\int_\Theta \Pr(S=1|X_S,\theta,\psi) \left\{ \iint f(E_1|X_E,\theta,\psi) f(Y_1|X_Y,\theta,\psi)\,\mathrm{d}F_{E_1}(E_1)\,\mathrm{d}F_{Y_1}(Y_1) \right\}$$
$$\times \prod_{k=1}^{K} f(M_k|X_M,\theta,\psi)\,\mathrm{d}F_\theta(\theta)$$
$$= \int_\Theta \Pr(S=1|X_S,\theta,\psi) \prod_{k=1}^{K} f(M_k|X_M,\theta,\psi)\ \mathrm{d}F_\theta(\theta),$$

where $F_W(\cdot)$ represents the cdf of the corresponding random variable $W$. As a consequence, the parameters of the measurement system and of the schooling equation can be identified from the youth sample. However, due to the small sample size of youth who already earn a wage on the labor market, identification and estimation of the parameters of the labor market participation and wage equations from this sample is impossible.

In a similar fashion, consider a person without higher education from the adult sample, whose premarket locus of control is not observed. Her contribution to the likelihood is

$$\begin{aligned}
&\int_{\Theta} \Pr(S=0|X_S,\theta,\psi)\, f(E_0|X_E,\theta,\psi) f(Y_0|X_Y,\theta,\psi) \\
&\qquad\qquad \times \left\{ \prod_{k=1}^{K} \int f(M_k|X_M,\theta,\psi)\, \mathrm{d}F_{M_k}(M_k) \right\} \mathrm{d}F_\theta(\theta) \\
&= \int_{\Theta} \Pr(S=0|X_S,\theta,\psi)\, f(E_0|X_E,\theta,\psi) f(Y_0|X_Y,\theta,\psi)\ \mathrm{d}F_\theta(\theta),
\end{aligned}$$

and is obtained by integrating out the locus of control measures, which cannot be observed. Full identification of the model is clearly infeasible in this subsample, since no observations on premarket locus of control are available for the adults. However, since we are combining the two data sets and estimating the overall model simultaneously, the distribution of the latent factor is already identified from the youth sample.

Full identification of the model rests on the education equation, which is the only source of common information for most of the sample, and therefore the bridge between the two samples. Although our model can in theory be identified from two non-overlapping samples of youth and adults, in practice we found it helpful to use all available information—i.e., measurement, schooling and labor market information—for the small sample of individuals for whom both labor market outcomes and locus of control measurements are available.

### 2.4.2 Estimation

A fully Bayesian approach is used for the estimation of our model. Since the equations are independent once $\theta$ is conditioned on, the estimation can be divided into several pieces, and Markov chain Monte Carlo (MCMC) methods are particularly suited for this kind of problem. In the wake of Cunha et al. (2005), Carneiro et al. (2003) and Hansen et al. (2004), we use a Gibbs sampler that sequentially draws the parameters of interest from their respective conditional distributions, using flat priors to remain as general as possible.[4]

Data augmentation procedures (Tanner and Wong, 1987) make it possible to simulate the latent outcomes of the measurement system, of the schooling and labor market participation equations, as well as the latent factor $\theta$. Besides the practical convenience of the approach, augmenting the observed data with the latent variables has another major advantage in our case: the simulated latent factors and outcomes can be saved during the sampling process, and used for post-processing analyses, such as simulations.[5] In Section 2.5.2 for instance, these simulated variables are used to assess the fit of the model, and to conduct some formal tests.

Bayesian inference for ordinal variable models can be challenging. Slow convergence and high autocorrelation of the parameter chains are typical symptoms of the algorithm failing to cover the entire posterior distribution of the parameters. As noted by Cowles (1996), the high correlation between the cut-points and the latent response variable results in a poor mixing of the Markov chain for the parameters of Equation (2.6). In the end, this can lead to overinflated standard errors of the parameters, or even worse, to wrong estimates—in terms of bias—if the chain is not long enough to provide a representative sample of the conditional distribution. To remedy this problem, several technical improvements have been proposed.[6] We opt for the group transformation approach introduced by Liu and Sabatti (2000), which speeds up convergence and enhances the mixing of the chain, while being less computationally burdensome than other methods. We run a chain of 1,010,000 iterations for each gender. After a burn-in period of 10,000 iterations, 10,000 iterations are saved every $100^{\text{th}}$ sweep of the Gibbs sampler for post-processing inference. Such a long chain is actually not necessary, since we observe a fast convergence to the stationary distribution, and a good mixing of the chain thanks to the implementation of the group transformation.

### Sample construction

We draw a combined sample of 1,534 youth (age 17-24) and 1,192 'young adults' (age 26-35) from recent waves of the GSOEP. The special feature of the youth sample is that for these youth, a premarket measure of locus of control was administered

[4]For technical details on the Gibbs sampler in this framework, see Chapter 1 where all posterior distributions are derived.

[5]See Dyk and Meng (2001) for a review of data augmentation.

[6]Cowles (1996) introduces a Hastings-within-Gibbs step in the algorithm to draw the cut-points and the latent response variable simultaneously, while Nandram and Chen (1996) propose a simple reparameterization that proves to be particularly effective, especially in the three-category case.

when they were 17 years of age. In the German education system, individuals decide at around the age of 17 whether to finish their studies with a vocational high school certificate, or to continue their schooling with academic high school credentials. Only the latter entitles agents to attend higher education. Hence, our binary education variable reflects this choice of obtaining a vocational or an academic high school degree. Summary statistics of the education variable in the two samples are presented in Table 2.A.2. For a small part of our youth sample (about 280 individuals), no wage and employment information is available. However, because these individuals can be at most 24 years of age, most of them did not achieve higher education. Furthermore, separate estimations by gender and schooling considerably reduce the available sample size. Hence, to obtain more precise estimates, we augment the youth sample with a second sample of young adults, whose education and labor market outcomes can be assumed to be generated by the same data generating process. Summary statistics on wages and employment participation can be found in Table 2.A.3.[7] The table displays that males earn higher wages than females, and that the observed wage gap between high and low educated individuals is higher for males than for females. The low levels of labor market participation arise because many individuals still participate in education or training. To fully account for gender differences in the impact of locus of control on education decisions and outcomes, all estimates are obtained separately for males and females.

### Locus of control measurements

In the GSOEP youth questionnaire, locus of control is measured by a 10-item questionnaire. Each question is answered on a Likert scale ranging from 1 ("disagree completely") to 4 ("agree completely"). Table 2.1 gives an overview of the questions and items we use. We check whether, given these measurements, locus of control can indeed be represented by a single factor. Conducting a principal component analysis, and calculating the eigenvalues of the correlation matrix, we find two eigenvalues larger than 1. Hence, the Kaiser (1960) criterion (eigenvalue$<1$) is violated. However, the scree plot analysis displayed in Figure 2.A.1 reveals an early flattening of the curve, suggesting no more than one or two underlying factors. Furthermore, locus of control is usually conceptualized as referring to a unidimensional continuum, ranging from external to internal. Hence, we think that we are making a reasonable decision by extracting a single factor. A scatter plot of the respective factor loadings (Figure 2.A.3), with the first two principal factors on the

[7] Appendix 2.A provides a more detailed description of the sample construction.

**Table 2.1:** Item definition: locus of control

| Items | Questions |
|---|---|
| Q1 | My life's course depends on me |
| Q2 | I have not achieved what I deserve |
| Q3 | Success is a matter of fate or luck |
| Q4 | Others decide about my life |
| Q5 | Success is a matter of hard work |
| Q6 | In case of difficulties, doubt about own abilities |
| Q7 | Possibilities in life depend on social conditions |
| Q8 | Abilities are more important than effort |
| Q9 | Little control over what happens to me |
| Q10 | Social involvement can help influence social conditions |

axis, shows that some items load very highly on the extracted locus of control factor (factor 1), while some other items have a loading close to zero (Q1, Q5, Q8 and Q10). Furthermore, the items with a close to zero loading are items that capture an internal attitude, while the other items mostly capture the external dimension of locus of control. Consequently, we can draw two conclusions from this exploratory factor analysis. First, researchers who use an index, constructed for example as the standardized mean of the items, instead of a latent factor, force each of the measurement items to enter the index with an equal weight. Doing this yields a locus of control measure that is flawed by measurement error, and the coefficients are likely to be biased downward due to attenuation bias. Second, in this chapter we mostly capture the external attitude dimension of locus of control. For ease of interpretation, in our empirical application we normalize the model such that lower scores of the latent factor are associated with an external locus of control, and higher scores with an internal locus of control. To ensure that our results are not distorted by the inclusion of those items that have a low loading on the locus of control factor, we have conducted robustness checks using only those items loading highly on the first factor. It turned out that the use of the externalizing items only does not have a major impact on the results. The corresponding scree and loadings plots to this analysis are presented in Figures 2.A.2 and 2.A.4.

### Covariates

Table 2.2 summarizes the covariates used for our analysis, and also shows how the two samples are linked by the schooling equation. To account for family background,

**Table 2.2:** Samples and included covariates for the measurement system, education, employment and wage equations

| | Type[a] | Meas. | Educ. | Empl. | Wage |
|---|---|---|---|---|---|
| *Samples* | | | | | |
| Youth sample | | ✓ | ✓ | (✓)[b] | (✓)[b] |
| Adult sample | | — | ✓ | ✓ | ✓ |
| *Covariates* | | | | | |
| Number of siblings | D | ✓ | — | — | — |
| % of time in broken family | C | ✓ | ✓ | — | — |
| Father dropout | B | ✓ | ✓ | ✓ | ✓ |
| Father grammar school | B | ✓ | ✓ | ✓ | ✓ |
| Mother dropout | B | ✓ | ✓ | — | — |
| Mother grammar school | B | ✓ | ✓ | — | — |
| Region: North[c] | B | ✓ | ✓ | ✓ | ✓ |
| Region: South[c] | B | ✓ | ✓ | ✓ | ✓ |
| Childhood in large city[d] | B | ✓ | ✓ | ✓ | ✓ |
| Childhood in medium city[d] | B | ✓ | ✓ | ✓ | ✓ |
| Childhood in small city[d] | B | ✓ | ✓ | ✓ | ✓ |
| Track recommendation (highest)[e] | B | ✓ | — | — | — |
| Track recommendation (lowest)[e] | B | ✓ | — | — | — |
| Local unemployment rate | C | — | — | ✓ | ✓ |
| Local unemployment rate (edu)[f] | C | — | ✓ | — | — |
| Age of individual | C | — | — | ✓ | ✓ |
| Cohort 26/30 | B | — | ✓ | — | ✓ |
| Cohort 31/35 | B | — | ✓ | — | ✓ |
| Married | B | — | — | ✓ | ✓ |
| Number of Children | C | — | — | ✓ | ✓ |

[a] B = Binary, C = Continuous, D = Discrete. [b] Only a small subsample available for these equations. [c] Base category is *West Germany*. [d] Base category is *Childhood in countryside*. [e] Base category is *Recommendation for middle track*. [f] When the education decision is made.

socioeconomic status and labor market conditions, we control for a large range of background variables, as well as for local unemployment rates at the time of education decisions and labor market outcomes, respectively. In addition, Germany has an education system where tracking already takes place after the fourth grade. Hence, to proxy cognitive skills, and to account for the fact that these cognitive skills might affect the items revealing premarket locus of control, we include the primary school teacher track recommendation as a covariate in the measurement system. Because locus of control is estimated from the residual variance, net of covariates in the measurement system, covariates included in the measurement equation are a means to purge locus of control of their influence. However, the inclusion of track recommendation only proxies cognitive skills and the resulting track type. It cannot account for other conflicting effects such as school quality. Hence, locus of control, as identified in this chapter, only captures *premarket* locus of control, and not necessarily *pre-compulsory-school* locus of control.

Summary statistics of control variables in the measurement and outcome equations can be found in Tables 2.A.4 and 2.A.5. Most of these variables are dummy variables, with a low level of observed variability. This is one of the main reasons why nonparametric identification cannot be achieved, thus motivating the use of a fully parametric approach. Appendix 2.A provides a detailed description of the coding of all control variables.

## 2.5 Empirical results

The results are presented and discussed in two stages. We first provide a description of the main findings in Section 2.5.1, with an emphasis on the statistical significance of the impact of locus of control on the different outcomes, and on the fit of our model. Then, we gain more insights in Section 2.5.2 by conducting some simulations that make it possible to better grasp the magnitude of the impact of locus of control.

### 2.5.1 MCMC results

***Factor loadings.*** The factor loadings express how the different measurements and outcomes are affected by the latent factor. The larger the magnitude of the loadings, the higher the contribution of the corresponding items to the distribution of the latent factor. In the education, employment and wage equations, the loadings measure the impact of the factor on the respective outcomes. Cross-model comparisons should however be carefully done: the factor loadings of the different models cannot be directly compared, as their magnitude and their sign depend on the normalization retained to set the scale of the factor. We normalize the factor loading of the fourth indicator to $-1$ in all models, which is a way of anchoring the factor distribution in a real measurement (Cunha and Heckman, 2008).[8] However, contrary to Cunha and Heckman (2008), who anchor the factor in earnings, we cannot give an interpretable metric to the latent factor, because of the ordinal nature of the measurement. Moreover, the respective item of the questionnaire used for the normalization might be perceived differently by males and females, and gender comparisons are therefore not straightforward.

Table 2.3 summarizes the estimation results for the factor loadings of the different models. The results of the measurement system are in line with our expectations. Typical questions associated with an external locus of control such as ‘I have little control over what happens to me’ (Q9), ‘Success is a matter of fate or luck’ (Q3) or ‘I have not achieved what I deserve’ (Q2) have negative factor loadings, whereas the statement reflecting an internal locus of control, ‘My life’s course depends on me’ (Q1), has a positive factor loading. Also, the heterogeneity of these factor loadings is worth noting, as well as the fact that some of them are not significantly different from zero. For instance, in our application the item ‘Success is a matter of hard work’ (Q5) does not measure locus of control *per se*, but rather something related to diligence, and is evicted from the analysis since a non-significant factor loading is assigned to it.

In the outcome system of equations, the factor loading of the education equation is always significant and positive, indicating an actual impact of locus of control. When we do not control for education [columns (1) and (3)], wages appear to be affected by locus of control, whereas this impact vanishes when education is controlled for [columns (2) and (4)]. Hence, we can conclude that *locus of control matters for wages only through the channel of education.*

With respect to the theoretical framework laid out in Section 2.3.1, we can conclude that the impact of premarket locus of control on $w_t^0$ and $w_t^1$, denoted by $\alpha_0$ and $\alpha_1$ respectively, is zero. However, we find that locus of control does have an impact on education decisions ($P(S = 1)$), and thus on wages in the end. Hence, reverting to Equation (2.1), we can conclude that locus of control does not affect education decisions via higher expected wages ($\alpha_0$,$\alpha_1$), but instead through its impact on the cost of education $\alpha_C$.

[8]The fourth indicator is a typical externalizers’ statement, hence the normalization to a negative integer.

**Table 2.3:** Factor loadings of the model estimated by conditioning labor market outcomes on education [(2) and (4)] and without conditioning on education [(1) and (3)]

| | **Males** | | | | **Females** | | | |
|---|---|---|---|---|---|---|---|---|
| | (1) | | (2) | | (3) | | (4) | |
| Measurement system: Locus of control items | | | | | | | | |
| Q1 | 0.354*** | (0.087) | 0.364*** | (0.086) | 0.423*** | (0.095) | 0.440*** | (0.101) |
| Q2 | -0.735*** | (0.119) | -0.729*** | (0.116) | -0.895*** | (0.132) | -0.938*** | (0.143) |
| Q3 | -0.741*** | (0.118) | -0.743*** | (0.116) | -0.619*** | (0.107) | -0.650*** | (0.113) |
| Q4 | -1.000 | — | -1.000 | — | -1.000 | — | -1.000 | — |
| Q5 | 0.013 | (0.074) | 0.024 | (0.075) | 0.026 | (0.085) | 0.025 | (0.089) |
| Q6 | -0.640*** | (0.108) | -0.605*** | (0.102) | -0.890*** | (0.134) | -0.916*** | (0.139) |
| Q7 | -0.559*** | (0.099) | -0.565*** | (0.099) | -0.581*** | (0.105) | -0.617*** | (0.112) |
| Q8 | -0.195*** | (0.072) | -0.197*** | (0.072) | -0.107* | (0.078) | -0.112* | (0.082) |
| Q9 | -1.045*** | (0.175) | -1.035*** | (0.175) | -1.781*** | (0.309) | -1.858*** | (0.332) |
| Q10 | -0.122** | (0.067) | -0.140** | (0.068) | 0.143** | (0.078) | 0.146** | (0.080) |
| Education choice | | | | | | | | |
| $S$ | 0.634*** | (0.134) | 0.404*** | (0.118) | 0.444*** | (0.123) | 0.364*** | (0.127) |
| Labor market participation | | | | | | | | |
| $E$ | 0.055 | (0.136) | | | -0.021 | (0.131) | | |
| $E_0$ | | | 0.757*** | (0.287) | | | 0.357** | (0.222) |
| $E_1$ | | | -0.126 | (0.331) | | | -0.268 | (0.286) |
| log Wages | | | | | | | | |
| $Y$ | 0.181*** | (0.041) | | | 0.121*** | (0.048) | | |
| $Y_0$ | | | 0.007 | (0.060) | | | 0.058 | (0.064) |
| $Y_1$ | | | -0.072 | (0.086) | | | 0.020 | (0.087) |
| Variance of the latent factor | | | | | | | | |
| $\sigma^2_\theta$ | 0.635*** | (0.138) | 0.622*** | (0.135) | 0.446*** | (0.092) | 0.411*** | (0.088) |

**Notes:** Factor loading of item 4 (statement reflecting an external locus of control) fixed to -1 to set the scale of the latent factor. Standard errors in brackets. Significance check: */**/*** if zero lies outside the 90%/95%/99% confidence interval of the posterior distribution of the corresponding parameter.

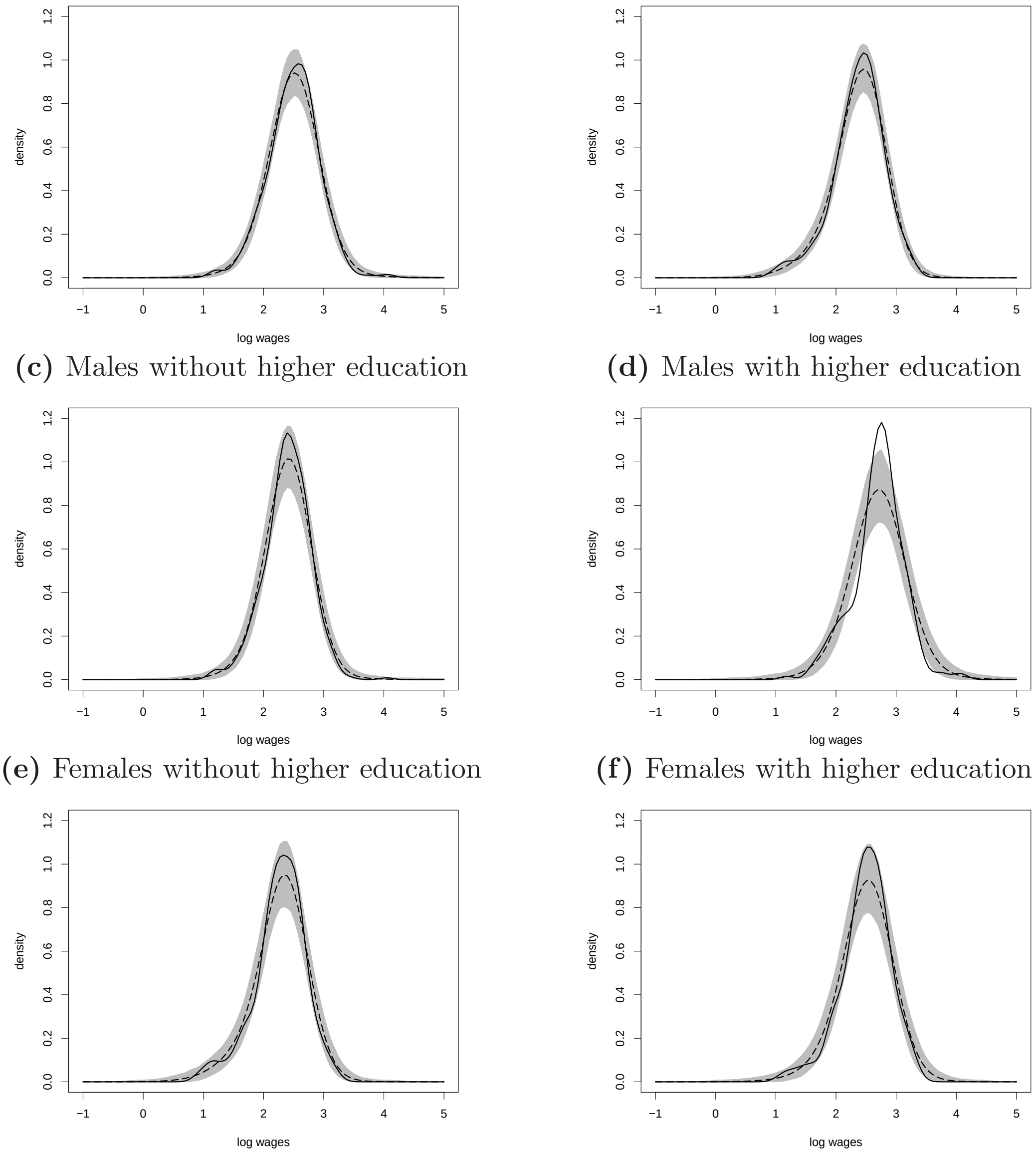
So far, no firm conclusions have been made as to the magnitude of the impact of locus of control on education decisions and overall wages. In the following Section 2.5.2, the simulations we conduct make it possible to unravel and quantify the actual impact of locus of control on the different outcomes of interest.

***Model fit to actual data.*** Our model provides a good fit to the data, and especially to the distribution of wages. Figure 2.1 displays the observed distribution of wages, along with their posterior predictive distribution for the different specifications. The actual distribution is quite well approximated by the posterior predictive distribution, particularly in the case where the two schooling groups are pooled for the estimation of the wage equation (panels 2.1a and 2.1b). When the wage equation is estimated by level of schooling (panels 2.1c, 2.1d , 2.1e and 2.1f), the fit is not as good. Nevertheless, the observed distribution is still contained in the 95% confidence interval of the predictive distribution, except for males with higher education. The smaller number of observations probably explains this deterioration. The Kolmogorov-Smirnov tests we conduct to compare the actual distribution and the posterior predictive distribution never reject the null hypothesis of equal distribution (Table 2.B.2). This result is in great part due to the use of normal mixtures for the error term, allowing for a flexible approximation of the true distribution.

To assess the goodness of fit to the education decision, Table 2.B.3 shows the proportion of correct predictions of education achievement for each decile of the latent factor distribution. The fit appears good overall, especially for the lower deciles of the distribution.

A general observation can be made regarding goodness-of-fit across models. When controlling for education, the fit always appears slightly deteriorated. To better understand why this happens, note that in this configuration, the outcome system is made up of five distinct equations, against three for the case where education is not controlled for. The more equations in the outcome system, the more contaminated the estimation of the latent factor distribution. Because the whole system is estimated simultaneously, the added noise from the extra equations of the outcome system makes it more difficult to extract the distribution of the factor from the ten-equation measurement system. However, this pernicious effect is relatively moderate in our case, and therefore not likely to affect our main conclusions.

**Figure 2.1:** Goodness-of-fit check for wages: posterior predictive (dashed) vs. actual distribution (solid) for the model estimated by conditioning labor market outcomes on education (panels 2.1c, 2.1d, 2.1e and 2.1f) and without conditioning on education (panels 2.1a and 2.1b).




**(a)** Males, overall

**(b)** Females, overall

**(c)** Males without higher education

**(d)** Males with higher education

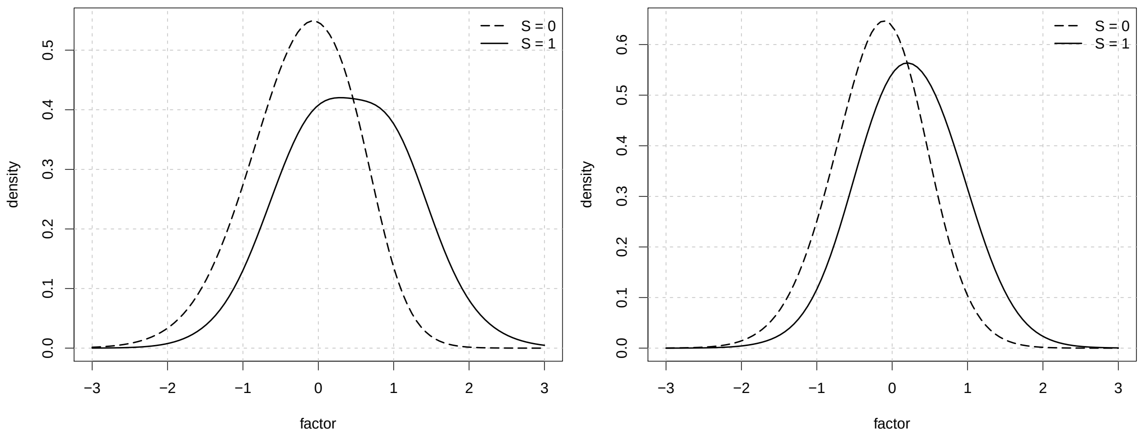
**(e)** Females without higher education

**(f)** Females with higher education

**Notes:** Kernel density estimation implemented using a Gaussian kernel with bandwidth selected using Silverman's rule of thumb (Silverman, 1986) with the variation proposed by Scott (1992). Wages predicted from their posterior distribution using 1,000 replications of the sample. Shaded area represents 95% confidence interval of posterior predictive distribution.

**Figure 2.2:** Latent factor distribution by levels of education: people with higher education ($S = 1$) and without higher education ($S = 0$).

**(a)** Males **(b)** Females




**Notes:** Simulation from the estimates of the model using 1,000 replications of the posterior sample. Model estimated without conditioning labor market outcomes on education. Predicted levels of education used ($\Pr(S = 1) > .5$). Kernel density estimation implemented using a Gaussian kernel with bandwidth selected using Silverman's rule of thumb (Silverman, 1986) with the variation proposed by Scott (1992).

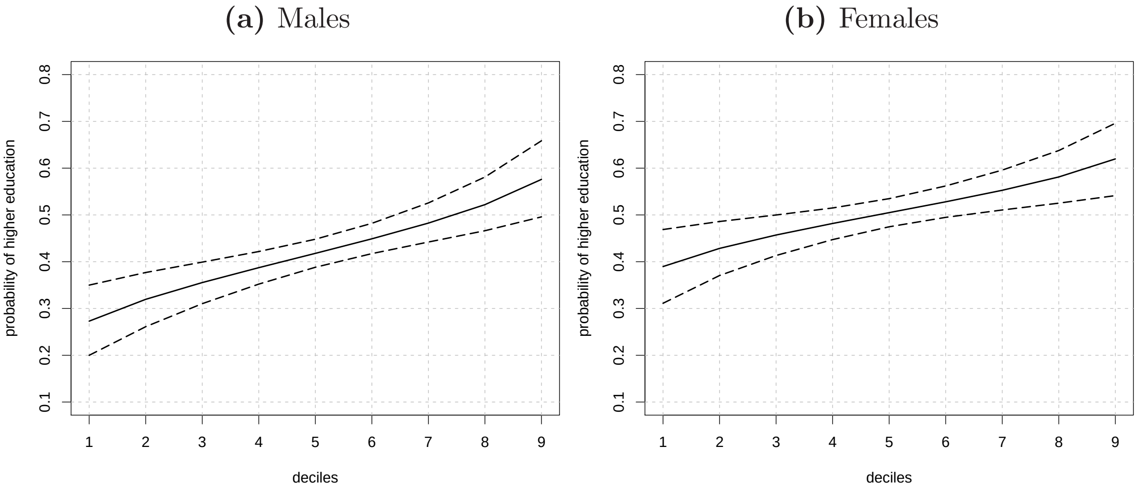
## 2.5.2 Simulation of the model

To shed more light on the implications of our model, we need to go beyond the mere interpretation of the factor loadings. Their statistical significance reveals an impact of locus of control on the outcomes, but is quite uninformative regarding the magnitude of this impact (McCloskey and Ziliak, 1996; Ziliak and McCloskey, 2004). Since the effects of premarket locus of control are intertwined and potentially operate through different channels on wages, the best way to understand our model is to simulate its main features.

Figure 2.2 plots the estimated posterior distribution of the latent factor by levels of education, and shows that people who achieve higher education have a more internal locus of control. For males, the gap between the two schooling groups is even wider, revealing some gender differences in the way locus of control influences education decision. The Kolmogorov-Smirnov test displayed in Table 2.B.1 confirms that the discrepancy between the two distributions is statistically significant for both genders.

To get more insight on the impact of premarket locus of control on later outcomes, various simulations can be carried out. It is for instance interesting, and of particular policy relevance, to know how the wage of a given individual would be affected if she were exogenously moved along the distribution of the latent factor, for a given set of observed characteristics $X_Y$ (Heckman et al., 2006). For this purpose, we compute the expected wage for different quantiles of the distribution of the

**Figure 2.3:** Probability of achieving higher education for each decile of the factor distribution




**Notes:** Simulation from the estimates of the model using 10,000 replications of the posterior sample. Model estimated conditioning labor market outcomes on education. 95% confidence band between dashed lines.

factor, conditional on a given set of covariates $X_Y$. The Gibbs algorithm we implement to estimate our model generates a sample of the model parameters from their conditional distribution. These simulated parameters can then be used as follows to approximate the expected wage for each quantile $q_\theta$ of the factor distribution:

$$\frac{1}{M}\sum_{m=1}^{M}\left(X_Y\beta_Y^{(m)} + q_\theta^{(m)}\alpha_Y^{(m)}\right),$$

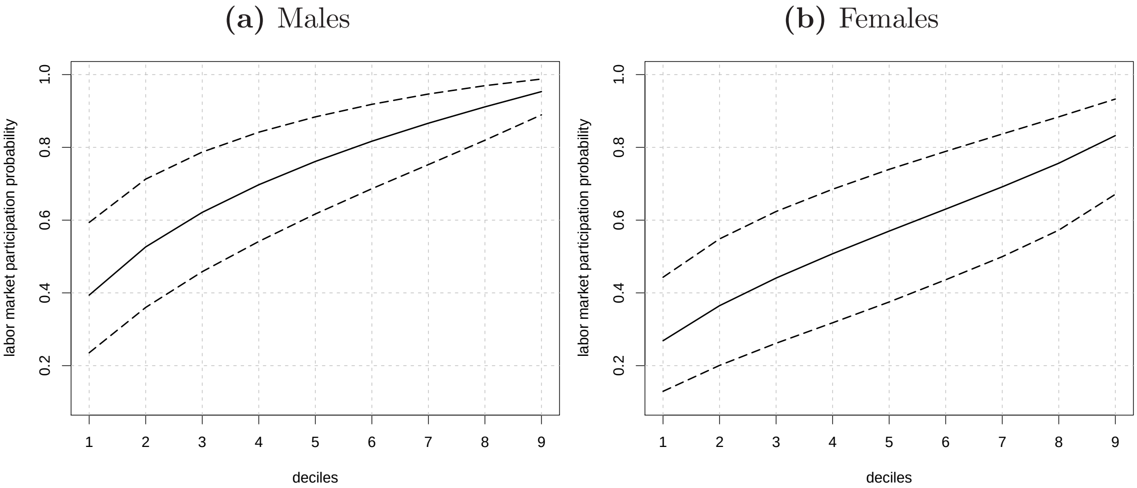
for a set of $M$ simulated parameters $(\beta_Y^{(1)}, \alpha_Y^{(1)}), \ldots, (\beta_Y^{(M)}, \alpha_Y^{(M)})$. The quantile of the latent factor $q_\theta^{(m)}$ also has a superscript $(m)$, since it depends on the variance of the factor $\sigma_\theta^{2(m)}$, and therefore varies during the MCMC sampling. Similarly, the schooling and labor market participation probabilities in the $q^{\text{th}}$ quantile of the latent factor distribution can be approximated by:

$$\frac{1}{M}\sum_{m=1}^{M}\Phi\left(X_S\beta_S^{(m)} + q_\theta^{(m)}\alpha_S^{(m)}\right), \qquad \frac{1}{M}\sum_{m=1}^{M}\Phi\left(X_E\beta_E^{(m)} + q_\theta^{(m)}\alpha_E^{(m)}\right),$$

respectively, where $\Phi(\cdot)$ denotes the cdf of the standard normal distribution. More specifically, the simulations we present rely on the deciles of the distribution. In the following, our simulations are performed for the mean individual of the corresponding sample.

From Figure 2.3, locus of control appears to have a large impact on the schooling decision, since moving the mean individual from the first to the last decile of the

**Figure 2.4:** Probability of labor market participation for people without higher education for each decile of the factor distribution




**Notes:** Simulation from the estimates of the model using 10,000 replications of the posterior sample. Model estimated conditioning labor market outcomes on education. 95% confidence band between dashed lines.

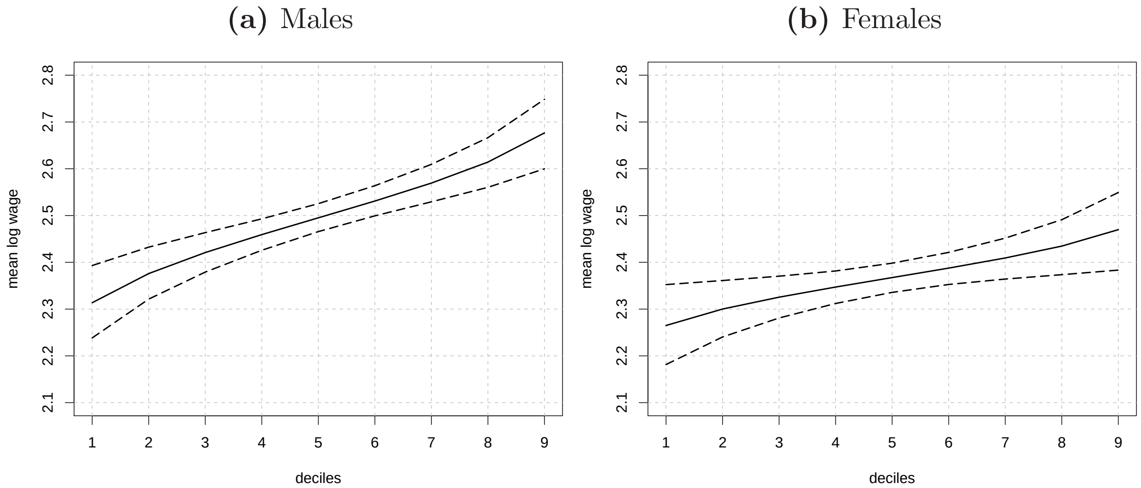
distribution results in a 0.30 point increase in the probability of achieving higher education for males, and a 0.23 point increase for females. Similarly, Figure 2.4 shows that in the group of people who did not achieve higher education, locus of control has a huge impact on labor market participation. This effect is more or less linear for females, whereas for males the concavity of the curve indicates that people in the low deciles are more affected than people in the higher deciles of the distribution. Concerning wages, Figure 2.5 shows that if the mean individual could be moved exogenously from the first to the $9^{th}$ decile of the locus of control distribution, this would corresponds to an increase in hourly wages of roughly 4.40 Euros for the mean male individual, and of roughly 2.20 Euros for the mean female individual.

At first sight, the effect of locus of control on education choice and labor market outcomes seems large. For instance, the mean male individual would earn 36% more in the last decile than in the first one. However, note that it is unrealistic to see an individual move all the way across the distribution. People are more likely to make small moves from one decile to the adjacent ones, and Figures 2.3 to 2.5 show that in the middle of the distribution, the locus of control effect is much smaller.

### 2.5.3 Some remarks on the results

In summary, we find an effect of locus of control on schooling probabilities, where males are more affected than females. Moving the mean individual in the distribution

**Figure 2.5:** Mean log wage for each decile of the factor distribution




**Notes:** Simulation from the estimates of the model using 10,000 replications of the posterior sample. Model estimated without conditioning labor market outcomes on education. 95% confidence band between dashed lines.

of the latent factor substantially changes her/his wage. However, this overall effect only operates through the channel of schooling.

These results might seem inconsistent with some of the literature, where a direct effect of noncognitive skills on wages has been found (Heckman et al., 2006; Heineck and Anger, 2010). Three different answers can be put forward to address this apparent contradiction. First, the term ‘noncognitive skills’ is very often used as a generic expression encompassing a lot of different personal abilities and traits, sometimes leading to confusion. A fair comparison of results can only be made if the same concept is used. For instance, Heckman et al. (2006) find a significant effect of noncognitive skills on wages. However, they use a single underlying factor for noncognitive skills constructed from two psychometric tests, namely the Rosenberg self-esteem scale and the Rotter scale. This composite factor thus captures a different dimension than our factor, especially since it loads more on the self-esteem scale than on the locus of control scale in their empirical study. It is therefore impossible to directly compare our results with theirs. Second, different data sets may yield different conclusions. For example, differences with respect to American data sets may arise from cultural differences in the way personal abilities influence outcomes. Third, and most importantly, we focus on *premarket* locus of control as a measure of locus of control that is independent of labor market experience. Furthermore, we only look at a sample of young labor market entrants. At this stage, wage setting is likely to be foremost a function of formal qualifications. Hence, only after individuals have entered the labor market, a complex dynamic interaction process begins. While working on-the-job, individuals learn about their abilities, while at the same time, employers adapt their knowledge about an individual's locus of control. The result may be a positive interdependence between locus of control and wages, as found for example by Heineck and Anger (2010).

## 2.6 Conclusion

This chapter establishes that an individual's *premarket* locus of control has an influence on schooling decisions. It also shows that this concept of locus of control influences wages through schooling, but that there is no direct impact on wages once schooling is controlled for. Thus, in a framework where schooling decisions depend on relative lifetime earnings returns for each schooling level, net of the costs of obtaining either level of education, we can infer from our results that *premarket* locus of control, as measured at the age of 17, is not directly rewarded as a skill on the labor market. Instead, it is a personality trait that merely influences non-pecuniary costs of education. This finding that *premarket* locus of control influences schooling is in line with Coleman and DeLeire (2003), although in their paper the mechanism through which locus of control affects schooling is different.

We find that a ceteris paribus increase in locus of control substantially raises the probability of choosing higher education for a representative individual. Furthermore, locus of control influences wages only via its effect on schooling, but this indirect effect through schooling can be large. Moreover, this positive effect on schooling is also likely to have a positive influence on other outcomes not considered in this chapter. Some examples are risky behaviors, crime, smoking, employment probabilities and various health outcomes.

Our findings are somewhat contrary to the results presented by Heineck and Anger (2010), who find a strong and significant impact of locus of control on wages, even after controlling for education. One reason could be that the authors do not estimate separate models by education level. More likely, however, the difference in results arises because of the use of contemporaneous measurements in their study, while we focus on the impact of *premarket* locus of control.

Although in our empirical analysis we find that early locus of control does not influence wages directly, we cannot rule out that it has an influence on late locus of control, and that late locus of control is directly rewarded on the labor market. We leave it for future research to find out whether there exists a constant and invariable component to noncognitive skills in general, and to locus of control in particular.

Such a component may be extracted using dynamic factor models, and would require repeated measurements of locus of control over large parts of the life-cycle.

# Bibliography

ANDRISANI, P. J. (1977): "Internal-External Attitudes, Personal Initiative, and the Labor Market Experience of Black and White Men," *Journal of Human Resources*, 12, 308–328.

——— (1981): "Internal-External Attitudes, Sense of Efficacy, and Labor Market Experience: A Reply to Duncan and Morgan," *Journal of Human Resources*, 16, 658–66.

BECKER, G. S. (1964): *Human Capital: A Theoretical and Empirical Analysis, With Special Reference to Education*, New York: National Bureau of Economic Research.

BORGHANS, L., A. L. DUCKWORTH, J. J. HECKMAN, AND B. TER WEEL (2008): "The Economics and Psychology of Personality Traits," *Journal of Human Resources*, 43, 972–1059.

BOWLES, S. AND H. GINTIS (1976): *Schooling in Capitalist America: Educational Reform and the Contradictions of Economic Life*, Routledge.

——— (2002): "Schooling in Capitalist America Revisited," *Sociology of Education*, 75, 1–18.

BOWLES, S., H. GINTIS, AND M. A. OSBORNE (2001a): "Incentive-Enhancing Preferences: Personality, Behavior, and Earnings," *American Economic Review*, 91, 155–158.

——— (2001b): "The Determinants of Earnings: A Behavioral Approach," *Journal of Economic Literature*, 39, 1137–1176.

CALIENDO, M., D. A. COBB-CLARK, AND A. UHLENDORFF (2010): "Locus of Control and Job Search Strategies," Institute for the Study of Labor (IZA) Discussion Paper No. 4750.

CARNEIRO, P., K. HANSEN, AND J. J. HECKMAN (2003): "Estimating Distributions of Treatment Effects with an Application to the Returns to Schooling and Measurement of the Effects of Uncertainty on College Choice," *International Economic Review*, 44, 361–422.

Cebi, M. (2007): "Locus of Control and Human Capital Investment Revisited," *Journal of Human Resources*, 42, 919–932.

Coleman, J. S. (1968): "Equality of Educational Opportunity," *Equity & Excellence in Education*, 6, 19–28.

Coleman, M. and T. DeLeire (2003): "An Economic Model of Locus of Control and the Human Capital Investment Decision," *Journal of Human Resources*, 38, 701–721.

Cowles, M. K. (1996): "Accelerating Monte Carlo Markov Chain Convergence for Cumulative-Link Generalized Linear Models," *Statistics and Computing*, 6, 101–111.

Cunha, F. and J. J. Heckman (2008): "Formulating, Identifying and Estimating the Technology of Cognitive and Noncognitive Skill Formation," *Journal of Human Resources*, 43, 738–782.

Cunha, F., J. J. Heckman, and S. Navarro (2005): "Separating Uncertainty from Heterogeneity in Life Cycle Earnings," *Oxford Economic Papers*, 57, 191–261, The 2004 Hicks Lecture.

Dohmen, T. J., A. Falk, D. Huffman, and U. Sunde (2007): "Are Risk Aversion and Impatience Related to Cognitive Ability?" Institute for the Study of Labor (IZA) Discussion Paper No. 2735.

Duncan, G. and J. Morgan (1981): "Sense of Efficacy and Subsequent Change in Earnings–A Replication," *Journal of Human Resources*, 16, 649–657.

Duncan, G. J. and R. Dunifon (1998): "Soft-Skills and Long-Run Labor Market Success," *Research in Labor Economics*, 17, 123–149.

Dyk, D. A. v. and X.-L. Meng (2001): "The Art of Data Augmentation," *Journal of Computational and Graphical Statistics*, 10, 1–50.

Escobar, M. D. and M. West (1995): "Bayesian Density Estimation and Inference Using Mixtures," *Journal of the American Statistical Association*, 90, 577–588.

Falk, A., D. Huffman, and U. Sunde (2006): "Self-Confidence and Search," Institute for the Study of Labor (IZA) Discussion Paper No. 2525.

Ferguson, T. S. (1983): "Bayesian Density Estimation by Mixtures of Normal Distributions," in *Recent Advances in Statistics: Papers in Honor of Herman Chernoff on his Sixtieth Birthday*, ed. by H. Chernoff, M. Rizvi, J. Rustagi, and D. Siegmund, New York: Academic Press, 287–302.

Flossmann, A. L., R. Piatek, and L. Wichert (2007): "Going Beyond Returns to Education: The Effect of Noncognitive Skills on Wages in Germany," Paper presented at the European Meeting of the Econometric Society, Budapest.

Frey, B. S. and S. Meier (2004): "Pro-Social Behavior, Reciprocity or Both?" *Journal of Economic Behavior and Organization*, 54, 65–68.

Hansen, K. T., J. J. Heckman, and K. J. Mullen (2004): "The Effect of Schooling and Ability on Achievement Test Scores," *Journal of Econometrics*, 121, 39–98.

Heckman, J. J. (1979): "Sample Selection Bias as a Specification Error," *Econometrica*, 47, 153–161.

Heckman, J. J., J. Stixrud, and S. Urzua (2006): "The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior," *Journal of Labor Economics*, 24, 411–482.

Heineck, G. and S. Anger (2010): "The Returns to Cognitive Abilities and Personality Traits in Germany," *Labour Economics*, 17, 535–546.

Hill, M. S. (1979): "The Wage Effects of Marital Status and Children," *Journal of Human Resources*, 14, 579–594.

Jöreskog, K. G. and I. Moustaki (2001): "Factor Analysis of Ordinal Variables: A Comparison of Three Approaches," *Multivariate Behavioral Research*, 36, 347–387.

Judge, T. A. and J. E. Bono (2001): "Relationship of Core Self-Evaluations Traits-Self-Esteem, Generalized Self-Efficacy, Locus of Control, and Emotional Stability-with Job Satisfaction and Job Performance: A Meta-Analysis," *Journal of Applied Psychology*, 86, 80–92.

Kaiser, H. F. (1960): "The Application of Electronic Computers to Factor Analysis," *Educational and Psychological Measurement*, 20, 141–151.

Kotlarski, I. (1967): "On Characterizing the Gamma and the Normal Distribution," *Pacific Journal of Mathematics*, 20, 69–76.

Liu, J. S. and C. Sabatti (2000): "Generalised Gibbs Sampler and Multigrid Monte Carlo for Bayesian Computation," *Biometrika*, 87, 353–369.

McCloskey, D. N. and S. T. Ziliak (1996): "The Standard Error of Regressions," *Journal of Economic Literature*, 34, 97–114.

Mincer, J. (1958): "Investment in Human Capital and Personal Income Distribution," *Journal of Political Economy*, 66, 281–302.

Mueller, G. and E. Plug (2006): "Estimating the Effect of Personality on Male and Female Earnings," *Industrial and Labor Relations Review*, 60, 3–22.

Nandram, B. and M.-H. Chen (1996): "Reparameterizing the Generalized Linear Model to Accelerate Gibbs Sampler Convergence," *Journal of Statistical Computation and Simulation*, 54, 129–144.

Nyhus, E. K. and E. Pons (2005): "The Effects of Personality on Earnings," *Journal of Economic Psychology*, 26, 363–384.

Osborne, M. A. (2000): "The Power of Personality: Labor Market Rewards and the Transmission of Earnings," Ph.D. thesis, University of Massachusetts.

Rotter, J. B. (1966): "Generalized Expectancies for Internal versus External Control of Reinforcement," *Psychological Monographs: General & Applied*, 80, 1–28.

Scott, D. W. (1992): *Multivariate Density Estimation: Theory, Practice, and Visualization*, Wiley Interscience.

Silverman, B. W. (1986): *Density Estimation for Statistics and Data Analysis*, CRC Press.

Tanner, M. A. and W. H. Wong (1987): "The Calculation of Posterior Distributions by Data Augmentation," *Journal of the American Statistical Association*, 82, 528–540.

Trzcinski, E. and E. Holst (2010): "Interrelationship among Locus of Control and Years in Management and Unemployment: Differences by Gender," Discussion Papers of DIW Berlin 974, DIW Berlin, German Institute for Economic Research.

Ziliak, S. T. and D. N. McCloskey (2004): "Size Matters: the Standard Error of Regressions in the American Economic Review," *Journal of Socio-Economics*, 33, 527 – 546.

# Appendix 2.A Data addendum

Our data come from the German Socioeconomic Panel (GSOEP), a representative longitudinal micro-dataset that contains a wide range of socio-economic information on individuals in Germany, comprising follow-ups for the years 1984-2008. Information was first collected from about 12,200 randomly selected adult respondents in West Germany in 1984. After German reunification in 1990, the GSOEP was extended to around 4,500 persons from East Germany, and subsequently supplemented and expanded by additional samples. The data are well-suited for our analysis in that they allow us to exploit information on a wide range of background variables, locus of control and wages, for a representative panel of individuals. Furthermore, the inclusion of a special youth survey, comprising information on 17-year-olds, allows us to obtain background variables and locus of control measures for individuals who have not yet entered the labor market.

## 2.A.1 Combining samples

Our focus is to analyze the impact of locus of control and to purge our estimates of measurement error and endogeneity problems. Hence, to investigate how locus of control affects schooling decisions and wages, respectively, we would ideally need a sample of individuals for whom locus of control measures are collected at several points in time: first, at the time when individuals make education decisions, and second, at a time just before they start the respective job for which labor market returns can be observed. In this way, we would obtain locus of control measures that are truly exogenous, and not influenced by previous on-the-job labor market experience. However, we only have access to one measure of what we term 'premarket' locus of control. This measure is taken when individuals are 17 years of age, just after compulsory schooling, but before they enter the labor market.[9] We then combine the sample of youth for which we have 'premarket' locus of control measures with a sample of young adults for whom we observe labor market outcomes. We draw our samples on the basis of selection criteria that are explained in the following.

[9]Locus of control measures have also been collected for a cross section of young adults in 2007, but we disregard this information, as we suspect it to be flawed by previous labor market experience.

**Youth sample**

Our youth sample is composed of 1,534 individuals born between 1984 and 1991, all of which are children of GSOEP panel members. A comprehensive set of background variables, schooling choices, as well as locus of control measures of these individuals, have been collected in the years 2001-2008, when the subjects were 17 years of age. After the first interview at age 17, all subjects are subsequently interviewed on a yearly basis until early adulthood. For example, in 2008, the oldest youth are 24 years of age. An exception to the age rule was made for the 2001 wave, such that some subjects were already 18 or 19 years of age when first completing the questionnaire. We exclude these individuals from our sample. Besides, to ensure that our results are not flawed by post 1991 schooling and labor market adjustments, all individuals who went to school in East Germany (the former German Democratic Republic) have been excluded. Last, we exclude all individuals with missing locus of control measures, missing schooling information, or missing information among the covariates.

**Adult sample**

The adult sample used for our analysis comprises information on 1,192 individuals, aged 26-35, who are drawn from all West German representative subsamples We construct a cross-section of individuals based on the most recent information available from the waves 2004-2008. Hence, most of our information on the adult sample stems from the 2008 wave. However, if some important pieces of information on certain individuals in that wave are missing, they are filled up with information from 2007. If the information in the 2007 wave is also missing, information from 2006 is used, and so on.

We want to ensure that labor market outcomes and cognitive measures are not related to language problems, post 1991 adjustments, or discrimination. Hence, we exclude non-German citizens, individuals who did not live in West Germany at the time of reunification, as well as individuals whose parents do not speak German as a mother tongue. We also exclude handicapped individuals and individuals in vocational training. Furthermore, we exclude individuals with missing schooling information, because the schooling equation is crucial as it links our two samples and ensures identification. Also, individuals with missings among the control variables are dropped from the sample.

### 2.A.2 'Premarket' locus of control

In the GSOEP, locus of control is measured by a 10-item questionnaire (see Table 2.1). However, the number of possible answers differs between the years 2001-2005, where a 7-point scale was used, and the years 2006-2008, where a 4-point item scale was used. To make the questionnaire comparable across samples, we transform the 7-point scale into a 4-point scale by assigning the middle category (4) either to category 2 or 3 of the 4-item scale, depending on the most probable answer. For example, if in the 2005 sample most youth answered ‘completely agree’, people who answered ‘indifferent’ in the 2006 sample are assumed to tend toward the ‘slightly agree’ answer. After transforming answers to have the same scale, each question is answered on a Likert scale ranging from 1 (‘completely disagree’) to 4 (‘completely agree’).

### 2.A.3 Schooling choice

We group schooling into two broad categories: higher education and lower education. Individuals are classified as being highly educated whenever they have some kind of academic qualification. That is, to qualify as highly educated, individuals need to have passed at least those exams that mark the completion of secondary schooling, and which are obtained in tracks with an academic orientation (German high school diploma—Abitur—obtained either at Gymnasium or Gesamtschule). To identify the level of schooling obtained, we use the international Comparative Analysis of Social Mobility in Industrial Nations (CASMIN) Classification, which is a generated variable available in the GSOEP. We define individuals as being highly educated when their attained education level corresponds to CASMIN categories (2c, 3a, 3b). Similarly, individuals are low-educated if their education status is classified according to CASMIN classification categories (1b, 1c, 2a, 2b). Furthermore, for a subsample of youth who have not completed their education at the time of the last interview, we replace their final education status with their aspired (planned) level of education.

### 2.A.4 Wage construction and labor market participation

Wages are constructed by using most recent wage information available from the GSOEP. Whenever occurring, missing wage information was substituted by wage information obtained in one of the earlier years. Wages have been inflation adjusted to match 2008 wage levels (inflation rates obtained from Eurostat). Wages are assigned a missing whenever the respective individual is indicating not to have a regular (full time or part time) job. We exclude other types of employment such as marginal employment, to ensure that we are not including typical student jobs.

Hourly wages have been constructed by dividing gross monthly wages by the actual number of hours worked in the last month before the interview. Log hourly wages are then obtained by taking the natural logarithm of the hourly wage variable. To account for outliers, we trim hourly wages below the first and above the ninety ninth percentiles. All individuals who indicate a positive wage and are full- or part-time employed are classified as labor market participants.

### 2.A.5 Covariates

In our measurements system, schooling equation and outcome equations, we control for a large set of background variables. The locus of control factor distribution is identified from the covariance structure of the unobservables of the model. Hence, any controls in the measurement system purge our measures of locus of control of any effects which are captured by the covariates. Thus, the covariates in place should be uncorrelated with the latent trait we want to capture, since in our model the latent factor has to be uncorrelated with these covariates by construction. In the following, a brief description of the different categories of covariates is provided.

#### Parental education and investment

*Parental education* variables have been constructed in the form of dummy variables for higher secondary degree (German Gymnasium), lower secondary degree (German Hauptschule or Realschule), dropout and other degree. This information was collected using the Biography Questionnaire, which every person answers when she is first interviewed in the GSOEP.

Apart from parental education, *Parental investment* is proxied by two variables: broken home and number of siblings. Our broken home variable reflects the percentage of childhood time spent in a broken home until the age of 15. This information was also obtained from the Biography Questionnaire. Last, the number of siblings is obtained for the youth by counting the number of siblings living in the household. If an individual has many brothers and sisters, this may indicate that parental time is spread among more individuals, and that overall parental investment is lower.

#### Region dummies and city size

Because school quality and availability, culture and incomes may vary between large and small municipalities, we control for the size of the city where agents spent most of their childhood. Hence, we specify dummy variables for large city, medium city, small city and countryside. Furthermore, we specify four region variables to represent the current region of residence. Hereby, the German Länder are classified as follows:

- North: Berlin, Bremen, Hamburg, Lower Saxony, Schleswig-Holstein,
- South: Bavaria, Baden-Württemberg,
- West: Hessen, North Rhine-Westphalia, Rhineland-Palatinate, Saarland,
- East: Brandenburg, Mecklenburg Western Pomerania, Saxony, Saxony-Anhalt, Thuringia.

### Unemployment rates

We construct unemployment rates at two different points in time. First, we use overall German unemployment at the time when individuals are 17, to have a rough measure of the business cycle when schooling decisions are made. Second, we use region (Länder) specific unemployment rates at the time when labor market outcomes are observed. The latter are important to explain the participation decision, as well as local wage rates. All local unemployment rates are obtained from the Federal Employment Office (Bundesagentur für Arbeit), and overall unemployment from the German Federal Statistical Agency (Bundesamt für Statistik).

### Marital status and number of children

We construct a dummy variable for whether someone is married by looking at her current marital status. Furthermore, we identify the number of dependent children by counting all children for which child benefit payments (Kindergeld) are received by the household. These variables are important, because previous studies show that being married and the number of dependent children have a positive impact on labor market participation and wages for males, and a negative one for females (see, e.g., Hill, 1979, among others).

### Track recommendation after elementary school

We acknowledge that both schooling decisions and locus of control measures may be correlated with cognitive skills. Hence, in order to proxy cognitive skills, and to account for the fact that schooling decisions may depend on prior track attendance, we include an individual's track recommendations after elementary school. In Germany, track recommendations are given to every student during $4^{th}$ grade by their elementary school teachers. In some of the German Länder, track recommendations are non-mandatory (but generally adhered to). In some other Länder, track recommendations are compulsory.

### 2.A.6 Descriptive statistics

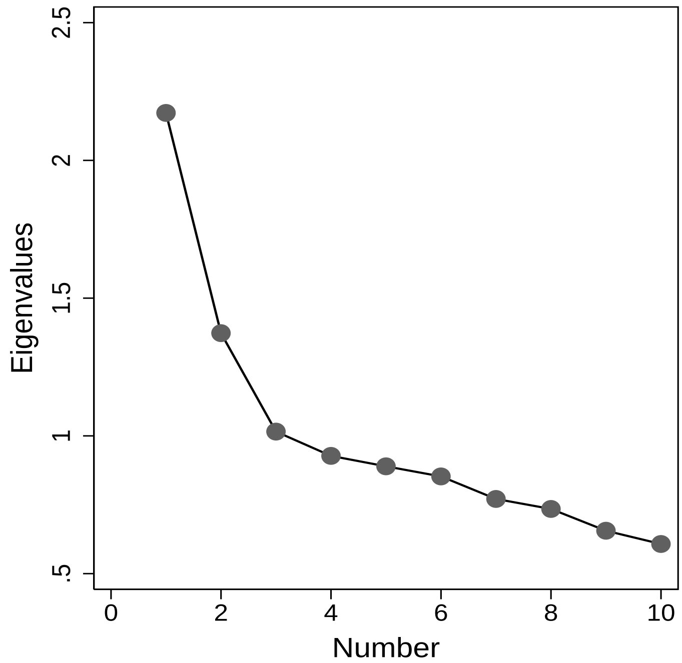



**Figure 2.A.1:** Scree plot: locus of control measurements (10 items)

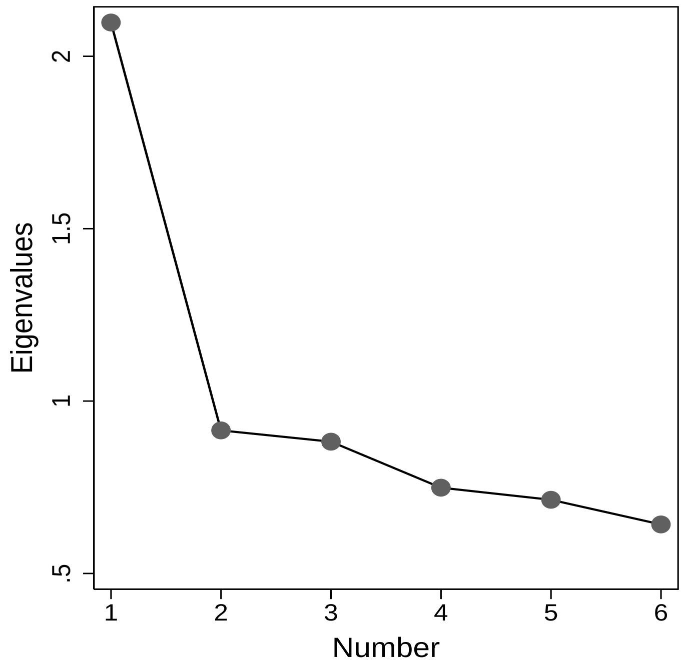



**Figure 2.A.2:** Scree plot: locus of control measurements (6 items with high loadings only, reflecting an external locus of control: Q2, Q3, Q4, Q6, Q7 and Q9)

**Table 2.A.1:** Locus of control, youth sample

| | | Males | | Females | |
|---|---|---|---|---|---|
| | **Variables** | **Mean** | **SD** | **Mean** | **SD** |
| Q1 | My life's course depends on me | 3.55 | 0.63 | 3.51 | 0.59 |
| Q2 | I have not achieved what I deserve | 2.05 | 0.85 | 1.92 | 0.79 |
| Q3 | Success is a matter of fate or luck | 2.22 | 0.81 | 2.29 | 0.77 |
| Q4 | Others decide about my life | 2.18 | 0.83 | 2.12 | 0.83 |
| Q5 | Success is a matter of hard work | 3.48 | 0.62 | 3.51 | 0.57 |
| Q6 | In case of difficulties, doubt about own abilities | 2.08 | 0.81 | 2.31 | 0.85 |
| Q7 | Possibilities in life depend on social conditions | 2.69 | 0.78 | 2.72 | 0.75 |
| Q8 | Abilities are more important than effort | 3.02 | 0.71 | 3.05 | 0.69 |
| Q9 | Little control over what happens to me | 1.92 | 0.75 | 1.95 | 0.76 |
| Q10 | Social involvement can help influence social conditions | 2.48 | 0.87 | 2.51 | 0.77 |
| | # Observations | 760 | | 774 | |

**Table 2.A.2:** Proportion of people with higher education (all samples)

| Variables | Mean | SD | N |
|---|---|---|---|
| Females (youth sample) | 0.518 | 0.500 | 774 |
| Males (youth sample) | 0.459 | 0.499 | 760 |
| Females (adult sample) | 0.461 | 0.499 | 592 |
| Males (adult sample) | 0.368 | 0.483 | 600 |

**Table 2.A.3:** Descriptive statistics: labor market outcomes by schooling

| | High education | | | Low education | | | |
|---|---|---|---|---|---|---|---|
| **Variables** | **Mean** | **SD** | **N** | **Mean** | **SD** | **N** | ***p*-value** |
| Labor market participation (males) | 0.49 | 0.50 | 472 | 0.71 | 0.45 | 617 | 0.00 |
| Hourly wage (males) | 16.03 | 7.16 | 228 | 11.58 | 4.67 | 435 | 0.00 |
| Labor market participation (females) | 0.49 | 0.50 | 553 | 0.58 | 0.49 | 558 | 0.00 |
| Hourly wage (females) | 12.89 | 4.86 | 269 | 10.35 | 4.00 | 316 | 0.00 |

**Source:** GSOEP, cross section using most recent information from the waves 2004-2008. Own calculations.
**Notes:** $p$-values of a two-sided $t$-test for differences in means are reported.

**Table 2.A.4:** Descriptive statistics: covariates in the measurement system

| | Males | | Females | |
|---|---|---|---|---|
| **Variables** | **Mean** | **SD** | **Mean** | **SD** |
| Childhood in large city | 0.20 | 0.40 | 0.22 | 0.42 |
| Childhood in medium city | 0.19 | 0.40 | 0.19 | 0.40 |
| Childhood in small city | 0.29 | 0.45 | 0.25 | 0.44 |
| North | 0.26 | 0.44 | 0.24 | 0.43 |
| South | 0.31 | 0.46 | 0.34 | 0.47 |
| Recommendation: grammar school | 0.39 | 0.49 | 0.45 | 0.50 |
| Recommendation: general secondary school | 0.17 | 0.38 | 0.13 | 0.34 |
| Number of siblings | 0.98 | 1.27 | 1.01 | 1.22 |
| Broken home | 0.24 | 0.43 | 0.24 | 0.43 |
| Father grammar school | 0.29 | 0.45 | 0.33 | 0.47 |
| Father dropout | 0.03 | 0.16 | 0.04 | 0.19 |
| Mother grammar school | 0.23 | 0.42 | 0.25 | 0.44 |
| Mothers dropout | 0.01 | 0.10 | 0.03 | 0.16 |
| # Observations | 760 | | 774 | |

**Source:** GSOEP, cross section using most recent information from the waves 2004-2008. Own calculations.
**Notes:** $p$-values of a two-sided $t$-test for differences in means are reported.

**Table 2.A.5:** Descriptive statistics: covariates in the outcome equations (by schooling)

| | **Males** | | | | | **Females** | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | High education | | Low education | | | High education | | Low eduction | | |
| **Variables** | **Mean** | **SD** | **Mean** | **SD** | *p*-**val** | **Mean** | **SD** | **Mean** | **SD** | *p*-**val** |
| Age | 24.96 | 5.88 | 26.52 | 5.53 | 0.00 | 25.31 | 5.86 | 25.82 | 5.50 | 0.14 |
| Broken home | 0.14 | 0.35 | 0.18 | 0.39 | 0.07 | 0.17 | 0.37 | 0.22 | 0.41 | 0.03 |
| Father grammar school | 0.44 | 0.50 | 0.07 | 0.26 | 0.00 | 0.41 | 0.49 | 0.08 | 0.28 | 0.00 |
| Father dropout | 0.01 | 0.09 | 0.03 | 0.16 | 0.03 | 0.01 | 0.07 | 0.04 | 0.20 | 0.00 |
| Mother grammar school | 0.33 | 0.47 | 0.08 | 0.27 | 0.00 | 0.30 | 0.46 | 0.08 | 0.27 | 0.00 |
| Mother dropout | 0.00 | 0.07 | 0.02 | 0.14 | 0.03 | 0.01 | 0.11 | 0.03 | 0.17 | 0.06 |
| Childhood in large city | 0.24 | 0.43 | 0.18 | 0.38 | 0.01 | 0.21 | 0.40 | 0.20 | 0.40 | 0.71 |
| Childhood in medium city | 0.21 | 0.41 | 0.20 | 0.40 | 0.58 | 0.23 | 0.42 | 0.18 | 0.38 | 0.03 |
| Childhood in small city | 0.26 | 0.44 | 0.24 | 0.42 | 0.29 | 0.25 | 0.44 | 0.25 | 0.43 | 0.77 |
| North | 0.24 | 0.43 | 0.21 | 0.41 | 0.25 | 0.23 | 0.42 | 0.20 | 0.40 | 0.27 |
| South | 0.31 | 0.46 | 0.33 | 0.47 | 0.58 | 0.28 | 0.45 | 0.34 | 0.47 | 0.06 |
| Unemployment at schooling decision | 9.01 | 1.30 | 8.93 | 1.37 | 0.35 | 9.03 | 1.39 | 9.03 | 1.34 | 0.97 |
| Unemployment | 7.47 | 2.90 | 7.72 | 3.25 | 0.18 | 7.70 | 3.06 | 7.52 | 3.12 | 0.33 |
| Married | 0.16 | 0.37 | 0.23 | 0.42 | 0.01 | 0.18 | 0.38 | 0.27 | 0.45 | 0.00 |
| Number of children | 1.03 | 1.12 | 0.79 | 1.01 | 0.00 | 0.96 | 1.17 | 0.92 | 1.12 | 0.55 |
| # Observations | 472 | | 617 | | | 553 | | 558 | | |

**Source:** GSOEP, cross section using most recent information from the waves 2004-2008. Own calculations.
**Notes:** *p*-values of a two-sided *t*-test for differences in means are reported.

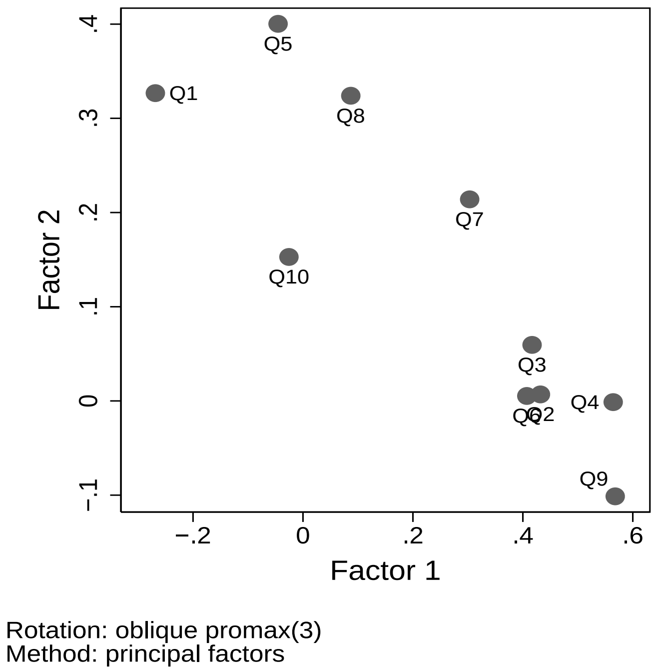



**Figure 2.A.3:** Scatterplot of loadings: locus of control measurements (10 items)

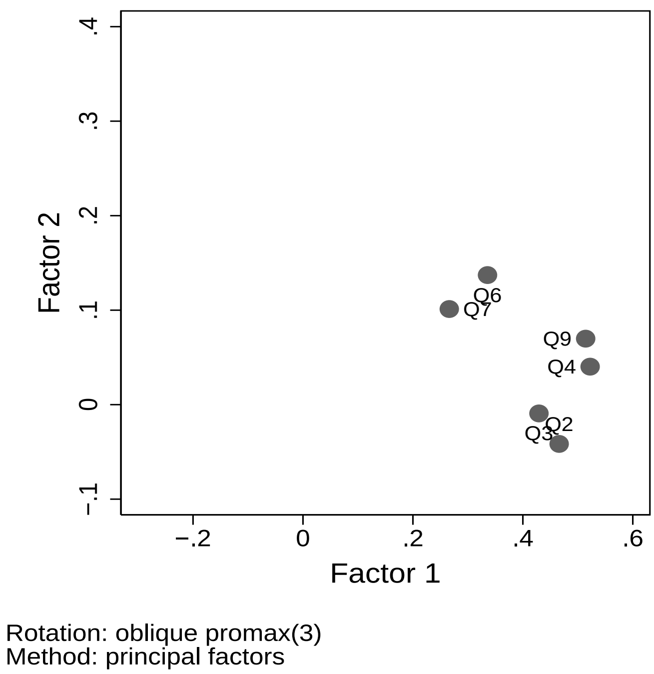



**Figure 2.A.4:** Scatterplot of loadings: locus of control measurements (6 items with high loadings only, reflecting an external locus of control: Q2, Q3, Q4, Q6, Q7 and Q9)

# Appendix 2.B Goodness-of-fit tests

**Table 2.B.1:** Test for equality of distributions of the latent factor across schooling groups

| | Males | | Females | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| Kolmogorov-Smirnov test | | | | |
| Stat. | 0.298 | 0.162 | 0.242 | 0.191 |
| ($p$–val) | (0.000) | (0.000) | (0.000) | (0.000) |

**Notes:** Model estimated by conditioning labor market outcomes on education [(2) and (4)] and without conditioning on education [(1) and (3)]. Two-sample K-S test with null hypothesis that the distribution of the latent factor is the same for the two education groups. Exact $p$-values could not be computed due to ties in the distributions.

**Table 2.B.2:** Goodness-of-fit test for log wages (Kolmogorov-Smirnov test)

| | Males | | Females | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| Kolmogorov-Smirnov test | | | | |
| overall | 0.026 | | 0.027 | |
| | (0.745) | | (0.799) | |
| $S = 0$ | | 0.085 | | 0.055 |
| | | (0.072) | | (0.398) |
| $S = 1$ | | 0.042 | | 0.047 |
| | | (0.418) | | (0.479) |

**Notes:** Model estimated by conditioning labor market outcomes on education [(2) and (4)] and without conditioning on education [(1) and (3)]. Two-sample K-S test with null hypothesis that the actual sample and the posterior predictive sample have the same distribution. $p$-values in brackets. Exact $p$-values could not be computed due to ties in the distribution of actual wages.

**Table 2.B.3:** Goodness-of-fit check: proportion of correct predictions of education achievement for each decile of the latent factor distribution

| | | Deciles of latent factor distribution | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 10% | 20% | 30% | 40% | 50% | 60% | 70% | 80% | 90% | 100% |
| Males | (1) | 0.827 | 0.816 | 0.804 | 0.785 | 0.764 | 0.737 | 0.708 | 0.673 | 0.645 | 0.699 |
| | | (0.033) | (0.031) | (0.032) | (0.033) | (0.035) | (0.035) | (0.035) | (0.038) | (0.039) | (0.044) |
| | (2) | 0.788 | 0.785 | 0.777 | 0.762 | 0.745 | 0.728 | 0.711 | 0.691 | 0.671 | 0.669 |
| | | (0.033) | (0.033) | (0.034) | (0.034) | (0.035) | (0.036) | (0.036) | (0.039) | (0.040) | (0.040) |
| Females | (3) | 0.771 | 0.746 | 0.728 | 0.714 | 0.702 | 0.689 | 0.677 | 0.667 | 0.667 | 0.692 |
| | | (0.033) | (0.035) | (0.036) | (0.037) | (0.037) | (0.036) | (0.037) | (0.038) | (0.041) | (0.040) |
| | (4) | 0.756 | 0.736 | 0.722 | 0.709 | 0.698 | 0.689 | 0.678 | 0.666 | 0.659 | 0.675 |
| | | (0.034) | (0.037) | (0.037) | (0.037) | (0.037) | (0.039) | (0.039) | (0.039) | (0.039) | (0.039) |

**Notes:** Model estimated by conditioning labor market outcomes on education [(2) and (4)] and without conditioning [(1) and (3)]. Proportions of correct predictions computed for each MCMC replication, corresponding means and standard errors (in brackets) are reported.

# Chapter 3

# Constructing Justified Aggregates

## An Application to the Early Origins of Health

## 3.1 Introduction

This chapter presents novel Bayesian econometric methods for reducing high-dimensional data into low-dimensional aggregates using factor models to examine the effect of early-life conditions and education on health. A burgeoning literature in Bayesian econometrics currently focuses on the estimation of factor models with an unknown number of latent factors. In this chapter, we implement the *Parsimonious Bayesian Factor Analysis* (PBFA) of Frühwirth-Schnatter and Lopes (2009). Their new Bayesian approach to factor analysis and its posterior estimation procedure goes beyond traditional approaches like preliminary exploratory factor analysis or scree plots that are widely used in practice. It represents an important step forward in probabilistically determining the latent structure of the model, as well as the posterior distribution for the number of common latent factors. In fact, the recent literature is rich on highly structured factor analysis, mostly due to increasing degree of complexity of data and/or applications in modern applied sciences; see Lopes and West (2004), Lopes and Carvalho (2007), Lopes et al. (2008) and Carvalho et al. (2008), among others.

We apply this modern modeling and estimation technique to the 1970 British Cohort Study (BCS) to analyze the effect of early-life cognition and psychosocial attributes on education and later-life health. These data are unusually rich in terms of both the quantity and the quality of measurements on early cognitive and personality traits. In this application, we focus on seven cognitive scales and five psychosocial scales collected from the cohort members when they were aged ten, and also from their mothers and from their teachers. The total number of items amounts to more than one hundred twenty, and covers a wide range of cognitive abilities, personal attitudes, and behaviors. However, this richness has rarely been fully exploited, and so far the available measurements have been aggregated and used in various ways by different researchers. We provide a method for exploiting the available information in a systematic way that does not rely on arbitrary a priori decisions on model structure and sub-scale construction. Our method has wide applicability to several other circumstances in which researchers face a data-rich environment but are agnostic as to the underlying latent structure.

This method is applied within a life course framework to analyze the effect of childhood cognitive ability and psychosocial traits on education and adult health. This constitutes the first instance in which the market and non-market returns to education are estimated in a modeling framework that simultaneously accounts for selection on gains and endogenous determination of the early factors driving the selection process. Previous work has established a significant role played by early-life conditions on adult health. We assess the sensitivity of these results to misspecification of the latent structure underlying the observable measurements used as proxies for early intelligence and personality traits. We show that not properly specifying the multiple common factors that drive the correlation between education and adult outcomes might lead to an incorrect assessment of the importance of early-life conditions. Both traditional principal component analysis and standard factor analysis approaches indeed indicate a less important role of cognitive and noncognitive traits in determining later health. In addition to this, we revisit the causal effect of education on health in a model where individuals are allowed to select into education on the basis of their idiosyncratic gains, and we show the importance of accounting for heterogeneity in returns the agents act upon, in order to reconcile the different findings of the education-health literature. We provide evidence that the average treatment effect of education for females at different margins of the distribution of the unobservables varies in a way that reconciles the contradictory findings reported in the literature.

This chapter is structured as follows. In the next section, we provide an overview of the literature we refer to. Section 3.3 describes the data and the rich measurements we use. Section 3.4 outlines the potential outcomes model, while Section 3.5 gives an overview of the steps for fully Bayesian posterior inference and model choice. Empirical implementation and results are shown in Section 3.6. In Section 3.7, we compare our results with those obtained using classical methods. Section 3.8 concludes.

## 3.2 Literature

With this empirical application, we join different strands of the literature in economics, epidemiology and psychology.

The first strand refers to the relationship between health and cognitive ability. While the importance of the 'ability bias' has long been recognized in labor economics (see, for example, Griliches, 1977), the effect of cognitive ability on health has received relatively less attention.[1] However, this topic has recently received considerable attention in the field of cognitive epidemiology: large epidemiological studies have found that intelligence in childhood predicts substantial differences in

[1] Grossman (1975), Hartog and Oosterbeek (1998), Auld and Sidhu (2005), Cutler and Lleras-Muney (2010), Kaestner (2009) are the only exceptions we know of.

adult morbidity and mortality (Whalley and Deary, 2001; Gottfredson and Deary, 2004; Batty et al., 2007).

The second strand refers to the relationship between personality traits and health. While there is already an established tradition in psychology on their importance (see, for example, Roberts et al., 2006; 2007; Hampson and Friedman, 2008), economists have just started to explore the effects of personality traits on health (Kaestner, 2009) and health-related behaviors (Heckman et al., 2006; Cutler and Lleras-Muney, 2010; Conti et al., 2010).

Our work constitutes an important methodological advancement in these different strands of literature. Previous research has not been able to fully exploit such a data-rich environment, and has usually made a priori choices of the measurements to include in the analysis. Moreover, it has used rather crude methods—such as the scree test—for selecting the number of factors. While in many cases researchers have been limited by the availability of the measurements present in the data (for example, Kaestner, 2009, uses the NLSY79, where only the Rosenberg's self-esteem and the Rotter's locus of control scales are available), in other cases they have focused the analysis only on a subset of the available measurements. For example, Gale et al. (2008), using the same British cohort data that we analyze in this chapter, focus on the locus of control scale, and find that a more internal locus of control is a protective factor for health later in life, and partially mediates the effect of IQ. Their measure of cognitive ability ($g$) is the first unrotated component from the principal component analysis of the four British Ability Scales, while they use the raw score (obtained by simply summing the scores on the individual items) for the locus of control scale. Another paper closely related to our work is Gale et al. (2009), where the authors use both the 1958 and 1970 British Cohort Studies to examine the effect of intelligence in childhood on risk of psychological distress in adulthood. Their measure of cognitive ability is the same as in Gale et al. (2008), while as measure of behavior problems in childhood they use three factors extracted after oblique rotation following a principal component analysis of the Rutter Parental 'A' Scale of Behavior Disorder, that they label 'Anti-social Behavior', 'Anxiety' and 'Attention Problems'. For both cognitive ability and behavior problems, the number of components is selected using the scree plot, and the factors extracted are significant determinants of psychological distress in adulthood. Finally, Stumm et al. (2009) have the closest paper to our study, as they employ structural equation modeling using full information maximum likelihood to analyze the effect of childhood intelligence (a '$g$' factor extracted as in Gale et al., 2008) and behavioral problems on intergenerational social mobility. In contrast to previous work, they use an extensive set of indicators of childhood behavior, such as the Child Developmental Scale, the Rutter and the Conners scales, and the Self-Esteem and Locus of Control scales; however, like all previous studies, they also conduct preliminary factor analysis (a principal axis factor analysis, PFA, in this case) to select the factors to include in their structural equation model. More specifically, they apply PFA separately to the Child Developmental Scale and to the Rutter and Conners scales. For the Child Developmental Scale, the scree plot suggests a four-factor solution after oblique rotation, and the four factors are labeled 'Anger', 'Anxiety', 'Concentration Difficulties', and 'Hand Skills'. For the Rutter and Conners scales, the scree plot also suggests a four-factor solution after oblique rotation, and the four factors are labeled 'Restlessness', 'Clumsiness', 'Aggression', and 'Attention Deficit'.[2] By contrast, we use all the individual items simultaneously in our analysis.

A few studies using the BCS data can also be found in the economic literature. Blanden et al. (2007) use principal component analysis to construct several noncognitive measures, such as a Rutter Internalizing and a Rutter Externalizing score from the Rutter scale, and an Hyperactive, Application and Anxious scale using selected items from the Child Developmental Scale, to analyze their role in explaining the rise in intergenerational income persistence across the 1958 and the 1970 cohorts.[3] Feinstein (2000) also uses principal component analysis to construct several indicators of psychological and behavioral development, and analyzes their effects on education and labor market outcomes. Murasko (2007), instead, uses the standardized raw scores from the locus of control and self-esteem scales (obtained by simply summing the scores on the single items), and shows that both are significant predictors of self-reported poor health at age 30.[4] Our analysis goes beyond all these previous studies, as we simultaneously identify the latent structure underlying *all* measurements, and we estimate the effect of the factors on both labor market and health outcomes.

The final strand of literature we refer to is that on the non-market returns to education. The positive correlation between education and health has long been recognized in the economic, epidemiologic and medical literature, and several attempts at disentangling correlation from causality have been made.[5] Our methodology allows us to disentangle the average fraction of the health gap by education that can be explained by these early factors from that which can be attributed to the causal

[2]In the construction of the aggregate scores for the locus of control and self-esteem scales, they exclude items with low inter-item correlations, whereas no items are discarded in our analysis.

[3]They do not use the Conners scale, which is included in our analysis.

[4]He does not use the Child Developmental Scale, nor the Rutter and the Conners scales.

[5]See Grossman (2006) for a review.

effect of education. More importantly, it allows us to shed some light on the differences in the findings that have emerged from various studies, where researchers have reached opposite conclusions on the effect of education on health. We refer in particular to two widely-cited studies. Currie and Moretti (2003), using college openings as an instrument for maternal education, find a strong positive effect of education on reduction in smoking during pregnancy. In contrast, McCrary and Royer (2010), using age-at-school entry policies to identify the effect of education on smoking during pregnancy, find much smaller and often statistically insignificant effects. They notice that their instrument affects women at risk of dropping out of high school, while college openings affect women with a higher level of education. Lindeboom et al. (2009) use the 1947 increase of the minimum school-leaving age in UK and obtain insignificant results, in line with McCrary and Royer (2010). In this chapter, we also show that understanding how the returns vary, at different margins of the unobservables driving the educational choice, is the key to reconcile these contradictory findings.

## 3.3 Data

We use data from the British Cohort Study (BCS), a survey of all babies born (alive or dead) after the 24$^{th}$ week of gestation from 00.01 hours on Sunday, 5$^{th}$ April to 24.00 hours on Saturday, 11$^{th}$ April, 1970 in England, Scotland, Wales and Northern Ireland.[6] There have been seven follow-ups so far to trace all members of the birth cohort: in 1975, 1980, 1986, 1996, 2000, 2004 and 2008. We draw information on background characteristics from the birth survey, on cognitive and noncognitive measurements from the second sweep (age 10) and on education and adult health from the fifth sweep (age 30). We remove children born with congenital abnormalities and non-whites (or those with missing information on ethnicity), and we delete responses with missing information on the covariates. We are left with a sample of 2,293 women and 2,398 men.

### 3.3.1 Schooling and post-schooling outcomes

The outcomes considered in our model are:

[6] The original name of the data was the British Births Survey (BBS), sponsored by the National Birthday Trust Fund in association with the Royal College of Obstetricians and Gynecologists.

**Table 3.1:** Health and wage outcomes by levels of education

| **Males** | | | | |
|---|---|---|---|---|
| | $d = 0$ | $d = 1$ | $\Delta$ | $p$-value |
| Self-reported Poor Health | 0.2083 | 0.1166 | -0.0917 | 0.0000 |
| Obesity | 0.1509 | 0.0968 | -0.0541 | 0.0010 |
| Daily Smoking | 0.4129 | 0.2072 | -0.2057 | 0.0000 |
| Wage | 1.7818 | 2.0086 | 0.2268 | 0.0000 |
| **Females** | | | | |
| | $d = 0$ | $d = 1$ | $\Delta$ | $p$-value |
| Self-reported Poor Health | 0.1765 | 0.1049 | -0.0716 | 0.0000 |
| Obesity | 0.1139 | 0.0894 | -0.0244 | 0.0904 |
| Daily Smoking | 0.3529 | 0.1786 | -0.1744 | 0.0000 |
| Wage | 1.6930 | 1.9058 | 0.2128 | 0.0000 |

**Notes:** For each row, the first two cells show the proportion of individuals in the sample who have that outcome among those with a low and a high level of education, respectively. The third cell shows the difference between the two, and the fourth cell shows the $p$-value of a two-sided test for the significance of that difference.

- Schooling. Our schooling measure is a dummy variable indicating whether or not the individual has achieved as the highest level of educational qualification an A-level or above (i.e., a Diploma or a Degree).[7]
- Post-Schooling Outcomes. We analyze three health outcomes, all measured at age 30: self-reported poor health, obesity and daily smoking.[8] We also analyze a labor market outcome: (log) hourly wages, also measured at age 30.

Outcomes by education are presented in Table 3.1. Despite the young age of the sample (30 years), big and significant differences across schooling groups already appear, especially in smoking behaviors, which are only slightly larger for males. This

[7] Under the category "Degree" the following qualifications are grouped: Higher degree or degree; National Vocational Qualification level 5 or 6; Postgraduate certificate in education. Under the category "Diploma" the following qualifications are grouped: Higher Education diploma; Nursing qualification; Higher National Certificate/Higher National Diploma; Technician Education Council: Higher Certificate/Higher National/Higher National Diploma; National Vocational Qualification Level 4; City and Guilds Full Technological Diploma; Full- and part- professional qualification.

[8] The variable "smoking" takes the value 1 if the individual reports smoking cigarettes every day. The variable "poor health" takes the value 1 if the individual reports his/her health to be generally "fair" or "poor". The variable "obesity" is constructed in the standard way as having a Body Mass Index (BMI) greater than 30, where the BMI is calculated as the weight in kilograms divided by height in meters squared.

underlines the importance of understanding the role played by early-life behavioral and cognitive factors to help preventing the emergence and the widening of health disparities in adulthood.

### 3.3.2 The measurement system

The measurement system includes more than one hundred twenty measurements of cognitive and noncognitive traits, and consists of both binary and continuous measurements.

***Cognitive Abilities.*** As indicators of cognitive ability, we use the following seven tests administered to the cohort members when they were aged ten:

1. The Picture Language Comprehension Test [PLCT] is a new test specifically developed for the BCS on the basis of the American Peabody Picture Vocabulary Test and the English Picture Vocabulary Test; it covers vocabulary, sequence and sentence comprehension.

2. The Friendly Math Test [FMT] is a new test specifically designed for the BCS; it covers arithmetic, fractions, algebra, geometry and statistics.

3. The Shortened Edinburgh Reading Test [SERT] is a shortened version of the Edinburgh Reading Test, which is a test of word recognition particularly designed to detect poor readers; it covers vocabulary, syntax, sequencing, comprehension, and retention.

4–7. The four British Ability Scales [BAS] measure a construct similar to IQ, and include two verbal scales (Word Definition [WD] and Word Similarities [WS]) and two non-verbal scales (Recall Digits [RD] and Matrices [M]).

***Noncognitive Abilities.*** As indicators of noncognitive ability, we use the items from the following five tests, all administered during the age ten sweep:[9]

1. The Rutter Parental 'A' Scale of Behavioral Disorder (Rutter et al., 1970) contains 19 items that are descriptions of behavior. This instrument was administered to the mother, who was asked to indicate whether each description 'does not apply', 'applies somewhat' or 'definitely applies' to the child, on a scale from 0 to 100.[10]

[9] The items from the various measurement scales are all listed in Appendix 3.B.

[10] A visual analogue scale was used: the mother had to draw a vertical line through the printed line to show how much a behavior applied (or not) to the child.

2. The Conners Hyperactivity Scale (Conners, 1969) contains 19 items and helps assess attention deficit/hyperactivity disorder and evaluate problem behavior in children and adolescents. This scale was also administered to the mother, who was asked to indicate whether each description applied to the child on a scale from 0 to 100.[11]

3. The Child Developmental Scale contains 53 items, answered by a teacher with knowledge of the child, to assess the child neurodevelopmental behavior against the 'average' behavior of most children of a similar age. Teachers were asked to indicate their level of agreement with each statement by bisecting a line, which was coded into a 47-point scale ranging from "Not at all" to "A great deal".[12]

4. The Locus of Control (Caraloc) Scale includes 16 items[13] that measure the child's perceived achievement control. It was constructed from several well known tests of locus of control (Gammage, 1975). It was administered by the teacher and completed by the child, who was asked to answer 'yes', 'no' or 'don't know', where the answer 'no' represents a more internal locus of control,[14] which is desirable and also referred to as "self-agency", "personal control", or "self-determination".

5. The Self-Esteem (Lawseq) Scale includes 12 items which measure the child's self-esteem with reference to teachers, peers and parents.[15] It was created by former Chief Educational Psychologist of Somerset LEA Lawrence (Lawrence, 1973; 1978). It was administered by the teacher and completed by the child, who was asked to answer 'yes', 'no' or 'don't know', where the answer 'no' represents a higher level of self-esteem.[14]

For use in our empirical application, all the items of the Rutter Parental 'A' Scale, the Conners Hyperactivity and the Child Developmental Scales have been standardized with mean zero and standard deviation one. All the answers of the Locus of Control and Self-Esteem scales have been recoded into binary measurements, by giving a value of 1 to all the 'no' answers, and a value of 0 to all the 'yes' and 'don't know' answers.

[11] A visual analogue was used like for the Rutter Scale.

[12] The items for the scale were taken mainly from the Conners Teachers Hyperactivity Rating Scale (Conners, 1969) and the Rutter Teacher Behavioral Scale B (Rutter, 1967), and questions from the Swansea Assessment Battery (with permission of Professor Maurice Chazan; see Butler et al., 1997).

[13] The total number of questions included is 20, but four of them are distractors.

[14] Only one question is reverse-scored, and we have recoded it accordingly.

[15] The total number of questions included is 16, but four of them are distractors.

### 3.3.3 Control variables

The following control variables are included both in the measurement system and in the outcome system ($\boldsymbol{x}$): mother's age at birth, mother's education at birth (a dummy variable for whether the mother continued education beyond the minimum school-leaving age), father's high social class at birth,[16] total gross family income at age 10,[17] an indicator for broken family (a dummy variable for whether the child lived with both parents since birth until age 10), the number of previous livebirths, and the number of children in the family at age 10.

## 3.4 A potential outcomes model for health outcomes and wages

We consider that individuals select into education, defined in our empirical analysis as achieving A-level or above, depending on some personal characteristics observed by the econometrician ($\boldsymbol{x}$), and on their cognitive abilities and personality traits. These traits cannot be directly observed, but are revealed through a series of psychometric tests performed during childhood, and captured in our model by the set of latent factors $\boldsymbol{f}$. Later in life, these individuals receive different health and wage outcomes that depend on their level of education, but also on these latent traits. Thus, cognitive skills and personality affect the outcomes both directly and indirectly through education. The latent variable model we rely on produces structure on the Neyman (1923), Fisher (1935), Cox (1958), Rubin (1974) model of potential outcomes, and enables us to disentangle the different channels through which the latent traits affect the outcomes of interest (Section 3.4.1). Within this framework, various treatment effects can be generated to gain more insight into the impact of education on health and wages, and on the heterogeneity of this impact due to the unobserved latent traits (Section 3.4.2).

[16]High Social Class comprises SCI, SCII and SCIIINM (Non-Manual) . The BCS uses the Registrar General's classification for measuring social class (SC). Social class I includes professionals, such as lawyers, architects and doctors; Social Class II includes intermediate workers, such as shopkeepers, farmers and teachers; Social Class III Non Manual includes skilled non-manual workers, such as shop assistants and clerical workers in offices.

[17]This is a categorical indicator taking the following values: 1=under £35 pw; 2=£35-49 pw; 3=£50-99 pw; 4=£100-149 pw; 5=£150-199 pw; 6=£200-249 pw; 7=£250 or more per week.

### 3.4.1 The model

The latent variable model used in this chapter is essentially similar to the one presented by Heckman and Vytlacil (1999) and Aakvik et al. (2005), among others. Each individual $i$ in $1, \ldots, n$ (2,293 women and 2,398 men) chooses the level of education that maximizes her/his utility. More precisely, we assume that the education outcome is a Bernoulli variable $d_i$ that is related to $v$ observed characteristics $\boldsymbol{x}_i$ and $k$ latent factors $\boldsymbol{f}_i$ via the unobservable latent continuous variable $d_i^\star$:

$$\begin{aligned} d_i^\star &= \boldsymbol{\beta}_d' \boldsymbol{x}_i + \boldsymbol{\alpha}_d' \boldsymbol{f}_i + \varepsilon_{di}, \\ d_i &= \begin{cases} 1 & \text{if } d_i^\star > 0, \\ 0 & \text{otherwise,} \end{cases} \end{aligned} \tag{3.1}$$

where $\boldsymbol{\beta}_d$ denotes the vector of parameters related to $\boldsymbol{x}_i$, $\boldsymbol{\alpha}_d$ represents the vector of factor loadings associated with the latent factors $\boldsymbol{f}_i$, and $\varepsilon_{di}$ is an idiosyncratic error term assumed to be independent of $\boldsymbol{x}_i$ and $\boldsymbol{f}_i$. In this chapter, we are interested in the impact of education on a set of $p$ health and wage outcomes $(y_{1i}, y_{2i}, \ldots, y_{pi})$, where cognitive abilities and personality traits measured during childhood capture the correlation between these outcomes and the education decision through the latent factors $\boldsymbol{f}_i$. We assume that a potential outcome is associated with each level of education. For individual (agent) $i$, the potential outcome $r$ in the participation state ($d_i = 1$) is denoted by $y_{r1i}$, for $r = 1, \ldots, p$, while in the non-participation state ($d_i = 0$) the corresponding potential outcome is denoted by $y_{r0i}$. The pair of potential outcomes $(y_{r1i}, y_{r0i})$ is assumed to exist for each individual, however only one of them can be observed, since it is impossible for a same person to be in both education groups. The measured outcome $y_{ri}$ can thus be expressed as:

$$y_{ri} = d_i \, y_{r1i} + (1 - d_i) \, y_{r0i}. \tag{3.2}$$

In our empirical application, both dichotomous and continuous outcomes are considered. We assume that each potential outcome $y_{rs}$ is generated by a latent outcome $y_{rs}^\star$, for $s = 0, 1$, through the following linear-in-parameter model:

$$\begin{aligned} y_{rsi}^\star &= \boldsymbol{\beta}_{rs}' \boldsymbol{x}_i + \boldsymbol{\alpha}_{rs}' \boldsymbol{f}_i + \varepsilon_{rsi}, \\ y_{rsi} &= \begin{cases} y_{rsi}^\star & \text{if } y_{rsi} \text{ is continuous,} \\ \mathbb{1}[y_{rsi}^\star > 0] & \text{if } y_{rsi} \text{ is dichotomous,} \end{cases} \end{aligned} \tag{3.3}$$

where $\mathbb{1}[\cdot]$ denotes the indicator function that is equal to 1 if the corresponding condition is fulfilled, and to 0 otherwise. The observed characteristics $\boldsymbol{x}_i$ influence the latent outcome through the vector of parameters $\boldsymbol{\beta}_{rs}$, and the latent factors $\boldsymbol{f}_i$ through the vector of associated factor loadings $\boldsymbol{\alpha}_{rs}$. The unobserved random term $\varepsilon_{rsi}$ is assumed to be independent of the latent factors and of the observed characteristics.

Finally, the latent factors are extracted from a set of $q$ measurement variables $m_1, \dots, m_q$ (cognitive and non-cognitive traits) that can be either continuous or dichotomous. We assume that each observed measurement $j$ in $1, \dots, q$ is determined by an underlying latent variable $m^\star_{ji}$ that linearly depends on the observed characteristics $\boldsymbol{x}_i$ and on the latent factors $\boldsymbol{f}_i$ for individual $i$:

$$
\begin{aligned}
m^\star_{ji} &= \boldsymbol{\beta}'_{m_j}\boldsymbol{x}_i + \boldsymbol{\alpha}'_{m_j}\boldsymbol{f}_i + \varepsilon_{m_j i}, \\
m_{ji} &= \begin{cases} m^\star_{ji} & \text{if } m_{ji} \text{ is continuous,} \\ \mathbb{1}[m^\star_{ji} > 0] & \text{if } m_{ji} \text{ is dichotomous,} \end{cases}
\end{aligned}
\tag{3.4}
$$

where $\boldsymbol{\beta}_{m_j}$ denotes the vector of parameters associated with $\boldsymbol{x}_i$, and $\boldsymbol{\alpha}_{m_j}$ the vector of factor loadings related to $\boldsymbol{f}_i$. The random variable $\varepsilon_{m_j i}$ captures the idiosyncratic components of the measurement that cannot be explained by the latent factors, nor by the observed characteristics, and is assumed to be independent of $\boldsymbol{x}_i$ and $\boldsymbol{f}_i$. Conditional on the observed characteristics $\boldsymbol{x}$, the latent factors $\boldsymbol{f}$ appear to be the only source of correlation between the education outcome, the potential outcomes and the measurements.

The overall model consisting of Equations (3.1) to (3.4) can be written in compact form as:

$$
\begin{pmatrix} m^\star_{1i} \\ \vdots \\ m^\star_{qi} \\ d^\star_i \\ y^\star_{10i} \\ y^\star_{11i} \\ \vdots \\ y^\star_{p0i} \\ y^\star_{p1i} \end{pmatrix}
=
\begin{pmatrix} \boldsymbol{\beta}'_{m_1} \\ \vdots \\ \boldsymbol{\beta}'_{m_q} \\ \boldsymbol{\beta}'_d \\ \boldsymbol{\beta}'_{10} \\ \boldsymbol{\beta}'_{11} \\ \vdots \\ \boldsymbol{\beta}'_{p0} \\ \boldsymbol{\beta}'_{p1} \end{pmatrix} \boldsymbol{x}_i +
\begin{pmatrix} \boldsymbol{\alpha}'_{m_1} \\ \vdots \\ \boldsymbol{\alpha}'_{m_q} \\ \boldsymbol{\alpha}'_d \\ \boldsymbol{\alpha}'_{10} \\ \boldsymbol{\alpha}'_{11} \\ \vdots \\ \boldsymbol{\alpha}'_{p0} \\ \boldsymbol{\alpha}'_{p1} \end{pmatrix} \boldsymbol{f}_i +
\begin{pmatrix} \varepsilon_{m_1 i} \\ \vdots \\ \varepsilon_{m_q i} \\ \varepsilon_{di} \\ \varepsilon_{10i} \\ \varepsilon_{11i} \\ \vdots \\ \varepsilon_{p0i} \\ \varepsilon_{p1i} \end{pmatrix},
\tag{3.5}
$$

which can simply be expressed as:

$$
\boldsymbol{y}^\star_i = \boldsymbol{\beta}\boldsymbol{x}_i + \boldsymbol{\alpha}\boldsymbol{f}_i + \boldsymbol{\varepsilon}_i,
\tag{3.6}
$$

where $\boldsymbol{\alpha}$ denotes the $(\rho \times k)$–dimensional matrix of factor loadings associated with the vector of latent factors $\boldsymbol{f}_i$, with $\rho = 2p + q + 1$ being the total number of measurement and outcome variables. The $(\rho \times v)$–dimensional matrix $\boldsymbol{\beta}$ contains the slope parameters related to the $v$ explanatory variables $\boldsymbol{x}_i$, and the $\rho$–dimensional vector $\boldsymbol{\varepsilon}_i$ contains the error terms. To complete the specification of the model, some distributional assumptions are required for the latent factors and the error terms. As usual in the factor analysis literature, we assume these unobserved variables to be independently and normally distributed (as in Lopes and West, 2004):

$$\boldsymbol{f}_i \sim \mathcal{N}_k(\mathbf{0}; \mathbf{I}_k)\,, \tag{3.7}$$

$$\boldsymbol{\varepsilon}_i \sim \mathcal{N}_\rho(\mathbf{0}; \boldsymbol{\Sigma})\,, \tag{3.8}$$

where $\boldsymbol{\Sigma} = \mathrm{diag}(\sigma^2_{m_1}, \ldots, \sigma^2_{m_q}, \sigma^2_d, \sigma^2_{10}, \sigma^2_{11}, \ldots, \sigma^2_{p0}, \sigma^2_{p1})$, and the idiosyncratic variances of the dichotomous variables are fixed to 1 to set the scale of the corresponding latent measurements or outcomes. Since the measurement responses and the outcome variables play the same role in the factor analytic model, no distinction will be made henceforth between these response variables and the subscripts will be dropped. Thus, $\alpha_{jl}$ will simply denote the loading on the $j^{\text{th}}$ row and $l^{\text{th}}$ column of the factor loading matrix $\boldsymbol{\alpha}$.

### 3.4.2 Treatment effects

To investigate the impact of education on health outcomes and wages, and the role played by early cognitive abilities and personality traits as a source of correlation between these different outcomes, our potential outcome model can be used to compute various treatment effects. We focus on the Average Treatment Effect (ATE), which measures the overall impact of education in a randomly selected subsample of the whole population, the Treatment effect on the Treated (TT), which measures the impact in the population that is actually selecting into education, and the Marginal Treatment Effect (MTE), which measures the effect of the intervention on those induced to select into education by the intervention. More exactly, these treatment effects are defined as follows, for each outcome variable $j$:

$$\mathrm{ATE}_j(\boldsymbol{x}; \boldsymbol{\Lambda}) = \mathrm{E}[y_{j1} - y_{j0} | \boldsymbol{x}, \boldsymbol{\Lambda}]\,, \tag{3.9}$$

$$\mathrm{TT}_j(\boldsymbol{x}, d = 1; \boldsymbol{\Lambda}) = \mathrm{E}[y_{j1} - y_{j0} | \boldsymbol{x}, d = 1, \boldsymbol{\Lambda}]\,, \tag{3.10}$$

$$\mathrm{MTE}_j(\boldsymbol{x}, u; \boldsymbol{\Lambda}) = \mathrm{E}[y_{j1} - y_{j0} | \boldsymbol{x}, u_d = u, \boldsymbol{\Lambda}]\,, \tag{3.11}$$

where $\boldsymbol{\Lambda}$ represents the set of model parameters, and $u_d = \boldsymbol{\alpha}_d' \boldsymbol{f} + \varepsilon_d$ for the MTE. Heckman (1997), Heckman and Smith (1998) and Heckman et al. (1999) address the economic questions underlying these treatment parameters. Heckman and Vytlacil (2000, 2001, 2007) discuss the relationships among these parameters, and more especially how they can all be derived from the marginal treatment effect parameter.

The Bayesian estimation approach we adopt generates a sample of the model parameters from their posterior distribution. Thereby, not only does it make it possible to compute point estimates for the treatment effects, like in a frequentist approach, but it also allows us to derive the distribution of these treatment parameters. As a consequence, various statistical tests can naturally be implemented using these posterior distributions.

### Treatment effects in the continuous case

In the continuous case, the treatment effects are derived as:

$$
\begin{aligned}
\mathrm{ATE}_j(\boldsymbol{x};\boldsymbol{\Lambda}) &= (\boldsymbol{\beta}_{j1} - \boldsymbol{\beta}_{j0})'\boldsymbol{x}, \\
\mathrm{TT}_j(\boldsymbol{x}, d=1;\boldsymbol{\Lambda}) &= \frac{1}{\Phi\big(\boldsymbol{\beta}_d'\boldsymbol{x}/\sqrt{1+\boldsymbol{\alpha}_d'\boldsymbol{\alpha}_d}\big)} \\
&\quad \times \int \left[(\boldsymbol{\beta}_{j1} - \boldsymbol{\beta}_{j0})'\boldsymbol{x} + (\boldsymbol{\alpha}_{j1} - \boldsymbol{\alpha}_{j0})'\boldsymbol{f}\right] \Phi(\boldsymbol{\beta}_d'\boldsymbol{x} + \boldsymbol{\alpha}_d'\boldsymbol{f})\, \phi(\boldsymbol{f})\, \mathrm{d}\boldsymbol{f}, \\
\mathrm{MTE}_j(\boldsymbol{x}, u;\boldsymbol{\Lambda}) &= \frac{\sqrt{1+\boldsymbol{\alpha}_d'\boldsymbol{\alpha}_d}}{\phi\big(u\big/\sqrt{1+\boldsymbol{\alpha}_d'\boldsymbol{\alpha}_d}\big)} \\
&\quad \times \int \left[(\boldsymbol{\beta}_{j1} - \boldsymbol{\beta}_{j0})'\boldsymbol{x} + (\boldsymbol{\alpha}_{j1} - \boldsymbol{\alpha}_{j0})'\boldsymbol{f}\right] \phi(u - \boldsymbol{\alpha}_d'\boldsymbol{f})\, \phi(\boldsymbol{f})\, \mathrm{d}\boldsymbol{f},
\end{aligned}
$$

where $\Phi(\cdot)$ and $\phi(\cdot)$ represent, respectively, the prior cumulative and probability density functions of the latent factors, in our case the cdf and pdf of the multivariate standard normal distribution.

### Treatment effects in the dichotomous case

In the dichotomous case, the treatment effects measure the difference in probabilities of getting the outcome between the treated and untreated groups, and are derived as:

$$
\begin{aligned}
\mathrm{ATE}_j(\boldsymbol{x};\boldsymbol{\Lambda}) &= \Pr(y_{j1} = 1|\boldsymbol{x}, \boldsymbol{\Lambda}) - \Pr(y_{j0} = 1|\boldsymbol{x}, \boldsymbol{\Lambda})\,, \\
&= \Phi\left(\frac{\boldsymbol{\beta}_{j1}'\boldsymbol{x}}{\sqrt{1+\boldsymbol{\alpha}_{j1}'\boldsymbol{\alpha}_{j1}}}\right) - \Phi\left(\frac{\boldsymbol{\beta}_{j0}'\boldsymbol{x}}{\sqrt{1+\boldsymbol{\alpha}_{j0}'\boldsymbol{\alpha}_{j0}}}\right),
\end{aligned}
$$

$$\begin{aligned}
\mathrm{TT}_j(\boldsymbol{x}, d=1; \boldsymbol{\Lambda}) &= \Pr(y_{j1}=1|\boldsymbol{x}, d=1, \boldsymbol{\Lambda}) - \Pr(y_{j0}=1|\boldsymbol{x}, d=1, \boldsymbol{\Lambda})\,, \\
&= \frac{1}{\Phi\big(\boldsymbol{\beta}_d'\boldsymbol{x}/\sqrt{1+\boldsymbol{\alpha}_d'\boldsymbol{\alpha}_d}\big)} \\
&\quad \times \int \big[\Phi\big(\boldsymbol{\beta}_{j1}'\boldsymbol{x}+\boldsymbol{\alpha}_{j1}'\boldsymbol{f}\big) - \Phi\big(\boldsymbol{\beta}_{j0}'\boldsymbol{x}+\boldsymbol{\alpha}_{j0}'\boldsymbol{f}\big)\big]\,\Phi(\boldsymbol{\beta}_d'\boldsymbol{x}+\boldsymbol{\alpha}_d'\boldsymbol{f})\,\phi(\boldsymbol{f})\,\mathrm{d}\boldsymbol{f}, \\
\mathrm{MTE}_j(\boldsymbol{x}, u; \boldsymbol{\Lambda}) &= \Pr(y_{j1}=1|\boldsymbol{x}, u_d=u, \boldsymbol{\Lambda}) - \Pr(y_{j0}=1|\boldsymbol{x}, u_d=u, \boldsymbol{\Lambda})\,, \\
&= \frac{\sqrt{1+\boldsymbol{\alpha}_d'\boldsymbol{\alpha}_d}}{\phi\big(u\,/\sqrt{1+\boldsymbol{\alpha}_d'\boldsymbol{\alpha}_d}\big)} \\
&\quad \times \int \big[\Phi\big(\boldsymbol{\beta}_{j1}'\boldsymbol{x}+\boldsymbol{\alpha}_{j1}'\boldsymbol{f}\big) - \Phi\big(\boldsymbol{\beta}_{j0}'\boldsymbol{x}+\boldsymbol{\alpha}_{j0}'\boldsymbol{f}\big)\big]\,\phi(u-\boldsymbol{\alpha}_d'\boldsymbol{f})\,\phi(\boldsymbol{f})\,\mathrm{d}\boldsymbol{f}.
\end{aligned}$$

### Distributional treatment effects

As an alternative to the traditional treatment effects presented in the previous section for the dichotomous case, Aakvik et al. (2005) introduce a distributional version of these parameters measuring the fraction of the population that benefits, is indifferent, or suffers from the treatment. Let us define $\Delta_j = y_{j1} - y_{j0}$ as the difference between the two potential outcomes $j$. Given that our binary response variables all measure negative health outcomes, an outcome is said to be *successful* if the corresponding potential outcome is equal to 0 (e.g., not smoking), and *unsuccessful* if it is equal to 1 (e.g., smoking). Three different cases can be considered:

- $\Delta_j = -1$ if the individual would benefit from the treatment: she would have a successful outcome if treated ($y_{j1} = 0$), and unsuccessful otherwise ($y_{j0} = 1$),
- $\Delta_j = 0$ if the individual would not be affected by the treatment: she would have a successful outcome whatever the treatment she receives ($y_{j1} = 0$ and $y_{j0} = 0$), or an unsuccessful outcome in either treatment state ($y_{j1} = 1$ and $y_{j0} = 1$),
- $\Delta_j = 1$ if the individual would suffer from the treatment: she would have a successful outcome if not treated ($y_{j0} = 0$) and an unsuccessful outcome if treated ($y_{j1} = 1$).

#### ***Distributional version of ATE***

$$\begin{aligned}
\Pr(\Delta_j = -1|\boldsymbol{x}, \boldsymbol{\Lambda}) &= \Pr(y_{j1}=0, y_{j0}=1|\boldsymbol{x}, \boldsymbol{\Lambda})\,, \\
&= \int [1-\Phi\big(\boldsymbol{\beta}_{j1}'\boldsymbol{x}+\boldsymbol{\alpha}_{j1}'\boldsymbol{f}\big)]\Phi\big(\boldsymbol{\beta}_{j0}'\boldsymbol{x}+\boldsymbol{\alpha}_{j0}'\boldsymbol{f}\big)\,\phi(\boldsymbol{f})\,\mathrm{d}\boldsymbol{f}, \\
\Pr(\Delta_j = 1|\boldsymbol{x}, \boldsymbol{\Lambda}) &= \Pr(y_{j1}=1, y_{j0}=0|\boldsymbol{x}, \boldsymbol{\Lambda})\,,
\end{aligned}$$

$$= \int \Phi\big(\boldsymbol{\beta}'_{j1}\boldsymbol{x} + \boldsymbol{\alpha}'_{j1}\boldsymbol{f}\big)\,[1 - \Phi\big(\boldsymbol{\beta}'_{j0}\boldsymbol{x} + \boldsymbol{\alpha}'_{j0}\boldsymbol{f}\big)]\phi(\boldsymbol{f})\,\mathrm{d}\boldsymbol{f},$$

$$\Pr(\Delta_j = 0|\boldsymbol{x}, \boldsymbol{\Lambda}) = 1 - \Pr(\Delta_j = 1|\boldsymbol{x}, \boldsymbol{\Lambda}) - \Pr(\Delta_j = -1|\boldsymbol{x}, \boldsymbol{\Lambda})\,.$$

***Distributional version of TT***

$$\begin{aligned}
\Pr(\Delta_j = -1|\boldsymbol{x}, d = 1, \boldsymbol{\Lambda}) &= \Pr(y_{j1} = 0, y_{j0} = 1|\boldsymbol{x}, d = 1, \boldsymbol{\Lambda})\,,\\
&= \frac{1}{\Phi\big(\boldsymbol{\beta}'_d\boldsymbol{x}/\sqrt{1 + \boldsymbol{\alpha}'_d\boldsymbol{\alpha}_d}\big)}\\
&\times \int [1 - \Phi\big(\boldsymbol{\beta}'_{j1}\boldsymbol{x} + \boldsymbol{\alpha}'_{j1}\boldsymbol{f}\big)]\Phi\big(\boldsymbol{\beta}'_{j0}\boldsymbol{x} + \boldsymbol{\alpha}'_{j0}\boldsymbol{f}\big)\,\Phi(\boldsymbol{\beta}'_d\boldsymbol{x} + \boldsymbol{\alpha}'_d\boldsymbol{f})\,\phi(f)\,\mathrm{d}\boldsymbol{f},\\
\Pr(\Delta_j = 1|\boldsymbol{x}, d = 1, \boldsymbol{\Lambda}) &= \Pr(y_{j1} = 1, y_{j0} = 0|\boldsymbol{x}, d = 1, \boldsymbol{\Lambda})\,,\\
&= \frac{1}{\Phi\big(\boldsymbol{\beta}'_d\boldsymbol{x}/\sqrt{1 + \boldsymbol{\alpha}'_d\boldsymbol{\alpha}_d}\big)}\\
&\times \int \Phi\big(\boldsymbol{\beta}'_{j1}\boldsymbol{x} + \boldsymbol{\alpha}'_{j1}\boldsymbol{f}\big)\,[1 - \Phi\big(\boldsymbol{\beta}'_{j0}\boldsymbol{x} + \boldsymbol{\alpha}'_{j0}\boldsymbol{f}\big)]\Phi(\boldsymbol{\beta}'_d\boldsymbol{x} + \boldsymbol{\alpha}'_d\boldsymbol{f})\,\phi(f)\,\mathrm{d}\boldsymbol{f},\\
\Pr(\Delta_j = 0|\boldsymbol{x}, d = 1, \boldsymbol{\Lambda}) &= 1 - \Pr(\Delta_j = 1|\boldsymbol{x}, d = 1, \boldsymbol{\Lambda}) - \Pr(\Delta_j = -1|\boldsymbol{x}, d = 1, \boldsymbol{\Lambda})\,.
\end{aligned}$$

***Distributional version of MTE***

$$\begin{aligned}
\Pr(\Delta_j = -1|\boldsymbol{x}, u_d = u, \boldsymbol{\Lambda}) &= \Pr(y_{j1} = 0, y_{j0} = 1|\boldsymbol{x}, u_d = u, \boldsymbol{\Lambda})\,,\\
&= \frac{\sqrt{1 + \boldsymbol{\alpha}'_d\boldsymbol{\alpha}_d}}{\phi\big(u\,/\sqrt{1 + \boldsymbol{\alpha}'_d\boldsymbol{\alpha}_d}\big)}\\
&\times \int [1 - \Phi\big(\boldsymbol{\beta}'_{j1}\boldsymbol{x} + \boldsymbol{\alpha}'_{j1}\boldsymbol{f}\big)]\Phi\big(\boldsymbol{\beta}'_{j0}\boldsymbol{x} + \boldsymbol{\alpha}'_{j0}\boldsymbol{f}\big)\,\phi(u - \alpha_d f)\,\phi(\boldsymbol{f})\,\mathrm{d}\boldsymbol{f},\\
\Pr(\Delta_j = 1|\boldsymbol{x}, u_d = u, \boldsymbol{\Lambda}) &= \Pr(y_{j1} = 1, y_{j0} = 0|\boldsymbol{x}, u_d = u, \boldsymbol{\Lambda})\,,\\
&= \frac{\sqrt{1 + \boldsymbol{\alpha}'_d\boldsymbol{\alpha}_d}}{\phi\big(u\,/\sqrt{1 + \boldsymbol{\alpha}'_d\boldsymbol{\alpha}_d}\big)}\\
&\times \int \Phi\big(\boldsymbol{\beta}'_{j1}\boldsymbol{x} + \boldsymbol{\alpha}'_{j1}\boldsymbol{f}\big)\,[1 - \Phi\big(\boldsymbol{\beta}'_{j0}\boldsymbol{x} + \boldsymbol{\alpha}'_{j0}\boldsymbol{f}\big)]\phi(u - \alpha_d f)\,\phi(\boldsymbol{f})\,\mathrm{d}\boldsymbol{f},\\
\Pr(\Delta_j = 0|\boldsymbol{x}, u_d = u, \boldsymbol{\Lambda}) &= 1 - \Pr(\Delta_j = 1|\boldsymbol{x}, u_d = u, \boldsymbol{\Lambda})\\
&\quad - \Pr(\Delta_j = -1|\boldsymbol{x}, u_d = u, \boldsymbol{\Lambda})\,.
\end{aligned}$$

The latent factors $\boldsymbol{f}$ are integrated out in the above-defined treatment parameters. The resulting $k$–dimensional integrals have no closed-form solution, and therefore have to be approximated in practice. This can turn out to become cumbersome when $k$ is growing, and the accuracy of the estimated treatment parameters may highly depend on the quality of the approximation retained. We rely on Monte Carlo integration to deal with this problem, and test the sensitivity of our results against different configurations of our simulator. Full details are provided in Section 3.5.6.

## 3.5 Bayesian inference

In this section, we outline the sampling scheme used to perform fully Bayesian posterior inference for the unknowns of the model represented by Equations (3.1) to (3.4). The discreteness of some of the measurements, along with the parsimonious factor model representation through the prior of the factor loadings matrix, prevent closed form posterior inference, and the Markov chain Monte Carlo (MCMC) scheme is tailored to provide approximate but efficient posterior model assessment. The MCMC scheme we highlight in what follows is essentially the one introduced by Frühwirth-Schnatter and Lopes (2009), with minor modifications to accommodate the limited dependent variables that describe education ($d_i$), some decision outcomes ($y_{r0i}, y_{r1i}$), as well as a subset of binary measurements from Section 3.4.1.

### 3.5.1 Parsimonious Bayesian factor analysis to explore the latent structure of the model

In the data set used for our empirical application, a large number of measurements is available on a wide range of cognitive abilities and personality traits. Hence, the model specified in Equation (3.6) raises three questions:

1. How many latent factors $k$ can be extracted from the measurement system?
2. Are all measurement variables $j$ in $1, \dots, q$ relevant for our analysis, or should some of them be discarded?
3. Should restrictions be imposed on the structure of the factor loading matrix $\boldsymbol{\alpha}$? For example, should some measurements be dedicated to one or several specific latent factors?

In empirical studies relying on factor structure models, restrictions are commonly imposed on the factor loading matrix to allow for dedicated measurements. This approach is motivated by the fact that, usually, the items to be factor-analyzed come from psychometric tests specially designed to capture well-defined traits. For example, the popular Big Five taxonomy in psychology—'Openness', 'Conscientiousness', 'Extraversion', 'Agreeableness', and 'Neuroticism'—makes it natural to specify a five-factor model with dedicated measurements (Digman, 1990; Goldberg, 1990; McCrae and John, 1992). This type of specification requires a sound justification of the underlying structure of the model, and always relies on theoretical concepts that are well-established in psychology and/or sociology. Nevertheless, some items might reflect different traits, and the independence assumption of the latent factors might be questionable in many contexts. Some personal traits captured by different factors are indeed likely to be correlated, although specified as non-correlated (Heckman et al., 2006). Cunha and Heckman (2008), Cunha et al. (2010), Conti et al. (2010) overcome this problem by relaxing the independence assumption to allow for correlated factors.

As for factor selection, conventional approaches in factor analysis rely on scree plots (Cattell, 1966), eigenvalues analysis (Kaiser, 1960), or partial correlations (Velicer, 1976) to determine the number of latent factors. More sophisticated methods have recently been proposed in a Bayesian framework to address these problems. For instance, the Reversible Jump Markov chain Monte Carlo (RJMCMC) introduced by Green (1995) (see also Richardson and Green, 1997) is a transdimensional algorithm that makes it possible to directly deal with model uncertainty with respect to the dimension of the parameter space. Lopes and West (2004) adapt and implement this sampling scheme for factor models. Another recent development is the evolutionary model search algorithm, initiated by West (2003) and conceptualized by Carvalho (2006) and Carvalho et al. (2008) for high-dimensional sparse factor modeling, which tackles both factor selection and variable choice for inclusion. This novel stochastic search has been successfully applied in genomics, where huge data sets have to be dealt with.

To address the three aforementioned problems simultaneously, we implement and adapt the parsimonious Bayesian factor analysis (henceforth PBFA) of Frühwirth-Schnatter and Lopes (2009). More precisely, we start by assuming *a priori* that each factor loading of the matrix $\boldsymbol{\alpha}$ has probability $\omega$ of not being zero:

$$\alpha_{jl} \sim (1-\omega)\delta_0(\alpha_{jl}) + \omega\mathcal{N}\big(0;\, \sigma_\alpha^2\big)\,, \tag{3.12}$$

where $\delta_0(\alpha)$ is the Dirac function that is equal to 1 if $\alpha$ equals 0, and to 0 otherwise, and $\sigma_\alpha^2$ represents the prior variance of the loading in case it is not exactly zero. This prior distribution allows some response variables to be discarded from the analysis when the corresponding row of the factor loading matrix only contains zeros. As for the question of imposing dedicated measurement, the selection process of the latent factors and of the response variables naturally achieves these restrictions on the factor loading matrix. This sparsity inducing prior is also the starting point of the evolutionary search algorithm of Carvalho et al. (2008). But the comparison with their method stops there, insofar as they use hierarchical priors and implement a reordering of the items during the sampling procedure to achieve variable and factor selection, whereas Frühwirth-Schnatter and Lopes (2009) introduce a simple yet efficient stochastic search scheme in the PBFA that makes it possible to select the number of latent factors by allowing the factor loading matrix to have columns containing only zeros. However, such a factor loading matrix with zero columns is rank-deficient and therefore raises an identification issue (Geweke and Singleton, 1980). To tackle this problem, Frühwirth-Schnatter and Lopes (2009) prove a theorem that allows them to identify a set of full-rank matrices that are observationally equivalent to the original rank-deficient matrix with a lower-triangular form.

To apply the prior with a mass-point at zero shown in Equation (3.12), the set of model parameters is augmented with a matrix of indicators $\boldsymbol{\delta}$ of the same dimension as the factor loading matrix $\boldsymbol{\alpha}$. More specifically, an indicator $\delta_{jl}$ is associated with each factor loading $\alpha_{jl}$, for $j$ in $1, \ldots, \rho$ and $l$ in $1, \ldots, k$, that is equal to 1 if the corresponding loading is different from zero, and to 0 otherwise.

### 3.5.2 Identification issues

The model defined in Equation (3.6) is not identified without imposing some restrictions on the factor loading matrix.[18] For any orthogonal matrix $\mathbf{P}$ of dimension $k \times k$ such that $\mathbf{P}'\mathbf{P} = \mathbf{I}_k$, the transformation that assigns $\tilde{\boldsymbol{\alpha}} = \boldsymbol{\alpha}\mathbf{P}'$ and $\tilde{\boldsymbol{f}}_i = \mathbf{P}\boldsymbol{f}_i$, for all $i$ in $1, \ldots, n$, yields the same likelihood as the initial model. To prevent this kind of rotation, the standard approach consists of assuming the following lower-diagonal structure of the factor loading matrix (Geweke and Zhou, 1996):

$$\boldsymbol{\alpha} = \begin{pmatrix} \alpha_{11} & 0 & 0 & \cdots & 0 \\ \alpha_{21} & \alpha_{22} & 0 & \cdots & 0 \\ \alpha_{31} & \alpha_{32} & \alpha_{33} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \alpha_{k1} & \alpha_{k2} & \alpha_{k3} & \cdots & \alpha_{kk} \\ \alpha_{(k+1)1} & \alpha_{(k+1)2} & \alpha_{(k+1)3} & \cdots & \alpha_{(k+1)k} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \alpha_{\rho 1} & \alpha_{\rho 2} & \alpha_{\rho 3} & \cdots & \alpha_{\rho k} \end{pmatrix}, \tag{3.13}$$

where the total number of free factor loadings is $\kappa = \rho k - k(k-1)/2$.

[18]See Chapter 1, Section 1.2.2, for further details on identification issues.

One last problem remains, since this configuration of the factor loading matrix does not prevent a sign switch of the latent factors and of their associated loadings. The usual solution to this problem consists of imposing a sign constraint on the diagonal elements of $\boldsymbol{\alpha}$, thus ensuring the identifiability of the model. However, sampling from a truncated distribution can turn out to be cumbersome in practice, and we therefore prefer to leave the sign of the latent factors and their loadings unidentified. As a consequence, the posterior distribution of the factor loadings is bimodal. Frühwirth-Schnatter and Lopes (2009) suggest to insert an additional step in the Gibbs sampler where a random sign switch is performed, to make sure the algorithm navigates through all the modes of the posterior distribution. Then, a simple sign-switch can be performed as a post-processing step to recover the unimodal posterior distribution of the loadings.

### 3.5.3 Prior specification

***Prior on the indicator matrix.*** Each indicator $\delta_{jl}$ is assumed to follow a Bernoulli distribution with parameter $\omega$. For a given $\omega$, this assumption is quite informative on the number $d_{\boldsymbol{\delta}}$ of non-zero factor loadings since $d_{\boldsymbol{\delta}} \sim \mathcal{B}ino\,(\kappa;\,\omega)$. To circumvent this problem, a noninformative hyperprior is placed on $\omega$ (Smith and Kohn, 2002; Tüchler, 2008; Frühwirth-Schnatter and Tüchler, 2008; Frühwirth-Schnatter and Lopes, 2009). Assuming that $\omega$ is i.i.d. uniformly distributed on $[0,1]$, it can be shown that:

$$\begin{aligned} p(\boldsymbol{\delta}) &= \int p(\boldsymbol{\delta}|\omega)p(\omega)\,\mathrm{d}\,\omega, \\ &= \mathcal{B}(d_{\boldsymbol{\delta}}+1;\,\kappa - d_{\boldsymbol{\delta}}+1)\,, \end{aligned}$$

where $\mathcal{B}(\cdot;\cdot)$ is the Beta function and $d_{\boldsymbol{\delta}} = \sum_{j=1}^{\rho}\sum_{l=1}^{k}\delta_{jl}$.

***Prior on the factor loadings.*** Prior independence between the rows of the factor loading matrix is assumed, conditional on the latent factors $\boldsymbol{f}$. A normal prior with covariance matrix proportional to the idiosyncratic variance $\sigma_j^2$ is placed on each row of unconstrained elements of the factor loading matrix:

$$\alpha_j^{\boldsymbol{\delta}} \sim \mathcal{N}_k\big(\boldsymbol{a}_{j0}^{\boldsymbol{\delta}};\,\mathbf{A}_{j0}^{\boldsymbol{\delta}}\sigma_j^2\big)\,, \tag{3.14}$$

for $j$ in $1,\ldots,\rho$, where $\alpha_j^{\boldsymbol{\delta}}$ denotes the column vector containing the non-zero elements of the $j^{\text{th}}$ row of the factor loading matrix $\boldsymbol{\alpha}$. The superscript $\boldsymbol{\delta}$ shows that only the unconstrained factor loadings, i.e., those for which the corresponding indicator is equal to 1, are considered. Prior dependence on $\sigma_j^2$ will allow to draw the factor loadings jointly with the idiosyncratic variance, and more importantly, to sample the indicators without conditioning on the parameters of the latent structure of the model.

As an alternative to the usual normal prior, Frühwirth-Schnatter and Lopes (2009) entertain a fractional-type prior, based on an idea introduced by O'Hagan (1995). This consists of considering the auxiliary regression:

$$\tilde{\boldsymbol{y}}_j^\star = \mathbf{F}_j^{\boldsymbol{\delta}} \boldsymbol{\alpha}_j^{\boldsymbol{\delta}} + \varepsilon_j,$$

where $\tilde{\boldsymbol{y}}_j^\star = \boldsymbol{y}_j^\star - \mathbf{X}\beta_j$, with $\beta_j$ denoting the column vector containing the $j^{\text{th}}$ row of the matrix of slope parameters $\boldsymbol{\beta}$, and $\mathbf{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)'$ contains the observed covariates for all individuals. The matrix $\mathbf{F} = (\boldsymbol{f}_1, \ldots, \boldsymbol{f}_n)'$ contains the latent factors for all individuals, and the matrix of regressors $\mathbf{F}_j^{\boldsymbol{\delta}}$ associated with $\boldsymbol{\alpha}_j^{\boldsymbol{\delta}}$ contains the factors corresponding to the non-zero indicators $\delta_{jl}$ in its columns, for $l = 1, \ldots, k$. Only a fraction $b$ of the information contained in this auxiliary regression is then used to build up the required prior as follows:

$$\begin{aligned} p(\boldsymbol{\alpha}_j^{\boldsymbol{\delta}} | \sigma_j^2, b) &\propto p(\tilde{\boldsymbol{y}}_j^\star | \boldsymbol{\alpha}_j^{\boldsymbol{\delta}}, \sigma_j^2)^b, \\ &\propto \exp\left\{ -\frac{b}{2\sigma_j^2} (\tilde{\boldsymbol{y}}_j^\star - \mathbf{F}_j^{\boldsymbol{\delta}} \boldsymbol{\alpha}_j^{\boldsymbol{\delta}})' (\tilde{\boldsymbol{y}}_j^\star - \mathbf{F}_j^{\boldsymbol{\delta}} \boldsymbol{\alpha}_j^{\boldsymbol{\delta}}) \right\}, \end{aligned}$$

where $0 < b \leqslant 1$. In our empirical study, we use fractional priors for the continuous measurements, and normal priors for the outcomes and the dichotomous measurements. By choosing the tuning parameter $b$ appropriately, the fractional prior helps to impose more sparsity on the measurement system (see Section 3.6.1).

***Prior on the idiosyncratic variances.*** We assume a proper inverse Gamma prior distribution for each idiosyncratic variance $\sigma_j^2$, for $j = 1, \ldots, \rho$:

$$\sigma_j^2 \sim \mathcal{G}^{-1}(c_0;\, C_{j0}), \qquad C_{j0} = \frac{c_0 - 1}{(\mathbf{S}^{-1})_{jj}}, \tag{3.15}$$

where $(\mathbf{S}^{-1})_{jj}$ represents the $j^{\text{th}}$ diagonal element of the inverse sample covariance matrix $\mathbf{S}$ of the observed continuous response variables. The prior parameters are chosen to bound the prior away from zero, for instance, with the shape parameter

$c_0 = 2.5$. The scale parameter $C_{j0}$ is selected to reduce the occurrence probability of a Heywood case (Frühwirth-Schnatter and Lopes, 2009).[19]

***Prior on the slope parameters.*** Assuming prior independence between the rows of the matrix of slope parameters $\boldsymbol{\beta}$, a normal prior is placed on each row $j$ in $1, \ldots, p$:

$$\boldsymbol{\beta}_j \sim \mathcal{N}_v(\boldsymbol{b}_{j0}; \mathbf{B}_{j0}) \,, \tag{3.16}$$

where $v$ represents the number of covariates in $\boldsymbol{\beta}$. To remain as general as possible, a flat prior with mean $\boldsymbol{b}_{j0} = \mathbf{0}$ and infinite covariance matrix such that $\mathbf{B}_{j0}^{-1} = \mathbf{0}$ is assumed in our empirical application.

### 3.5.4 Sampling scheme

The Gibbs sampler updates the parameters sequentially from their conditional distributions. For a given matrix of indicators $\boldsymbol{\delta}$, the model is a standard factor structure model and the sampling procedure is therefore identical to the one presented in Chapter 1. Since our specification is slightly different from the one used by Frühwirth-Schnatter and Lopes (2009), because of the inclusion of dichotomous response variables and of covariates, we provide full details of the Gibbs sweep in Appendix 3.A.

The drawing of the components of the indicator matrix $\boldsymbol{\delta}$ represents the most involved step of the algorithm. It is also the centerpiece of the procedure, since it defines and characterizes the stochastic search scheme behind the parsimonious Bayesian factor analysis. After each sweep of the Gibbs sampler, performed conditionally on $\boldsymbol{\delta}$, the indicator matrix is in turn updated column-wise, where for each $l$ in $1, \ldots, k$, the components are sampled as follows:

(i) Update the diagonal indicator $\delta_{ll}$ if $l > 1$, otherwise set $\delta_{11} = 1$;

(ii) Update the lower-diagonal elements of $\boldsymbol{\delta}$: if $\delta_{ll} = 0$, the $l^{\text{th}}$ column of the factor loading matrix is not identified, therefore set $\delta_{jl} = 0$ for $j = l+1, \ldots, \rho$. Otherwise, sample each indicator $\delta_{jl} = 0$, $j = l+1, \ldots, \rho$, from its conditional distribution;

(iii) Correct a spurious factor: if $\delta_{ll} = 1$ and $\delta_{jl} = 0$ for $j = l+1, \ldots, \rho$, set $\delta_{ll} = 0$.

[19] A Heywood case is a peculiarity in factor analysis that happens when at least one idiosyncratic variance is negative, i.e., when the solution lies outside the admissible parameter space (see Bartholomew, 1987, Section 3.6).

Different sampling schemes can be adopted to update the indicators $\delta_{jl}$, for $j = 1, \ldots, \rho$ and $l = 1, \ldots, k$. We opt for the procedure introduced by Smith and Kohn (2002), and used by Tüchler (2008), Frühwirth-Schnatter and Tüchler (2008), among others, which turns out to be far less computationally demanding than alternative approaches like the simple Gibbs scheme used in Frühwirth-Schnatter and Lopes (2009). After drawing $u \sim \mathcal{U}(0; 1)$, it consists of sampling the new indicator $\delta_{jl}^{\text{new}}$ as follows:

- If $\delta_{jl}^{\text{old}} = 1$ and $u > \pi(\delta_{jl} = 0)$, then set $\delta_{jl}^{\text{new}} = 1$.
- If $\delta_{jl}^{\text{old}} = 1$ and $u \leqslant \pi(\delta_{jl} = 0)$, then draw $e \sim \mathcal{U}(0; 1)$. Set $\delta_{jl}^{\text{new}} = 0$ if $e \leqslant l(\delta_{jl} = 0)/(l(\delta_{jl} = 0) + l(\delta_{jl} = 1))$ and $\delta_{jl}^{\text{new}} = 1$ otherwise.
- If $\delta_{jl}^{\text{old}} = 0$ and $u > \pi(\delta_{jl} = 1)$, then set $\delta_{jl}^{\text{new}} = 0$.
- If $\delta_{jl}^{\text{old}} = 0$ and $u \leqslant \pi(\delta_{jl} = 1)$, then draw $e \sim \mathcal{U}(0; 1)$. Set $\delta_{jl}^{\text{new}} = 1$ if $e \leqslant l(\delta_{jl} = 1)/(l(\delta_{jl} = 0) + l(\delta_{jl} = 1))$ and $\delta_{jl}^{\text{new}} = 0$ otherwise.

The conditional priors $\pi(\delta_{jl} = t)$, $t = 0, 1$, can be derived and computed at virtually no cost, making the whole procedure fast to implement. They are defined as:

$$\pi(\delta_{jl} = 0) = (\kappa - d_{\boldsymbol{\delta}} + 1)/(\kappa + 1) \qquad \text{if } \delta_{jl}^{\text{old}} = 1, \tag{3.17}$$

$$\pi(\delta_{jl} = 1) = (d_{\boldsymbol{\delta}} + 1)/(\kappa + 1) \qquad \text{if } \delta_{jl}^{\text{old}} = 0, \tag{3.18}$$

and are approximately equal to the proportion of free elements that are equal to zero in the factor loadings matrix (Equation 3.17) and to the proportion of non-zero free elements (Equation 3.18).

As for the likelihood $l(\delta_{jl} = t)$, $t = 0, 1$, it is equal to the marginal likelihood $p(\tilde{\boldsymbol{y}}_j^\star | \boldsymbol{f}_1, ..., \boldsymbol{f}_n, \boldsymbol{\delta}_{-jk}, \delta_{jk} = t)$, where the auxiliary latent response variable is defined as $\tilde{\boldsymbol{y}}_j^\star = \boldsymbol{y}_j^\star - \mathbf{X}\boldsymbol{\beta}_j$. The marginal distribution of $\tilde{\boldsymbol{y}}_j^\star$ depends on the type of the response variable, and on the prior placed on the factor loadings. If a normal prior is assumed, the marginal in the continuous response variable case is equal to:

$$\begin{aligned} p(\tilde{\boldsymbol{y}}_j | \boldsymbol{f}_1, ..., \boldsymbol{f}_n, \boldsymbol{\delta}) &= \frac{p(\tilde{\boldsymbol{y}}_j | \boldsymbol{\alpha}_j, \boldsymbol{f}_1, ..., \boldsymbol{f}_n, \boldsymbol{\delta}) p(\boldsymbol{\alpha}_j | \boldsymbol{\delta})}{p(\boldsymbol{\alpha}_j | \tilde{\boldsymbol{y}}_j, \boldsymbol{f}_1, ..., \boldsymbol{f}_n, \boldsymbol{\delta})}, \\ &= (2\pi)^{-n/2} \frac{|\mathbf{A}_{jn}|^{1/2}}{|\mathbf{A}_{j0}^{\boldsymbol{\delta}}|^{1/2}} \frac{\Gamma(c_n)(C_{j0})^{c_0}}{\Gamma(c_0)(C_{jn})^{c_n}}, \end{aligned}$$

where $\Gamma(\cdot)$ represents the Gamma function, $\mathbf{A}_{jn}$ is the posterior covariance matrix of $p(\boldsymbol{\alpha}_j | \tilde{\boldsymbol{y}}_j^\star, \boldsymbol{f}_1, ..., \boldsymbol{f}_n, \sigma_j^2, \boldsymbol{\delta})$ and $c_n$ and $C_{jn}$ are the posterior shape and scale parameters of $p(\sigma_j|\tilde{\boldsymbol{y}}_j^\star, \boldsymbol{f}_1, ..., \boldsymbol{f}_n, \boldsymbol{\delta})$ defined in Appendix 3.A. If the response variable $\boldsymbol{y}_j$ is dichotomous, the marginal distribution of the auxiliary response variable is:

$$p(\tilde{\boldsymbol{y}}_j^\star|\boldsymbol{f}_1, ..., \boldsymbol{f}_n, \boldsymbol{\delta}) = \\ (2\pi)^{-n/2}\frac{|\mathbf{A}_{jn}|^{1/2}}{|\mathbf{A}_{j0}^{\boldsymbol{\delta}}|^{1/2}} \exp\left\{-\frac{1}{2}\left(\tilde{\boldsymbol{y}}_j^{\star\prime}\tilde{\boldsymbol{y}}_j^\star + (\boldsymbol{a}_{j0}^{\boldsymbol{\delta}})'(\mathbf{A}_{j0}^{\boldsymbol{\delta}})^{-1}\boldsymbol{a}_{j0}^{\boldsymbol{\delta}} - \boldsymbol{a}_{jn}^{\boldsymbol{\delta}\,\prime}\mathbf{A}_{jn}^{-1}\boldsymbol{a}_{jn}^{\boldsymbol{\delta}}\right)\right\},$$

where $\boldsymbol{a}_{jn}^{\boldsymbol{\delta}}$ and $\mathbf{A}_{jn}$ are the posterior mean and covariance of $p(\boldsymbol{\alpha}_j|\tilde{\boldsymbol{y}}_j^\star, \boldsymbol{f}_1, ..., \boldsymbol{f}_n, \boldsymbol{\delta})$ defined in Appendix 3.A. When a fractional prior is placed on the factor loadings, the marginal likelihood of $\tilde{\boldsymbol{y}}_j$ in the continuous case is given by:

$$p(\tilde{\boldsymbol{y}}_j|\boldsymbol{f}_1, ..., \boldsymbol{f}_n, \boldsymbol{\delta}) = \frac{b^{z_j/2}\Gamma(c_n)(C_{j0})^{c_0}}{(2\pi)^{n(1-b)/2}\Gamma(c_0)(C_{jn})^{c_n}},$$

where $z_j = \sum_{l=1}^k \delta_{jl}$ is the number of non-zero loadings of $\boldsymbol{\alpha}_j$, and $c_n$ and $C_{jn}$ are the posterior shape and scale parameters defined in Appendix 3.A.

### 3.5.5 Assessing the number of latent factors

The parsimonious Bayesian factor analysis offers a natural framework to select the number of latent factors, which can be calculated as the number of non-zero columns of the factor loading matrix, i.e., $\eta = \sum_{l=1}^k \delta_{ll}$. The MCMC sampling scheme generates the posterior distribution of $\eta$, which can then be used for posterior inference. When the algorithm visits models with different number of factors, the question of model selection arises, and a criterion has to be chosen to assess the dimension of the latent structure of the model. Shannon (1948) introduced a measure of uncertainty that proves to be useful when a choice has to be made among a set of $k$ possible events with different occurrence probabilities $\pi_l$, for $l = 1, \ldots, k$. A measure of the uncertainty involved in the decision of selecting one of these events can be represented by the entropy associated with $\pi_1, \ldots, \pi_k$, which is defined as:

$$\mathcal{E}(\pi) = -\sum_{l=1}^{k} \pi_l \log(\pi_l),$$

where in our case $\pi_l$ represents the posterior probability that the number of factors is equal to $l$, i.e., $\Pr(\eta = l)$. The smaller the entropy, the stronger the evidence in favor of the number of factors with the highest posterior probability. On the contrary, when the entropy is larger, it becomes harder to discriminate between the different models and to assess the number of underlying latent factors.[20]

### 3.5.6 Computing the treatment effects from the MCMC chains

The MCMC sampler produces a sample of parameters from the posterior distribution $\boldsymbol{\Lambda} \sim p(\boldsymbol{\Lambda}|\mathbf{Y})$ that can be used to compute the treatment effects derived in Section 3.4.2. The computation of treatment effects in a Bayesian context is relatively rare compared to frequentist approaches. For some recent examples, see Chib and Hamilton (2000, 2002), Li et al. (2004), Li and Tobias (2007).

Assuming, for instance, that the interest is in computing the average treatment effect:

$$\mathrm{ATE}_j = \mathrm{E}_{\boldsymbol{\Lambda}|\mathbf{Y}}\mathrm{E}_{\mathbf{X}}\left[\mathrm{ATE}_j(\boldsymbol{x};\boldsymbol{\Lambda})\right],$$

then a Monte Carlo estimator can be easily obtained from the MCMC output as:

$$\widehat{\mathrm{ATE}}_j = \frac{1}{nM}\sum_{m=1}^{M}\sum_{i=1}^{n}\mathrm{ATE}(\boldsymbol{x}_i;\boldsymbol{\Lambda}^{(m)}),$$

where $\boldsymbol{\Lambda}^{(1)}, \ldots, \boldsymbol{\Lambda}^{(M)}$ are MCMC draws from the posterior distribution of $\boldsymbol{\Lambda}$. The covariates $\boldsymbol{x}$ are integrated out by averaging over the population. Similarly, the other treatment effects are estimated as follows:

$$\widehat{\mathrm{TT}}_j = \frac{1}{nM}\sum_{m=1}^{M}\sum_{i=1}^{n}\mathrm{TT}(\boldsymbol{x}_i, d(\boldsymbol{x}_i) = 1;\boldsymbol{\Lambda}^{(m)}),$$
$$\widehat{\mathrm{MTE}}_j(u) = \frac{1}{nM}\sum_{m=1}^{M}\sum_{i=1}^{n}\mathrm{MTE}(\boldsymbol{x}_i, u;\boldsymbol{\Lambda}^{(m)}).$$

The distributional treatment effects presented in Section 3.4.2 can be estimated in the same way by averaging over the MCMC draws.

In each of the treatment effect parameters, the latent factors $\boldsymbol{f}$ are integrated out. We opt for a Monte Carlo integration to numerically approximate these integrals. This choice is motivated by the fact that the factors are a priori standard normal distributed and independent, which makes it easy to draw from their prior distribution. For instance, the TT parameter can be computed by averaging over $B$

[20] Frühwirth-Schnatter and Lopes (2009) also use the entropy to determine the number of factors.

independent random draws as follows:

$$\widehat{\mathrm{TT}}_j = \frac{1}{nMB}\sum_{m=1}^{M}\sum_{i=1}^{n}\sum_{b=1}^{B}\frac{\left[(\boldsymbol{\beta}_{j1}^{(m)}-\boldsymbol{\beta}_{j0}^{(m)})'\boldsymbol{x}_i+(\boldsymbol{\alpha}_{j1}^{(m)}-\boldsymbol{\alpha}_{j0}^{(m)})'\boldsymbol{f}^{(b)}\right]\Phi\left(\boldsymbol{\beta}_d^{(m)\prime}\boldsymbol{x}_i+\boldsymbol{\alpha}_d^{(m)\prime}\boldsymbol{f}^{(b)}\right)}{\Phi\left(\boldsymbol{\beta}_d^{(m)\prime}\boldsymbol{x}_i/\sqrt{1+\boldsymbol{\alpha}_d^{(m)\prime}\boldsymbol{\alpha}_d^{(m)}}\right)},$$

where $\boldsymbol{f}^{(1)},\ldots,\boldsymbol{f}^{(B)}$ are sampled from the prior distribution of the latent factors. The same draws are used for each individual and for each MCMC iteration, to make sure that differences in computed parameters, for each $i=1,\ldots,n$ and $m=1,\ldots,M$, are due to differences in observed characteristics $\boldsymbol{x}$ or to differences in parameters over the MCMC iterations, and not due to different draws of the latent factors (Lee, 1992). The parameters computed for each MCMC iteration $m$ are saved, so as to obtain an $M$-sample of the treatment parameters of interest that can be used for posterior inference. With these simulated distributions, it is for instance possible to test for the equality of ATE and TT by computing the test statistic $\xi = \mathrm{ATE} - \mathrm{TT}$ (see Table 3.7).

In the framework of the parsimonious factor analysis, only a subset of the latent factors actually has an impact on the outcomes. Hence, to compute the treatment parameters, we only need to sample the factors for the corresponding non-zero loadings, i.e., for which $\delta_{jl}=1$, $l=1,\ldots,k$. This considerably reduces the computational burden of the procedure.

## 3.6 Empirical application

### 3.6.1 Implementing the parsimonious Bayesian factor analysis

We implement the MCMC scheme described in Section 3.5.1 to extract a set of relevant latent factors from the measurement system, in order to measure their impact on various health outcomes and wages. Because of the large number of items in the measurement system, the number of potential latent factors turns out to be very large.[21] To allow for more sparsity on this system, we specify a fractional prior on the factor loadings of the continuous measurement variables among $(m_1,\ldots,m_q)$, which can be tailored to control parsimony through the tuning parameter $b$. For the dichotomous measurements, the fractional prior would be harder to accommodate. For this reason, a normal prior is specified, which does not seem to affect

[21] A total of 98 continuous and 28 binary items is available.

the parsimony in the end. The choice of the tuning parameter $b$ of the fractional prior is not straightforward. Too small a value will lead to an underestimation of the true number of latent factors, while too large a value will not make it possible to discriminate between the latent factors, because of the lack of sparsity on the factor loading matrix. As a rule of thumb, Frühwirth-Schnatter and Lopes (2009) advocate to select $b = 1/(n\rho)$, which would in our case be of the order $3 \times 10^{-6}$. Rather, we test the sensitivity of the results against different values of $b$, ranging from $10^{-1}$ to $10^{-6}$, and find out that $b = 10^{-5}$ offers a good compromise with respect to the parsimony of the factor loading matrix and the interpretability of the resulting latent factors.

In the outcome system $(y_1, \ldots, y_p)$, a normal prior is assumed in order to be less restrictive than in the measurement system, and to better capture the impact of the factors on the outcomes. To enhance the mixing of the chain, we implement a parameter expansion of the model, following Ghosh and Dunson (2009) (see Appendix 3.A), and specify an noninformative prior for the variances of the latent factors. More precisely, we select the shape and rate parameters of the inverse Gamma distribution, respectively $g_l$ and $h_l$ for $l = 1, \ldots, k$, to be equal to 2.5. A summary of the prior parameters and hyperparameters is provided in Table 3.2.

Another point worth underlying concerns the order of the response variables, and the number of latent factors specified for the PBFA. A maximum number of 120 potential factors can be specified in our case, which is hardly feasible because of

**Table 3.2:** Prior parameter specification

| **Prior parameters** | | **Values** |
|---|---|---|
| $\boldsymbol{b}_{j0}$ | $j = 1, \ldots, \rho$ | $\mathbf{0}_v$ |
| $\mathbf{B}_{j0}$ | $j = 1, \ldots, \rho$ | $+\infty$ |
| $\boldsymbol{a}_{j0}$ | $j = 1, \ldots, \rho$ | $\mathbf{0}_k$ |
| $\mathbf{A}_{j0}$ | $j = 1, \ldots, \rho$ | $\mathbf{I}_k$ |
| $b$ | | $10^{-5}$ |
| $c_0$ | | 2.5 |
| $C_{j0}$ | $j = 1, \ldots, \rho$ | $1.5/(\mathbf{S}^{-1})_{jj}{}^{*}$ |
| $g_l$ | $l = 1, \ldots, k$ | 2.5 |
| $h_l$ | $l = 1, \ldots, k$ | 2.5 |

* $(\mathbf{S}^{-1})_{jj}$ denotes the $j^{\text{th}}$ diagonal element of the inverse sample covariance matrix $S$ of the observed continuous response variables.

the computational cost of the sampling procedure.[22] Moreover, such a huge number of latent factors would clearly be in opposition to our initial goal of constructing aggregates. When the number of true underlying latent factors is smaller than the maximum number of factors specified, the factor loading matrix will contain zero columns. These columns containing only zeros can appear anywhere in $\boldsymbol{\alpha}$, and reflect a poor ordering of the variables when they are not in the last columns. Frühwirth-Schnatter and Lopes (2009) thus suggest to reestimate the model, after an appropriate reordering of the variables and a reduction of the number of latent factors to the number of non-zero columns. To bypass this problem, we use prior information from a standard factor analysis (SFA) like the one presented in Section 3.7, and choose to put the items with the highest factor loadings on specific factors in the first rows of the vector $\boldsymbol{y}$ for our PBFA. Although the conventional SFA suggests to retain 11 latent factors, we specify $k = 20$ in a first attempt, to make sure we are not unreasonably reducing the number of potential factors. This strategy turns out to be successful, since the last columns of the resulting factor loading matrix contain only zeros. We then rerun the model with 10 factors and present the results in the next section.[23]

The model is estimated separately for males and females, using a chain with a total number of 220,000 iterations for each gender. After a burn-in period of 20,000 iterations, $M = 20{,}000$ iterations are saved every 10$^{\text{th}}$ step and used for posterior inference, e.g., for the computation of the treatment effects. The first 10,000 iterations of the chain are run without parsimonious factor analysis, so as to generate sensible values for the latent factors and the model parameters when the variable selection is launched.

### 3.6.2 Empirical results

The PBFA not only produces estimates of the factor loadings, but also posterior probabilities that these loadings are different from zero. This represents an important improvement over conventional factor analysis approaches, where the analyst usually has to decide which factor loadings are large enough to consider the items

[22]Frühwirth-Schnatter and Lopes (2009) show that the upper-bound for the number $k$ of potential factors is $k \leqslant \rho + 1.5 - \sqrt{2\rho + 2.25}$.

[23]Actually, the first PBFA attempt with 20 factors suggested more than 10 latent factors, but the last ones always appeared very unstable and redundant compared to the other factors, and the outcomes of interest did not load on them. On the contrary, the first 10 factors always appeared very stable and robust across the different specifications and MCMC runs with different starting values we performed. We therefore believe we are making a reasonable choice by fixing $k = 10$ with this ordering of the variables. Some results from this preliminary analysis are presented in Appendix 3.A.

as associated with the corresponding latent factors. This variable selection and the resulting interpretation of the factors can turn out to be tricky in these traditional approaches, especially when some factor loadings have a small magnitude but are actually significantly different from zero.

Figures 3.3 and 3.4 display the posterior probabilities of being different from zero for all factor loadings, in both the measurement and the outcome systems. The variables are ordered as in Equation (3.5), where the education, health and wage outcomes are at the bottom of the picture. The black cells of the table make it possible to easily identify the latent factors and their associated measurements. To make the interpretation more accurate, these pictures should be read jointly with Tables 3.9 and 3.10 that show the magnitude of the factor loadings, where only those with a posterior probability larger than 0.75 are presented. As becomes clear from these figures and tables, many factor loadings are actually extremely small, but still significantly different from zero. In a classical approach, they would simply be ignored because of their small magnitude, whereas the PBFA brings to light the fact that the corresponding items do matter for the latent factors, which can in the end lead to a slightly different interpretation.

While the measurement system appears remarkably stable across genders (only small differences in the loadings of the measurements on each factor are present, and will be noticed in the detailed description below), there are marked differences in the effects of the factors on the outcomes. The following interpretation of the latent factors comes out of Figures 3.3 and 3.4, as well as from Tables 3.9 and 3.10:

- Factor 1 can be defined as a composite factor, since all the cognitive tests and almost all the items of the child developmental scale load on it. This comes as no surprise since, given that the latter is administered to the teachers, it reflects the child neurodevelopment at school, which is likely to be correlated with her performance. Interestingly, a few items from the Rutter and Conners scale—which were administered to the mother—also load on the first factor: exactly those which use the same wording as employed in the items from the child developmental scale. Lastly, specific items from the locus of control and self-esteem scales also load on the first factor; they reflect the child's relationship with the school, the teacher, and the friends. However, if we consider also the magnitude of the loadings, we can identify Factor 1 with the traits of Disorganized Activity—indicating inability to concentrate—and

Antisocial Behavior (two subscales of the Child Developmental Scale with loadings larger than 0.4).[24]

- Factor 2 reflects attention-deficit, hyperactivity and behavioral disorders, as all the items of the Rutter and the Conners scale load on it, for both genders. Interestingly, while this is the same type of problems as reflected in the Child Developmental scale, differently from Factor 1, the cognitive ability tests do not load on Factor 2. The simplest explanation for this difference in patterns is the fact that the Rutter and Conners scales were administered to the mother, rather than to the teacher, hence are likely to reflect a less educational-oriented behavior. A closer look at the magnitude of the loadings allows us to identify more precisely Factor 2 with one of the Big Five personality traits, namely Neuroticism.[25]

- Factor 3 can also be interpreted as a composite factor, as all the cognitive tests and the locus of control and self-esteem scales load on it. In this case, it also appears that the same factor reflects both cognitive ability and personality; however, differently from Factor 1, here the personality measures were administered to the children, not to the teachers. Nonetheless, the items with the higher loadings on the Locus of Control scale are all related to study, tests, teachers, and marks. It is also important to remark that the items of the locus of control scale have higher loadings on this factor for females than for males.

- Factor 4 can easily be identified, for both genders, with the trait of Neuroticism/Anxiety, as the items of the Child Developmental Scale that constitute this subscale have the highest loadings. It is worth noticing that we find significant loadings also on the measurements of the subscales of Clumsiness, and Introversion/Extraversion.[24]

- Factor 5 can also be easily identified, for both genders, with the trait of Clumsiness, and for males also of Poor Hand-Eye Coordination, as defined by the respective subscales that group the items of the Child Developmental Scale loading on this factor. Significant loadings are also present for the measurements that define the subscales of Hyperkinesis and Behavioral Trauma.[24]

- Factor 6 can be interpreted as Poor Coordination, as reflected in the items of the Rutter and Conners scale loading on it.

[24]We use the original names of the subscales of the Child Developmental Scale as in Butler et al. (1997).

[25]We thank Angela Duckworth for her help in mapping these items to the Big Five Personality Traits.

**Table 3.3:** Factor loadings and corresponding posterior probabilities in the outcome system (Females)

| | $f_1$ | $f_2$ | $f_3$ | $f_4$ | $f_5$ | $f_6$ | $f_7$ | $f_8$ | $f_9$ | $f_{10}$ |
|---|---|---|---|---|---|---|---|---|---|---|
| Education | | | | | | | | | | |
| | -0.254 | | 0.398 | | | | | | | |
| | (1.000) | | (1.000) | | | | | | | |
| Self-reported poor health | | | | | | | | | | |
| $\boldsymbol{y}_0$ | 0.131 | 0.143 | -0.188 | | | | | | | |
| | (0.769) | (0.864) | (0.926) | | | | | | | |
| $\boldsymbol{y}_1$ | | | | | | | | | | |
| Obesity | | | | | | | | | | |
| $\boldsymbol{y}_0$ | 0.217 | | -0.132 | -0.088 | | | | | | -0.121 |
| | (0.995) | | (0.388) | (0.224) | | | | | | (0.331) |
| $\boldsymbol{y}_1$ | 0.119 | | -0.135 | | 0.112 | 0.111 | | | | |
| | (0.430) | | (0.373) | | (0.303) | (0.236) | | | | |
| Daily smoking | | | | | | | | | | |
| $\boldsymbol{y}_0$ | 0.152 | 0.087 | | -0.107 | -0.139 | | | 0.087 | 0.110 | |
| | (0.969) | (0.347) | | (0.482) | (0.825) | | | (0.304) | (0.570) | |
| $\boldsymbol{y}_1$ | 0.227 | | -0.183 | | -0.103 | | | | 0.076 | |
| | (0.998) | | (0.896) | | (0.316) | | | | (0.209) | |
| Wages | | | | | | | | | | |
| $\boldsymbol{y}_0$ | -0.026 | | 0.050 | | | | | | | |
| | (0.371) | | (0.900) | | | | | | | |
| $\boldsymbol{y}_1$ | -0.050 | | 0.076 | | | | | 0.031 | | -0.029 |
| | (0.976) | | (1.000) | | | | | (0.380) | | (0.210) |

**Notes:** Mean of the non-zero factor loadings in each cell, with the posterior probability that the corresponding loading is different from zero in brackets, i.e., $\Pr(\delta_{jl} = 1)$. Only loadings with posterior probabilities larger than 0.20 are displayed.

- Factor 7 can be identified with the presence of Attention Deficit, especially for males (for whom the corresponding items on the Rutter and Conners scale have a bigger magnitude).

- Factors 8 and 9 can be interpreted as general noncognitive traits, as the Child Developmental Scale, and, especially for males, the locus of control and self-esteem scales, almost entirely load on them. However, taking into account the magnitude of the loadings, we can more clearly identify Factor 8 as Extraversion/Intraversion, and Factor 9 as Antisocial Behavior (both subscales of the Child Developmental Scale).[24]

- Factor 10, finally, is loaded by many items of the Rutter and Conners scales associated with Clumsiness in girls, and Aggression in boys.

**Table 3.4:** Factor loadings and corresponding posterior probabilities in the outcome system (Males)

| | $f_1$ | $f_2$ | $f_3$ | $f_4$ | $f_5$ | $f_6$ | $f_7$ | $f_8$ | $f_9$ | $f_{10}$ |
|---|---|---|---|---|---|---|---|---|---|---|
| Education | | | | | | | | | | |
| | -0.313 | | 0.460 | 0.097 | | | | | | |
| | (1.000) | | (1.000) | (0.758) | | | | | | |
| Self-reported poor health | | | | | | | | | | |
| $\boldsymbol{y}_0$ | 0.166 | 0.100 | -0.095 | | | | | | | |
| | (0.939) | (0.355) | (0.266) | | | | | | | |
| $\boldsymbol{y}_1$ | 0.098 | 0.094 | | | | | | -0.129 | 0.146 | |
| | (0.394) | (0.286) | | | | | | (0.542) | (0.606) | |
| Obesity | | | | | | | | | | |
| $\boldsymbol{y}_0$ | 0.119 | | | -0.119 | | | | -0.136 | 0.149 | -0.154 |
| | (0.542) | | | (0.428) | | | | (0.604) | (0.755) | (0.487) |
| $\boldsymbol{y}_1$ | | 0.123 | | | 0.183 | | | 0.072 | 0.081 | |
| | | (0.407) | | | (0.838) | | | (0.224) | (0.239) | |
| Daily smoking | | | | | | | | | | |
| $\boldsymbol{y}_0$ | 0.159 | | | -0.096 | | | | | 0.138 | 0.205 |
| | (0.957) | | | (0.384) | | | | | (0.850) | (0.921) |
| $\boldsymbol{y}_1$ | 0.189 | 0.106 | | | | | 0.135 | | 0.144 | |
| | (0.988) | (0.422) | | | | | (0.504) | | (0.679) | |
| Wages | | | | | | | | | | |
| $\boldsymbol{y}_0$ | -0.019 | -0.027 | 0.066 | -0.030 | -0.040 | | | 0.047 | 0.021 | |
| | (0.232) | (0.325) | (0.997) | (0.431) | (0.822) | | | (0.927) | (0.212) | |
| $\boldsymbol{y}_1$ | -0.042 | -0.025 | 0.065 | -0.023 | | -0.031 | -0.050 | 0.035 | | 0.039 |
| | (0.867) | (0.243) | (0.999) | (0.233) | | (0.414) | (0.839) | (0.664) | | (0.445) |

**Notes:** Mean of the non-zero factor loadings in each cell, with the posterior probability that the corresponding loading is different from zero in brackets, i.e., $\Pr(\delta_{jl} = 1)$. Only loadings with posterior probabilities larger than 0.20 are displayed.

The results of the effects of the factors on the outcomes are reported in Tables 3.3 and 3.4. It can be noticed that both factors loading on the cognitive tests scores (Factor 1 and Factor 3) are significant determinants of the educational choice (see Tables 3.3 and 3.4, where both posterior probabilities are equal to 1 for both genders). Interestingly, while they are the only two significant factors for women, for men also the trait of Neuroticism/Anxiety (as reflected in Factor 4) exerts a relevant influence on the educational decision (notice the posterior probability of Factor 4 is 0.758). A similar observation can be made for wages: while for females the only two relevant factors are the cognitively-loaded Factors 1 and 3, for males many noncognitive traits are significant determinants of wages, at least for one education group (although some posterior probabilities are quite low).[26]

[26] For the high educated females, also the traits of Extraversion and Clumsiness are significant determinants of wages, although the posterior probabilities are much smaller in these cases.

Coming to the discussion on the health outcomes and starting with smoking, it appears that the various factors reflecting attention disorders, problem behavior and neuroticism are significant early predictors of the probability of becoming a daily smoker at age 30, for both males and females. Interestingly, for females only, the cognitively-loaded Factor 3 (on which also the locus of control scale loads heavily) exerts a strong influence on the probability of being a daily smoker at age 30. We also notice that girls who experienced relational problems in childhood (as reflected in the traits of Clumsiness—Factor 5—and Extraversion/Introversion—Factor 8) are more likely to be daily smokers in adulthood. A few comments on the early origins of poor health can also be made. On the one hand, the presence of behavioral problems in childhood, in particular associated with the personality trait of Neuroticism (Factor 2), is a significant predictor of the probability of being in poor health in adulthood; on the other hand, cognitive ability (Factor 3) acts as a protective factor. It is worth noticing that noncognitive skills developed in childhood exert a bigger impact on health later in life in boys than in girls (Factors 8 and 9 are significant for males only). Finally, we see that the bigger role played by the cognitively-loaded Factor 3 in predicting later life health in females, and the greater importance of behavioral traits for males, is confirmed in relation to obesity.

In sum, these findings point to the fact that, while the structure of personality and cognition, as proxied by early childhood cognitive and personality tests, is remarkably similar across genders, the importance of these factors for adult outcomes varies considerably.

The treatment parameters, estimated as explained in Section 3.5.6, are reported in Tables 3.5 and 3.6. A few main points are worth underlining. First, even if our parsimonious Bayesian factor model allows us to identify a rich latent structure, only a smaller subset of factors drives the correlation between the choice and the outcomes, leaving a significant role to education in accounting for the observed adult disparities. Second, it turns out that the role played by selection in accounting for the health disparities varies significantly by gender: while for males approximately three quarters of the observed difference can be attributed to education, for females it only accounts for half of the observed disparity in self-reported poor health and daily smoking, and plays no role in explaining differences in obesity rates. On the contrary, selection on early-life factors and observed characteristics accounts for 55% of the wage gap between high- and low-educated for both males and females.[27] These findings are consistent with previous work (Conti et al., 2010). Third, the proportion

[27]Note we do not decompose selection into the part due to observables $\mathbf{X}$ and the part due to unobservables $\boldsymbol{f}$.

**Table 3.5:** Mean and distributional treatment parameter estimates (Females)

| **ATE** | **TT** | **MTE for $u_d = -2$** | **MTE for $u_d = 0$** | **MTE for $u_d = 2$** |
|---|---|---|---|---|
| Self-reported poor health *[Observed diff.: -0.072]* | | | | |
| ATE = -0.038 (0.017) | TT = -0.022 (0.019) | MTE(-2) = -0.072 (0.022) | MTE(0) = -0.036 (0.017) | MTE(2) = -0.008 (0.023) |
| $\Pr(\Delta = t) =$ | $\Pr(\Delta = t \mid d = 1) =$ | $\Pr(\Delta = t \mid u_d = -2) =$ | $\Pr(\Delta = t \mid u_d = 0) =$ | $\Pr(\Delta = t \mid u_d = 2) =$ |
| $[t = -1]$ 0.133 (0.011) | $[t = -1]$ 0.120 (0.014) | $[t = -1]$ 0.164 (0.015) | $[t = -1]$ 0.131 (0.011) | $[t = -1]$ 0.108 (0.017) |
| $[t = 0]$ 0.772 (0.012) | $[t = 0]$ 0.783 (0.015) | $[t = 0]$ 0.743 (0.015) | $[t = 0]$ 0.774 (0.012) | $[t = 0]$ 0.791 (0.018) |
| $[t = 1]$ 0.095 (0.010) | $[t = 1]$ 0.098 (0.010) | $[t = 1]$ 0.092 (0.012) | $[t = 1]$ 0.095 (0.010) | $[t = 1]$ 0.101 (0.011) |
| Obesity *[Observed diff.: -0.024]* | | | | |
| ATE = 0.005 (0.016) | TT = 0.010 (0.016) | MTE(-2) = -0.005 (0.025) | MTE(0) = 0.006 (0.016) | MTE(2) = 0.015 (0.018) |
| $\Pr(\Delta = t) =$ | $\Pr(\Delta = t \mid d = 1) =$ | $\Pr(\Delta = t \mid u_d = -2) =$ | $\Pr(\Delta = t \mid u_d = 0) =$ | $\Pr(\Delta = t \mid u_d = 2) =$ |
| $[t = -1]$ 0.087 (0.009) | $[t = -1]$ 0.080 (0.011) | $[t = -1]$ 0.103 (0.012) | $[t = -1]$ 0.086 (0.009) | $[t = -1]$ 0.074 (0.012) |
| $[t = 0]$ 0.822 (0.012) | $[t = 0]$ 0.830 (0.013) | $[t = 0]$ 0.799 (0.018) | $[t = 0]$ 0.823 (0.012) | $[t = 0]$ 0.837 (0.015) |
| $[t = 1]$ 0.092 (0.011) | $[t = 1]$ 0.090 (0.010) | $[t = 1]$ 0.098 (0.018) | $[t = 1]$ 0.091 (0.011) | $[t = 1]$ 0.089 (0.012) |
| Daily smoking *[Observed diff.: -0.174]* | | | | |
| ATE = -0.094 (0.023) | TT = -0.113 (0.023) | MTE(-2) = -0.054 (0.035) | MTE(0) = -0.094 (0.023) | MTE(2) = -0.130 (0.026) |
| $\Pr(\Delta = t) =$ | $\Pr(\Delta = t \mid d = 1) =$ | $\Pr(\Delta = t \mid u_d = -2) =$ | $\Pr(\Delta = t \mid u_d = 0) =$ | $\Pr(\Delta = t \mid u_d = 2) =$ |
| $[t = -1]$ 0.237 (0.014) | $[t = -1]$ 0.243 (0.015) | $[t = -1]$ 0.229 (0.017) | $[t = -1]$ 0.237 (0.014) | $[t = -1]$ 0.248 (0.018) |
| $[t = 0]$ 0.619 (0.012) | $[t = 0]$ 0.626 (0.013) | $[t = 0]$ 0.596 (0.012) | $[t = 0]$ 0.621 (0.011) | $[t = 0]$ 0.634 (0.016) |
| $[t = 1]$ 0.144 (0.011) | $[t = 1]$ 0.131 (0.010) | $[t = 1]$ 0.175 (0.020) | $[t = 1]$ 0.142 (0.011) | $[t = 1]$ 0.118 (0.011) |
| Wages *[Observed diff.: 0.213]* | | | | |
| ATE = 0.118 (0.021) | TT = 0.138 (0.024) | MTE(-2) = 0.080 (0.028) | MTE(0) = 0.118 (0.021) | MTE(2) = 0.162 (0.031) |

**Notes:** MCMC standard errors in brackets. Integration of the latent factors using 1,000 draws from their prior distribution.

**Table 3.6:** Mean and distributional treatment parameter estimates (Males)

| **ATE** | **TT** | **MTE for $u_d = -2$** | **MTE for $u_d = 0$** | **MTE for $u_d = 2$** |
|---|---|---|---|---|
| Self-reported poor health *[Observed diff.: -0.092]* | | | | |
| ATE = -0.070 (0.020) | TT = -0.060 (0.023) | MTE(-2) = -0.091 (0.024) | MTE(0) = -0.069 (0.021) | MTE(2) = -0.050 (0.027) |
| $\Pr(\Delta = t) =$ | $\Pr(\Delta = t \mid d = 1) =$ | $\Pr(\Delta = t \mid u_d = -2) =$ | $\Pr(\Delta = t \mid u_d = 0) =$ | $\Pr(\Delta = t \mid u_d = 2) =$ |
| $[t = -1]$ 0.172 (0.015) | $[t = -1]$ 0.165 (0.017) | $[t = -1]$ 0.190 (0.017) | $[t = -1]$ 0.170 (0.015) | $[t = -1]$ 0.158 (0.021) |
| $[t = 0]$ 0.726 (0.015) | $[t = 0]$ 0.731 (0.017) | $[t = 0]$ 0.712 (0.016) | $[t = 0]$ 0.728 (0.015) | $[t = 0]$ 0.733 (0.020) |
| $[t = 1]$ 0.102 (0.010) | $[t = 1]$ 0.105 (0.010) | $[t = 1]$ 0.099 (0.012) | $[t = 1]$ 0.101 (0.010) | $[t = 1]$ 0.108 (0.011) |
| Obesity *[Observed diff.: -0.054]* | | | | |
| ATE = -0.037 (0.018) | TT = -0.034 (0.019) | MTE(-2) = -0.041 (0.019) | MTE(0) = -0.036 (0.018) | MTE(2) = -0.033 (0.023) |
| $\Pr(\Delta = t) =$ | $\Pr(\Delta = t \mid d = 1) =$ | $\Pr(\Delta = t \mid u_d = -2) =$ | $\Pr(\Delta = t \mid u_d = 0) =$ | $\Pr(\Delta = t \mid u_d = 2) =$ |
| $[t = -1]$ 0.127 (0.013) | $[t = -1]$ 0.127 (0.014) | $[t = -1]$ 0.130 (0.013) | $[t = -1]$ 0.126 (0.013) | $[t = -1]$ 0.129 (0.017) |
| $[t = 0]$ 0.782 (0.014) | $[t = 0]$ 0.780 (0.015) | $[t = 0]$ 0.781 (0.014) | $[t = 0]$ 0.783 (0.014) | $[t = 0]$ 0.775 (0.018) |
| $[t = 1]$ 0.091 (0.010) | $[t = 1]$ 0.093 (0.010) | $[t = 1]$ 0.089 (0.010) | $[t = 1]$ 0.090 (0.010) | $[t = 1]$ 0.096 (0.011) |
| Daily smoking *[Observed diff.: -0.206]* | | | | |
| ATE = -0.155 (0.025) | TT = -0.155 (0.026) | MTE(-2) = -0.154 (0.029) | MTE(0) = -0.154 (0.025) | MTE(2) = -0.155 (0.030) |
| $\Pr(\Delta = t) =$ | $\Pr(\Delta = t \mid d = 1) =$ | $\Pr(\Delta = t \mid u_d = -2) =$ | $\Pr(\Delta = t \mid u_d = 0) =$ | $\Pr(\Delta = t \mid u_d = 2) =$ |
| $[t = -1]$ 0.289 (0.017) | $[t = -1]$ 0.289 (0.018) | $[t = -1]$ 0.295 (0.018) | $[t = -1]$ 0.288 (0.017) | $[t = -1]$ 0.288 (0.022) |
| $[t = 0]$ 0.575 (0.012) | $[t = 0]$ 0.577 (0.014) | $[t = 0]$ 0.565 (0.011) | $[t = 0]$ 0.577 (0.012) | $[t = 0]$ 0.578 (0.017) |
| $[t = 1]$ 0.135 (0.010) | $[t = 1]$ 0.134 (0.010) | $[t = 1]$ 0.141 (0.013) | $[t = 1]$ 0.135 (0.010) | $[t = 1]$ 0.133 (0.011) |
| Wages *[Observed diff.: 0.227]* | | | | |
| ATE = 0.125 (0.022) | TT = 0.138 (0.024) | MTE(-2) = 0.110 (0.031) | MTE(0) = 0.127 (0.022) | MTE(2) = 0.151 (0.031) |

**Notes:** MCMC standard errors in brackets. Integration of the latent factors using 1,000 draws from their prior distribution.

**Table 3.7:** Testing for the equality of ATE and TT

| $\xi$ = ATE − TT | Mean | SE | [90% conf. int.][a] | | $p$-val[b] |
|---|---|---|---|---|---|
| Females | | | | | |
| Bad health | -0.016 | (0.008) | [-0.027 | -0.002] | 0.034 |
| Obesity | -0.006 | (0.008) | [-0.018 | 0.008] | 0.210 |
| Daily smoking | 0.019 | (0.010) | [0.001 | 0.035] | 0.961 |
| Wages | -0.021 | (0.011) | [-0.039 | -0.004] | 0.022 |
| Males | | | | | |
| Bad health | -0.010 | (0.008) | [-0.024 | 0.002] | 0.081 |
| Obesity | -0.003 | (0.005) | [-0.012 | 0.005] | 0.321 |
| Daily smoking | -0.000 | (0.007) | [-0.012 | 0.012] | 0.475 |
| Wages | -0.013 | (0.011) | [-0.030 | 0.004] | 0.106 |

**Notes:** MCMC standard errors in brackets. [a] 90% confidence interval of the sample statistic $\xi$. [b] Tail-area probabilities $\Pr(\xi \geqslant 0)$.

of people who do not change their behavior as a consequence of the treatment is between 70% and 80% for males, and between 75% and 85% for females for the cases of self-reported health or obesity. For smoking behavior, the proportion of people affected by the treatment is much higher than for self-reported health and obesity (about 38% for females and 43% for males, see the distributional treatment parameters in Tables 3.5 and 3.6). This can clearly be seen in Figures 3.5 and 3.6, which also display the proportion of people who benefit and who lose from the treatment.

Tables 3.5 and 3.6 also report estimates of the TT and of the MTE, as defined in Section 3.4.2. Formal tests for the equality of ATE and TT, and for the flatness of the MTE, are reported in Tables 3.7 and 3.8. There are two main results we want to emphasize. First, we find significant evidence of selection on gains for females, and some evidence for males. Both the tests for equality of ATE and TT, and for the flatness of the MTE, report, respectively, high probabilities of a non-zero difference between ATE and TT and between different portions of the MTE, for the case of poor health, smoking and wages for females, and—although to a smaller extent—for poor health and wages for males. Second, the results we obtain for females are able to rationalize some of the contradictory findings present in the education-health literature. Indeed, we find significant evidence of positive selection on gains for smoking: the women more likely to acquire A-levels or above possess certain traits that make them also less likely to smoke, as effect of their educational choice. Now, the fact that women at different margins enjoy different health returns from education reconciles the contradictory findings reported in the literature, whereby

**Table 3.8:** Testing for the flatness of the MTEs: $\xi(u, u') = \text{MTE}(u) - \text{MTE}(u')$

| | $\xi(2,0)$ | | | $\xi(0,-2)$ | | | $\xi(2,-2)$ | | |
|---|---|---|---|---|---|---|---|---|---|
| | $\text{E}[\xi]$ | SE | $p$-val[a] | $\text{E}[\xi]$ | SE | $p$-val[a] | $\text{E}[\xi]$ | SE | $p$-val[a] |
| Females | | | | | | | | | |
| Poor health | 0.028 | (0.013) | 0.966 | 0.036 | (0.017) | 0.970 | 0.064 | (0.029) | 0.968 |
| Obesity | 0.009 | (0.013) | 0.777 | 0.011 | (0.017) | 0.773 | 0.020 | (0.029) | 0.776 |
| Daily smoking | -0.036 | (0.017) | 0.023 | -0.041 | (0.021) | 0.035 | -0.077 | (0.038) | 0.028 |
| Wages | 0.044 | (0.021) | 0.987 | 0.039 | (0.021) | 0.975 | 0.083 | (0.041) | 0.982 |
| Males | | | | | | | | | |
| Poor health | 0.019 | (0.014) | 0.908 | 0.022 | (0.016) | 0.932 | 0.041 | (0.031) | 0.919 |
| Obesity | 0.003 | (0.011) | 0.560 | 0.006 | (0.010) | 0.641 | 0.009 | (0.021) | 0.596 |
| Daily smoking | -0.001 | (0.015) | 0.478 | 0.001 | (0.015) | 0.531 | -0.001 | (0.030) | 0.504 |
| Wages | 0.024 | (0.021) | 0.869 | 0.018 | (0.021) | 0.802 | 0.042 | (0.042) | 0.839 |

**Notes:** MCMC standard errors in brackets. [a] Tail-area probabilities $\Pr(\xi(u, u') \geqslant 0)$.

the women more likely to acquire higher education ($u_d = 2$, closer to the margin in Currie and Moretti, 2003) enjoy significantly higher returns (more than double) than the women at a lower margin ($u_d = -2$, closer to McCrary and Royer, 2010). Along the same lines, we also find significant evidence of positive selection on gains in the labor market, whereby women select into education on the basis of the returns in terms of higher wages (notice that the same evidence is present for men, although the respective tail-area probabilities are bigger for them). Finally, we find an apparently surprising result for the case of poor health: that individuals less likely to acquire a higher level of education are actually more likely to benefit from it, in terms of health returns.[28] Two remarks are worth making here. First, this result is consistent with previous work (Auld and Sidhu, 2005) that has tried to test for the presence of selection on gains by means of Wooldridge (2003) and Garen (1984) correlated random coefficient models. Second, for individuals at the top of the distribution of unobservables, the presence of higher wages make the choice to continue still profitable, despite the smaller returns to health. However, we cannot rule out the possibility that the self-reported nature of the health variable might have an effect on the results. Unfortunately, more objective health measures are not available in the BCS data, so we defer the answer to this question to another occasion.

## 3.7 Classical methods

In this section, we compare the results obtained with the PBFA to those obtained using traditional methods, namely principal component analysis (PCA) and standard factor analysis (SFA) estimated with the method of maximum likelihood. It is worth remarking that, while previous research has relied on a priori decisions on model structure and subscale construction to various degrees (as detailed in Section 3.2), here we are agnostic with regard to the latent structure. Hence, we apply both PCA and SFA to all measurements, and an orthogonal rotation is performed after having determined the number of components/factors to extract, in order to make the analysis as comparable as possible to the one performed using the Bayesian method.

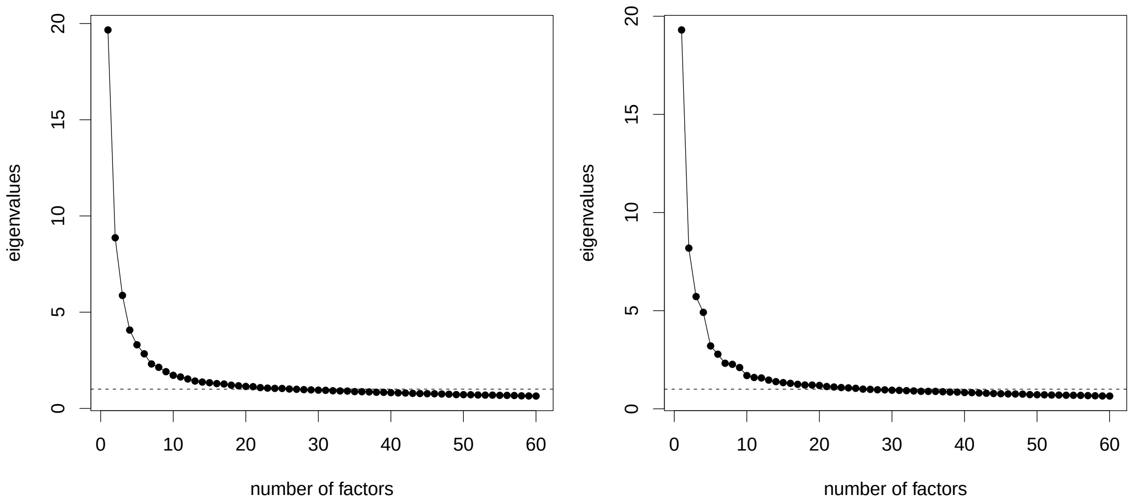
[28]Women at the top of the distribution of the unobservables actually have zero health benefits.

### 3.7.1 Selecting the number of factors

Selecting the number of factors is always a tricky issue in conventional factor analysis approaches. Many different criteria have been proposed in the literature, sometimes providing contradictory results (Linn, 1968; Tucker et al., 1969; Hakstian et al., 1982). The most widely used one is probably the Kaiser (1960) rule, which retains only the factors with eigenvalues greater than 1. This criterion suggests 26 components for PCA and 11 factors for SFA for both genders. Another popular approach is the scree test introduced by Cattell (1966). This graphical method consists of retaining the number of factors where an elbow appears on the curve plotting the eigenvalues values in decreasing order. The scree plots in Figures 3.1 and 3.2 show how difficult it is in our case to make any conclusion based on this approach, since no clear elbows appear in these graphs. These figures also show that many eigenvalues are actually very close to 1, especially for PCA.[29] Thus, the Kaiser criterion is not as clear-cut as it first appears, and might therefore be questionable. Another method, relying on matrices of partial correlations, was proposed by Velicer (1976). It gives an exact stopping rule to select the number of components, and suggests to retain 13 components for both genders in our case.

**Figure 3.1:** Scree plot, Principal Component Analysis on the 126 items

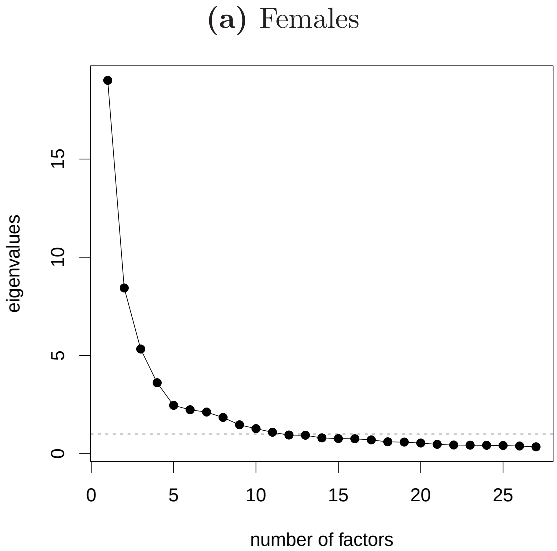
**(a)** Females **(b)** Males

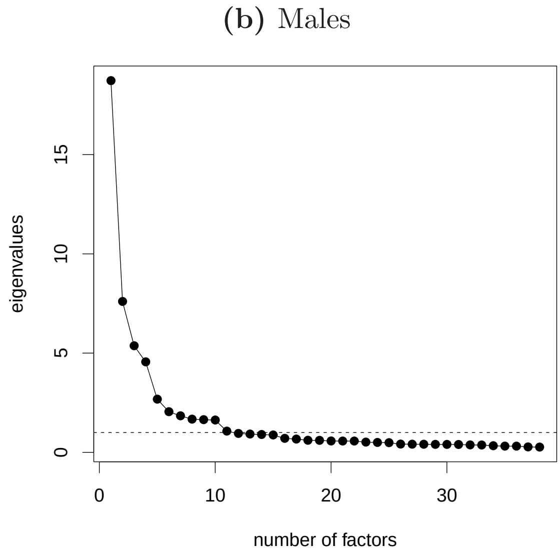



**Notes:** Cattell (1966) suggested to select the number of factors where an elbow appears on the curve. According to the Kaiser (1960) criterion, only factors with eigenvalues greater than 1 (above the horizontal dashed line) should be retained—26 factors for both genders.

[29]Shortcomings in the use of PCA for selecting the number of components and constructing cognitive ability measures are also pointed out in Conti and Pudney (2009), who develop an "eliminant method" as alternative approach.

**Figure 3.2:** Scree plot, Standard Factor Analysis on the 126 items




**Notes:** Cattell (1966) suggested to select the number of factors where an elbow appears on the curve. According to the Kaiser (1960) criterion, only factors with eigenvalues greater than 1 (above the horizontal dashed line) should be retained—11 factors for both genders.

The disparity of the answers provided by the different criteria reveals the difficulty inherent to the selection of the number of factors with the conventional approaches. For reasons of parsimony, we decided to extract the smaller number of components, and we performed an orthogonal rotation to make the interpretation of the results easier.[30] Differently from PBFA, much greater sparsity is attained with PCA, with basically each component identifying a different trait. As noticed in PBFA, it can also be confirmed that the latent structure underlying the measurements is gender-invariant. Concerning the standard factor analysis, the factors underlying the structure of the measurements are comparable to the ones obtained using PBFA.

### 3.7.2 Extracted components/factors and their impact on the outcomes

After the extraction of the components, we compute individual scores in a second step and include them in ordinary least squares (for wages) and probit regressions (for the dichotomous health outcomes), alongside with the observed covariates $\boldsymbol{x}$. The main results are reported in Appendix 3.B and can be summarized as follows. It turns out that the PCA approach fails to capture the impact of cognitive ability on health: while in the PBFA approach cognitive ability is a significant determinant

[30]Detailed results are presented in Appendix 3.B.

of the health outcomes for both males and females, in the PCA approach it only achieves statistical significance in the case of poor health for females in the low education group, although it is a strong and significant determinant of education and labor market outcomes for both genders.

Concerning the standard factor analysis approach, some striking facts can be underlined. Despite the similarities between SFA and PBFA for the measurement system, there are differences in the effects of the factors on the outcomes.[30] As in the PBFA, the two cognitively-loaded factors are significant determinants of the educational choice. However, while they are the only two significant factors for females in the PBFA (for males, also the trait of Neuroticism/Anxiety plays a significant role), in SFA many purely noncognitive factors are also important determinants of the educational choice, especially the ones related to Attention and Problem Behaviors. For males, overall the pattern of results looks quite different between the two approaches. The main difference emerges on the role played by noncognitive skills: for wages, health and obesity, the PBFA uncovers a bigger role for them than the SFA—this is especially true for the traits of Neuroticism and Antisocial Behavior. For females, instead, there is no general pattern in the difference between the two approaches: for most outcomes, the majority of the significant loadings is common to both approaches, and in general one or two—purely noncognitive—factors are relevant only in one approach. In particular, the traditional SFA attributes a greater role to Attention Problems in childhood.

The parsimonious Bayesian factor analysis therefore provides a different picture of the importance of cognitive ability and personality traits in explaining health outcomes compared to conventional approaches, not only because it makes it possible to determine the number of factors, with the uncertainty associated to this number, but also because the extracted latent factors play different roles in explaining health.

## 3.8 Conclusions

In this chapter, we present novel Bayesian econometric methods for reducing high-dimensional data into low-dimensional aggregates. We implement the *Parsimonious Bayesian Factor Analysis* (PBFA) of Frühwirth-Schnatter and Lopes (2009) to the 1970 British Cohort Study within a life course framework to analyze the effect of childhood cognitive ability and psychosocial traits on education and adult health. For the first time, the market and non-market returns to education are estimated in a modeling framework that simultaneously accounts for selection on gains and endogenous determination of the early factors driving the selection process. Several conclusions can be reached.

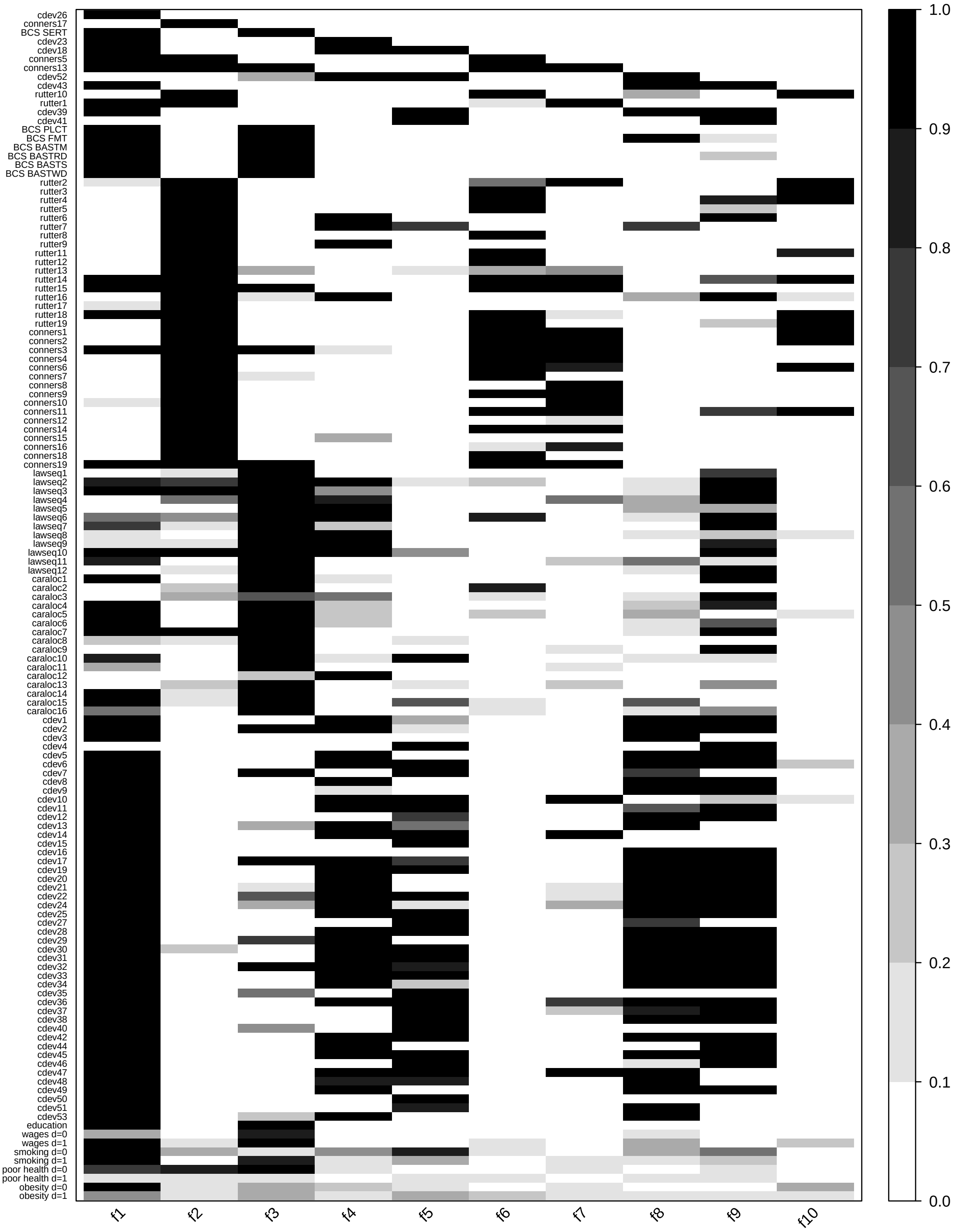
First, while the structure of personality and cognition is gender-invariant, the importance of these factors for later outcomes varies considerably across genders. Noncognitive factors developed by age ten are more important determinants of adult outcomes for males than for females.

Second, we find that the importance of education in explaining adult disparities also varies considerably by gender: for males, it accounts for three quarters of the observed health disparities, but only for a half or less of them for females.

Third, we show that individuals select into education on the basis of both market and non-market gains, and that the average effect of education for females at different margins of the distribution of the unobservables varies in a way that rationalizes the findings reported in the literature.

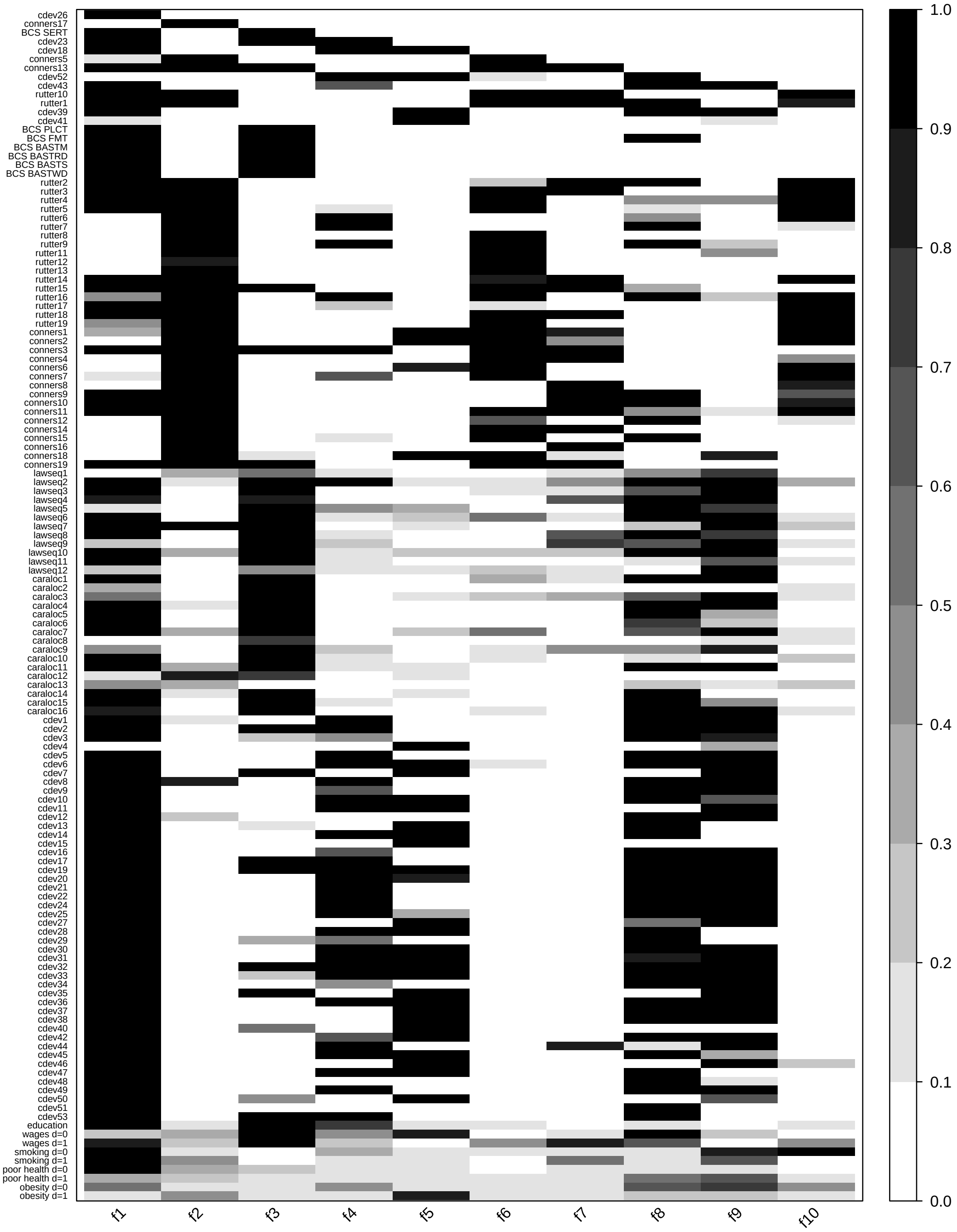
Finally, we provide evidence that a misspecification of the latent factor structure leads to an incorrect assessment of the importance of early-life conditions in influencing both education and later-life health: while principal component analysis (PCA) fails to capture the importance of cognitive ability for later-life health, standard factor analysis (SFA) fails to capture the role played by noncognitive skills in determining adult health for males.

**Figure 3.3:** Factor loadings posterior probabilities from PBFA (Females)



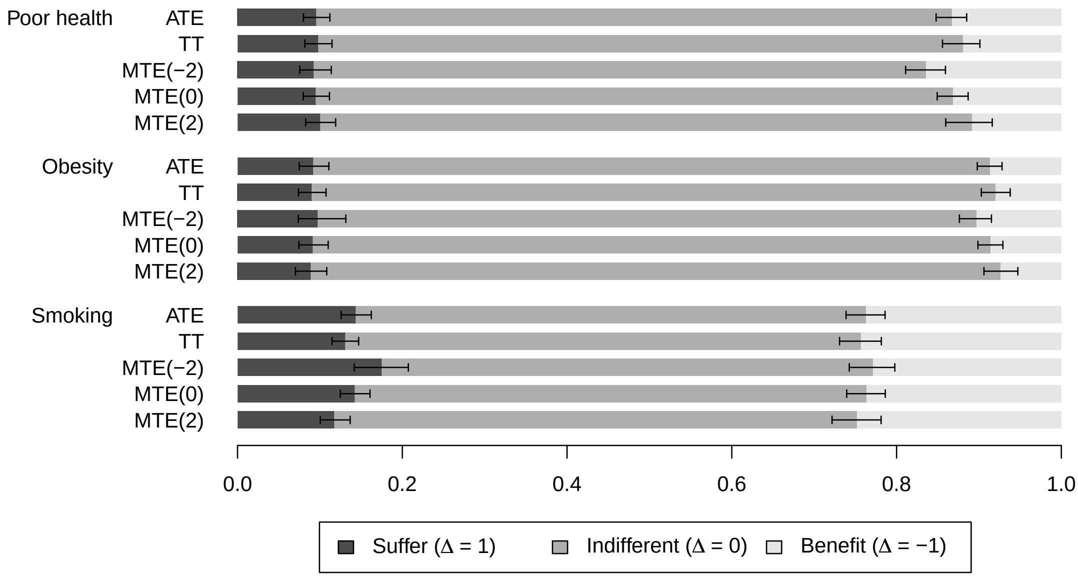

**Notes:** Each column represents a latent factor and shows the items/outcomes that significantly load on it (dark cells). Items: BCS = Cognitive Test Scores, conners = Conners Hyperactivity Scale, rutter = Rutter Parental 'A' Scale of Behavioral Disorder, lawseq = Self-Esteem Scale, caraloc = Locus of Control Scale, cdev = Child Development Scale. Education outcome, wages and health outcomes by levels of education at the bottom of the picture.

**Figure 3.4:** Factor loadings posterior probabilities from PBFA (Males)



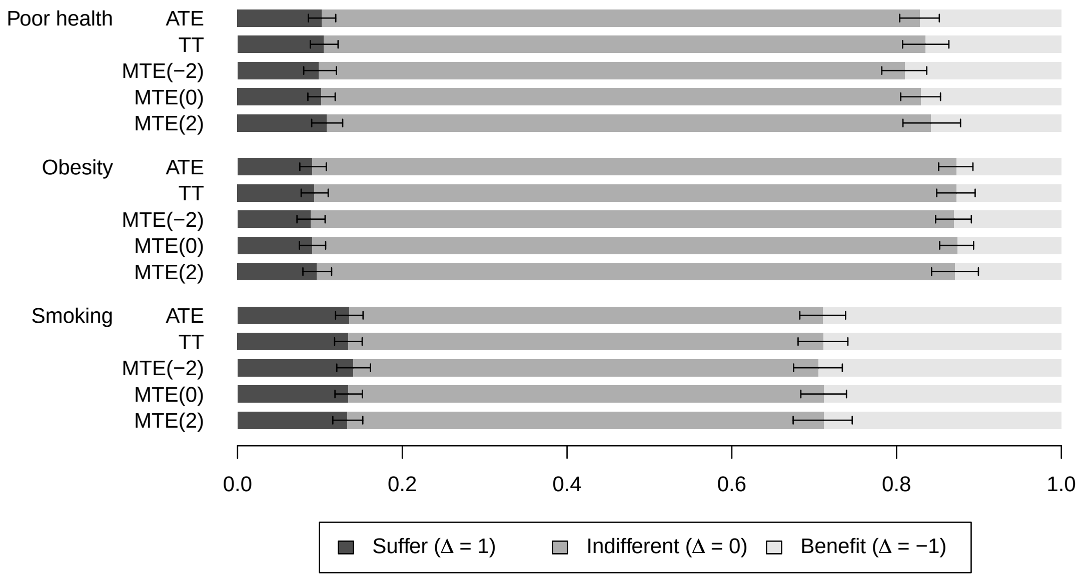

**Notes:** Each column represents a latent factor and shows the items/outcomes that significantly load on it (dark cells). Items: BCS = Cognitive Test Scores, conners = Conners Hyperactivity Scale, rutter = Rutter Parental 'A' Scale of Behavioral Disorder, lawseq = Self-Esteem Scale, caraloc = Locus of Control Scale, cdev = Child Development Scale. Education outcome, wages and health outcomes by levels of education at the bottom of the picture.

**Figure 3.5:** Distributional treatment effects (Females)




**Notes:** Distributional versions of the Average Treatment Effect (ATE), Treatment effect on the Treated (TT) and Marginal Treatment Effect (MTE) for different levels of unobservables (-2,0,2). Each bar displays the proportion of the population that suffers, is indifferent and benefits from the treatment. 90% confidence intervals are also displayed.

**Figure 3.6:** Distributional treatment effects (Males)




**Notes:** Distributional versions of the Average Treatment Effect (ATE), Treatment effect on the Treated (TT) and Marginal Treatment Effect (MTE) for different levels of unobservables (-2,0,2). Each bar displays the proportion of the population that suffers, is indifferent and benefits from the treatment. 90% confidence intervals are also displayed.

**Table 3.9:** Factor loadings in the measurement system from PBFA (Females)

| Items | $f_1$ | $f_2$ | $f_3$ | $f_4$ | $f_5$ | $f_6$ | $f_7$ | $f_8$ | $f_9$ | $f_{10}$ |
|---|---|---|---|---|---|---|---|---|---|---|
| cdev26 | 0.805 | | | | | | | | | |
| conners17 | | 0.603 | | | | | | | | |
| BCS SERT | -0.382 | | 0.512 | | | | | | | |
| cdev23 | 0.386 | | | 0.758 | | | | | | |
| cdev18 | 0.304 | | | 0.078 | 0.601 | | | | | |
| conners5 | -0.066 | 0.272 | | | | 0.566 | | | | |
| conners13 | 0.149 | 0.343 | -0.097 | | | 0.166 | 0.350 | | | |
| cdev52 | | | | -0.424 | -0.139 | | | 0.593 | | |
| cdev43 | 0.285 | | | | | | | 0.154 | 0.580 | |
| rutter10 | | 0.292 | | | | 0.438 | | | | 0.169 |
| rutter1 | 0.076 | 0.299 | | | | | 0.531 | | | |
| cdev39 | 0.301 | | | | 0.201 | | | 0.138 | 0.216 | |
| cdev41 | | | | | 0.306 | | | | 0.199 | |
| BCS PLCT | -0.172 | | 0.423 | | | | | | | |
| BCS FMT | -0.356 | | 0.501 | | | | | 0.060 | | |
| BCS BASTM | -0.306 | | 0.404 | | | | | | | |
| BCS BASTRD | -0.168 | | 0.351 | | | | | | | |
| BCS BASTS | -0.223 | | 0.512 | | | | | | | |
| BCS BASTWD | -0.248 | | 0.494 | | | | | | | |
| rutter2 | | 0.370 | | | | | 0.469 | | | -0.096 |
| rutter3 | | 0.271 | | | | 0.396 | | | | 0.106 |
| rutter4 | | 0.362 | | | | 0.225 | | | 0.057 | 0.143 |
| rutter5 | | 0.307 | | | | 0.213 | | | | |
| rutter6 | | 0.426 | | 0.135 | | | | | -0.111 | |
| rutter7 | | 0.257 | | 0.095 | | | | | | |
| rutter8 | | 0.623 | | | | -0.115 | | | | |
| rutter9 | | 0.602 | | 0.084 | | | | | | |
| rutter11 | | 0.186 | | | | 0.321 | | | | 0.071 |
| rutter12 | | 0.127 | | | | 0.096 | | | | |
| rutter13 | | 0.245 | | | | | | | | |
| rutter14 | 0.080 | 0.485 | | | | 0.112 | 0.091 | | | 0.106 |
| rutter15 | 0.113 | 0.351 | -0.092 | | | 0.143 | 0.367 | | | |
| rutter16 | | 0.378 | | 0.160 | | | | | -0.146 | |
| rutter17 | | 0.340 | | | | | | | | |
| rutter18 | 0.092 | 0.417 | | | | 0.288 | | | | 0.152 |
| rutter19 | | 0.431 | | | | 0.380 | | | | 0.186 |
| conners1 | | 0.333 | | | | 0.392 | 0.145 | | | -0.514 |
| conners2 | | 0.342 | | | | 0.385 | 0.132 | | | -0.532 |
| conners3 | 0.266 | 0.360 | -0.106 | | | 0.107 | 0.320 | | | |
| conners4 | | 0.318 | | | | 0.151 | 0.205 | | | |
| conners6 | | 0.310 | | | | 0.534 | 0.074 | | | -0.383 |
| conners7 | | 0.469 | | | | 0.109 | | | | |
| conners8 | | 0.625 | | | | | 0.091 | | | |
| conners9 | | 0.429 | | | | 0.088 | 0.440 | | | |
| conners10 | | 0.477 | | | | | 0.312 | | | |
| conners11 | | 0.410 | | | | 0.264 | 0.137 | | 0.050 | 0.126 |
| conners12 | | 0.570 | | | | | | | | |
| conners14 | | 0.294 | | | | 0.185 | 0.200 | | | |
| conners15 | | 0.579 | | | | | | | | |
| conners16 | | 0.694 | | | | | 0.074 | | | |
| conners18 | | 0.231 | | | | 0.437 | | | | |

continued on next page...

Table 3.9 – continued from previous page

| Items | $f_1$ | $f_2$ | $f_3$ | $f_4$ | $f_5$ | $f_6$ | $f_7$ | $f_8$ | $f_9$ | $f_{10}$ |
|---|---|---|---|---|---|---|---|---|---|---|
| conners19 | 0.154 | 0.332 | -0.091 | | | 0.218 | 0.261 | | | |
| lawseq1 | | | 0.197 | | | | | | | |
| lawseq2 | -0.093 | | 0.225 | -0.141 | | | | | -0.170 | |
| lawseq3 | -0.101 | -0.109 | 0.342 | | | | | | -0.172 | |
| lawseq4 | | | 0.226 | -0.082 | | | | | -0.202 | |
| lawseq5 | | | 0.250 | -0.140 | | | | | | |
| lawseq6 | | | 0.369 | -0.154 | | -0.090 | | | -0.145 | |
| lawseq7 | | | 0.271 | | | | | | -0.142 | |
| lawseq8 | | | 0.347 | -0.185 | | | | | | |
| lawseq9 | | | 0.302 | -0.098 | | | | | -0.089 | |
| lawseq10 | -0.114 | -0.109 | 0.287 | -0.137 | | | | | -0.163 | |
| lawseq11 | -0.114 | | 0.294 | | | | | | | |
| lawseq12 | | | 0.316 | | | | | | -0.202 | |
| caraloc1 | -0.100 | | 0.573 | | | | | | -0.143 | |
| caraloc2 | | | 0.136 | | | 0.078 | | | | |
| caraloc3 | | | | | | | | | 0.118 | |
| caraloc4 | -0.224 | | 0.635 | | | | | | -0.097 | |
| caraloc5 | -0.147 | | 0.560 | | | | | | | |
| caraloc6 | -0.189 | | 0.476 | | | | | | | |
| caraloc7 | -0.116 | -0.122 | 0.295 | | | | | | -0.141 | |
| caraloc8 | | | 0.163 | | | | | | | |
| caraloc9 | | | 0.276 | | | | | | -0.088 | |
| caraloc10 | -0.074 | | 0.354 | | -0.094 | | | | | |
| caraloc11 | | | 0.583 | | | | | | | |
| caraloc12 | | | | -0.111 | | | | | | |
| caraloc13 | | | 0.160 | | | | | | | |
| caraloc14 | -0.135 | | 0.433 | | | | | | | |
| caraloc15 | -0.259 | | 0.596 | | | | | | | |
| caraloc16 | | | 0.389 | | | | | | | |
| cdev1 | 0.632 | | | 0.081 | | | | -0.213 | -0.066 | |
| cdev2 | 0.400 | | -0.095 | 0.503 | | | | -0.249 | -0.132 | |
| cdev3 | 0.608 | | | | | | | -0.123 | | |
| cdev4 | | | | | 0.293 | | | | 0.137 | |
| cdev5 | 0.463 | | | 0.170 | | | | 0.206 | 0.393 | |
| cdev6 | 0.298 | | | 0.149 | 0.427 | | | 0.172 | 0.162 | |
| cdev7 | -0.381 | | 0.106 | | -0.226 | | | | | |
| cdev8 | 0.306 | | | 0.142 | | | | 0.206 | 0.526 | |
| cdev9 | 0.341 | | | | | | | 0.194 | 0.555 | |
| cdev10 | 0.387 | | | 0.173 | 0.480 | | -0.108 | | | |
| cdev11 | 0.381 | | | 0.402 | 0.178 | | | | 0.247 | |
| cdev12 | 0.655 | | | | 0.048 | | | -0.112 | 0.135 | |
| cdev13 | -0.676 | | | 0.073 | | | | 0.151 | | |
| cdev14 | 0.294 | | | 0.179 | 0.471 | | -0.118 | | | |
| cdev15 | -0.151 | | | | -0.259 | | | | | |
| cdev16 | 0.478 | | | | | | | 0.224 | 0.427 | |
| cdev17 | 0.646 | | -0.134 | 0.282 | | | | -0.121 | -0.119 | |
| cdev19 | 0.358 | | | 0.628 | 0.133 | | | -0.127 | -0.103 | |
| cdev20 | 0.381 | | | 0.497 | | | | 0.274 | 0.113 | |
| cdev21 | 0.454 | | | 0.281 | | | | 0.265 | 0.447 | |
| cdev22 | 0.432 | | | 0.166 | -0.067 | | | 0.505 | 0.230 | |
| cdev24 | 0.472 | | | 0.174 | | | | 0.370 | 0.216 | |
| cdev25 | 0.628 | | | 0.067 | 0.071 | | | 0.252 | 0.130 | |

continued on next page...

Table 3.9 – continued from previous page

| Items | $f_1$ | $f_2$ | $f_3$ | $f_4$ | $f_5$ | $f_6$ | $f_7$ | $f_8$ | $f_9$ | $f_{10}$ |
|---|---|---|---|---|---|---|---|---|---|---|
| cdev27 | -0.242 | | | | -0.372 | | | 0.065 | | |
| cdev28 | 0.414 | | | 0.141 | 0.489 | | | 0.143 | 0.069 | |
| cdev29 | -0.741 | | 0.045 | 0.097 | | | | 0.128 | -0.069 | |
| cdev30 | 0.373 | | | 0.449 | 0.197 | | | -0.123 | 0.335 | |
| cdev31 | 0.419 | | | 0.378 | 0.169 | | | 0.148 | 0.239 | |
| cdev32 | 0.710 | | -0.113 | 0.110 | 0.065 | | | -0.098 | -0.062 | |
| cdev33 | 0.121 | | | 0.381 | 0.178 | | | -0.360 | 0.196 | |
| cdev34 | 0.465 | | | 0.126 | | | | 0.172 | 0.563 | |
| cdev35 | -0.335 | | | | -0.401 | | | | | |
| cdev36 | 0.507 | | | 0.152 | 0.243 | | | -0.292 | 0.071 | |
| cdev37 | 0.280 | | | | 0.195 | | | 0.061 | 0.444 | |
| cdev38 | 0.253 | | | | 0.182 | | | 0.115 | 0.210 | |
| cdev40 | 0.294 | | | | 0.368 | | | | | |
| cdev42 | 0.294 | | | 0.170 | 0.388 | | | 0.116 | 0.204 | |
| cdev44 | 0.410 | | | 0.230 | | | | | 0.555 | |
| cdev45 | 0.174 | | | 0.111 | 0.235 | | | 0.085 | 0.146 | |
| cdev46 | 0.143 | | | | 0.210 | | | | 0.141 | |
| cdev47 | 0.314 | | | 0.319 | 0.424 | | -0.107 | -0.208 | | |
| cdev48 | -0.616 | | | 0.054 | -0.067 | | | 0.179 | | |
| cdev49 | 0.448 | | | 0.181 | | | | 0.189 | 0.333 | |
| cdev50 | -0.288 | | | | -0.304 | | | | | |
| cdev51 | 0.617 | | | | 0.070 | | | -0.150 | | |
| cdev53 | 0.280 | | | 0.694 | | | | -0.227 | | |

**Notes:** Only factor loadings with a posterior probability greater than 0.75 are displayed. White cell if $|\alpha_{jl}| < 0.4$, light gray cell if $0.4 \leqslant |\alpha_{jl}| < 0.7$ and dark gray cell if $|\alpha_{jl}| \geqslant 0.7$. Items: BCS = Cognitive Test Scores, conners = Conners Hyperactivity Scale, rutter = Rutter Parental 'A' Scale of Behavioral Disorder, lawseq = Self-Esteem Scale, caraloc = Locus of Control Scale, cdev = Child Development Scale.

**Table 3.10:** Factor loadings in the measurement system from PBFA (Males)

| Items | $f_1$ | $f_2$ | $f_3$ | $f_4$ | $f_5$ | $f_6$ | $f_7$ | $f_8$ | $f_9$ | $f_{10}$ |
|---|---|---|---|---|---|---|---|---|---|---|
| cdev26 | 0.889 | | | | | | | | | |
| conners17 | | 0.743 | | | | | | | | |
| BCS SERT | -0.374 | | 0.636 | | | | | | | |
| cdev23 | 0.320 | | -0.105 | 0.827 | | | | | | |
| cdev18 | 0.360 | | | 0.112 | 0.702 | | | | | |
| conners5 | | 0.228 | | | | 0.640 | | | | |
| conners13 | 0.269 | 0.253 | -0.104 | | | 0.168 | 0.606 | | | |
| cdev52 | | | | -0.456 | -0.118 | | | 0.584 | | |
| cdev43 | 0.456 | | | | | | | 0.285 | 0.583 | |
| rutter10 | 0.109 | 0.372 | | | | 0.386 | 0.133 | | | 0.405 |
| rutter1 | 0.118 | 0.362 | | | | -0.086 | 0.443 | 0.148 | | -0.106 |
| cdev39 | 0.558 | | | | 0.194 | | | 0.334 | 0.250 | |
| cdev41 | | | | | 0.233 | | | | | |
| BCS PLCT | -0.148 | | 0.486 | | | | | | | |
| BCS FMT | -0.383 | | 0.601 | | | | | 0.085 | | |
| BCS BASTM | -0.307 | | 0.516 | | | | | | | |
| BCS BASTRD | -0.220 | | 0.358 | | | | | | | |
| BCS BASTS | -0.198 | | 0.579 | | | | | | | |
| BCS BASTWD | -0.220 | | 0.613 | | | | | | | |
| rutter2 | 0.129 | 0.362 | | | | | 0.411 | 0.118 | | -0.153 |
| rutter3 | 0.103 | 0.432 | | | | 0.351 | 0.091 | | | 0.319 |

continued on next page...

Table 3.10 – continued from previous page

| Items | $f_1$ | $f_2$ | $f_3$ | $f_4$ | $f_5$ | $f_6$ | $f_7$ | $f_8$ | $f_9$ | $f_{10}$ |
|---|---|---|---|---|---|---|---|---|---|---|
| rutter4 | 0.125 | 0.400 | | | | 0.209 | | | | 0.247 |
| rutter5 | 0.079 | 0.289 | | | | 0.190 | | | | 0.100 |
| rutter6 | | 0.357 | | 0.146 | | | | | | -0.195 |
| rutter7 | | 0.263 | | 0.099 | | | | -0.136 | | |
| rutter8 | | 0.696 | | | | -0.120 | | | | |
| rutter9 | | 0.508 | | 0.073 | | 0.128 | | -0.072 | | |
| rutter11 | | 0.234 | | | | 0.311 | | | | |
| rutter12 | | 0.067 | | | | 0.169 | | | | |
| rutter13 | | 0.166 | | | | 0.102 | | | | |
| rutter14 | 0.141 | 0.534 | | | | 0.082 | 0.164 | | | 0.215 |
| rutter15 | 0.216 | 0.275 | -0.117 | | | 0.131 | 0.597 | | | |
| rutter16 | | 0.251 | | 0.132 | | 0.104 | | -0.110 | | -0.167 |
| rutter17 | -0.099 | 0.321 | | | | | | | | -0.107 |
| rutter18 | 0.135 | 0.434 | | | | 0.275 | 0.150 | | | 0.352 |
| rutter19 | | 0.449 | | | | 0.346 | | | | 0.294 |
| conners1 | | 0.238 | | | 0.122 | 0.539 | 0.074 | | | -0.281 |
| conners2 | | 0.257 | | | 0.125 | 0.512 | | | | -0.232 |
| conners3 | 0.346 | 0.255 | -0.114 | -0.071 | | 0.146 | 0.514 | | | |
| conners4 | | 0.365 | | | | 0.199 | 0.291 | | | |
| conners6 | | 0.230 | | | 0.059 | 0.608 | | | | -0.202 |
| conners7 | | 0.473 | | | | 0.206 | | | | -0.196 |
| conners8 | | 0.653 | | | | | 0.129 | | | -0.093 |
| conners9 | 0.098 | 0.533 | | | | | 0.462 | 0.169 | | |
| conners10 | 0.085 | 0.524 | | | | | 0.310 | 0.135 | | -0.112 |
| conners11 | 0.095 | 0.470 | | | | 0.277 | 0.138 | | | 0.182 |
| conners12 | | 0.518 | | | | | | -0.110 | | |
| conners14 | | 0.348 | | | | 0.255 | 0.204 | | | |
| conners15 | | 0.462 | | | | 0.107 | | -0.101 | | |
| conners16 | | 0.683 | | | | | 0.084 | | | |
| conners18 | | 0.168 | | | 0.087 | 0.528 | | | -0.072 | |
| conners19 | 0.235 | 0.206 | -0.101 | | | 0.215 | 0.533 | | | |
| lawseq1 | | | | | | | | | | |
| lawseq2 | -0.104 | | 0.139 | -0.136 | | | | 0.238 | -0.245 | |
| lawseq3 | -0.187 | | 0.249 | | | | | | -0.318 | |
| lawseq4 | -0.067 | | 0.072 | | | | | 0.106 | -0.228 | |
| lawseq5 | | | 0.143 | | | | | 0.100 | -0.070 | |
| lawseq6 | -0.123 | | 0.154 | | | | | 0.167 | -0.142 | |
| lawseq7 | -0.128 | -0.090 | 0.137 | | | | | | -0.161 | |
| lawseq8 | -0.112 | | 0.208 | | | | | 0.129 | -0.068 | |
| lawseq9 | | | 0.182 | | | | | | -0.139 | |
| lawseq10 | -0.209 | | 0.217 | | | | | 0.210 | -0.189 | |
| lawseq11 | -0.160 | | 0.265 | | | | | | | |
| lawseq12 | | | | | | | | | -0.236 | |
| caraloc1 | -0.146 | | 0.330 | | | | | 0.128 | -0.181 | |
| caraloc2 | | | 0.214 | | | | | | | |
| caraloc3 | | | 0.234 | | | | | | 0.125 | |
| caraloc4 | -0.324 | | 0.490 | | | | | 0.142 | -0.146 | |
| caraloc5 | -0.205 | | 0.392 | | | | | 0.168 | | |
| caraloc6 | -0.207 | | 0.346 | | | | | | | |
| caraloc7 | -0.193 | | 0.163 | | | | | | -0.316 | |
| caraloc8 | | | 0.063 | | | | | | | |
| caraloc9 | | | 0.103 | | | | | | -0.072 | |

continued on next page...

Table 3.10 – continued from previous page

| Items | $f_1$ | $f_2$ | $f_3$ | $f_4$ | $f_5$ | $f_6$ | $f_7$ | $f_8$ | $f_9$ | $f_{10}$ |
|---|---|---|---|---|---|---|---|---|---|---|
| caraloc10 | -0.120 | | 0.377 | | | | | | | |
| caraloc11 | -0.134 | | 0.351 | | | | | 0.167 | -0.103 | |
| caraloc12 | | -0.070 | | | | | | | | |
| caraloc13 | | | | | | | | | | |
| caraloc14 | -0.289 | | 0.381 | | | | | 0.213 | | |
| caraloc15 | -0.325 | | 0.509 | | | | | 0.183 | | |
| caraloc16 | -0.088 | | 0.231 | | | | | 0.133 | -0.151 | |
| cdev1 | 0.630 | | | 0.141 | | | | -0.268 | -0.091 | |
| cdev2 | 0.320 | | -0.170 | 0.568 | | | | -0.201 | -0.117 | |
| cdev3 | 0.620 | | | | | | | -0.095 | -0.063 | |
| cdev4 | | | | | 0.197 | | | | | |
| cdev5 | 0.453 | | | 0.161 | | | | 0.237 | 0.452 | |
| cdev6 | 0.394 | | | 0.113 | 0.576 | | | 0.211 | 0.123 | |
| cdev7 | -0.373 | | 0.120 | | -0.346 | | | | 0.131 | |
| cdev8 | 0.438 | 0.049 | | 0.125 | | | | 0.306 | 0.653 | |
| cdev9 | 0.481 | | | | | | | 0.397 | 0.572 | |
| cdev10 | 0.388 | | | 0.198 | 0.573 | | | -0.118 | | |
| cdev11 | 0.263 | | | 0.330 | 0.184 | | | | 0.232 | |
| cdev12 | 0.812 | | | | | | | -0.126 | 0.138 | |
| cdev13 | -0.787 | | | | 0.071 | | | 0.156 | | |
| cdev14 | 0.218 | | | 0.187 | 0.535 | | | -0.130 | | |
| cdev15 | -0.209 | | | | -0.310 | | | | | |
| cdev16 | 0.672 | | | | | | | 0.311 | 0.430 | |
| cdev17 | 0.633 | | -0.262 | 0.326 | | | | -0.110 | -0.140 | |
| cdev19 | 0.363 | | -0.128 | 0.710 | 0.127 | | | -0.068 | -0.106 | |
| cdev20 | 0.215 | | | 0.528 | 0.066 | | | 0.107 | 0.140 | |
| cdev21 | 0.505 | | | 0.280 | | | | 0.294 | 0.488 | |
| cdev22 | 0.475 | | | 0.148 | | | | 0.618 | 0.215 | |
| cdev24 | 0.654 | | | 0.197 | | | | 0.567 | 0.203 | |
| cdev25 | 0.794 | | | 0.083 | | | | 0.345 | 0.119 | |
| cdev27 | -0.228 | | | | -0.509 | | | | 0.116 | |
| cdev28 | 0.439 | | | 0.112 | 0.565 | | | 0.140 | | |
| cdev29 | -0.779 | | | | | | | 0.159 | | |
| cdev30 | 0.331 | | | 0.425 | 0.215 | | | -0.156 | 0.331 | |
| cdev31 | 0.298 | | | 0.427 | 0.163 | | | 0.063 | 0.224 | |
| cdev32 | 0.715 | | -0.193 | 0.181 | 0.073 | | | -0.108 | -0.117 | |
| cdev33 | 0.152 | | | 0.427 | 0.174 | | | -0.333 | 0.206 | |
| cdev34 | 0.553 | | | | | | | 0.237 | 0.559 | |
| cdev35 | -0.366 | | 0.091 | | -0.565 | | | | 0.135 | |
| cdev36 | 0.545 | | | 0.197 | 0.179 | | | -0.385 | 0.099 | |
| cdev37 | 0.501 | | | | 0.190 | | | 0.239 | 0.493 | |
| cdev38 | 0.562 | | | | 0.216 | | | 0.376 | 0.258 | |
| cdev40 | 0.429 | | | | 0.478 | | | | | |
| cdev42 | 0.428 | | | | 0.403 | | | 0.192 | 0.148 | |
| cdev44 | 0.459 | | | 0.198 | | | -0.062 | | 0.548 | |
| cdev45 | 0.318 | | | 0.229 | 0.378 | | | 0.183 | | |
| cdev46 | 0.181 | | | | 0.167 | | | | 0.169 | |
| cdev47 | 0.231 | | | 0.277 | 0.416 | | | -0.281 | | |
| cdev48 | -0.760 | | | | | | | 0.191 | | |
| cdev49 | 0.536 | | | 0.218 | | | | 0.279 | 0.350 | |
| cdev50 | -0.362 | | | | -0.422 | | | | | |
| cdev51 | 0.785 | | | | | | | -0.195 | | |

continued on next page...

Table 3.10 – continued from previous page

| Items | $f_1$ | $f_2$ | $f_3$ | $f_4$ | $f_5$ | $f_6$ | $f_7$ | $f_8$ | $f_9$ | $f_{10}$ |
|---|---|---|---|---|---|---|---|---|---|---|
| cdev53 | 0.244 | | -0.073 | 0.762 | | | | -0.139 | | |

**Notes:** Only factor loadings with a posterior probability greater than 0.75 are displayed. White cell if $|\alpha_{jl}| < 0.4$, light gray cell if $0.4 \leqslant |\alpha_{jl}| < 0.7$ and dark gray cell if $|\alpha_{jl}| \geqslant 0.7$. Items: BCS = Cognitive Test Scores, conners = Conners Hyperactivity Scale, rutter = Rutter Parental 'A' Scale of Behavioral Disorder, lawseq = Self-Esteem Scale, caraloc = Locus of Control Scale, cdev = Child Development Scale.

# Bibliography

Aakvik, A., J. J. Heckman, and E. J. Vytlacil (2005): "Estimating Treatment Effects for Discrete Outcomes When Responses to Treatment Vary: An Application to Norwegian Vocational Rehabilitation Programs," *Journal of Econometrics*, 125, 15–51.

Auld, M. C. and N. Sidhu (2005): "Schooling, Cognitive Ability and Health," *Health Economics*, 14, 1019–1034.

Bartholomew, D. J. (1987): *Latent Variable Models and Factors Analysis*, London: Charles Griffin.

Batty, G. D., I. J. Deary, I. Schoon, and C. R. Gale (2007): "Mental Ability across Childhood in Relation to Risk Factors for Premature Mortality in Adult Life: The 1970 British Cohort Study," *Journal of Epidemiology and Community Health*, 61, 997–1003.

Blanden, J., P. Gregg, and L. Macmillan (2007): "Accounting for Intergenerational Income Persistence: Noncognitive Skills, Ability and Education," *Economic Journal*, 117, C43–C60.

Butler, N., S. Despotidou, and P. Shepherd (1997): "British Cohort Study (BCS70) Ten-year Follow-up: A Guide to the BCS70 10-year Data Available at the Economic and Social Research Unit Data Archive," Tech. rep., London: Social Statistics Research Unit, City University.

Carvalho, C. M. (2006): "Structure and Sparsity in High-Dimensional Multivariate Analysis," Ph.D. thesis, Duke University.

Carvalho, C. M., J. Chang, J. E. Lucas, J. R. Nevins, Q. Wang, and M. West (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.

Chib, S. and B. H. Hamilton (2000): "Bayesian Analysis of Cross-Section and Clustered Data Treatment Models," *Journal of Econometrics*, 97, 25–50.

——— (2002): "Semiparametric Bayes Analysis of Longitudinal Data Treatment Models," *Journal of Econometrics*, 110, 67–89.

Conners, C. K. (1969): "A Teacher Rating Scale for Use in Drug Studies with Children," *American Journal of Psychiatry*, 126, 884–888.

Conti, G., J. J. Heckman, and S. Urzua (2010): "The Education-Health Gradient," *American Economic Review Papers and Proceedings*, 100, 234–238.

Conti, G. and S. Pudney (2009): "The Dynamics of Cognitive Development," Unpublished Manuscript, University of Essex, Institute for Social and Economic Research.

Cox, D. R. (1958): *Planning of Experiments*, New York: Wiley.

Cunha, F. and J. J. Heckman (2008): "Formulating, Identifying and Estimating the Technology of Cognitive and Noncognitive Skill Formation," *Journal of Human Resources*, 43, 738–782.

Cunha, F., J. J. Heckman, and S. M. Schennach (2010): "Estimating the Technology of Cognitive and Noncognitive Skill Formation," *Econometrica*, 78, 883–931.

Currie, J. and E. Moretti (2003): "Mother's Education and the Intergenerational Transmission of Human Capital: Evidence from College Openings," *Quarterly Journal of Economics*, 118, 1495–1532.

Cutler, D. M. and A. Lleras-Muney (2010): "Understanding Differences in Health Behaviors by Education," *Journal of Health Economics*, 29, 1–28.

Digman, J. M. (1990): "Personality Structure: Emergence of the Five-Factor Model," *Annual Review of Psychology*, 41, 417–440.

Feinstein, L. (2000): "The Relative Economic Importance of Academic, Psychological and Behavioural Attributes Developed on Chilhood," Discussion Paper, Centre for Economic Performance, London.

Fisher, R. A. (1935): *The Design of Experiments*, London: Oliver and Boyd.

Frühwirth-Schnatter, S. and H. F. Lopes (2009): "Parsimonious Bayesian Factor Analysis when the Number of Factors is Unknown," Unpublished Technical Report.

Frühwirth-Schnatter, S. and R. Tüchler (2008): "Bayesian Parsimonious Covariance Estimation for Hierarchical Linear Mixed Models," *Statistics and Computing*, 18, 1–13.

Gale, C. R., G. D. Batty, and I. J. Deary (2008): "Locus of Control at Age 10 Years and Health Outcomes and Behaviors at Age 30 Years: the 1970 British Cohort Study," *Psychosomatic Medicine*, 70, 397–403.

Gale, C. R., S. L. Hatch, G. D. Batty, and I. J. Deary (2009): "Intelligence in Childhood and Risk of Psychological Distress in Adulthood: The 1958 National Child Development Survey and the 1970 British Cohort Study," *Intelligence*, 37, 592–599.

Gammage, P. (1975): "Socialization, Schooling and Locus of Control," Ph.D. thesis, University of Bristol.

Garen, J. (1984): "The Returns to Schooling: A Selectivity Bias Approach with a Continuous Choice Variable," *Econometrica*, 52, 1199–1218.

Geweke, J. and G. Zhou (1996): "Measuring the Price of the Arbitrage Pricing Theory," *The Review of Financial Studies*, 9, 557–587.

Geweke, J. F. and K. J. Singleton (1980): "Interpreting the Likelihood Ratio Statistic in Factor Models when Sample Size is Small," *Journal of the American Statistical Association*, 75, 133–137.

Ghosh, J. and D. B. Dunson (2009): "Default Prior Distributions and Efficient Posterior Computation in Bayesian Factor Analysis," *Journal of Computational and Graphical Statistics*, 18, 306–320.

Goldberg, L. R. (1990): "An Alternative 'Description of Personality': The Big-Five Factor Structure," *Journal of Personality and Social Psychology*, 59, 1216–1229.

Gottfredson, L. S. and I. J. Deary (2004): "Intelligence Predicts Health and Longevity, but Why?" *Current Directions in Psychological Science*, 13, 1–4.

Green, P. J. (1995): "Reversible Jump Markov Chain Monte Carlo Computation and Bayesian Model Determination," *Biometrika*, 82, 711–732.

Griliches, Z. (1977): "Estimating the Returns to Schooling: Some Econometric Problems," *Econometrica*, 45, 1–22.

Grossman, M. (1975): "The Correlation Between Health and Schooling," in *Household Production and Consumption*, ed. by N. E. Terleckyj, New York: Columbia University Press, 147–211.

——— (2006): "Education and Nonmarket Outcomes," in *Handbook of the Economics of Education*, ed. by E. Hanushek and F. Welch, Amsterdam: Elsevier, vol. 1, chap. 10, 577–633.

Hakstian, A. R., W. T. Rogers, and R. B. Cattell (1982): "The Behavior of Number-of-Factors Rules with Simulated Data," *Multivariate Behavioral Research*, 17, 193–219.

Hampson, S. E. and H. S. Friedman (2008): "Personality and Health: A Lifespan Perspective," in *The Handbook of Personality: Theory and Research*, ed. by O. P. John, R. Robins, and L. Pervin, New York: Guilford, 770–794, third ed.

Hartog, J. and H. Oosterbeek (1998): "Health, Wealth and Happiness: Why Pursue a Higher Education?" *Economics of Education Review*, 17, 245–256.

Heckman, J. J. (1997): "Instrumental Variables: A Study of Implicit Behavioral Assumptions Used in Making Program Evaluations," *Journal of Human Resources*, 32, 441–462, addendum published vol. 33 no. 1 (Winter 1998).

Heckman, J. J., R. J. LaLonde, and J. A. Smith (1999): "The Economics and Econometrics of Active Labor Market Programs," in *Handbook of Labor Economics*, ed. by O. Ashenfelter and D. Card, New York: North-Holland, vol. 3A, chap. 31, 1865–2097.

Heckman, J. J. and J. A. Smith (1998): "Evaluating the Welfare State," in *Econometrics and Economic Theory in the Twentieth Century: The Ragnar Frisch Centennial Symposium*, ed. by S. Strom, New York: Cambridge University Press, 241–318.

Heckman, J. J., J. Stixrud, and S. Urzua (2006): "The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior," *Journal of Labor Economics*, 24, 411–482.

Heckman, J. J. and E. J. Vytlacil (1999): "Local Instrumental Variables and Latent Variable Models for Identifying and Bounding Treatment Effects," *Proceedings of the National Academy of Sciences*, 96, 4730–4734.

——— (2000): "The Relationship Between Treatment Parameters Within a Latent Variable Framework," *Economics Letters*, 66, 33–39.

——— (2001): "Local Instrumental Variables," in *Nonlinear Statistical Modeling: Proceedings of the Thirteenth International Symposium in Economic Theory and Econometrics: Essays in Honor of Takeshi Amemiya*, ed. by C. Hsiao, K. Morimune, and J. L. Powell, New York: Cambridge University Press, 1–46.

——— (2007): "Econometric Evaluation of Social Programs, Part II: Using the Marginal Treatment Effect to Organize Alternative Economic Estimators to Evaluate Social Programs and to Forecast Their Effects in New Environments," in *Handbook of Econometrics*, ed. by J. Heckman and E. Leamer, Amsterdam: Elsevier, vol. 6B, 4875–5144.

Kaestner, R. (2009): "Adolescent Cognitive and Non-cognitive Correlates of Adult Health," National Bureau of Economic Research (NBER) Working Paper No. 14924.

Kaiser, H. F. (1960): "The Application of Electronic Computers to Factor Analysis," *Educational and Psychological Measurement*, 20, 141–151.

Lawrence, D. (1973): *Improved Reading through Counselling*, London: Ward Lock Educational.

——— (1978): *Counselling Students with Reading Difficulties: A Handbook for Tutors and Organisors*, Good Reading Limited.

Lee, L. F. (1992): "On Efficiency of Methods of Simulated Moments and Maximum Simulated Likelihood Estimation of Discrete Response Models," *Econometric Theory*, 8, 518–552.

Li, M., D. J. Poirier, and J. L. Tobias (2004): "Do Dropouts Suffer from Dropping out? Estimation and Prediction of Outcome Gains in Generalized Selection Models," *Journal of Applied Econometrics*, 19, 203–225.

Li, M. and J. L. Tobias (2007): "Bayesian Analysis of Treatment Effects in an Ordered Potential Outcomes Model," *Advances in Econometrics*, 21, 57–91.

Lindeboom, M., A. Llena-Nozal, and B. van der Klaauw (2009): "Parental Education and Child Health: Evidence from a Schooling Reform," *Journal of Health Economics*, 28, 109–131.

Linn, R. L. (1968): "A Monte Carlo Approach to the Number of Factors Problem," *Psychometrika*, 33, 37–71.

Lopes, H. F. and C. M. Carvalho (2007): "Factor Stochastic Volatility with Time Varying Loadings and Markov Switching Regimes," *Journal of Statistical Planning and Inference*, 137, 3082–3091.

Lopes, H. F., E. Salazar, and D. Gamerman (2008): "Spatial Dynamic Factor Models," *Bayesian Analysis*, 3, 759–992.

Lopes, H. F. and M. West (2004): "Bayesian Model Assessment in Factor Analysis," *Statistica Sinica*, 14, 41–67.

McCrae, R. R. and O. P. John (1992): "An Introduction to the Five-Factor Model and its Applications." *Journal of personality*, 60, 175–215.

McCrary, J. and H. Royer (2010): "The Effect of Female Education on Fertility and Infant Health: Evidence from School Entry Policies Using Exact Date of Birth," *American Economic Review*, forthcoming.

Murasko, J. E. (2007): "A Lifecourse Study on Education and Health: The Relationship between Childhood Psychosocial Resources and Outcomes in Adolescence and Young Adulthood," *Social Science Research*, 36, 1348–1370.

Neyman, J. (1923): "Statistical Problems in Agricultural Experiments," *Journal of the Royal Statistical Society*, 2 (Supplement), 107–180.

O'Hagan, A. (1995): "Fractional Bayes Factors for Model Comparison," *Journal of the Royal Statistical Society. Series B (Methodological)*, 57, 99–138.

Richardson, S. and P. J. Green (1997): "On Bayesian Analysis of Mixtures with an Unknown Number of Components," *Journal of the Royal Statistical Society. Series B (Methodological)*, 59, 731–792.

Roberts, B. W., P. Harms, J. L. Smith, D. Wood, and M. Webb (2006): "Using Multiple Methods in Personality Psychology," in *Handbook of Multimethod Measurement in Psychology*, ed. by M. Eid and E. Diener, Washington, D.C.: American Psychological Association, 321–335.

Roberts, B. W., N. R. Kuncel, R. L. Shiner, A. Caspi, and L. R. Goldberg (2007): "The Power of Personality: The Comparative Validity of Personality Traits, Socioeconomic Status, and Cognitive Ability for Predicting Important Life Outcomes," *Perspectives in Psychological Science*, 2, 313–345.

ROSENBERG, M. (1965): *Society and the Adolescent Self-Image*, Princeton, NJ: Princeton University Press.

ROTTER, J. B. (1966): "Generalized Expectancies for Internal versus External Control of Reinforcement," *Psychological Monographs: General & Applied*, 80, 1–28.

RUBIN, D. B. (1974): "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies," *Journal of Educational Psychology*, 66, 688–701.

RUTTER, M. (1967): "A Children's Behaviour Questionnaire for Completion by Teachers: Preliminary Findings," *Journal of Child Psychology and Psychiatry*, 8, 1–11.

RUTTER, M., J. TIZARD, AND K. WHITMORE (1970): *Education, health and behaviour*, London, UK: Longmans.

SHANNON, C. E. (1948): *The Mathematical Theory of Communication*, vol. 27, Bell System Tech. J.

SMITH, M. AND R. KOHN (2002): "Parsimonious Covariance Matrix Estimation for Longitudinal Data," *Journal of the American Statistical Association*, 97, 1141–1153.

STUMM, S. V., C. R. GALE, G. D. BATTYA, AND I. J. DEARY (2009): "Childhood Intelligence, Locus of Control and Behaviour Disturbance as Determinants of Intergenerational Social Mobility: British Cohort Study 1970," *Intelligence*, 37, 329–340.

TÜCHLER, R. (2008): "Bayesian Variable Selection for Logistic Models Using Auxiliary Mixture Sampling," *Journal of Computational and Graphical Statistics*, 17, 76–94.

TUCKER, L. R., R. F. KOOPMAN, AND R. L. LINN (1969): "Evaluation of Factor Analytic Research Procedures by Means of Simulated Correlation Matrices," *Psychometrika*, 34, 421–459.

VELICER, W. F. (1976): "Determining the Number of Components from the Matrix of Partial Correlations," *Psychometrika*, 41, 321–327.

WEST, M. (2003): "Bayesian Factor Regression Models in the 'Large p, Small n' Paradigm," *Bayesian Statistics*, 7, 723–732.

Whalley, L. J. and I. J. Deary (2001): "Longitudinal Cohort Study of Childhood IQ and Survival up to Age 76," *British Medical Journal*, 322, 819–822.

Wooldridge, J. M. (2003): "Further Results on Instrumental Variables Estimation of Average Treatment Effects in the Correlated Random Coefficient Model," *Economics Letters*, 79, 185–191.

## Appendix 3.A Details of the Gibbs sweep

We outline the steps of the MCMC algorithm for one sweep of the Gibbs sampler. Full details can be found in Chapter 1, as well as in Frühwirth-Schnatter and Lopes (2009).

***Parameter expanded model.*** To enhance the performance of the MCMC algorithm, Ghosh and Dunson (2009) suggest to expand the original model by relaxing the restrictions on the variance of the latent factors in Equation (3.7). More precisely, the expanded model is specified as $\boldsymbol{y}_i^\star = \boldsymbol{\beta}\boldsymbol{x}_i + \tilde{\boldsymbol{\alpha}}\tilde{\boldsymbol{f}}_i + \boldsymbol{\varepsilon}_i$ and $\tilde{\boldsymbol{f}}_i \sim \mathcal{N}_k(\mathbf{0};\, \boldsymbol{\Psi})$, where $\boldsymbol{\Psi} = \text{diag}\,(\Psi_1, \dots, \Psi_k)$. The parameters of the original model can be recovered through the following transformation: $\boldsymbol{\alpha} = \tilde{\boldsymbol{\alpha}}\boldsymbol{\Psi}^{1/2}$ and $\boldsymbol{f}_i = \boldsymbol{\Psi}^{-1/2}\tilde{\boldsymbol{f}}_i$. The factor loadings of the expanded model $\tilde{\boldsymbol{\alpha}}$ are sampled with the same prior as specified in Section 3.5.3, and an additional step is inserted in the MCMC scheme to draw the variances of the latent factors. To do so, each $\Psi_l$, for $l = 1, \dots, k$, is equipped with a conjugate inverse Gamma prior, i.e., $\Psi_l \sim \mathcal{G}^{-1}(g_l;\, h_l)$.

***Sampling the slope parameters.*** Given the conjugate normal prior assumed for the slope parameters in Equation (3.16) and standard Bayesian linear regression updates, the conditional distributional of each row $j$ of the matrix of slope parameters $\boldsymbol{\beta}$, i.e. $\boldsymbol{\beta}_j$, is $\mathcal{N}_v(\boldsymbol{b}_{jn};\, \mathbf{B}_{jn})$, where $\mathbf{B}_{jn} = \left(\sigma_j^{-2}\mathbf{X}'\mathbf{X} + \mathbf{B}_{j0}^{-1}\right)^{-1}$ and $\boldsymbol{b}_{jn} = \mathbf{B}_{jn}\left(\sigma_j^{-2}\mathbf{X}'(\boldsymbol{y}_j^\star - \tilde{\mathbf{F}}\tilde{\boldsymbol{\alpha}}_j) + \mathbf{B}_{j0}^{-1}\boldsymbol{b}_{j0}\right)$. Here $\mathbf{X} = (\boldsymbol{x}_1, \dots, \boldsymbol{x}_n)'$ contains the covariates for all individuals.

***Sampling the latent factors and their variances.*** The conditional distribution of $\tilde{\boldsymbol{f}}_1, \dots, \tilde{\boldsymbol{f}}_n$ can be factorized into the product of $n$ independent $k$-variate normal distributions:

$$\tilde{\boldsymbol{f}}_i \sim \mathcal{N}_k\big((\tilde{\boldsymbol{\alpha}}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\alpha} + \boldsymbol{\Psi}^{-1})^{-1}\tilde{\boldsymbol{\alpha}}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{y}_i^\star - \boldsymbol{\beta}\boldsymbol{x}_i);\, (\tilde{\boldsymbol{\alpha}}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\alpha} + \boldsymbol{\Psi}^{-1})^{-1}\big)\,.$$

It is straightforward to show that the conditional posterior for $\Psi_l$ is:

$$\Psi_l \sim \mathcal{G}^{-1}\left(g_l + \frac{n}{2};\; h_l + \frac{1}{2}\sum_{i=1}^{n} \tilde{\boldsymbol{f}}_{il}\right).$$

***Sampling the factor loadings and the idiosyncratic variances.*** Similarly, the joint sampling of the idiosyncratic variance and of the factor loadings is performed in two steps following standard Bayesian multivariate regression updates: $\sigma_j^2 \sim \mathcal{G}^{-1}(c_n;\, C_{jn})$ and $\tilde{\boldsymbol{\alpha}}_j|\sigma_j^2 \sim \mathcal{N}_v\big(\boldsymbol{a}_{jn};\, \mathbf{A}_{jn}\sigma_j^2\big)$, where the parameters of these two conditional distributions depend on the type of prior placed on the factor loadings, and on the type of the response variable $\boldsymbol{y}_j$. If response variable $j$ is dichotomous, the idiosyncratic variance $\sigma_j^2$ is fixed to 1 for identification purpose, and only the factor loadings $\tilde{\boldsymbol{\alpha}}_j$ have to be sampled from the normal conditional distribution with parameters $\mathbf{A}_{jn} = (\tilde{\mathbf{F}}_j^{\boldsymbol{\delta}\prime}\tilde{\mathbf{F}}_j^{\boldsymbol{\delta}} + (\mathbf{A}_{j0}^{\boldsymbol{\delta}})^{-1})^{-1}$ and $\boldsymbol{a}_{jn} = \mathbf{A}_{jn}(\tilde{\mathbf{F}}_j^{\boldsymbol{\delta}\prime}(\boldsymbol{y}_j^\star - \mathbf{X}\boldsymbol{\beta}_j) + (\mathbf{A}_{j0}^{\boldsymbol{\delta}})^{-1}\boldsymbol{a}_{j0}^{\boldsymbol{\delta}})$. In the continuous case with normal prior on the factor loadings, the shape and scale parameters of the conditional distribution of the idiosyncratic variances are $c_n = c_0 + n/2$ and $C_{jn} = C_{j0} + 0.5\,\big((\boldsymbol{y}_j^\star - \mathbf{X}\boldsymbol{\beta}_j)'(\boldsymbol{y}_j^\star - \mathbf{X}\boldsymbol{\beta}_j) + (\boldsymbol{a}_{j0}^{\boldsymbol{\delta}})'(\mathbf{A}_{j0}^{\boldsymbol{\delta}})^{-1}(\boldsymbol{a}_{j0}^{\boldsymbol{\delta}}) - \boldsymbol{a}_{jn}'\mathbf{A}_{jn}^{-1}\boldsymbol{a}_{jn}\big)$, with $\tilde{\boldsymbol{\alpha}}_j$ then sampled as in the dichotomous case. If a fractional prior is assumed, the parameters of the conditional distribution of the idiosyncratic variance and of the factor loadings are $c_n = c_0 + (1-b)n/2$, $C_{jn} = C_{j0} + 0.5(1-b)n(\boldsymbol{y}_j^\star - \mathbf{X}\boldsymbol{\beta}_j - \tilde{\mathbf{F}}_j^{\boldsymbol{\delta}}\tilde{\boldsymbol{\alpha}}_j^{\boldsymbol{\delta}})'(\boldsymbol{y}_j^\star - \mathbf{X}\boldsymbol{\beta}_j - \tilde{\mathbf{F}}_j^{\boldsymbol{\delta}}\tilde{\boldsymbol{\alpha}}_j^{\boldsymbol{\delta}})$, $\mathbf{A}_{jn} = (\tilde{\mathbf{F}}_j^{\boldsymbol{\delta}\prime}\tilde{\mathbf{F}}_j^{\boldsymbol{\delta}})^{-1}$ and $\boldsymbol{a}_{jn} = \mathbf{A}_{jn}(\tilde{\mathbf{F}}_j^{\boldsymbol{\delta}\prime}(\boldsymbol{y}_j^\star - \mathbf{X}\boldsymbol{\beta}_j))$.

***Sampling the latent response variables.*** If $y_{ji}$ is dichotomous, $j$ in $1,\ldots,\rho$, the latent latent response variable $y_{ji}^\star$ is sampled from a truncated normal distribution, sequentially for each individual $i$ in $1,\ldots,n$, i.e., $y_{ji}^\star \sim \mathcal{TN}_{(0;+\infty)}(\boldsymbol{\beta}_j'\boldsymbol{x}_i + \alpha_j{}'\tilde{\boldsymbol{f}}_i, 1)$ if $y_{ji} > 0$, and $y_{ji}^\star \sim \mathcal{TN}_{(-\infty;0]}(\boldsymbol{\beta}_j'\boldsymbol{x}_i + \alpha_j{}'\tilde{\boldsymbol{f}}_i, 1)$ if $y_{ji} \leqslant 0$.

# Appendix 3.B Additional material

## 3.B.1 Classical Methods

### Principal Component Analysis

**Table 3.B.1:** Component loadings from Principal Component Analysis (Females)

| Items | $pc_1$ | $pc_2$ | $pc_3$ | $pc_4$ | $pc_5$ | $pc_6$ | $pc_7$ | $pc_8$ | $pc_9$ | $pc_{10}$ | $pc_{11}$ | $pc_{12}$ | $pc_{13}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| cdev26 | 0.030 | 0.292 | -0.020 | 0.066 | 0.022 | 0.010 | 0.002 | -0.022 | -0.005 | -0.037 | -0.003 | 0.107 | -0.030 |
| conners17 | 0.035 | -0.004 | 0.020 | -0.037 | 0.155 | 0.175 | 0.023 | 0.093 | -0.039 | -0.090 | 0.085 | -0.090 | -0.036 |
| BCS SERT | -0.001 | -0.049 | -0.003 | 0.014 | 0.011 | -0.010 | 0.373 | -0.001 | -0.012 | -0.003 | 0.026 | 0.005 | 0.004 |
| cdev23 | -0.010 | -0.013 | -0.027 | 0.402 | 0.007 | 0.003 | 0.007 | 0.036 | 0.014 | -0.024 | 0.009 | -0.046 | -0.027 |
| cdev18 | -0.008 | -0.040 | 0.346 | -0.013 | 0.011 | -0.002 | 0.010 | -0.001 | -0.004 | 0.006 | 0.014 | 0.014 | 0.065 |
| conners5 | -0.031 | -0.010 | -0.012 | 0.038 | 0.232 | -0.178 | 0.004 | 0.048 | -0.034 | 0.282 | 0.044 | 0.064 | 0.043 |
| conners13 | -0.029 | 0.092 | -0.039 | 0.004 | -0.076 | 0.187 | 0.019 | 0.062 | -0.013 | 0.210 | -0.045 | 0.041 | 0.018 |
| cdev52 | 0.041 | -0.005 | 0.011 | -0.097 | -0.002 | 0.019 | 0.018 | 0.029 | 0.006 | -0.030 | 0.014 | 0.448 | -0.037 |
| cdev43 | 0.321 | -0.024 | -0.018 | -0.086 | 0.016 | -0.017 | -0.028 | -0.034 | -0.016 | 0.022 | 0.001 | -0.017 | 0.051 |
| rutter10 | -0.001 | 0.044 | -0.039 | -0.015 | 0.343 | -0.096 | -0.005 | -0.028 | 0.032 | 0.077 | -0.013 | -0.030 | 0.005 |
| rutter1 | -0.016 | -0.019 | -0.018 | 0.023 | -0.052 | 0.389 | -0.043 | -0.107 | 0.019 | 0.004 | 0.016 | 0.037 | 0.034 |
| cdev39 | 0.083 | 0.033 | 0.068 | 0.017 | 0.002 | 0.004 | -0.001 | -0.003 | 0.012 | -0.006 | -0.014 | 0.119 | 0.223 |
| cdev41 | -0.016 | -0.017 | -0.023 | -0.002 | -0.005 | 0.012 | 0.001 | 0.018 | -0.009 | -0.003 | 0.010 | -0.021 | 0.518 |
| BCS PLCT | -0.012 | 0.068 | -0.023 | -0.002 | -0.024 | -0.019 | 0.344 | 0.012 | 0.025 | 0.035 | -0.015 | -0.003 | -0.012 |
| BCS FMT | 0.002 | -0.050 | 0.004 | 0.004 | 0.006 | -0.003 | 0.355 | -0.004 | -0.010 | -0.010 | 0.035 | 0.019 | 0.023 |
| BCS BASTM | -0.006 | -0.019 | -0.019 | 0.017 | -0.008 | 0.010 | 0.344 | -0.007 | -0.067 | 0.003 | 0.021 | -0.020 | 0.023 |
| BCS BASTRD | 0.010 | 0.025 | -0.029 | -0.007 | 0.009 | -0.001 | 0.257 | -0.023 | 0.007 | 0.003 | 0.010 | -0.019 | 0.060 |
| BCS BASTS | -0.010 | 0.029 | 0.018 | -0.007 | -0.018 | -0.000 | 0.376 | 0.002 | 0.041 | 0.034 | -0.023 | 0.017 | -0.018 |
| BCS BASTWD | -0.011 | 0.025 | 0.021 | -0.025 | -0.025 | -0.003 | 0.368 | 0.009 | 0.049 | 0.023 | -0.019 | 0.006 | -0.030 |
| rutter2 | -0.025 | -0.016 | 0.006 | 0.016 | -0.053 | 0.349 | 0.016 | -0.065 | 0.007 | 0.095 | -0.013 | 0.004 | 0.022 |
| rutter3 | -0.012 | 0.019 | -0.039 | 0.023 | 0.326 | -0.068 | -0.014 | -0.047 | 0.033 | 0.110 | -0.013 | 0.005 | 0.049 |
| rutter4 | 0.014 | -0.016 | 0.036 | -0.032 | 0.308 | 0.042 | -0.026 | -0.030 | 0.002 | -0.089 | -0.002 | 0.004 | 0.003 |
| rutter5 | 0.051 | -0.003 | 0.032 | -0.059 | 0.143 | -0.039 | -0.010 | 0.090 | 0.064 | 0.064 | -0.067 | -0.068 | -0.047 |
| rutter6 | -0.028 | 0.024 | -0.015 | 0.042 | -0.082 | -0.050 | 0.006 | 0.442 | 0.021 | 0.015 | -0.046 | 0.049 | 0.027 |
| rutter7 | 0.033 | -0.002 | 0.013 | -0.044 | -0.027 | -0.050 | 0.003 | 0.263 | 0.004 | 0.007 | -0.028 | -0.107 | 0.041 |
| rutter8 | 0.031 | -0.007 | 0.042 | -0.048 | 0.107 | 0.171 | 0.030 | 0.175 | -0.048 | -0.178 | 0.060 | -0.071 | -0.062 |
| rutter9 | 0.040 | -0.009 | 0.012 | -0.034 | 0.085 | 0.054 | 0.018 | 0.252 | -0.017 | -0.048 | -0.015 | -0.090 | -0.027 |
| rutter11 | -0.055 | -0.012 | -0.014 | 0.071 | 0.209 | -0.114 | -0.023 | 0.040 | 0.032 | 0.126 | -0.012 | 0.097 | 0.030 |
| rutter12 | -0.049 | 0.036 | 0.017 | 0.030 | 0.079 | -0.036 | 0.067 | 0.048 | -0.017 | 0.005 | -0.022 | 0.082 | -0.0171 |
| rutter13 | -0.033 | 0.003 | -0.033 | 0.054 | 0.060 | 0.029 | -0.041 | 0.101 | 0.009 | -0.011 | -0.053 | 0.089 | 0.001 |
| rutter14 | 0.022 | 0.020 | 0.002 | -0.027 | 0.225 | 0.181 | 0.033 | -0.018 | 0.004 | -0.121 | -0.005 | -0.031 | -0.020 |
| rutter15 | -0.008 | 0.044 | -0.038 | 0.005 | -0.036 | 0.218 | -0.031 | 0.030 | 0.021 | 0.151 | -0.044 | 0.023 | 0.022 |
| rutter16 | -0.049 | 0.006 | -0.014 | 0.074 | -0.061 | -0.050 | 0.006 | 0.358 | -0.019 | 0.069 | -0.008 | 0.031 | -0.000 |
| rutter17 | -0.013 | -0.041 | 0.021 | 0.023 | 0.020 | -0.050 | -0.040 | 0.338 | 0.007 | -0.084 | 0.004 | 0.109 | 0.045 |
| rutter18 | 0.015 | 0.034 | -0.022 | -0.018 | 0.287 | 0.053 | -0.017 | -0.029 | 0.017 | -0.028 | -0.011 | -0.049 | -0.002 |
| rutter19 | 0.006 | -0.021 | 0.023 | -0.010 | 0.351 | -0.010 | -0.020 | -0.027 | -0.001 | -0.017 | 0.014 | 0.006 | -0.042 |
| conners1 | 0.020 | -0.018 | 0.030 | -0.022 | -0.030 | 0.085 | 0.028 | -0.020 | -0.016 | 0.402 | -0.019 | -0.048 | -0.046 |
| conners2 | 0.026 | -0.025 | 0.028 | -0.029 | -0.028 | 0.089 | 0.035 | -0.030 | -0.048 | 0.393 | 0.002 | -0.064 | -0.030 |
| conners3 | -0.026 | 0.130 | -0.022 | -0.019 | -0.043 | 0.205 | -0.012 | 0.051 | -0.004 | 0.124 | -0.045 | 0.049 | 0.012 |
| conners4 | -0.029 | -0.033 | 0.011 | 0.011 | 0.087 | 0.184 | -0.009 | -0.028 | -0.007 | 0.061 | 0.016 | 0.037 | 0.040 |
| conners6 | 0.012 | -0.023 | 0.005 | 0.009 | 0.069 | -0.018 | 0.013 | -0.016 | -0.003 | 0.413 | 0.006 | -0.021 | 0.003 |
| conners7 | 0.017 | -0.012 | 0.008 | -0.018 | 0.071 | -0.002 | 0.005 | 0.277 | 0.029 | 0.012 | 0.007 | 0.007 | 0.041 |
| conners8 | 0.006 | 0.002 | 0.015 | -0.016 | 0.066 | 0.185 | 0.028 | 0.192 | -0.010 | -0.076 | 0.037 | -0.026 | -0.035 |
| conners9 | -0.018 | -0.026 | -0.012 | 0.024 | 0.007 | 0.320 | -0.007 | -0.019 | 0.037 | 0.069 | -0.001 | 0.028 | 0.015 |
| conners10 | -0.004 | -0.009 | 0.023 | -0.016 | 0.013 | 0.274 | 0.013 | 0.054 | 0.020 | -0.002 | -0.016 | 0.030 | 0.001 |
| conners11 | 0.018 | -0.008 | 0.001 | -0.016 | 0.240 | 0.103 | 0.002 | -0.032 | 0.010 | 0.005 | -0.010 | 0.024 | -0.035 |
| conners12 | 0.061 | -0.004 | -0.007 | -0.063 | 0.088 | 0.083 | 0.004 | 0.191 | -0.055 | -0.041 | 0.015 | -0.104 | -0.043 |
| conners14 | -0.047 | -0.026 | -0.019 | 0.065 | 0.093 | 0.143 | -0.012 | -0.013 | -0.004 | 0.082 | 0.025 | 0.078 | 0.051 |
| conners15 | 0.044 | -0.021 | 0.002 | -0.026 | 0.055 | 0.076 | 0.010 | 0.238 | -0.078 | -0.030 | 0.057 | -0.067 | -0.017 |
| conners16 | 0.042 | -0.021 | 0.012 | -0.037 | 0.083 | 0.207 | 0.008 | 0.159 | -0.052 | -0.073 | 0.058 | -0.074 | -0.035 |
| conners18 | -0.023 | -0.003 | 0.017 | 0.040 | 0.188 | -0.173 | -0.027 | 0.086 | -0.003 | 0.247 | 0.018 | 0.096 | -0.020 |
| conners19 | -0.017 | 0.067 | -0.023 | 0.015 | -0.029 | 0.101 | -0.035 | 0.095 | 0.018 | 0.221 | -0.014 | 0.051 | 0.026 |
| lawseq1 | -0.001 | 0.013 | -0.025 | -0.004 | -0.081 | -0.029 | -0.044 | 0.064 | 0.266 | 0.065 | -0.073 | -0.010 | 0.010 |
| lawseq2 | -0.019 | -0.021 | -0.010 | 0.035 | 0.021 | 0.003 | -0.027 | -0.014 | -0.054 | -0.043 | 0.352 | 0.077 | -0.005 |
| lawseq3 | -0.013 | 0.003 | -0.001 | 0.025 | -0.005 | -0.001 | 0.049 | -0.032 | -0.019 | 0.021 | 0.344 | -0.008 | 0.020 |
| lawseq4 | -0.022 | 0.014 | -0.031 | 0.017 | -0.022 | 0.044 | -0.073 | -0.049 | 0.080 | 0.034 | 0.279 | -0.022 | 0.040 |
| lawseq5 | 0.077 | 0.005 | 0.011 | -0.091 | -0.085 | -0.015 | -0.050 | 0.060 | 0.121 | 0.086 | 0.142 | -0.013 | -0.062 |
| lawseq6 | 0.006 | 0.024 | 0.002 | -0.016 | -0.008 | 0.034 | 0.046 | -0.022 | -0.033 | -0.027 | 0.379 | -0.002 | 0.001 |
| lawseq7 | -0.007 | -0.008 | -0.015 | 0.020 | -0.003 | -0.057 | -0.047 | 0.066 | 0.105 | 0.019 | 0.208 | 0.038 | 0.036 |

continued on next page...

Table 3.B.1 – continued from previous page

| Items | $pc_1$ | $pc_2$ | $pc_3$ | $pc_4$ | $pc_5$ | $pc_6$ | $pc_7$ | $pc_8$ | $pc_9$ | $pc_{10}$ | $pc_{11}$ | $pc_{12}$ | $pc_{13}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| lawseq8 | 0.083 | 0.040 | -0.030 | -0.113 | -0.094 | -0.012 | -0.015 | 0.067 | 0.107 | 0.103 | 0.203 | -0.074 | -0.005 |
| lawseq9 | 0.083 | 0.003 | -0.031 | -0.074 | -0.045 | -0.027 | -0.088 | 0.029 | 0.166 | 0.081 | 0.180 | -0.095 | -0.040 |
| lawseq10 | -0.029 | -0.002 | 0.028 | 0.009 | 0.033 | -0.037 | 0.032 | -0.016 | -0.054 | -0.008 | 0.369 | 0.049 | 0.022 |
| lawseq11 | 0.076 | -0.011 | -0.018 | -0.064 | -0.088 | -0.025 | -0.030 | 0.061 | 0.087 | 0.112 | 0.094 | -0.109 | -0.049 |
| lawseq12 | -0.006 | 0.018 | 0.009 | 0.014 | -0.018 | 0.006 | -0.054 | 0.006 | 0.184 | 0.007 | 0.192 | -0.033 | -0.039 |
| caraloc1 | -0.022 | 0.020 | 0.012 | 0.012 | -0.012 | 0.000 | 0.060 | 0.040 | 0.263 | -0.011 | 0.051 | 0.027 | -0.016 |
| caraloc2 | 0.018 | 0.011 | 0.019 | -0.017 | 0.099 | 0.036 | 0.048 | -0.116 | 0.030 | -0.005 | 0.062 | -0.062 | -0.083 |
| caraloc3 | 0.015 | 0.020 | -0.024 | -0.011 | 0.029 | 0.019 | 0.059 | -0.033 | 0.144 | -0.008 | -0.183 | -0.077 | 0.023 |
| caraloc4 | -0.019 | -0.014 | 0.012 | 0.008 | -0.003 | -0.002 | 0.067 | 0.031 | 0.225 | -0.012 | 0.047 | 0.033 | 0.013 |
| caraloc5 | 0.006 | -0.001 | 0.025 | -0.025 | 0.072 | -0.003 | 0.078 | -0.036 | 0.286 | -0.034 | -0.036 | -0.000 | -0.005 |
| caraloc6 | -0.025 | -0.024 | 0.018 | 0.004 | 0.018 | -0.016 | 0.017 | 0.016 | 0.259 | -0.012 | -0.014 | 0.041 | 0.002 |
| caraloc7 | -0.009 | -0.025 | -0.014 | 0.026 | -0.024 | -0.017 | -0.055 | 0.005 | 0.128 | 0.010 | 0.226 | -0.009 | 0.037 |
| caraloc8 | -0.016 | 0.006 | -0.002 | -0.005 | -0.046 | -0.014 | 0.017 | 0.037 | 0.170 | 0.015 | -0.109 | 0.002 | 0.080 |
| caraloc9 | -0.025 | -0.014 | -0.005 | 0.042 | 0.017 | 0.055 | -0.066 | -0.021 | 0.286 | -0.050 | -0.000 | 0.009 | 0.006 |
| caraloc10 | 0.046 | 0.041 | -0.043 | -0.032 | 0.016 | -0.033 | 0.101 | 0.021 | 0.120 | -0.015 | 0.032 | -0.017 | -0.027 |
| caraloc11 | -0.000 | 0.029 | 0.022 | -0.004 | 0.029 | 0.039 | 0.093 | -0.040 | 0.294 | -0.024 | -0.029 | -0.029 | -0.026 |
| caraloc12 | -0.006 | 0.058 | 0.032 | -0.053 | 0.010 | -0.018 | 0.080 | 0.015 | -0.077 | 0.009 | 0.107 | 0.032 | -0.043 |
| caraloc13 | -0.032 | -0.006 | 0.006 | 0.020 | 0.036 | 0.074 | -0.085 | -0.032 | 0.278 | -0.081 | -0.095 | -0.005 | -0.005 |
| caraloc14 | 0.026 | -0.007 | 0.006 | -0.020 | 0.004 | 0.000 | 0.044 | -0.012 | 0.140 | -0.002 | 0.073 | -0.045 | 0.006 |
| caraloc15 | 0.039 | -0.056 | -0.013 | 0.012 | 0.018 | 0.002 | 0.033 | -0.002 | 0.272 | -0.032 | 0.035 | -0.010 | -0.029 |
| caraloc16 | -0.047 | -0.009 | -0.006 | 0.049 | 0.071 | -0.022 | -0.011 | -0.004 | 0.237 | -0.054 | -0.061 | 0.058 | 0.042 |
| cdev1 | -0.051 | 0.296 | -0.022 | 0.060 | 0.003 | 0.001 | 0.034 | 0.013 | -0.010 | -0.025 | 0.025 | -0.039 | 0.029 |
| cdev2 | -0.085 | 0.102 | -0.026 | 0.264 | -0.019 | -0.019 | -0.032 | 0.021 | -0.009 | 0.021 | 0.006 | -0.123 | 0.011 |
| cdev3 | -0.015 | 0.271 | -0.004 | 0.018 | 0.003 | -0.004 | 0.011 | 0.010 | 0.008 | -0.022 | -0.022 | 0.004 | -0.012 |
| cdev4 | -0.025 | -0.005 | -0.009 | -0.000 | -0.015 | -0.001 | 0.009 | 0.025 | -0.008 | 0.008 | 0.001 | -0.022 | 0.451 |
| cdev5 | 0.228 | 0.023 | -0.023 | 0.082 | -0.020 | -0.011 | -0.010 | 0.022 | -0.004 | 0.004 | -0.004 | 0.068 | -0.043 |
| cdev6 | 0.067 | -0.085 | 0.251 | 0.063 | -0.004 | 0.029 | 0.014 | -0.037 | 0.008 | 0.036 | 0.010 | 0.047 | 0.036 |
| cdev7 | 0.007 | -0.099 | -0.206 | 0.085 | 0.046 | -0.005 | 0.070 | -0.006 | -0.028 | -0.018 | 0.047 | 0.028 | 0.123 |
| cdev8 | 0.308 | -0.040 | -0.010 | 0.001 | -0.012 | -0.003 | 0.004 | -0.005 | -0.012 | 0.014 | -0.020 | -0.013 | -0.001 |
| cdev9 | 0.312 | 0.012 | -0.015 | -0.089 | 0.007 | -0.030 | -0.020 | 0.000 | -0.017 | 0.000 | 0.001 | 0.029 | 0.008 |
| cdev10 | 0.011 | -0.002 | 0.321 | 0.027 | -0.011 | -0.018 | 0.035 | 0.008 | -0.020 | 0.022 | 0.000 | -0.054 | -0.093 |
| cdev11 | 0.105 | -0.038 | 0.056 | 0.191 | -0.013 | -0.002 | 0.004 | 0.020 | 0.008 | -0.007 | -0.024 | -0.027 | 0.040 |
| cdev12 | 0.059 | 0.281 | -0.002 | -0.008 | 0.025 | -0.010 | 0.032 | -0.021 | -0.024 | -0.020 | 0.011 | -0.017 | 0.029 |
| cdev13 | -0.020 | -0.315 | 0.030 | 0.046 | 0.003 | 0.011 | 0.024 | -0.005 | -0.015 | -0.003 | 0.026 | -0.002 | 0.040 |
| cdev14 | -0.042 | -0.016 | 0.306 | 0.049 | -0.005 | -0.016 | 0.041 | 0.015 | -0.026 | -0.019 | 0.037 | -0.024 | -0.070 |
| cdev15 | -0.001 | -0.034 | -0.167 | 0.105 | 0.067 | 0.011 | -0.005 | -0.071 | -0.009 | 0.001 | 0.033 | 0.015 | -0.076 |
| cdev16 | 0.245 | 0.063 | 0.013 | -0.027 | 0.022 | 0.005 | 0.001 | -0.033 | -0.004 | -0.012 | -0.002 | 0.085 | -0.029 |
| cdev17 | -0.061 | 0.171 | 0.028 | 0.192 | 0.010 | 0.018 | -0.079 | -0.034 | -0.012 | 0.001 | 0.038 | -0.003 | -0.038 |
| cdev19 | -0.069 | 0.004 | 0.029 | 0.356 | 0.011 | 0.015 | -0.010 | -0.008 | -0.010 | 0.009 | 0.043 | -0.093 | 0.014 |
| cdev20 | 0.073 | -0.046 | 0.014 | 0.311 | -0.044 | 0.007 | 0.029 | 0.051 | 0.019 | -0.005 | -0.001 | 0.128 | -0.075 |
| cdev21 | 0.263 | -0.006 | 0.001 | 0.107 | -0.005 | -0.009 | 0.021 | -0.002 | -0.006 | -0.006 | -0.002 | 0.025 | -0.051 |
| cdev22 | 0.171 | -0.002 | 0.011 | 0.139 | -0.019 | 0.040 | 0.019 | 0.006 | 0.035 | -0.003 | -0.008 | 0.252 | -0.084 |
| cdev24 | 0.152 | 0.037 | 0.026 | 0.141 | -0.012 | 0.045 | 0.026 | 0.007 | 0.010 | -0.025 | 0.008 | 0.198 | -0.012 |
| cdev25 | 0.082 | 0.143 | 0.048 | 0.101 | 0.017 | 0.035 | 0.022 | -0.030 | -0.010 | -0.035 | 0.020 | 0.189 | 0.004 |
| cdev27 | 0.041 | -0.035 | -0.301 | 0.103 | 0.014 | 0.006 | 0.025 | -0.013 | -0.009 | 0.009 | -0.000 | 0.028 | 0.061 |
| cdev28 | 0.007 | -0.010 | 0.278 | 0.060 | 0.004 | 0.019 | 0.018 | -0.017 | 0.007 | 0.004 | -0.010 | 0.076 | 0.042 |
| cdev29 | -0.034 | -0.310 | -0.006 | 0.055 | -0.010 | 0.016 | 0.026 | 0.000 | 0.006 | 0.021 | -0.008 | -0.023 | 0.009 |
| cdev30 | 0.144 | -0.010 | 0.033 | 0.135 | 0.014 | 0.012 | 0.013 | -0.006 | 0.011 | -0.008 | -0.037 | -0.211 | 0.044 |
| cdev31 | 0.114 | -0.036 | 0.067 | 0.201 | -0.021 | 0.006 | 0.004 | 0.026 | 0.004 | -0.008 | -0.006 | 0.034 | 0.007 |
| cdev32 | -0.030 | 0.231 | 0.035 | 0.094 | 0.008 | 0.018 | -0.064 | -0.028 | -0.011 | -0.014 | 0.023 | 0.018 | -0.027 |
| cdev33 | 0.096 | -0.015 | 0.024 | 0.045 | -0.005 | 0.002 | 0.012 | -0.026 | 0.014 | 0.016 | -0.041 | -0.411 | 0.009 |
| cdev34 | 0.309 | 0.017 | -0.034 | -0.005 | -0.001 | -0.010 | -0.028 | 0.000 | 0.004 | 0.013 | -0.037 | -0.017 | -0.043 |
| cdev35 | 0.037 | -0.046 | -0.303 | 0.065 | 0.028 | -0.008 | 0.045 | 0.018 | -0.036 | -0.050 | 0.035 | -0.000 | 0.013 |
| cdev36 | 0.011 | 0.185 | 0.087 | 0.021 | 0.026 | -0.004 | 0.027 | -0.039 | -0.026 | -0.026 | 0.042 | -0.201 | 0.045 |
| cdev37 | 0.212 | 0.012 | 0.023 | -0.076 | 0.015 | -0.008 | -0.014 | -0.033 | 0.001 | 0.011 | -0.003 | -0.038 | 0.235 |
| cdev38 | 0.103 | 0.032 | 0.053 | -0.005 | 0.003 | -0.009 | -0.005 | -0.001 | 0.021 | 0.008 | -0.011 | 0.085 | 0.229 |
| cdev40 | -0.009 | 0.008 | 0.268 | -0.041 | 0.026 | -0.023 | -0.039 | -0.003 | 0.007 | 0.002 | -0.018 | 0.039 | 0.000 |
| cdev42 | 0.070 | -0.096 | 0.206 | 0.072 | 0.020 | 0.023 | -0.035 | -0.045 | 0.015 | 0.010 | -0.004 | 0.028 | 0.075 |
| cdev44 | 0.291 | 0.018 | -0.026 | 0.015 | 0.006 | -0.014 | 0.013 | -0.024 | -0.022 | 0.013 | -0.003 | -0.151 | -0.027 |
| cdev45 | 0.030 | -0.046 | 0.094 | 0.092 | 0.036 | 0.022 | 0.023 | -0.023 | -0.042 | -0.033 | 0.061 | 0.056 | 0.212 |
| cdev46 | -0.004 | 0.059 | -0.006 | -0.047 | 0.019 | 0.023 | -0.013 | -0.001 | 0.023 | -0.052 | 0.010 | -0.049 | 0.390 |
| cdev47 | -0.046 | 0.010 | 0.210 | 0.111 | 0.026 | -0.032 | 0.022 | -0.004 | -0.017 | -0.004 | 0.038 | -0.154 | 0.016 |
| cdev48 | 0.015 | -0.317 | -0.013 | 0.052 | 0.008 | 0.017 | -0.008 | -0.022 | -0.013 | -0.012 | -0.003 | 0.022 | -0.032 |
| cdev49 | 0.203 | 0.035 | 0.000 | 0.085 | -0.013 | -0.008 | 0.000 | 0.005 | 0.019 | 0.031 | -0.000 | 0.043 | -0.017 |
| cdev50 | 0.020 | -0.063 | -0.219 | 0.064 | 0.027 | 0.020 | 0.030 | -0.042 | -0.038 | -0.047 | 0.014 | -0.004 | -0.016 |
| cdev51 | -0.004 | 0.312 | 0.003 | -0.022 | 0.010 | -0.018 | 0.026 | 0.007 | -0.006 | -0.003 | 0.010 | -0.014 | 0.044 |
| cdev53 | -0.004 | -0.005 | -0.061 | 0.327 | -0.013 | -0.002 | -0.014 | 0.030 | 0.018 | 0.021 | -0.025 | -0.195 | -0.000 |

**Notes:** The cells display rotated component loadings obtained from principal component analysis of all the measurements. For ease of visualization, the measurements have been ordered to make them comparable with the PBFA results. Type of rotation performed: orthogonal. White cell if $|\alpha_{jl}| < 0.2$, light gray cell if $0.2 \leqslant |\alpha_{jl}| < 0.3$ and dark gray cell if $|\alpha_{jl}| \geqslant 0.3$. Items: BCS = Cognitive Test Scores, conners = Conners Hyperactivity Scale, rutter = Rutter Parental 'A' Scale of Behavioral Disorder, lawseq = Self-Esteem Scale, caraloc = Locus of Control Scale, cdev = Child Development Scale.

**Table 3.B.2:** Component loadings from Principal Component Analysis (Males)

| Items | $pc_1$ | $pc_2$ | $pc_3$ | $pc_4$ | $pc_5$ | $pc_6$ | $pc_7$ | $pc_8$ | $pc_9$ | $pc_{10}$ | $pc_{11}$ | $pc_{12}$ | $pc_{13}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| cdev26 | 0.040 | 0.301 | -0.014 | -0.006 | 0.009 | 0.013 | 0.020 | 0.011 | -0.010 | 0.084 | -0.004 | 0.006 | -0.046 |
| conners17 | 0.035 | -0.047 | 0.046 | 0.249 | -0.021 | -0.019 | 0.038 | 0.156 | -0.136 | -0.050 | 0.026 | -0.001 | -0.029 |
| BCS SERT | 0.002 | -0.040 | -0.009 | 0.008 | -0.009 | 0.354 | -0.015 | 0.002 | -0.009 | -0.007 | 0.015 | 0.009 | -0.003 |
| cdev23 | -0.015 | -0.013 | -0.031 | 0.017 | 0.423 | 0.005 | -0.008 | 0.008 | -0.008 | 0.000 | 0.013 | -0.004 | -0.023 |
| cdev18 | 0.005 | -0.038 | 0.330 | -0.031 | 0.007 | 0.006 | -0.006 | 0.022 | 0.027 | 0.019 | 0.017 | -0.015 | 0.066 |
| conners5 | -0.014 | -0.008 | -0.010 | 0.096 | 0.000 | -0.029 | -0.087 | 0.016 | 0.394 | 0.029 | 0.018 | -0.015 | 0.010 |
| conners13 | -0.036 | 0.063 | -0.021 | -0.015 | 0.010 | -0.014 | 0.337 | -0.061 | 0.069 | -0.026 | -0.040 | 0.031 | 0.011 |
| cdev52 | 0.027 | -0.023 | 0.004 | -0.005 | -0.141 | -0.007 | 0.008 | 0.005 | 0.023 | 0.369 | 0.023 | 0.022 | -0.007 |
| cdev43 | 0.298 | 0.001 | -0.004 | 0.005 | -0.107 | -0.049 | -0.029 | 0.001 | 0.015 | 0.017 | 0.024 | -0.002 | 0.041 |
| rutter10 | -0.026 | 0.036 | -0.032 | 0.303 | 0.022 | 0.024 | -0.009 | -0.126 | 0.110 | 0.023 | -0.003 | -0.001 | 0.051 |
| rutter1 | -0.001 | -0.041 | 0.007 | -0.007 | 0.004 | -0.007 | 0.360 | 0.005 | -0.114 | 0.028 | 0.026 | -0.006 | 0.034 |
| cdev39 | 0.092 | 0.067 | 0.064 | 0.004 | 0.026 | 0.060 | 0.001 | 0.009 | 0.006 | 0.190 | -0.016 | -0.017 | 0.114 |
| cdev41 | -0.022 | 0.001 | -0.010 | -0.009 | -0.011 | -0.003 | 0.018 | 0.003 | 0.010 | -0.000 | 0.010 | 0.014 | 0.565 |
| BCS PLCT | -0.025 | 0.050 | -0.003 | -0.023 | 0.013 | 0.352 | 0.024 | 0.015 | -0.005 | -0.006 | -0.024 | 0.013 | -0.031 |
| BCS FMT | 0.008 | -0.060 | -0.005 | 0.025 | -0.006 | 0.342 | -0.018 | -0.029 | -0.001 | 0.010 | 0.028 | -0.011 | -0.022 |
| BCS BASTM | -0.015 | -0.012 | -0.028 | -0.000 | 0.016 | 0.334 | -0.015 | 0.017 | -0.004 | 0.001 | 0.006 | -0.055 | 0.020 |
| BCS BASTRD | -0.006 | -0.024 | 0.023 | 0.033 | -0.028 | 0.217 | 0.002 | -0.029 | -0.018 | -0.006 | 0.043 | -0.030 | 0.015 |
| BCS BASTS | -0.023 | 0.046 | 0.001 | -0.004 | -0.004 | 0.376 | -0.007 | 0.027 | -0.002 | -0.003 | -0.014 | 0.022 | -0.006 |
| BCS BASTWD | -0.029 | 0.032 | 0.012 | -0.008 | 0.003 | 0.385 | 0.003 | 0.020 | -0.009 | 0.020 | -0.017 | 0.034 | -0.024 |
| rutter2 | 0.010 | -0.023 | 0.022 | -0.032 | -0.008 | 0.047 | 0.326 | 0.041 | -0.003 | 0.016 | 0.020 | -0.042 | -0.015 |
| rutter3 | -0.026 | 0.029 | -0.009 | 0.298 | 0.021 | 0.016 | -0.015 | -0.093 | 0.084 | 0.000 | -0.013 | 0.005 | -0.008 |
| rutter4 | 0.024 | -0.010 | 0.005 | 0.298 | 0.006 | -0.017 | -0.026 | -0.088 | 0.009 | 0.024 | 0.017 | -0.026 | -0.002 |
| rutter5 | -0.015 | 0.029 | 0.004 | 0.171 | 0.031 | 0.021 | -0.044 | 0.036 | 0.067 | -0.006 | -0.018 | -0.006 | -0.009 |
| rutter6 | -0.031 | 0.018 | -0.000 | -0.075 | 0.030 | 0.031 | -0.007 | 0.422 | 0.010 | 0.050 | -0.045 | 0.017 | 0.019 |
| rutter7 | 0.045 | 0.021 | 0.013 | -0.001 | -0.052 | 0.028 | -0.037 | 0.227 | 0.023 | -0.122 | -0.045 | 0.001 | -0.019 |
| rutter8 | 0.024 | -0.037 | 0.032 | 0.210 | -0.026 | -0.008 | 0.034 | 0.217 | -0.187 | -0.029 | 0.007 | 0.001 | 0.001 |
| rutter9 | 0.032 | 0.020 | 0.003 | 0.177 | -0.016 | 0.017 | -0.063 | 0.227 | -0.004 | -0.055 | -0.031 | 0.003 | -0.026 |
| rutter11 | -0.064 | 0.040 | -0.020 | 0.086 | 0.041 | 0.032 | -0.031 | 0.076 | 0.178 | 0.082 | 0.073 | -0.054 | 0.042 |
| rutter12 | -0.011 | 0.030 | -0.043 | 0.008 | 0.017 | -0.011 | -0.047 | 0.051 | 0.153 | 0.029 | -0.056 | 0.060 | 0.018 |
| rutter13 | -0.055 | -0.031 | -0.007 | 0.075 | 0.032 | -0.028 | -0.003 | 0.047 | 0.038 | 0.123 | -0.055 | 0.033 | 0.062 |
| rutter14 | 0.015 | -0.000 | 0.008 | 0.266 | -0.012 | -0.002 | 0.099 | -0.034 | -0.078 | -0.020 | 0.005 | 0.000 | 0.001 |
| rutter15 | -0.033 | 0.014 | 0.000 | 0.002 | 0.020 | -0.029 | 0.346 | -0.068 | 0.040 | 0.005 | -0.008 | 0.012 | 0.009 |
| rutter16 | -0.034 | 0.043 | -0.036 | -0.114 | 0.037 | 0.015 | -0.019 | 0.379 | 0.094 | 0.023 | 0.039 | -0.007 | 0.045 |
| rutter17 | 0.034 | -0.024 | -0.030 | -0.040 | -0.007 | -0.028 | -0.027 | 0.326 | 0.040 | 0.004 | 0.072 | -0.014 | 0.004 |
| rutter18 | -0.008 | 0.029 | -0.040 | 0.303 | 0.009 | 0.006 | 0.038 | -0.115 | 0.032 | -0.003 | -0.011 | 0.000 | 0.046 |
| rutter19 | -0.020 | -0.007 | 0.014 | 0.334 | 0.025 | -0.005 | -0.086 | -0.046 | 0.095 | 0.059 | -0.003 | 0.018 | -0.023 |
| conners1 | 0.036 | -0.002 | 0.019 | -0.045 | -0.038 | 0.027 | 0.087 | 0.033 | 0.361 | -0.066 | -0.033 | 0.016 | -0.032 |
| conners2 | 0.047 | -0.024 | 0.022 | -0.012 | -0.045 | -0.006 | 0.066 | 0.028 | 0.349 | -0.080 | -0.027 | 0.027 | -0.006 |
| conners3 | -0.011 | 0.094 | -0.027 | -0.022 | -0.034 | -0.015 | 0.311 | -0.032 | 0.057 | -0.010 | -0.017 | -0.019 | 0.018 |
| conners4 | 0.036 | -0.029 | -0.013 | 0.006 | -0.021 | -0.006 | 0.209 | 0.035 | 0.096 | -0.030 | 0.004 | -0.038 | -0.059 |
| conners6 | 0.003 | -0.006 | 0.001 | 0.002 | 0.002 | 0.005 | 0.006 | 0.058 | 0.398 | 0.022 | 0.010 | -0.032 | -0.003 |
| conners7 | 0.020 | -0.011 | -0.007 | 0.012 | -0.003 | -0.003 | 0.011 | 0.300 | 0.105 | -0.007 | 0.027 | 0.006 | -0.030 |
| conners8 | 0.022 | -0.035 | 0.019 | 0.106 | -0.006 | -0.009 | 0.123 | 0.233 | -0.098 | -0.021 | 0.024 | 0.013 | -0.025 |
| conners9 | 0.014 | -0.061 | 0.010 | 0.046 | 0.020 | 0.017 | 0.337 | 0.053 | -0.073 | 0.037 | 0.018 | -0.002 | 0.003 |
| conners10 | 0.019 | -0.057 | 0.029 | 0.050 | -0.004 | -0.007 | 0.259 | 0.104 | -0.061 | 0.028 | 0.001 | 0.018 | 0.001 |
| conners11 | 0.014 | -0.019 | -0.008 | 0.238 | 0.031 | -0.000 | 0.052 | -0.023 | 0.072 | 0.022 | -0.027 | 0.041 | -0.028 |
| conners12 | 0.040 | 0.019 | -0.017 | 0.210 | -0.047 | -0.024 | -0.042 | 0.159 | -0.052 | -0.098 | -0.015 | 0.008 | 0.002 |
| conners14 | -0.001 | -0.039 | 0.008 | 0.051 | -0.009 | 0.001 | 0.134 | 0.040 | 0.114 | 0.020 | -0.007 | -0.021 | -0.044 |
| conners15 | 0.008 | 0.020 | 0.020 | 0.145 | -0.029 | 0.017 | -0.019 | 0.194 | -0.005 | -0.065 | 0.011 | -0.036 | -0.004 |
| conners16 | 0.023 | -0.031 | 0.018 | 0.212 | -0.034 | -0.012 | 0.084 | 0.158 | -0.118 | -0.058 | 0.012 | -0.001 | -0.014 |
| conners18 | -0.048 | -0.028 | 0.023 | 0.052 | 0.035 | -0.033 | -0.021 | -0.008 | 0.351 | 0.048 | 0.017 | 0.008 | 0.011 |
| conners19 | -0.030 | 0.065 | -0.032 | -0.023 | 0.004 | -0.024 | 0.295 | -0.054 | 0.109 | -0.027 | -0.014 | 0.012 | 0.024 |
| lawseq1 | -0.042 | 0.035 | 0.016 | -0.021 | -0.010 | -0.030 | -0.008 | 0.083 | -0.022 | 0.063 | -0.124 | 0.323 | 0.017 |
| lawseq2 | -0.029 | -0.006 | 0.018 | 0.039 | 0.022 | -0.007 | -0.002 | -0.037 | -0.018 | 0.084 | 0.333 | -0.069 | -0.012 |
| lawseq3 | -0.050 | 0.032 | 0.008 | -0.017 | 0.011 | 0.043 | 0.001 | 0.060 | -0.006 | 0.021 | 0.325 | -0.005 | 0.038 |
| lawseq4 | -0.009 | 0.022 | -0.008 | -0.023 | 0.016 | -0.086 | -0.030 | 0.057 | 0.024 | 0.020 | 0.270 | 0.111 | 0.017 |
| lawseq5 | 0.061 | -0.009 | 0.041 | -0.018 | -0.047 | 0.011 | 0.021 | -0.027 | 0.041 | -0.063 | 0.215 | 0.012 | -0.074 |
| lawseq6 | 0.033 | -0.007 | -0.007 | 0.010 | -0.007 | -0.018 | 0.008 | 0.011 | -0.030 | -0.011 | 0.335 | -0.011 | 0.000 |
| lawseq7 | 0.004 | 0.006 | -0.030 | -0.020 | 0.021 | -0.043 | -0.001 | -0.031 | 0.023 | -0.036 | 0.230 | 0.071 | 0.032 |
| lawseq8 | 0.065 | -0.025 | 0.014 | 0.009 | -0.028 | -0.000 | -0.001 | -0.037 | 0.010 | -0.067 | 0.206 | 0.079 | -0.033 |
| lawseq9 | 0.044 | 0.008 | 0.031 | 0.025 | -0.025 | -0.011 | -0.018 | -0.049 | 0.025 | -0.072 | 0.215 | 0.088 | -0.090 |
| lawseq10 | -0.005 | -0.012 | -0.008 | -0.014 | 0.023 | 0.031 | -0.001 | 0.021 | 0.001 | 0.061 | 0.366 | -0.078 | 0.022 |
| lawseq11 | 0.002 | 0.005 | -0.031 | 0.027 | 0.027 | 0.036 | 0.025 | -0.061 | -0.011 | -0.069 | 0.130 | 0.055 | 0.038 |
| lawseq12 | 0.008 | 0.029 | 0.004 | -0.068 | -0.006 | -0.093 | 0.009 | 0.057 | 0.016 | -0.073 | 0.184 | 0.186 | 0.019 |
| caraloc1 | -0.030 | 0.014 | 0.020 | 0.018 | 0.009 | 0.031 | -0.010 | -0.007 | -0.028 | 0.023 | 0.077 | 0.290 | 0.047 |
| caraloc2 | 0.052 | 0.031 | -0.032 | 0.032 | -0.037 | 0.092 | 0.001 | -0.053 | 0.017 | -0.091 | 0.146 | -0.073 | 0.024 |
| caraloc3 | 0.070 | 0.054 | -0.036 | -0.050 | -0.059 | 0.104 | 0.048 | -0.021 | 0.055 | -0.117 | -0.078 | 0.057 | 0.024 |
| caraloc4 | -0.010 | -0.019 | -0.012 | 0.017 | 0.018 | 0.080 | 0.019 | -0.034 | 0.012 | -0.017 | 0.082 | 0.228 | 0.043 |
| caraloc5 | 0.055 | -0.037 | -0.006 | 0.018 | -0.001 | 0.038 | 0.015 | -0.057 | 0.024 | -0.055 | 0.031 | 0.288 | -0.026 |
| caraloc6 | 0.010 | -0.005 | -0.025 | -0.007 | -0.003 | 0.039 | -0.003 | -0.017 | 0.033 | -0.016 | -0.030 | 0.275 | 0.016 |

continued on next page...

Table 3.B.2 – continued from previous page

| Items | $pc_1$ | $pc_2$ | $pc_3$ | $pc_4$ | $pc_5$ | $pc_6$ | $pc_7$ | $pc_8$ | $pc_9$ | $pc_{10}$ | $pc_{11}$ | $pc_{12}$ | $pc_{13}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| caraloc7 | -0.043 | -0.021 | 0.024 | -0.046 | 0.038 | -0.051 | 0.023 | 0.020 | -0.008 | 0.012 | 0.178 | 0.152 | 0.019 |
| caraloc8 | -0.044 | 0.037 | -0.019 | -0.036 | 0.010 | 0.026 | 0.002 | 0.060 | -0.005 | 0.048 | -0.139 | 0.227 | 0.097 |
| caraloc9 | -0.007 | -0.002 | -0.016 | -0.015 | -0.014 | -0.082 | 0.031 | -0.020 | -0.018 | 0.004 | -0.039 | 0.276 | 0.050 |
| caraloc10 | 0.014 | 0.039 | -0.013 | 0.050 | -0.024 | 0.138 | -0.005 | -0.038 | -0.037 | -0.047 | 0.052 | 0.095 | 0.004 |
| caraloc11 | 0.028 | 0.016 | -0.010 | -0.002 | -0.020 | 0.049 | -0.005 | 0.025 | 0.004 | -0.009 | 0.031 | 0.294 | -0.041 |
| caraloc12 | 0.025 | 0.031 | -0.048 | -0.012 | -0.009 | 0.051 | 0.023 | -0.066 | 0.038 | -0.059 | 0.155 | -0.104 | 0.065 |
| caraloc13 | 0.023 | -0.011 | -0.006 | 0.058 | 0.002 | -0.091 | -0.020 | -0.041 | -0.016 | 0.001 | -0.110 | 0.230 | -0.003 |
| caraloc14 | 0.058 | -0.062 | 0.026 | 0.018 | -0.016 | 0.057 | 0.001 | -0.024 | 0.009 | -0.047 | 0.060 | 0.168 | -0.054 |
| caraloc15 | 0.060 | -0.051 | -0.006 | -0.017 | -0.025 | 0.068 | 0.010 | -0.027 | 0.036 | -0.046 | 0.068 | 0.237 | -0.010 |
| caraloc16 | -0.061 | 0.002 | 0.029 | 0.058 | 0.047 | -0.005 | -0.040 | -0.006 | -0.038 | 0.094 | -0.045 | 0.275 | -0.049 |
| cdev1 | -0.058 | 0.300 | -0.007 | -0.009 | 0.051 | 0.068 | 0.010 | 0.004 | -0.025 | -0.062 | -0.001 | -0.015 | 0.027 |
| cdev2 | -0.066 | 0.069 | -0.016 | -0.005 | 0.298 | -0.066 | -0.009 | 0.030 | -0.002 | -0.066 | 0.047 | -0.013 | -0.008 |
| cdev3 | -0.028 | 0.255 | 0.002 | -0.013 | 0.028 | -0.013 | -0.009 | 0.026 | 0.015 | 0.038 | 0.011 | 0.020 | 0.013 |
| cdev4 | -0.007 | -0.013 | -0.015 | -0.014 | -0.009 | -0.008 | 0.008 | 0.021 | -0.001 | -0.020 | -0.008 | -0.009 | 0.515 |
| cdev5 | 0.253 | -0.009 | -0.005 | -0.024 | 0.045 | 0.003 | 0.012 | -0.005 | 0.008 | 0.021 | -0.011 | -0.005 | -0.050 |
| cdev6 | 0.069 | -0.050 | 0.265 | -0.044 | 0.022 | 0.049 | 0.016 | 0.003 | 0.062 | 0.075 | -0.006 | -0.034 | 0.008 |
| cdev7 | 0.056 | -0.056 | -0.257 | -0.015 | 0.030 | 0.059 | -0.010 | 0.033 | 0.008 | 0.006 | 0.023 | -0.032 | 0.124 |
| cdev8 | 0.322 | -0.043 | -0.027 | 0.010 | -0.006 | -0.024 | -0.009 | 0.005 | -0.008 | 0.002 | -0.001 | -0.017 | 0.001 |
| cdev9 | 0.294 | 0.004 | -0.022 | -0.010 | -0.073 | -0.022 | -0.007 | 0.010 | 0.011 | 0.076 | 0.016 | 0.010 | 0.001 |
| cdev10 | 0.066 | 0.007 | 0.263 | -0.056 | 0.012 | 0.030 | 0.010 | 0.002 | 0.064 | -0.148 | -0.029 | 0.009 | -0.066 |
| cdev11 | 0.122 | -0.047 | 0.037 | -0.014 | 0.141 | 0.006 | 0.012 | -0.003 | -0.002 | -0.080 | -0.021 | -0.011 | 0.088 |
| cdev12 | 0.090 | 0.276 | -0.028 | -0.014 | -0.036 | -0.009 | -0.003 | -0.011 | 0.010 | -0.041 | -0.002 | 0.004 | 0.002 |
| cdev13 | -0.017 | -0.316 | 0.022 | -0.003 | 0.038 | 0.007 | -0.013 | -0.014 | 0.017 | 0.007 | 0.002 | 0.003 | 0.025 |
| cdev14 | 0.041 | -0.037 | 0.252 | -0.035 | 0.014 | 0.020 | -0.004 | 0.021 | 0.038 | -0.155 | -0.044 | 0.021 | 0.007 |
| cdev15 | 0.001 | -0.018 | -0.169 | -0.003 | 0.042 | -0.012 | 0.013 | -0.021 | 0.015 | 0.022 | 0.015 | -0.007 | -0.088 |
| cdev16 | 0.236 | 0.100 | -0.001 | -0.008 | -0.063 | 0.001 | 0.015 | -0.001 | 0.001 | 0.084 | -0.016 | 0.021 | -0.053 |
| cdev17 | -0.057 | 0.178 | 0.016 | 0.008 | 0.186 | -0.120 | 0.000 | 0.013 | -0.010 | 0.023 | 0.032 | 0.014 | -0.046 |
| cdev19 | -0.072 | 0.023 | 0.033 | 0.010 | 0.377 | 0.002 | 0.008 | 0.010 | -0.007 | 0.005 | 0.029 | -0.017 | 0.009 |
| cdev20 | 0.077 | -0.079 | -0.016 | -0.044 | 0.311 | 0.027 | 0.011 | 0.023 | 0.031 | 0.028 | -0.035 | 0.043 | -0.037 |
| cdev21 | 0.250 | -0.016 | -0.014 | 0.008 | 0.109 | -0.002 | -0.016 | -0.006 | -0.014 | 0.051 | -0.020 | 0.002 | -0.035 |
| cdev22 | 0.138 | -0.012 | 0.015 | -0.008 | 0.126 | 0.036 | 0.028 | 0.004 | -0.014 | 0.292 | -0.016 | 0.033 | -0.066 |
| cdev24 | 0.129 | 0.048 | 0.020 | 0.007 | 0.135 | 0.044 | 0.035 | -0.021 | -0.014 | 0.253 | 0.013 | 0.011 | -0.044 |
| cdev25 | 0.084 | 0.164 | 0.035 | 0.009 | 0.068 | 0.055 | 0.024 | -0.009 | -0.015 | 0.196 | 0.002 | -0.006 | -0.019 |
| cdev27 | 0.057 | 0.014 | -0.336 | -0.040 | 0.038 | 0.017 | 0.011 | 0.022 | 0.042 | 0.039 | -0.022 | -0.008 | 0.059 |
| cdev28 | 0.016 | -0.010 | 0.281 | 0.012 | 0.040 | 0.026 | -0.013 | -0.009 | -0.005 | 0.079 | 0.011 | -0.014 | 0.036 |
| cdev29 | -0.037 | -0.297 | -0.002 | -0.002 | 0.059 | 0.020 | -0.006 | -0.006 | 0.004 | 0.028 | -0.007 | 0.009 | 0.016 |
| cdev30 | 0.154 | -0.018 | 0.032 | 0.009 | 0.145 | 0.013 | -0.006 | -0.027 | -0.013 | -0.195 | -0.036 | -0.033 | 0.095 |
| cdev31 | 0.107 | -0.061 | 0.021 | -0.008 | 0.239 | 0.007 | 0.012 | -0.011 | -0.018 | -0.004 | -0.030 | 0.019 | 0.048 |
| cdev32 | -0.047 | 0.225 | 0.048 | 0.018 | 0.104 | -0.075 | -0.007 | 0.011 | -0.026 | 0.036 | 0.022 | -0.008 | -0.023 |
| cdev33 | 0.086 | -0.001 | 0.012 | 0.000 | 0.130 | 0.050 | 0.003 | -0.014 | 0.003 | -0.296 | -0.096 | 0.020 | 0.010 |
| cdev34 | 0.300 | 0.030 | -0.017 | 0.007 | -0.032 | -0.026 | -0.012 | -0.016 | -0.005 | 0.004 | -0.016 | -0.002 | -0.049 |
| cdev35 | 0.060 | -0.017 | -0.342 | -0.019 | 0.042 | 0.043 | -0.010 | 0.038 | 0.005 | 0.007 | -0.011 | -0.014 | 0.056 |
| cdev36 | 0.052 | 0.193 | 0.033 | 0.026 | 0.013 | 0.010 | 0.003 | -0.041 | -0.014 | -0.244 | 0.013 | -0.020 | 0.048 |
| cdev37 | 0.218 | 0.009 | 0.040 | 0.012 | -0.046 | -0.002 | 0.018 | -0.023 | 0.003 | 0.034 | -0.014 | 0.007 | 0.149 |
| cdev38 | 0.105 | 0.046 | 0.083 | 0.002 | 0.025 | 0.045 | 0.012 | 0.012 | 0.007 | 0.192 | -0.019 | 0.002 | 0.077 |
| cdev40 | -0.017 | 0.048 | 0.260 | 0.005 | -0.026 | -0.005 | 0.001 | -0.011 | 0.009 | 0.012 | 0.000 | -0.014 | 0.017 |
| cdev42 | 0.056 | -0.004 | 0.187 | 0.001 | 0.024 | 0.032 | -0.011 | -0.004 | 0.006 | 0.113 | 0.012 | -0.045 | 0.083 |
| cdev44 | 0.291 | 0.035 | -0.040 | 0.012 | -0.003 | -0.037 | -0.051 | -0.006 | -0.003 | -0.162 | 0.012 | -0.024 | 0.005 |
| cdev45 | -0.004 | -0.026 | 0.131 | 0.023 | 0.120 | 0.027 | -0.024 | 0.004 | 0.018 | 0.126 | 0.012 | -0.007 | 0.189 |
| cdev46 | 0.054 | 0.015 | 0.007 | 0.062 | -0.023 | -0.031 | -0.029 | -0.023 | -0.038 | -0.022 | 0.011 | 0.007 | 0.395 |
| cdev47 | 0.042 | 0.003 | 0.157 | -0.047 | 0.059 | 0.006 | 0.020 | -0.008 | 0.045 | -0.251 | 0.000 | 0.002 | 0.028 |
| cdev48 | 0.021 | -0.318 | -0.030 | -0.012 | 0.037 | -0.014 | 0.009 | -0.020 | 0.011 | 0.006 | -0.002 | -0.017 | -0.010 |
| cdev49 | 0.210 | 0.013 | -0.007 | -0.027 | 0.093 | -0.020 | 0.005 | 0.018 | 0.017 | 0.071 | 0.009 | -0.001 | -0.067 |
| cdev50 | 0.047 | -0.050 | -0.275 | -0.010 | 0.038 | 0.020 | 0.022 | -0.003 | 0.027 | -0.033 | 0.001 | -0.010 | 0.006 |
| cdev51 | -0.007 | 0.322 | -0.005 | 0.023 | -0.019 | 0.034 | -0.006 | 0.010 | -0.010 | -0.013 | -0.009 | 0.000 | 0.019 |
| cdev53 | -0.006 | -0.017 | -0.036 | 0.017 | 0.376 | 0.008 | 0.008 | -0.004 | -0.014 | -0.090 | -0.008 | 0.001 | -0.037 |

**Notes:** The cells display rotated component loadings obtained from principal component analysis of all the measurements. For ease of visualization, the measurements have been ordered to make them comparable with the PBFA results. Type of rotation performed: orthogonal. White cell if $|\alpha_{jl}| < 0.2$, light gray cell if $0.2 \leqslant |\alpha_{jl}| < 0.3$ and dark gray cell if $|\alpha_{jl}| \geqslant 0.3$. Items: BCS = Cognitive Test Scores, conners = Conners Hyperactivity Scale, rutter = Rutter Parental 'A' Scale of Behavioral Disorder, lawseq = Self-Esteem Scale, caraloc = Locus of Control Scale, cdev = Child Development Scale.

## Interpretation of the principal components:

pc1 — Antisocial Behavior (subscale of the Child Developmental Scale).

pc2 — Disorganized Activity (subscale of the Child Developmental Scale).

pc3 — Clumsiness and Poor Hand-Eye Coordination (subscales of the Child Developmental Scale).

pc4 — Agreeableness for males,[31] Neuroticism/Anxiety for females (subscale of the Child Developmental Scale).

pc5 — Neuroticism/Anxiety for males (subscale of the Child Developmental Scale), Agreeableness for females.

pc6 — Cognitive Ability for males, Restlessness for females.

pc7 — Restlessness for males, Cognitive Ability for females.

pc8 — Neuroticism.[31]

pc9 — Coordination Problems for males, Locus of Control for females.

pc10 — Hyperkinesis and Introversion/Extraversion (subscales of the Child Developmental Scale) for males, Coordination Problems for females.

pc11 — Self-Esteem.

pc12 — Locus of Control for males, Introversion/Extraversion (subscale of the Child Developmental Scale) for females.

pc13 — Behavioral Trauma.

[31] We thank Angela Duckworth for her help in mapping these items to the Big Five Personality Traits.

**Table 3.B.3:** Effect of the principal components (PCA) on the outcomes ($p$-values)

| | $pc_1$ | $pc_2$ | $pc_3$ | $pc_4$ | $pc_5$ | $pc_6$ | $pc_7$ | $pc_8$ | $pc_9$ | $pc_{10}$ | $pc_{11}$ | $pc_{12}$ | $pc_{13}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Males** | | | | | | | | | | | | | |
| Education | 0.771 | 0.025 | 0.938 | 0.271 | 0.246 | 0.000 | 0.197 | 0.265 | 0.405 | 0.061 | 0.472 | 0.186 | 0.728 |
| Poor health $d=0$ | 0.445 | 0.414 | 0.330 | 0.536 | 0.296 | 0.381 | 0.191 | 0.712 | 0.835 | 0.530 | 0.984 | 0.577 | 0.565 |
| Poor health $d=1$ | 0.284 | 0.305 | 0.852 | 0.109 | 0.563 | 0.843 | 0.654 | 0.298 | 0.112 | 0.049 | 0.087 | 0.631 | 0.070 |
| Obesity $d=0$ | 0.017 | 0.217 | 0.941 | 0.092 | 0.014 | 0.463 | 0.301 | 0.596 | 0.092 | 0.010 | 0.041 | 0.789 | 0.943 |
| Obesity $d=1$ | 0.559 | 0.140 | 0.018 | 0.456 | 0.829 | 0.292 | 0.496 | 0.289 | 0.456 | 0.466 | 0.389 | 0.404 | 0.501 |
| Smoking $d=0$ | 0.119 | 0.052 | 0.508 | 0.014 | 0.065 | 0.267 | 0.943 | 0.080 | 0.021 | 0.964 | 0.570 | 0.017 | 0.041 |
| Smoking $d=1$ | 0.012 | 0.014 | 0.100 | 0.708 | 0.286 | 0.852 | 0.040 | 0.757 | 0.475 | 0.109 | 0.847 | 0.496 | 0.381 |
| Wages $d=0$ | 0.062 | 0.675 | 0.011 | 0.207 | 0.172 | 0.002 | 0.174 | 0.349 | 0.652 | 0.000 | 0.760 | 0.212 | 0.691 |
| Wages $d=1$ | 0.651 | 0.405 | 0.877 | 0.139 | 0.413 | 0.010 | 0.004 | 0.690 | 0.036 | 0.181 | 0.463 | 0.022 | 0.301 |
| **Females** | | | | | | | | | | | | | |
| | $pc_1$ | $pc_2$ | $pc_3$ | $pc_4$ | $pc_5$ | $pc_6$ | $pc_7$ | $pc_8$ | $pc_9$ | $pc_{10}$ | $pc_{11}$ | $pc_{12}$ | $pc_{13}$ |
| Education | 0.955 | 0.100 | 0.259 | 0.359 | 0.549 | 0.088 | 0.000 | 0.189 | 0.031 | 0.257 | 0.215 | 0.803 | 0.138 |
| Poor health $d=0$ | 0.375 | 0.596 | 0.936 | 0.568 | 0.667 | 0.059 | 0.029 | 0.988 | 0.730 | 0.746 | 0.209 | 0.383 | 0.387 |
| Poor health $d=1$ | 0.593 | 0.391 | 0.884 | 0.770 | 0.624 | 0.867 | 0.715 | 0.273 | 0.659 | 0.878 | 0.029 | 0.679 | 0.887 |
| Obesity $d=0$ | 0.783 | 0.340 | 0.005 | 0.183 | 0.098 | 0.005 | 0.542 | 0.085 | 0.195 | 0.421 | 0.251 | 0.886 | 0.055 |
| Obesity $d=1$ | 0.274 | 0.505 | 0.038 | 0.636 | 0.292 | 0.767 | 0.108 | 0.058 | 0.746 | 0.666 | 0.247 | 0.604 | 0.036 |
| Smoking $d=0$ | 0.016 | 0.031 | 0.033 | 0.147 | 0.207 | 0.118 | 0.829 | 0.154 | 0.529 | 0.690 | 0.052 | 0.291 | 0.417 |
| Smoking $d=1$ | 0.289 | 0.013 | 0.176 | 0.544 | 0.272 | 0.081 | 0.119 | 0.009 | 0.132 | 0.552 | 0.534 | 0.897 | 0.902 |
| Wages $d=0$ | 0.513 | 0.616 | 0.832 | 0.900 | 0.201 | 0.749 | 0.010 | 0.597 | 0.186 | 0.654 | 0.569 | 0.566 | 0.707 |
| Wages $d=1$ | 0.329 | 0.053 | 0.841 | 0.548 | 0.854 | 0.653 | 0.000 | 0.077 | 0.400 | 0.203 | 0.942 | 0.852 | 0.732 |

**Notes:** Each cell of the table reports the $p$-value that the corresponding loading on the principal component is different from 0. The outcome equations include the same covariates as in the PBFA. Standard errors obtained by bootstrapping (100 replications).

## Standard Factor Analysis

**Table 3.B.4:** Factor loadings from Standard Factor Analysis (Females)

| Items | $f_1$ | $f_2$ | $f_3$ | $f_4$ | $f_5$ | $f_6$ | $f_7$ | $f_8$ | $f_9$ | $f_{10}$ | $f_{11}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| cdev26 | 0.737 | 0.055 | -0.169 | 0.139 | 0.046 | 0.026 | -0.053 | 0.171 | 0.412 | -0.001 | -0.022 |
| conners17 | 0.063 | 0.649 | -0.050 | 0.029 | 0.048 | 0.154 | 0.032 | -0.044 | 0.074 | -0.023 | -0.006 |
| BCS SERT | -0.337 | -0.080 | 0.762 | -0.095 | -0.060 | -0.012 | 0.034 | 0.004 | -0.100 | -0.055 | -0.005 |
| cdev23 | 0.149 | 0.037 | -0.058 | 0.811 | 0.049 | -0.002 | -0.015 | 0.141 | 0.215 | -0.020 | -0.031 |
| cdev18 | 0.166 | 0.056 | -0.051 | 0.115 | 0.654 | 0.037 | -0.007 | 0.013 | 0.261 | 0.017 | 0.168 |
| conners5 | -0.020 | 0.189 | -0.006 | 0.008 | 0.043 | 0.603 | -0.012 | -0.001 | -0.025 | 0.217 | 0.011 |
| conners13 | 0.205 | 0.426 | -0.085 | 0.025 | -0.015 | 0.066 | -0.084 | 0.067 | 0.062 | 0.393 | 0.057 |
| cdev52 | -0.067 | 0.012 | 0.086 | -0.546 | -0.120 | -0.003 | 0.035 | 0.472 | 0.218 | 0.036 | -0.047 |
| cdev43 | 0.071 | 0.036 | -0.058 | -0.091 | 0.022 | 0.055 | 0.002 | -0.132 | 0.785 | 0.017 | 0.117 |
| rutter10 | 0.080 | 0.309 | -0.038 | -0.007 | -0.040 | 0.575 | -0.016 | -0.049 | 0.052 | 0.023 | 0.018 |
| rutter1 | 0.103 | 0.435 | -0.141 | -0.045 | -0.077 | -0.107 | 0.042 | 0.098 | 0.120 | 0.276 | 0.115 |
| cdev39 | 0.209 | 0.036 | -0.039 | 0.027 | 0.197 | 0.029 | -0.024 | 0.118 | 0.467 | 0.034 | 0.373 |
| cdev41 | -0.006 | 0.049 | -0.014 | 0.030 | 0.149 | 0.013 | -0.009 | -0.062 | 0.160 | 0.031 | 0.489 |
| BCS PLCT | -0.126 | -0.085 | 0.600 | -0.059 | -0.057 | -0.028 | 0.046 | -0.010 | -0.060 | 0.020 | -0.038 |
| BCS FMT | -0.320 | -0.080 | 0.726 | -0.115 | -0.055 | -0.017 | 0.048 | 0.029 | -0.078 | -0.058 | 0.042 |
| BCS BASTM | -0.239 | -0.058 | 0.607 | -0.054 | -0.061 | -0.025 | -0.035 | -0.007 | -0.095 | -0.041 | 0.037 |
| BCS BASTRD | -0.136 | -0.059 | 0.429 | -0.075 | -0.038 | -0.010 | 0.051 | -0.031 | 0.003 | -0.019 | 0.028 |
| BCS BASTS | -0.181 | -0.072 | 0.709 | -0.095 | -0.017 | -0.025 | 0.054 | 0.015 | -0.049 | 0.036 | -0.034 |
| BCS BASTWD | -0.203 | -0.079 | 0.714 | -0.115 | -0.021 | -0.045 | 0.072 | 0.003 | -0.084 | 0.018 | -0.052 |
| rutter2 | 0.087 | 0.465 | -0.050 | -0.010 | -0.017 | -0.021 | -0.008 | 0.064 | 0.079 | 0.356 | 0.084 |
| rutter3 | 0.068 | 0.317 | -0.050 | 0.014 | -0.022 | 0.576 | -0.021 | 0.016 | 0.075 | 0.074 | 0.052 |
| rutter4 | 0.063 | 0.425 | -0.093 | -0.046 | 0.033 | 0.362 | -0.049 | 0.016 | 0.134 | -0.064 | 0.058 |
| rutter5 | 0.023 | 0.305 | -0.010 | 0.016 | 0.070 | 0.261 | -0.021 | -0.070 | 0.090 | 0.073 | -0.047 |
| rutter6 | 0.006 | 0.404 | 0.007 | 0.169 | 0.061 | 0.004 | -0.057 | 0.014 | -0.069 | 0.072 | -0.030 |
| rutter7 | -0.017 | 0.261 | 0.010 | 0.118 | 0.097 | 0.037 | -0.041 | -0.093 | -0.023 | 0.002 | 0.029 |
| rutter8 | 0.034 | 0.642 | -0.041 | 0.040 | 0.059 | 0.032 | -0.009 | -0.015 | 0.051 | -0.102 | -0.025 |
| rutter9 | 0.009 | 0.569 | -0.025 | 0.106 | 0.069 | 0.136 | -0.069 | -0.055 | 0.031 | -0.003 | -0.028 |
| rutter11 | -0.012 | 0.184 | -0.029 | 0.027 | 0.002 | 0.378 | -0.011 | 0.033 | -0.015 | 0.108 | -0.010 |
| rutter12 | 0.021 | 0.096 | 0.049 | 0.008 | 0.005 | 0.102 | -0.049 | 0.033 | -0.010 | 0.034 | -0.023 |
| rutter13 | 0.041 | 0.252 | -0.087 | 0.042 | -0.047 | 0.092 | -0.083 | 0.080 | 0.017 | 0.061 | -0.011 |
| rutter14 | 0.110 | 0.561 | -0.028 | -0.028 | -0.012 | 0.212 | -0.040 | 0.034 | 0.147 | -0.017 | 0.029 |
| rutter15 | 0.152 | 0.459 | -0.135 | 0.018 | -0.033 | 0.067 | -0.043 | 0.056 | 0.100 | 0.344 | 0.072 |
| rutter16 | -0.017 | 0.333 | -0.022 | 0.197 | 0.048 | 0.055 | -0.053 | -0.004 | -0.120 | 0.112 | -0.052 |
| rutter17 | -0.057 | 0.345 | -0.039 | 0.054 | 0.051 | 0.040 | -0.017 | 0.042 | -0.004 | -0.037 | 0.014 |
| rutter18 | 0.125 | 0.455 | -0.092 | 0.015 | -0.026 | 0.388 | -0.039 | -0.015 | 0.119 | 0.007 | 0.035 |
| rutter19 | 0.031 | 0.425 | -0.078 | -0.029 | 0.013 | 0.489 | -0.029 | -0.008 | 0.089 | -0.025 | -0.005 |
| conners1 | 0.031 | 0.255 | -0.004 | 0.003 | 0.099 | 0.315 | -0.024 | -0.002 | 0.027 | 0.544 | -0.038 |
| conners2 | 0.027 | 0.254 | -0.002 | 0.000 | 0.100 | 0.310 | -0.039 | -0.018 | 0.025 | 0.518 | -0.021 |
| conners3 | 0.322 | 0.460 | -0.158 | 0.004 | 0.003 | 0.043 | -0.081 | 0.062 | 0.121 | 0.338 | 0.056 |
| conners4 | 0.035 | 0.403 | -0.056 | -0.043 | 0.002 | 0.164 | -0.010 | 0.034 | 0.048 | 0.196 | 0.063 |
| conners6 | 0.007 | 0.212 | -0.002 | 0.023 | 0.079 | 0.484 | 0.013 | -0.011 | 0.030 | 0.493 | -0.003 |
| conners7 | -0.018 | 0.476 | 0.015 | 0.046 | 0.061 | 0.182 | 0.016 | -0.027 | 0.035 | 0.068 | 0.029 |
| conners8 | 0.045 | 0.669 | -0.023 | 0.050 | 0.035 | 0.059 | 0.009 | 0.012 | 0.033 | 0.044 | -0.018 |
| conners9 | 0.059 | 0.556 | -0.059 | -0.018 | -0.052 | 0.041 | 0.039 | 0.057 | 0.087 | 0.322 | 0.065 |
| conners10 | 0.075 | 0.565 | -0.039 | -0.039 | 0.011 | 0.015 | -0.006 | 0.049 | 0.106 | 0.216 | 0.042 |
| conners11 | 0.056 | 0.470 | -0.045 | -0.060 | -0.029 | 0.335 | -0.034 | 0.028 | 0.131 | 0.081 | 0.012 |
| conners12 | 0.026 | 0.551 | -0.068 | 0.045 | 0.034 | 0.134 | -0.075 | -0.062 | 0.048 | 0.005 | -0.032 |
| conners14 | 0.035 | 0.362 | -0.057 | 0.000 | -0.026 | 0.193 | -0.001 | 0.077 | 0.032 | 0.190 | 0.061 |
| conners15 | -0.005 | 0.554 | -0.065 | 0.078 | 0.060 | 0.101 | -0.049 | -0.033 | 0.032 | 0.017 | -0.031 |

continued on next page...

Table 3.B.4 – continued from previous page

| Items | $f_1$ | $f_2$ | $f_3$ | $f_4$ | $f_5$ | $f_6$ | $f_7$ | $f_8$ | $f_9$ | $f_{10}$ | $f_{11}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| conners16 | 0.034 | 0.717 | -0.076 | 0.036 | 0.042 | 0.071 | -0.017 | -0.029 | 0.080 | 0.029 | -0.019 |
| conners18 | 0.016 | 0.190 | -0.048 | 0.024 | 0.072 | 0.499 | -0.011 | 0.038 | 0.010 | 0.189 | -0.048 |
| conners19 | 0.180 | 0.396 | -0.124 | 0.046 | 0.026 | 0.154 | -0.020 | 0.050 | 0.081 | 0.344 | 0.048 |
| lawseq1 | -0.056 | -0.057 | 0.108 | 0.012 | -0.028 | -0.049 | 0.232 | 0.014 | -0.034 | 0.054 | -0.050 |
| lawseq2 | -0.071 | -0.062 | -0.002 | -0.087 | -0.053 | -0.021 | 0.344 | 0.059 | -0.097 | -0.090 | 0.024 |
| lawseq3 | -0.069 | -0.096 | 0.124 | -0.041 | -0.010 | -0.010 | 0.395 | 0.008 | -0.107 | -0.048 | 0.063 |
| lawseq4 | 0.002 | -0.040 | -0.024 | -0.033 | -0.042 | -0.026 | 0.418 | 0.007 | -0.099 | 0.011 | 0.065 |
| lawseq5 | -0.028 | -0.019 | 0.066 | -0.112 | -0.002 | -0.026 | 0.315 | 0.031 | 0.010 | 0.025 | -0.036 |
| lawseq6 | -0.037 | -0.044 | 0.100 | -0.098 | -0.020 | -0.055 | 0.421 | 0.007 | -0.079 | -0.074 | 0.063 |
| lawseq7 | -0.071 | -0.049 | 0.046 | -0.033 | -0.022 | 0.011 | 0.363 | 0.029 | -0.072 | -0.029 | 0.030 |
| lawseq8 | -0.025 | -0.006 | 0.102 | -0.126 | -0.017 | -0.031 | 0.385 | -0.037 | -0.016 | 0.027 | 0.005 |
| lawseq9 | -0.037 | -0.028 | 0.030 | -0.061 | -0.030 | 0.006 | 0.411 | -0.018 | -0.027 | -0.017 | -0.024 |
| lawseq10 | -0.075 | -0.103 | 0.078 | -0.095 | 0.018 | 0.022 | 0.378 | 0.031 | -0.118 | -0.079 | 0.072 |
| lawseq11 | -0.067 | -0.014 | 0.066 | -0.008 | -0.012 | -0.014 | 0.220 | -0.057 | -0.038 | 0.029 | -0.020 |
| lawseq12 | -0.013 | -0.040 | 0.049 | 0.006 | 0.011 | -0.039 | 0.449 | 0.002 | -0.067 | -0.018 | -0.050 |
| caraloc1 | -0.088 | -0.028 | 0.275 | -0.046 | -0.016 | -0.045 | 0.395 | 0.023 | -0.049 | 0.010 | -0.068 |
| caraloc2 | -0.003 | 0.025 | 0.066 | -0.028 | -0.009 | 0.079 | 0.108 | -0.022 | 0.018 | 0.003 | -0.027 |
| caraloc3 | -0.012 | 0.026 | 0.134 | 0.030 | -0.014 | 0.017 | -0.034 | -0.047 | 0.060 | 0.025 | -0.028 |
| caraloc4 | -0.150 | -0.034 | 0.276 | -0.069 | -0.013 | -0.032 | 0.345 | 0.028 | -0.053 | 0.000 | -0.037 |
| caraloc5 | -0.128 | -0.025 | 0.298 | -0.088 | -0.020 | 0.027 | 0.321 | 0.002 | 0.028 | -0.004 | -0.042 |
| caraloc6 | -0.136 | -0.044 | 0.220 | -0.065 | -0.024 | -0.007 | 0.297 | 0.026 | -0.051 | 0.013 | -0.052 |
| caraloc7 | -0.098 | -0.096 | 0.062 | -0.020 | -0.023 | -0.037 | 0.424 | -0.018 | -0.090 | -0.035 | 0.019 |
| caraloc8 | -0.048 | -0.041 | 0.127 | -0.008 | 0.023 | -0.046 | 0.080 | -0.023 | 0.002 | 0.034 | 0.007 |
| caraloc9 | -0.039 | 0.016 | 0.083 | 0.006 | -0.031 | -0.038 | 0.313 | 0.034 | -0.011 | 0.006 | -0.025 |
| caraloc10 | -0.081 | -0.040 | 0.234 | -0.071 | -0.077 | -0.004 | 0.200 | -0.003 | 0.030 | -0.036 | -0.051 |
| caraloc11 | -0.067 | -0.012 | 0.298 | -0.031 | -0.013 | -0.025 | 0.340 | -0.019 | 0.034 | 0.036 | -0.060 |
| caraloc12 | 0.041 | 0.000 | 0.062 | -0.086 | 0.022 | 0.025 | 0.043 | 0.005 | -0.025 | -0.010 | 0.000 |
| caraloc13 | 0.009 | 0.059 | 0.023 | 0.012 | -0.026 | -0.031 | 0.184 | 0.018 | 0.005 | 0.007 | -0.034 |
| caraloc14 | -0.096 | -0.041 | 0.176 | -0.037 | -0.002 | -0.022 | 0.263 | -0.014 | 0.001 | -0.014 | -0.009 |
| caraloc15 | -0.211 | -0.049 | 0.275 | -0.051 | -0.072 | -0.031 | 0.395 | 0.013 | 0.010 | -0.030 | -0.072 |
| caraloc16 | -0.065 | -0.023 | 0.139 | -0.004 | -0.034 | 0.032 | 0.191 | 0.041 | -0.017 | -0.016 | -0.010 |
| cdev1 | 0.626 | 0.063 | -0.127 | 0.241 | 0.095 | -0.002 | -0.019 | -0.022 | 0.138 | -0.001 | 0.047 |
| cdev2 | 0.315 | -0.010 | -0.184 | 0.629 | 0.083 | -0.005 | -0.039 | -0.052 | 0.000 | 0.034 | 0.014 |
| cdev3 | 0.572 | 0.057 | -0.149 | 0.148 | 0.083 | 0.000 | -0.047 | 0.007 | 0.216 | 0.016 | -0.008 |
| cdev4 | 0.009 | 0.037 | -0.006 | 0.040 | 0.170 | 0.007 | -0.020 | -0.047 | 0.125 | 0.019 | 0.392 |
| cdev5 | 0.205 | 0.041 | -0.054 | 0.129 | 0.046 | -0.016 | -0.021 | 0.052 | 0.653 | 0.021 | -0.071 |
| cdev6 | 0.098 | 0.036 | -0.012 | 0.143 | 0.483 | 0.028 | 0.008 | 0.079 | 0.396 | 0.072 | 0.110 |
| cdev7 | -0.338 | -0.029 | 0.193 | -0.063 | -0.306 | 0.039 | 0.033 | 0.058 | -0.133 | -0.056 | 0.083 |
| cdev8 | 0.054 | 0.054 | -0.009 | 0.057 | 0.060 | -0.003 | -0.032 | -0.052 | 0.768 | 0.018 | 0.004 |
| cdev9 | 0.127 | 0.042 | -0.052 | -0.118 | 0.023 | 0.030 | -0.007 | -0.073 | 0.781 | -0.014 | 0.046 |
| cdev10 | 0.244 | 0.031 | -0.042 | 0.218 | 0.629 | 0.004 | -0.044 | -0.066 | 0.244 | 0.020 | -0.108 |
| cdev11 | 0.136 | 0.036 | -0.039 | 0.388 | 0.221 | -0.022 | -0.048 | 0.004 | 0.437 | 0.008 | 0.019 |
| cdev12 | 0.663 | 0.060 | -0.118 | 0.115 | 0.120 | 0.039 | -0.047 | -0.053 | 0.399 | -0.010 | 0.054 |
| cdev13 | -0.672 | -0.061 | 0.202 | -0.061 | -0.019 | -0.015 | 0.035 | 0.047 | -0.261 | -0.052 | 0.060 |
| cdev14 | 0.174 | 0.004 | -0.021 | 0.204 | 0.553 | -0.022 | -0.006 | -0.029 | 0.115 | -0.028 | -0.062 |
| cdev15 | -0.116 | -0.026 | 0.029 | -0.004 | -0.258 | 0.053 | 0.024 | 0.047 | -0.112 | -0.020 | -0.123 |
| cdev16 | 0.277 | 0.056 | -0.052 | -0.061 | 0.072 | 0.040 | -0.012 | 0.058 | 0.734 | 0.001 | -0.011 |
| cdev17 | 0.544 | 0.025 | -0.305 | 0.429 | 0.148 | 0.022 | -0.005 | 0.058 | 0.164 | 0.050 | -0.039 |
| cdev19 | 0.208 | 0.020 | -0.119 | 0.749 | 0.170 | 0.027 | -0.001 | 0.049 | 0.087 | 0.015 | 0.055 |
| cdev20 | 0.089 | -0.003 | 0.030 | 0.444 | 0.088 | -0.055 | -0.007 | 0.250 | 0.402 | 0.027 | -0.122 |
| cdev21 | 0.150 | 0.031 | -0.008 | 0.221 | 0.090 | -0.003 | -0.019 | 0.079 | 0.756 | -0.012 | -0.063 |
| cdev22 | 0.140 | 0.052 | 0.021 | 0.078 | 0.018 | -0.014 | 0.018 | 0.407 | 0.650 | 0.061 | -0.091 |

continued on next page...

Table 3.B.4 – continued from previous page

| Items | $f_1$ | $f_2$ | $f_3$ | $f_4$ | $f_5$ | $f_6$ | $f_7$ | $f_8$ | $f_9$ | $f_{10}$ | $f_{11}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| cdev24 | 0.246 | 0.079 | -0.014 | 0.153 | 0.091 | -0.015 | -0.004 | 0.361 | 0.650 | 0.024 | 0.048 |
| cdev25 | 0.483 | 0.056 | -0.076 | 0.105 | 0.142 | 0.021 | -0.028 | 0.308 | 0.552 | 0.012 | 0.082 |
| cdev27 | -0.208 | -0.032 | 0.111 | -0.048 | -0.425 | 0.014 | -0.009 | 0.072 | -0.079 | -0.025 | -0.017 |
| cdev28 | 0.233 | 0.048 | -0.050 | 0.168 | 0.538 | 0.026 | -0.029 | 0.124 | 0.350 | 0.045 | 0.138 |
| cdev29 | -0.710 | -0.056 | 0.216 | -0.038 | -0.085 | -0.021 | 0.024 | 0.030 | -0.331 | -0.008 | -0.014 |
| cdev30 | 0.168 | 0.099 | -0.040 | 0.463 | 0.210 | 0.006 | -0.057 | -0.193 | 0.440 | -0.020 | 0.066 |
| cdev31 | 0.144 | 0.041 | -0.043 | 0.366 | 0.215 | -0.030 | -0.032 | 0.089 | 0.481 | 0.013 | 0.038 |
| cdev32 | 0.641 | 0.051 | -0.292 | 0.266 | 0.159 | 0.015 | -0.019 | 0.053 | 0.236 | 0.037 | -0.014 |
| cdev33 | 0.077 | 0.039 | -0.016 | 0.435 | 0.180 | -0.009 | -0.037 | -0.374 | 0.116 | -0.042 | 0.038 |
| cdev34 | 0.189 | 0.077 | -0.079 | 0.063 | 0.020 | 0.022 | -0.041 | -0.067 | 0.793 | 0.014 | -0.056 |
| cdev35 | -0.293 | -0.028 | 0.146 | -0.084 | -0.491 | -0.011 | 0.010 | 0.031 | -0.174 | -0.106 | -0.056 |
| cdev36 | 0.499 | 0.046 | -0.117 | 0.303 | 0.298 | 0.011 | -0.002 | -0.225 | 0.201 | -0.051 | 0.088 |
| cdev37 | 0.143 | 0.038 | -0.047 | -0.023 | 0.149 | 0.049 | 0.000 | -0.127 | 0.634 | 0.021 | 0.400 |
| cdev38 | 0.190 | 0.031 | -0.032 | 0.011 | 0.193 | 0.043 | -0.004 | 0.080 | 0.467 | 0.031 | 0.369 |
| cdev40 | 0.211 | 0.042 | -0.132 | 0.063 | 0.434 | 0.052 | -0.032 | 0.021 | 0.213 | 0.018 | 0.092 |
| cdev42 | 0.088 | 0.038 | -0.079 | 0.169 | 0.394 | 0.041 | -0.004 | 0.035 | 0.392 | 0.043 | 0.171 |
| cdev44 | 0.181 | 0.054 | -0.030 | 0.193 | 0.071 | 0.013 | -0.029 | -0.218 | 0.687 | -0.015 | -0.025 |
| cdev45 | 0.072 | 0.051 | -0.016 | 0.138 | 0.243 | 0.023 | 0.002 | 0.058 | 0.309 | -0.008 | 0.286 |
| cdev46 | 0.140 | 0.074 | -0.054 | 0.007 | 0.142 | 0.003 | 0.024 | -0.091 | 0.180 | -0.016 | 0.384 |
| cdev47 | 0.209 | 0.013 | -0.065 | 0.397 | 0.472 | 0.026 | 0.000 | -0.152 | 0.080 | -0.039 | 0.019 |
| cdev48 | -0.654 | -0.069 | 0.157 | -0.086 | -0.121 | -0.015 | 0.001 | 0.068 | -0.205 | -0.057 | -0.066 |
| cdev49 | 0.223 | 0.041 | -0.040 | 0.167 | 0.095 | 0.021 | 0.015 | 0.067 | 0.624 | 0.048 | -0.018 |
| cdev50 | -0.244 | -0.057 | 0.108 | -0.072 | -0.358 | -0.023 | -0.020 | 0.029 | -0.154 | -0.082 | -0.058 |
| cdev51 | 0.659 | 0.063 | -0.146 | 0.116 | 0.122 | 0.025 | -0.018 | -0.044 | 0.252 | 0.032 | 0.090 |
| cdev53 | 0.132 | 0.038 | -0.082 | 0.785 | 0.016 | -0.002 | -0.043 | -0.098 | 0.103 | 0.021 | -0.010 |

**Notes:** The cells display rotated factor loadings obtained from factor analysis of all the measurements. For ease of visualization, the factors and the measurements have been ordered to make them comparable with the PBFA results. Method used for FA: maximum likelihood. Type of rotation performed: orthogonal. White cell if $|\alpha_{jl}| < 0.4$, light gray cell if $0.4 \leqslant |\alpha_{jl}| < 0.7$ and dark gray cell if $|\alpha_{jl}| \geqslant 0.7$. Items: BCS = Cognitive Test Scores, conners = Conners Hyperactivity Scale, rutter = Rutter Parental 'A' Scale of Behavioral Disorder, lawseq = Self-Esteem Scale, caraloc = Locus of Control Scale, cdev = Child Development Scale.

**Table 3.B.5:** Factor loadings from Standard Factor Analysis (Males)

| Items | $f_1$ | $f_2$ | $f_3$ | $f_4$ | $f_5$ | $f_6$ | $f_7$ | $f_8$ | $f_9$ | $f_{10}$ | $f_{11}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| cdev26 | 0.731 | 0.031 | -0.161 | 0.080 | 0.060 | 0.017 | 0.104 | 0.136 | 0.453 | -0.024 | -0.037 |
| conners17 | 0.025 | 0.714 | -0.083 | 0.036 | 0.020 | 0.043 | 0.101 | -0.028 | 0.111 | 0.083 | 0.014 |
| BCS SERT | -0.304 | -0.060 | 0.774 | -0.080 | -0.072 | -0.028 | -0.104 | -0.003 | -0.103 | -0.008 | 0.049 |
| cdev23 | 0.119 | 0.017 | -0.058 | 0.840 | 0.062 | 0.011 | -0.008 | 0.106 | 0.174 | 0.004 | 0.002 |
| cdev18 | 0.138 | -0.016 | -0.067 | 0.171 | 0.689 | 0.053 | 0.004 | 0.032 | 0.211 | -0.017 | 0.000 |
| conners5 | -0.017 | 0.130 | -0.069 | -0.020 | 0.011 | 0.643 | 0.034 | -0.012 | -0.005 | 0.206 | 0.007 |
| conners13 | 0.210 | 0.169 | -0.117 | 0.030 | 0.037 | 0.135 | 0.680 | -0.031 | 0.100 | 0.070 | -0.026 |
| cdev52 | -0.098 | -0.057 | 0.006 | -0.535 | -0.136 | 0.032 | 0.068 | 0.394 | 0.274 | 0.021 | 0.076 |
| cdev43 | 0.108 | 0.073 | -0.093 | -0.083 | 0.059 | 0.022 | 0.012 | -0.135 | 0.732 | 0.061 | 0.041 |
| rutter10 | 0.096 | 0.292 | -0.064 | -0.045 | -0.033 | 0.255 | 0.179 | -0.019 | 0.134 | 0.474 | -0.018 |
| rutter1 | 0.050 | 0.323 | -0.076 | -0.040 | 0.011 | -0.043 | 0.475 | 0.102 | 0.138 | -0.081 | 0.014 |
| cdev39 | 0.216 | 0.003 | -0.007 | 0.033 | 0.221 | 0.008 | 0.061 | 0.162 | 0.541 | 0.084 | -0.076 |
| cdev41 | -0.007 | -0.018 | -0.012 | 0.057 | 0.214 | -0.008 | 0.003 | -0.064 | 0.068 | 0.107 | -0.016 |
| BCS PLCT | -0.108 | -0.071 | 0.628 | -0.003 | -0.022 | -0.010 | -0.011 | -0.005 | -0.061 | -0.063 | 0.017 |
| BCS FMT | -0.329 | -0.076 | 0.740 | -0.104 | -0.081 | -0.024 | -0.106 | 0.024 | -0.085 | 0.033 | 0.054 |
| BCS BASTM | -0.234 | -0.066 | 0.632 | -0.024 | -0.072 | -0.020 | -0.091 | -0.004 | -0.099 | 0.002 | -0.033 |

continued on next page...

Table 3.B.5 – continued from previous page

| Items | $f_1$ | $f_2$ | $f_3$ | $f_4$ | $f_5$ | $f_6$ | $f_7$ | $f_8$ | $f_9$ | $f_{10}$ | $f_{11}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| BCS BASTRD | -0.172 | -0.023 | 0.387 | -0.088 | -0.010 | -0.023 | -0.060 | 0.003 | -0.060 | 0.014 | 0.050 |
| BCS BASTS | -0.140 | -0.060 | 0.709 | -0.032 | -0.022 | -0.002 | -0.054 | -0.015 | -0.070 | -0.043 | 0.033 |
| BCS BASTWD | -0.169 | -0.074 | 0.742 | -0.045 | -0.018 | -0.017 | -0.046 | 0.024 | -0.071 | -0.050 | 0.049 |
| rutter2 | 0.062 | 0.301 | 0.002 | -0.021 | 0.047 | 0.109 | 0.451 | 0.082 | 0.137 | -0.077 | -0.029 |
| rutter3 | 0.101 | 0.337 | -0.077 | -0.017 | -0.018 | 0.248 | 0.150 | -0.008 | 0.102 | 0.379 | -0.021 |
| rutter4 | 0.076 | 0.359 | -0.124 | -0.045 | -0.021 | 0.155 | 0.078 | 0.016 | 0.203 | 0.315 | -0.014 |
| rutter5 | 0.074 | 0.257 | -0.049 | 0.070 | 0.018 | 0.194 | 0.072 | -0.031 | 0.063 | 0.174 | -0.040 |
| rutter6 | -0.030 | 0.338 | 0.022 | 0.180 | 0.010 | 0.161 | 0.078 | 0.002 | -0.050 | -0.186 | -0.064 |
| rutter7 | 0.020 | 0.261 | 0.030 | 0.114 | 0.044 | 0.130 | 0.001 | -0.109 | -0.025 | -0.059 | -0.072 |
| rutter8 | 0.010 | 0.697 | -0.070 | 0.039 | -0.001 | -0.021 | 0.078 | 0.007 | 0.084 | 0.004 | -0.014 |
| rutter9 | 0.056 | 0.521 | -0.034 | 0.100 | -0.004 | 0.212 | 0.038 | -0.060 | 0.078 | 0.068 | -0.074 |
| rutter11 | 0.027 | 0.169 | -0.030 | 0.016 | -0.014 | 0.295 | 0.078 | 0.048 | -0.021 | 0.128 | 0.018 |
| rutter12 | 0.006 | 0.035 | -0.010 | 0.027 | -0.029 | 0.190 | 0.050 | -0.028 | 0.006 | 0.061 | -0.010 |
| rutter13 | -0.040 | 0.137 | -0.047 | -0.038 | -0.028 | 0.101 | 0.087 | 0.053 | 0.013 | 0.085 | -0.038 |
| rutter14 | 0.107 | 0.491 | -0.096 | -0.034 | -0.010 | 0.067 | 0.234 | 0.002 | 0.175 | 0.252 | -0.015 |
| rutter15 | 0.135 | 0.196 | -0.145 | 0.008 | 0.053 | 0.099 | 0.657 | 0.019 | 0.111 | 0.080 | -0.001 |
| rutter16 | -0.031 | 0.211 | 0.009 | 0.177 | -0.019 | 0.202 | 0.069 | -0.042 | -0.117 | -0.147 | 0.013 |
| rutter17 | -0.078 | 0.307 | -0.021 | 0.073 | -0.050 | 0.164 | 0.010 | -0.017 | -0.019 | -0.114 | 0.045 |
| rutter18 | 0.120 | 0.366 | -0.091 | -0.044 | -0.059 | 0.181 | 0.207 | -0.013 | 0.154 | 0.413 | -0.037 |
| rutter19 | 0.043 | 0.392 | -0.087 | -0.040 | -0.019 | 0.292 | 0.062 | 0.019 | 0.124 | 0.393 | 0.008 |
| conners1 | 0.044 | 0.138 | 0.044 | 0.018 | 0.100 | 0.633 | 0.185 | -0.012 | 0.035 | -0.072 | -0.030 |
| conners2 | 0.019 | 0.176 | -0.002 | 0.012 | 0.105 | 0.624 | 0.154 | -0.040 | 0.039 | -0.033 | -0.012 |
| conners3 | 0.279 | 0.191 | -0.158 | -0.036 | 0.020 | 0.136 | 0.588 | 0.013 | 0.160 | 0.041 | -0.064 |
| conners4 | 0.033 | 0.287 | -0.067 | -0.013 | -0.020 | 0.223 | 0.329 | 0.017 | 0.092 | -0.001 | -0.043 |
| conners6 | 0.010 | 0.132 | -0.024 | 0.026 | 0.048 | 0.683 | 0.111 | 0.039 | 0.026 | 0.041 | -0.025 |
| conners7 | -0.014 | 0.429 | -0.003 | 0.086 | -0.018 | 0.328 | 0.085 | 0.013 | -0.005 | -0.115 | 0.008 |
| conners8 | 0.019 | 0.612 | -0.047 | 0.057 | -0.005 | 0.090 | 0.213 | 0.014 | 0.070 | -0.089 | 0.021 |
| conners9 | 0.010 | 0.456 | -0.031 | -0.008 | 0.008 | 0.049 | 0.509 | 0.095 | 0.179 | -0.036 | 0.009 |
| conners10 | 0.022 | 0.475 | -0.057 | -0.011 | 0.027 | 0.096 | 0.385 | 0.085 | 0.153 | -0.081 | 0.008 |
| conners11 | 0.050 | 0.412 | -0.057 | -0.001 | -0.028 | 0.260 | 0.216 | 0.016 | 0.173 | 0.255 | -0.002 |
| conners12 | 0.081 | 0.545 | -0.081 | 0.042 | -0.040 | 0.133 | 0.030 | -0.089 | 0.056 | 0.102 | -0.053 |
| conners14 | -0.002 | 0.274 | -0.032 | -0.036 | -0.008 | 0.261 | 0.240 | 0.043 | 0.048 | 0.049 | -0.035 |
| conners15 | 0.068 | 0.466 | -0.045 | 0.076 | 0.025 | 0.179 | 0.069 | -0.059 | 0.018 | 0.053 | -0.056 |
| conners16 | 0.033 | 0.687 | -0.076 | 0.008 | -0.015 | 0.070 | 0.168 | -0.015 | 0.069 | 0.041 | -0.014 |
| conners18 | -0.037 | 0.067 | -0.067 | 0.024 | 0.066 | 0.529 | 0.118 | 0.029 | -0.040 | 0.163 | 0.032 |
| conners19 | 0.192 | 0.130 | -0.129 | 0.023 | 0.021 | 0.176 | 0.605 | -0.050 | 0.086 | 0.100 | -0.016 |
| lawseq1 | -0.006 | 0.016 | 0.091 | -0.034 | 0.016 | -0.011 | 0.051 | 0.017 | -0.022 | -0.039 | 0.166 |
| lawseq2 | -0.060 | -0.039 | 0.015 | -0.114 | -0.039 | -0.051 | -0.045 | 0.095 | -0.050 | 0.041 | 0.315 |
| lawseq3 | -0.070 | -0.025 | 0.144 | -0.050 | -0.023 | -0.030 | -0.048 | 0.061 | -0.161 | -0.014 | 0.358 |
| lawseq4 | -0.033 | -0.025 | -0.009 | -0.026 | -0.027 | -0.004 | -0.053 | 0.032 | -0.075 | 0.001 | 0.397 |
| lawseq5 | -0.037 | -0.021 | 0.095 | -0.068 | 0.038 | 0.021 | -0.018 | 0.017 | 0.004 | -0.015 | 0.244 |
| lawseq6 | -0.083 | 0.005 | 0.052 | -0.069 | -0.052 | -0.064 | -0.044 | 0.031 | -0.026 | 0.012 | 0.359 |
| lawseq7 | -0.059 | -0.088 | 0.044 | -0.017 | -0.040 | -0.035 | -0.036 | -0.012 | -0.089 | 0.023 | 0.315 |
| lawseq8 | -0.084 | -0.008 | 0.126 | -0.056 | -0.008 | -0.015 | -0.066 | 0.008 | -0.006 | -0.004 | 0.306 |
| lawseq9 | -0.026 | -0.019 | 0.081 | -0.053 | 0.008 | 0.006 | -0.062 | -0.006 | -0.025 | 0.000 | 0.325 |
| lawseq10 | -0.136 | -0.060 | 0.104 | -0.074 | -0.061 | -0.052 | -0.055 | 0.077 | -0.060 | 0.015 | 0.348 |
| lawseq11 | -0.052 | -0.017 | 0.126 | 0.012 | -0.030 | -0.037 | -0.016 | -0.021 | -0.062 | 0.025 | 0.186 |
| lawseq12 | -0.002 | -0.024 | 0.005 | 0.028 | 0.030 | -0.015 | -0.021 | -0.026 | -0.102 | -0.061 | 0.371 |
| caraloc1 | -0.090 | -0.040 | 0.224 | -0.055 | 0.006 | -0.047 | -0.040 | 0.024 | -0.063 | -0.012 | 0.397 |
| caraloc2 | -0.024 | -0.002 | 0.140 | -0.024 | -0.026 | -0.011 | -0.032 | -0.040 | 0.015 | 0.065 | 0.092 |
| caraloc3 | 0.036 | -0.017 | 0.184 | 0.014 | 0.017 | 0.043 | 0.045 | -0.088 | 0.067 | -0.006 | -0.027 |
| caraloc4 | -0.169 | -0.030 | 0.300 | -0.051 | -0.029 | 0.001 | -0.012 | 0.002 | -0.085 | -0.002 | 0.346 |

continued on next page...

Table 3.B.5 – continued from previous page

| Items | $f_1$ | $f_2$ | $f_3$ | $f_4$ | $f_5$ | $f_6$ | $f_7$ | $f_8$ | $f_9$ | $f_{10}$ | $f_{11}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| caraloc5 | -0.150 | -0.019 | 0.260 | -0.032 | -0.032 | 0.018 | -0.008 | -0.022 | 0.009 | 0.002 | 0.366 |
| caraloc6 | -0.124 | -0.054 | 0.237 | -0.045 | -0.041 | 0.016 | -0.003 | -0.028 | -0.040 | 0.009 | 0.260 |
| caraloc7 | -0.087 | -0.062 | 0.075 | -0.007 | 0.017 | -0.039 | -0.022 | 0.042 | -0.142 | -0.062 | 0.344 |
| caraloc8 | -0.015 | -0.017 | 0.118 | 0.008 | 0.013 | 0.000 | 0.022 | 0.020 | -0.027 | -0.039 | 0.060 |
| caraloc9 | -0.035 | -0.024 | 0.040 | -0.055 | -0.016 | -0.045 | 0.055 | -0.006 | -0.026 | -0.001 | 0.203 |
| caraloc10 | -0.056 | -0.001 | 0.269 | -0.061 | -0.035 | -0.042 | -0.042 | -0.036 | -0.006 | 0.012 | 0.175 |
| caraloc11 | -0.097 | 0.003 | 0.259 | -0.063 | -0.053 | 0.007 | -0.016 | 0.018 | -0.006 | -0.038 | 0.358 |
| caraloc12 | -0.012 | -0.072 | 0.050 | -0.011 | -0.031 | -0.011 | -0.009 | -0.018 | -0.005 | 0.042 | 0.074 |
| caraloc13 | 0.016 | 0.043 | -0.035 | -0.007 | -0.004 | -0.001 | 0.019 | -0.019 | 0.075 | 0.053 | 0.078 |
| caraloc14 | -0.175 | 0.005 | 0.244 | -0.047 | -0.005 | 0.009 | -0.037 | -0.004 | -0.001 | -0.010 | 0.264 |
| caraloc15 | -0.206 | -0.048 | 0.335 | -0.090 | -0.031 | 0.020 | -0.040 | -0.021 | -0.023 | -0.025 | 0.354 |
| caraloc16 | -0.051 | -0.012 | 0.115 | -0.042 | -0.017 | -0.031 | -0.019 | 0.052 | -0.040 | 0.007 | 0.231 |
| cdev1 | 0.615 | -0.009 | -0.074 | 0.229 | 0.142 | -0.031 | 0.048 | -0.044 | 0.113 | -0.011 | -0.064 |
| cdev2 | 0.261 | 0.027 | -0.215 | 0.624 | 0.102 | 0.019 | -0.012 | -0.026 | -0.008 | -0.038 | 0.031 |
| cdev3 | 0.545 | 0.002 | -0.181 | 0.128 | 0.116 | 0.038 | 0.048 | 0.047 | 0.211 | -0.008 | 0.001 |
| cdev4 | -0.016 | 0.003 | -0.017 | 0.078 | 0.182 | -0.011 | -0.024 | -0.062 | 0.059 | 0.074 | -0.058 |
| cdev5 | 0.139 | 0.047 | -0.013 | 0.137 | 0.074 | 0.022 | 0.046 | -0.060 | 0.661 | -0.048 | -0.025 |
| cdev6 | 0.100 | -0.031 | 0.005 | 0.133 | 0.550 | 0.097 | 0.049 | 0.076 | 0.369 | -0.034 | -0.045 |
| cdev7 | -0.295 | 0.007 | 0.189 | -0.087 | -0.395 | -0.017 | -0.066 | -0.031 | -0.071 | -0.009 | -0.014 |
| cdev8 | 0.052 | 0.117 | -0.048 | 0.088 | 0.032 | 0.006 | 0.022 | -0.123 | 0.790 | 0.014 | -0.030 |
| cdev9 | 0.099 | 0.059 | -0.043 | -0.085 | 0.011 | 0.016 | 0.047 | -0.043 | 0.776 | 0.025 | 0.042 |
| cdev10 | 0.227 | -0.015 | 0.000 | 0.277 | 0.584 | 0.107 | 0.032 | -0.138 | 0.199 | -0.091 | -0.032 |
| cdev11 | 0.071 | 0.036 | -0.010 | 0.367 | 0.219 | 0.010 | 0.010 | -0.112 | 0.323 | -0.020 | -0.060 |
| cdev12 | 0.645 | 0.004 | -0.164 | 0.089 | 0.087 | 0.018 | 0.058 | -0.091 | 0.415 | 0.014 | -0.030 |
| cdev13 | -0.711 | -0.045 | 0.199 | -0.064 | -0.052 | 0.004 | -0.107 | 0.030 | -0.303 | -0.022 | 0.035 |
| cdev14 | 0.118 | 0.023 | 0.005 | 0.263 | 0.566 | 0.080 | -0.011 | -0.140 | 0.099 | -0.072 | -0.043 |
| cdev15 | -0.122 | -0.005 | 0.036 | -0.070 | -0.320 | 0.002 | -0.013 | 0.048 | -0.110 | -0.033 | 0.015 |
| cdev16 | 0.322 | 0.048 | -0.060 | -0.064 | 0.054 | 0.018 | 0.089 | 0.018 | 0.745 | 0.007 | 0.000 |
| cdev17 | 0.527 | 0.035 | -0.382 | 0.408 | 0.138 | 0.010 | 0.063 | 0.069 | 0.154 | -0.008 | 0.041 |
| cdev19 | 0.198 | 0.006 | -0.110 | 0.751 | 0.199 | 0.007 | 0.025 | 0.089 | 0.091 | 0.016 | 0.004 |
| cdev20 | -0.008 | -0.012 | 0.069 | 0.522 | 0.091 | 0.054 | 0.012 | 0.047 | 0.286 | -0.106 | -0.005 |
| cdev21 | 0.127 | 0.059 | -0.036 | 0.252 | 0.064 | -0.006 | 0.012 | 0.001 | 0.732 | 0.005 | -0.039 |
| cdev22 | 0.094 | 0.010 | 0.021 | 0.076 | 0.022 | 0.008 | 0.074 | 0.398 | 0.676 | -0.046 | 0.003 |
| cdev24 | 0.238 | 0.009 | -0.001 | 0.145 | 0.073 | 0.003 | 0.091 | 0.381 | 0.704 | 0.011 | -0.001 |
| cdev25 | 0.470 | 0.018 | -0.049 | 0.095 | 0.136 | 0.004 | 0.092 | 0.296 | 0.615 | 0.020 | -0.047 |
| cdev27 | -0.160 | -0.017 | 0.099 | -0.097 | -0.537 | 0.024 | -0.008 | 0.026 | -0.021 | -0.030 | -0.049 |
| cdev28 | 0.190 | -0.006 | -0.051 | 0.157 | 0.581 | 0.010 | 0.011 | 0.094 | 0.311 | 0.036 | -0.015 |
| cdev29 | -0.690 | -0.045 | 0.218 | -0.050 | -0.106 | -0.016 | -0.089 | 0.062 | -0.320 | -0.028 | 0.033 |
| cdev30 | 0.143 | 0.058 | -0.017 | 0.495 | 0.263 | -0.021 | -0.020 | -0.265 | 0.355 | 0.050 | -0.116 |
| cdev31 | 0.062 | 0.022 | -0.005 | 0.443 | 0.198 | -0.018 | 0.009 | -0.033 | 0.375 | -0.024 | -0.042 |
| cdev32 | 0.616 | 0.044 | -0.333 | 0.278 | 0.193 | -0.005 | 0.052 | 0.073 | 0.213 | 0.002 | -0.005 |
| cdev33 | 0.102 | 0.071 | 0.077 | 0.485 | 0.195 | 0.021 | -0.030 | -0.315 | 0.073 | -0.021 | -0.124 |
| cdev34 | 0.210 | 0.078 | -0.082 | 0.044 | 0.041 | 0.005 | 0.035 | -0.119 | 0.791 | 0.015 | -0.025 |
| cdev35 | -0.242 | 0.018 | 0.173 | -0.089 | -0.577 | -0.015 | -0.067 | -0.010 | -0.080 | -0.041 | -0.042 |
| cdev36 | 0.490 | 0.061 | -0.106 | 0.303 | 0.248 | -0.016 | 0.022 | -0.286 | 0.158 | 0.055 | -0.045 |
| cdev37 | 0.139 | 0.055 | -0.050 | 0.022 | 0.208 | 0.002 | 0.084 | -0.088 | 0.652 | 0.117 | -0.025 |
| cdev38 | 0.196 | 0.023 | -0.025 | 0.028 | 0.239 | 0.017 | 0.081 | 0.171 | 0.564 | 0.064 | -0.052 |
| cdev40 | 0.266 | 0.000 | -0.123 | 0.107 | 0.500 | 0.031 | 0.047 | 0.007 | 0.188 | 0.033 | -0.023 |
| cdev42 | 0.152 | -0.015 | -0.049 | 0.095 | 0.404 | 0.021 | 0.018 | 0.079 | 0.386 | 0.028 | -0.047 |
| cdev44 | 0.205 | 0.089 | -0.084 | 0.222 | 0.069 | -0.012 | -0.052 | -0.313 | 0.605 | 0.046 | -0.031 |
| cdev45 | 0.064 | -0.014 | -0.026 | 0.215 | 0.345 | 0.018 | 0.009 | 0.104 | 0.268 | 0.104 | -0.020 |
| cdev46 | 0.080 | 0.055 | -0.087 | 0.060 | 0.193 | -0.047 | -0.034 | -0.093 | 0.229 | 0.145 | -0.017 |

continued on next page...

Table 3.B.5 – continued from previous page

| Items | $f_1$ | $f_2$ | $f_3$ | $f_4$ | $f_5$ | $f_6$ | $f_7$ | $f_8$ | $f_9$ | $f_{10}$ | $f_{11}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| cdev47 | 0.175 | 0.008 | -0.009 | 0.373 | 0.443 | 0.059 | 0.007 | -0.251 | 0.023 | -0.065 | -0.015 |
| cdev48 | -0.687 | -0.025 | 0.169 | -0.085 | -0.157 | -0.009 | -0.079 | 0.039 | -0.244 | -0.044 | 0.008 |
| cdev49 | 0.192 | 0.049 | -0.073 | 0.197 | 0.064 | 0.046 | 0.039 | 0.058 | 0.629 | -0.061 | -0.008 |
| cdev50 | -0.247 | 0.020 | 0.139 | -0.078 | -0.466 | 0.017 | -0.016 | -0.023 | -0.129 | -0.024 | -0.015 |
| cdev51 | 0.697 | 0.044 | -0.140 | 0.107 | 0.117 | 0.014 | 0.072 | -0.035 | 0.269 | 0.052 | -0.047 |
| cdev53 | 0.099 | 0.033 | -0.034 | 0.790 | 0.052 | -0.003 | 0.010 | -0.045 | 0.096 | 0.006 | -0.018 |

**Notes:** The cells display rotated factor loadings obtained from factor analysis of all the measurements. For ease of visualization, the factors and the measurements have been ordered to make them comparable with the PBFA results. Method used for FA: maximum likelihood. Type of rotation performed: orthogonal. White cell if $|\alpha_{jl}| < 0.4$, light gray cell if $0.4 \leqslant |\alpha_{jl}| < 0.7$ and dark gray cell if $|\alpha_{jl}| \geqslant 0.7$. Items: BCS = Cognitive Test Scores, conners = Conners Hyperactivity Scale, rutter = Rutter Parental 'A' Scale of Behavioral Disorder, lawseq = Self-Esteem Scale, caraloc = Locus of Control Scale, cdev = Child Development Scale.

**Interpretation of the latent factors:**

f1 — Can be identified with the trait of Disorganized Activity; however, the loadings on the measurements for the Antisocial Behavior trait are not as high in the traditional factor analysis as in the PBFA.

f2 — Can be identified with the trait of Neuroticism.

f3 — Can be clearly identified with a cognitive factor; notice that, differently from the PBFA, the loadings on the Locus of Control scale are smaller in magnitude in the traditional factor analysis.

f4 — Highly loaded by the measurements for Neuroticism/Anxiety (subscale of the Child Developmental Scale).

f5 — Can be identified with Clumsiness, and, in case of males, also with Poor Hand-Eye Coordination.

f6 — Highly loaded by measurements of the Conners Scale which denote Poor Co-ordination.

f7 — Displays high loadings on measurements of Attention Problems for males, and of Self-Esteem for females.

f8 — Can be identified with the trait of Extraversion/Intraversion, although the loadings are smaller in the traditional factor analysis than in the Bayesian FA.

f9 — Can be identified with the trait of Antisocial Behavior.

f10 — Highly loaded by measurements of Disruptive Behavior for boys, and of Clumsiness for girls.

f11 — Highly loaded by the Self-Esteem scale for males, while it cannot be easily identified with a factor for females.

**Table 3.B.6:** Effect of the factors (SFA) on the outcomes ($p$-values)

| | $f_1$ | $f_2$ | $f_3$ | $f_4$ | $f_5$ | $f_6$ | $f_7$ | $f_8$ | $f_9$ | $f_{10}$ | $f_{11}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Males** | | | | | | | | | | | |
| Education | 0.000 | 0.135 | 0.000 | 0.332 | 0.718 | 0.286 | 0.001 | 0.347 | 0.000 | 0.055 | 0.717 |
| Poor health $d = 0$ | 0.044 | 0.142 | 0.098 | 0.897 | 0.181 | 0.936 | 0.135 | 0.939 | 0.051 | 0.780 | 0.769 |
| Poor health $d = 1$ | 0.103 | 0.108 | 0.986 | 0.639 | 0.479 | 0.367 | 0.272 | 0.027 | 0.199 | 0.244 | 0.438 |
| Obesity $d = 0$ | 0.139 | 0.441 | 0.396 | 0.290 | 0.341 | 0.187 | 0.256 | 0.001 | 0.175 | 0.191 | 0.240 |
| Obesity $d = 1$ | 0.169 | 0.110 | 0.664 | 0.883 | 0.022 | 0.238 | 0.938 | 0.375 | 0.323 | 0.981 | 0.123 |
| Smoking $d = 0$ | 0.025 | 0.404 | 0.901 | 0.102 | 0.658 | 0.016 | 0.522 | 0.444 | 0.000 | 0.005 | 0.006 |
| Smoking $d = 1$ | 0.006 | 0.042 | 0.782 | 0.948 | 0.451 | 0.743 | 0.019 | 0.040 | 0.002 | 0.689 | 0.908 |
| Wages $d = 0$ | 0.048 | 0.046 | 0.000 | 0.001 | 0.000 | 0.769 | 0.112 | 0.141 | 0.016 | 0.516 | 0.292 |
| Wages $d = 1$ | 0.007 | 0.187 | 0.000 | 0.007 | 0.543 | 0.005 | 0.000 | 0.576 | 0.645 | 0.260 | 0.011 |
| **Females** | | | | | | | | | | | |
| | $f_1$ | $f_2$ | $f_3$ | $f_4$ | $f_5$ | $f_6$ | $f_7$ | $f_8$ | $f_9$ | $f_{10}$ | $f_{11}$ |
| Education | 0.000 | 0.025 | 0.000 | 0.256 | 0.195 | 0.920 | 0.002 | 0.862 | 0.071 | 0.824 | 0.364 |
| Poor health $d = 0$ | 0.017 | 0.001 | 0.005 | 0.949 | 0.777 | 0.854 | 0.126 | 0.685 | 0.173 | 0.503 | 0.593 |
| Poor health $d = 1$ | 0.244 | 0.289 | 0.678 | 0.756 | 0.750 | 0.437 | 0.076 | 0.687 | 0.736 | 0.808 | 0.495 |
| Obesity $d = 0$ | 0.010 | 0.248 | 0.069 | 0.260 | 0.012 | 0.165 | 0.065 | 0.894 | 0.137 | 0.024 | 0.004 |
| Obesity $d = 1$ | 0.329 | 0.809 | 0.062 | 0.788 | 0.040 | 0.400 | 0.409 | 0.654 | 0.173 | 0.587 | 0.072 |
| Smoking $d = 0$ | 0.006 | 0.062 | 0.344 | 0.024 | 0.014 | 0.270 | 0.032 | 0.337 | 0.000 | 0.753 | 0.438 |
| Smoking $d = 1$ | 0.000 | 0.430 | 0.003 | 0.702 | 0.154 | 0.743 | 0.062 | 0.457 | 0.017 | 0.766 | 0.416 |
| Wages $d = 0$ | 0.043 | 0.769 | 0.001 | 0.457 | 0.628 | 0.290 | 0.597 | 0.570 | 0.822 | 0.398 | 0.303 |
| Wages $d = 1$ | 0.001 | 0.007 | 0.000 | 0.253 | 0.648 | 0.424 | 0.377 | 0.158 | 0.841 | 0.500 | 0.986 |

**Notes:** Each cell of the table reports the $p$-value that the corresponding loading on the factor is different from 0. The outcome equations include the same covariates as in the PBFA. Standard errors obtained by bootstrapping (100 replications).

### 3.B.2 Items used for the Noncognitive part of the Measurement System

The single items of the scales presented in Section 3.3.2 for the noncognitive part of the measurement system are detailed below.

**Rutter Scale**

rutter1 — Very restless. Often running or jumping up and down. Hardly ever still.

rutter2 — Is squirmy or fidgety.

rutter3 — Often destroys own or others' belongings.

rutter4 — Frequently fights with other children.

rutter5 — Not much liked by other children.

rutter6 — Often worried, worries about many things.

rutter7 — Tends to do things on his/her own, rather solitary.

rutter8 — Irritable. Is quick to 'fly off the handle'.

rutter9 — Often appears miserable, unhappy, tearful or distressed.

rutter10 — Sometimes takes things belonging to others.

rutter11 — Has twitches, mannerisms or tics of the face or body.

rutter12 — Frequently sucks thumb or finger.

rutter13 — Frequently bites nails or fingers.

rutter14 — Is often disobedient.

rutter15 — Cannot settle to do anything for more than a few moments.

rutter16 — Tends to be fearful or afraid of new things or new situation.

rutter17 — Is fussy or over-particular.

rutter18 — Often tells lies.

rutter19 — Bullies other children.

**Conners Hyperactivity Scale**

conners1 — Is noticeably clumsy.

conners2 — Trips or falls easily or bumps into objects or other children.

conners3 — Inattentive, easily distracted.

conners4 — Hums or makes other odd noises at inappropriate times.

conners5 — Has difficulty picking up small objects.

conners6 — Drops things which are being carried.

conners7 — Becomes obsessional about unimportant things.

conners8 — Requests must be met immediately, easily frustrated.

conners9 — Shows restless or over-active behavior.

conners10 — Is impulsive, excitable.

conners11 — Interferes with the activity of other children.

conners12 — Is sullen or sulky.

conners13 — Fails to finish things he/she starts, short attention span.

conners14 — Given to rhythmic tapping or kicking.

conners15 — Cries for little cause.

conners16 — Changes mood quickly and drastically.

conners17 — Displays outbursts of temper, explosive or unpredictable behavior.

conners18 — Has difficulty using scissors.

conners19 — Has difficulty concentrating on any particular task though may return to it frequently.

**Child Developmental Scale**

cdev1 — Is given to daydreaming.

cdev2 — Is fearful or afraid of new things or situations.

cdev3 — Cannot concentrate on any particular task, even though the child may return to it frequently.

cdev4 — Has problems with wetting pants during class.

cdev5 — Complains about things.

cdev6 — Trips or falls easily or bumps into objects or other children.

cdev7 — Works deftly with his or her hands.

cdev8 — Displays outbursts of temper, explosive or unpredictable behaviour.

cdev9 — Teases other children to excess.

cdev10 — Is noticeably clumsy in formal or informal games.

cdev11 — Cries for little cause.

cdev12 — Becomes bored during class.

cdev13 — Shows perseverance; persists with difficult or routine work.

cdev14 — Finds it difficult to kick a ball forward.

cdev15 — Dresses and undresses competently (e.g., for P.E.).

cdev16 — Interferes with the activities of other children.

cdev17 — Becomes confused or hesitant when given a complex task.

cdev18 — Shows difficulty when picking up small objects.

cdev19 — Behaves 'nervously'.

cdev20 — Is fussy or over-particular.

cdev21 — Changes mood quickly and drastically.

cdev22 — Is excitable, impulsive.

cdev23 — Is worried and anxious about many things.

cdev24 — Shows restless or over-active behavior.

cdev25 — Is squirmy and fidgety.

cdev26 — Is easily distracted.

cdev27 — Manipulates small objects easily with his/her hands.

cdev28 — Drops things which are being carried.

cdev29 — Pays attention to what is being explained in class.

cdev30 — In relations with others appears to be miserable, unhappy, tearful or distressed.

cdev31 — Becomes obsessional about unimportant tasks.

cdev32 — Is forgetful when given a complex task.

cdev33 — Tends to do things on his or her own, is rather solitary.

cdev34 — Quarrels with other children.

cdev35 — Can use scissors and similar manipulative equipment competently.

cdev36 — Shows lethargic and listless behavior.

cdev37 — Destroys own or other children's belongings.

cdev38 — Hums or makes other odd vocal noises at inappropriate times.

cdev39 — Given to rhythmic tapping or rhythmic kicking during class.

cdev40 — Shows inadequate control when handling a pencil or paint brush.

cdev41 — Has problems of soiling pants during class.

cdev42 — Experiences classroom or playground accidents.

cdev43 — Bullies other children.

cdev44 — Is sullen or sulky.

cdev45 — Has twitches, mannerisms or tics of the face or body.

cdev46 — Truants from school.

cdev47 — Fearful in movements, requires much encouragement to move faster.

cdev48 — Completes tasks which are started.

cdev49 — Request must be satisfied immediately - is easily frustrated.

cdev50 — Holds writing and drawing instruments appropriately.

cdev51 — Fails to finish things he starts

Please use your knowledge of the study child to assess his/her disposition or temperament:

cdev52 — Introvert/Extrovert

cdev53 — Anxious/Unworried

**Locus of Control (CARALOC) Scale**

caraloc1 — Do you feel that most of the time it's not worth trying hard because things never turn out right anyway?

caraloc2 — Do you feel that wishing can make good things happen?

caraloc3 — Are people good to you no matter how you act towards them?

caraloc4 — Do you usually feel that it's almost useless to try in school because most children are cleverer than you?

caraloc5 — Is a high mark just a matter of "luck" for you?

caraloc6 — Are tests just a lot of guesswork for you?

caraloc7 — Are you often blamed for things which just aren't your fault?

caraloc8 — Are you the kind of person who believes that planning ahead makes things turn out better?

caraloc9 — When bad things happen to you, is it usually someone else's fault?

caraloc10 — When someone is very angry with you, is it impossible to make him your friend again?

caraloc11 — When nice things happen to you is it only good luck?

caraloc12 — Do you feel sad when it is time to leave school each day?

caraloc13 — When you get into an argument is it usually the other person's fault?

caraloc14 — Are you surprised when your teacher says you've done well?

caraloc15 — Do you usually get low marks, even when you study hard?

caraloc16 — Do you think studying for tests is a waste of time?

**Lawrence Self-Esteem (LAWSEQ) Scale**

lawseq1 — Do you think that your parents usually like to hear about your ideas?

lawseq2 — Do you often feel lonely at school?

lawseq3 — Do other children often break friends or fall out with you?

lawseq4 — Do you think that other children often say nasty things about you?

lawseq5 — When you have to say things in front of teachers, do you usually feel shy?

lawseq6 — Do you often feel sad because you have nobody to play with at school?

lawseq7 — Are there lots of things about yourself you would like to change?

lawseq8 — When you have to say things in front of other children, do you usually feel foolish?

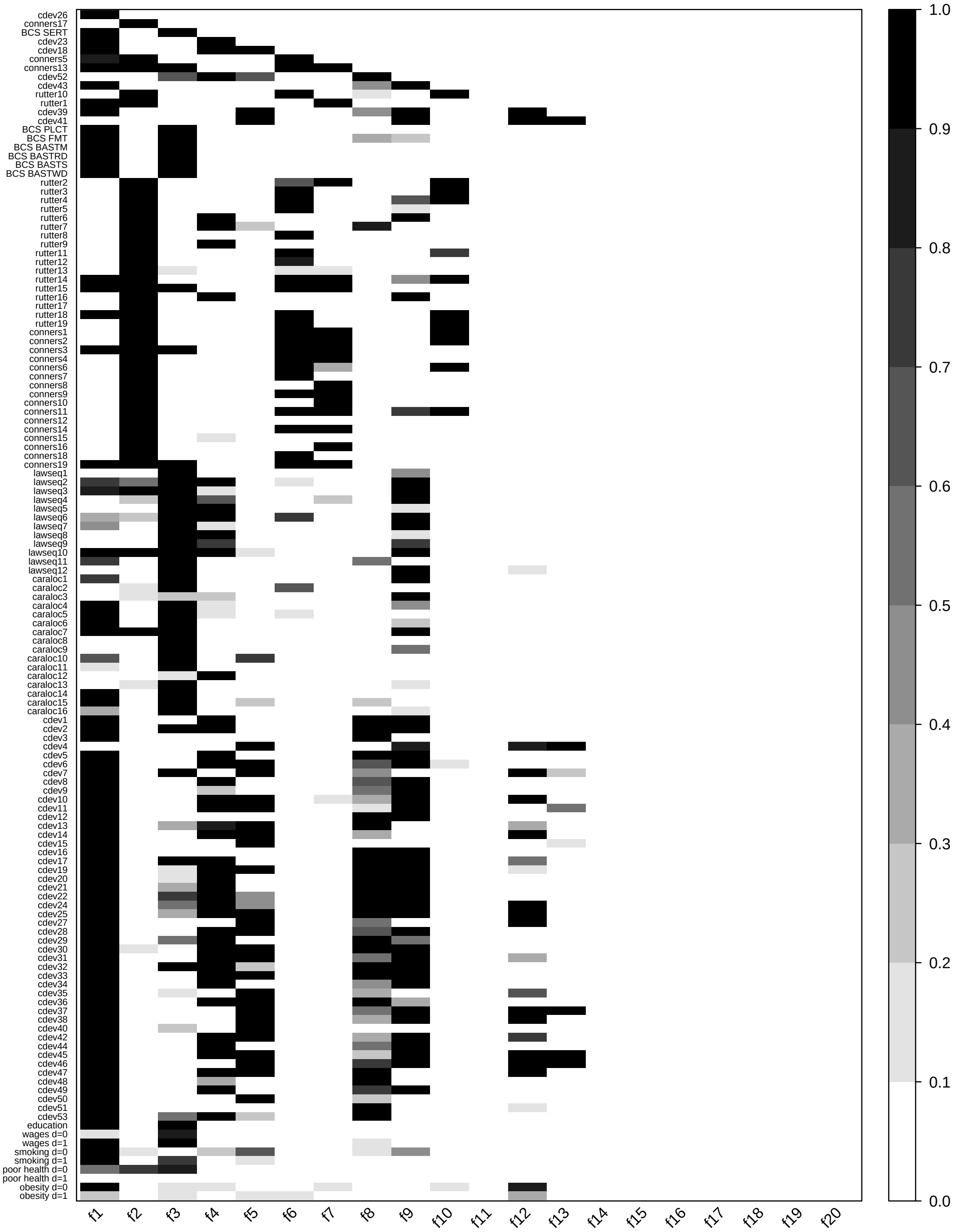
lawseq9 — When you want to tell a teacher something, do you usually feel foolish?

lawseq10 — Do you often have to find new friends because your old friends are playing with somebody else?

lawseq11 — Do you usually feel foolish when you talk to your parents?

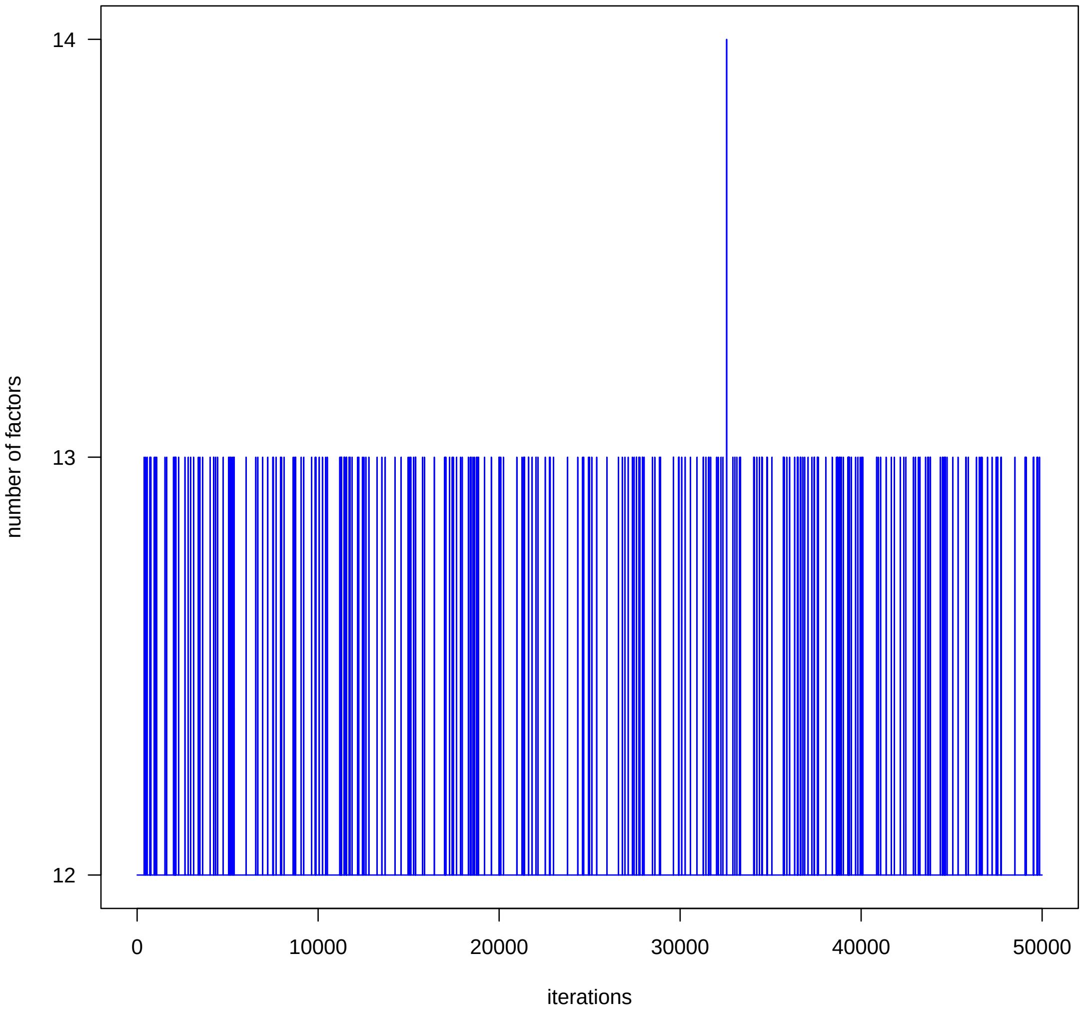
lawseq12 — Do other people often think that you tell lies?

**Figure 3.B.1:** Factor loadings posterior probabilities, 20-factor model (Females)



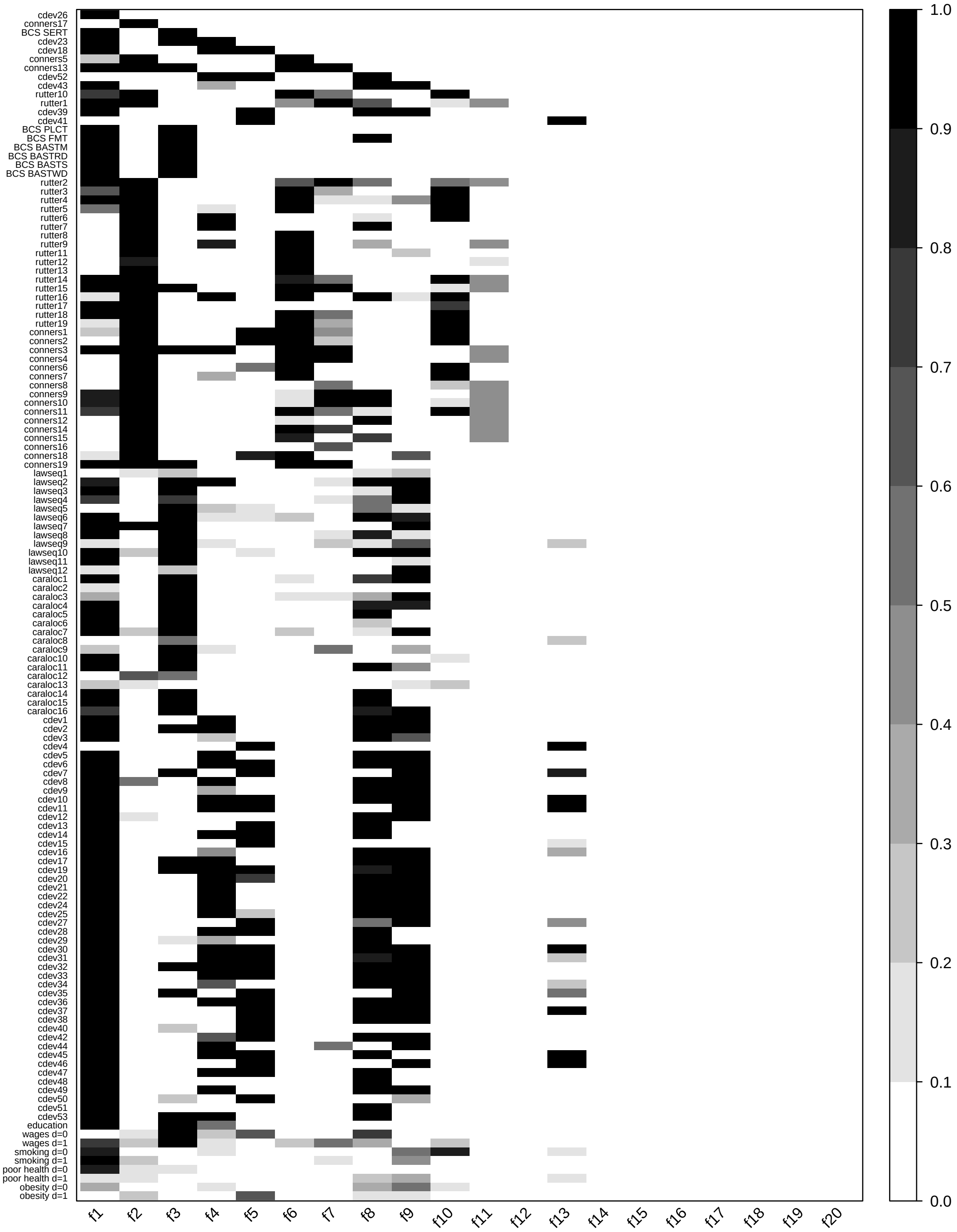

**Notes:** Each column represents a latent factor and shows the items/outcomes that significantly load on it (dark cells). Items: BCS = Cognitive Test Scores, conners = Conners Hyperactivity Scale, rutter = Rutter Parental 'A' Scale of Behavioral Disorder, lawseq = Self-Esteem Scale, caraloc = Locus of Control Scale, cdev = Child Development Scale. Education outcome, wages and health outcomes by levels of education at the bottom of the picture.

**Figure 3.B.2:** Posterior draws of the number of factors, 20-factor model (Females)



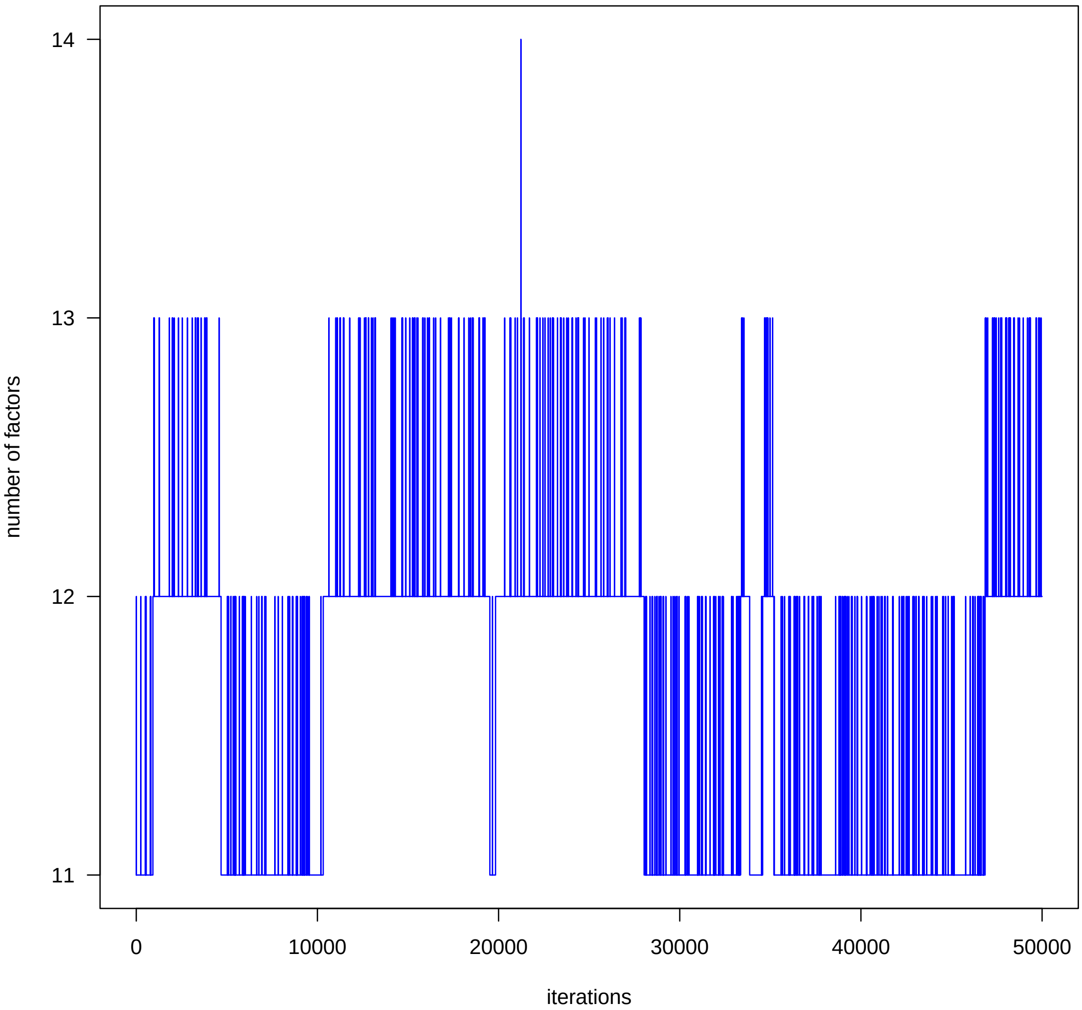

**Notes:** Total number of MCMC replications: 520,000. After a burn-in period of 20,000 iterations, 50,000 iterations were saved every 10 steps and are presented here. Parsimonious factor analysis started after 10,000 iterations. Only models containing between 12 and 14 latent factors were visited by the algorithm.

**Figure 3.B.3:** Factor loadings posterior probabilities, 20-factor model (Males)




**Notes:** Each column represents a latent factor and shows the items/outcomes that significantly load on it (dark cells). Items: BCS = Cognitive Test Scores, conners = Conners Hyperactivity Scale, rutter = Rutter Parental 'A' Scale of Behavioral Disorder, lawseq = Self-Esteem Scale, caraloc = Locus of Control Scale, cdev = Child Development Scale. Education outcome, wages and health outcomes by levels of education at the bottom of the picture.

**Figure 3.B.4:** Posterior draws of the number of factors, 20-factor model (Males)




**Notes:** Total number of MCMC replications: 520,000. After a burn-in period of 20,000 iterations, 50,000 iterations were saved every 10 steps and are presented here. Parsimonious factor analysis started after 10,000 iterations. Only models containing between 11 and 14 latent factors were visited by the algorithm.

# Complete Bibliography

Aakvik, A., J. J. Heckman, and E. J. Vytlacil (2005): "Estimating Treatment Effects for Discrete Outcomes When Responses to Treatment Vary: An Application to Norwegian Vocational Rehabilitation Programs," *Journal of Econometrics*, 125, 15–51.

Albert, J. H. and S. Chib (1993): "Bayesian Analysis of Binary and Polychotomous Response Data," *Journal of the American Statistical Association*, 88, 669–679.

Andrisani, P. J. (1977): "Internal-External Attitudes, Personal Initiative, and the Labor Market Experience of Black and White Men," *Journal of Human Resources*, 12, 308–328.

——— (1981): "Internal-External Attitudes, Sense of Efficacy, and Labor Market Experience: A Reply to Duncan and Morgan," *Journal of Human Resources*, 16, 658–66.

Auld, M. C. and N. Sidhu (2005): "Schooling, Cognitive Ability and Health," *Health Economics*, 14, 1019–1034.

Bartholomew, D. J. (1987): *Latent Variable Models and Factors Analysis*, London: Charles Griffin.

Batty, G. D., I. J. Deary, I. Schoon, and C. R. Gale (2007): "Mental Ability across Childhood in Relation to Risk Factors for Premature Mortality in Adult Life: The 1970 British Cohort Study," *Journal of Epidemiology and Community Health*, 61, 997–1003.

Becker, G. S. (1964): *Human Capital: A Theoretical and Empirical Analysis, With Special Reference to Education*, New York: National Bureau of Economic Research.

Blanden, J., P. Gregg, and L. Macmillan (2007): "Accounting for Intergenerational Income Persistence: Noncognitive Skills, Ability and Education," *Economic Journal*, 117, C43–C60.

Borghans, L., A. L. Duckworth, J. J. Heckman, and B. ter Weel (2008): "The Economics and Psychology of Personality Traits," *Journal of Human Resources*, 43, 972–1059.

Bowles, S. and H. Gintis (1976): *Schooling in Capitalist America: Educational Reform and the Contradictions of Economic Life*, Routledge.

——— (2002): "Schooling in Capitalist America Revisited," *Sociology of Education*, 75, 1–18.

Bowles, S., H. Gintis, and M. A. Osborne (2001a): "Incentive-Enhancing Preferences: Personality, Behavior, and Earnings," *American Economic Review*, 91, 155–158.

——— (2001b): "The Determinants of Earnings: A Behavioral Approach," *Journal of Economic Literature*, 39, 1137–1176.

Butler, N., S. Despotidou, and P. Shepherd (1997): "British Cohort Study (BCS70) Ten-year Follow-up: A Guide to the BCS70 10-year Data Available at the Economic and Social Research Unit Data Archive," Tech. rep., London: Social Statistics Research Unit, City University.

Caliendo, M., D. A. Cobb-Clark, and A. Uhlendorff (2010): "Locus of Control and Job Search Strategies," Institute for the Study of Labor (IZA) Discussion Paper No. 4750.

Carneiro, P., K. Hansen, and J. J. Heckman (2003): "Estimating Distributions of Treatment Effects with an Application to the Returns to Schooling and Measurement of the Effects of Uncertainty on College Choice," *International Economic Review*, 44, 361–422.

Carvalho, C. M. (2006): "Structure and Sparsity in High-Dimensional Multivariate Analysis," Ph.D. thesis, Duke University.

Carvalho, C. M., J. Chang, J. E. Lucas, J. R. Nevins, Q. Wang, and M. West (2008): "High-Dimensional Sparse Factor Modeling: Applications in Gene Expression Genomics," *Journal of the American Statistical Association*, 103, 1438–1456.

CASELLA, G. AND E. I. GEORGE (1992): "Explaining the Gibbs Sampler," *The American Statistician*, 46, 167–174.

CATTELL, R. B. (1966): "The Scree Test for the Number of Factors," *Multivariate Behavioral Research*, 1, 245–276.

CEBI, M. (2007): "Locus of Control and Human Capital Investment Revisited," *Journal of Human Resources*, 42, 919–932.

CHEN, M. H. AND D. K. DEY (2000): "Bayesian Analysis for Correlated Ordinal Data Models," in *Generalized Linear Models: A Bayesian Perspective*, ed. by D. Dey, S. Ghosh, and B. Mallick, CRC Press, 133–157.

CHIB, S. (1992): "Bayes Inference in the Tobit Censored Regression Model," *Journal of Econometrics*, 51, 79–99.

CHIB, S. AND B. H. HAMILTON (2000): "Bayesian Analysis of Cross-Section and Clustered Data Treatment Models," *Journal of Econometrics*, 97, 25–50.

——— (2002): "Semiparametric Bayes Analysis of Longitudinal Data Treatment Models," *Journal of Econometrics*, 110, 67–89.

COLEMAN, J. S. (1968): "Equality of Educational Opportunity," *Equity & Excellence in Education*, 6, 19–28.

COLEMAN, M. AND T. DELEIRE (2003): "An Economic Model of Locus of Control and the Human Capital Investment Decision," *Journal of Human Resources*, 38, 701–721.

CONNERS, C. K. (1969): "A Teacher Rating Scale for Use in Drug Studies with Children," *American Journal of Psychiatry*, 126, 884–888.

CONTI, G., J. J. HECKMAN, AND S. URZUA (2010): "The Education-Health Gradient," *American Economic Review Papers and Proceedings*, 100, 234–238.

CONTI, G. AND S. PUDNEY (2009): "The Dynamics of Cognitive Development," Unpublished Manuscript, University of Essex, Institute for Social and Economic Research.

COWLES, M. K. (1996): "Accelerating Monte Carlo Markov Chain Convergence for Cumulative-Link Generalized Linear Models," *Statistics and Computing*, 6, 101–111.

COX, D. R. (1958): *Planning of Experiments*, New York: Wiley.

Cunha, F. and J. J. Heckman (2007): "The Technology of Skill Formation," *American Economic Review*, 97, 31–47.

——— (2008): "Formulating, Identifying and Estimating the Technology of Cognitive and Noncognitive Skill Formation," *Journal of Human Resources*, 43, 738–782.

Cunha, F., J. J. Heckman, and S. Navarro (2005): "Separating Uncertainty from Heterogeneity in Life Cycle Earnings," *Oxford Economic Papers*, 57, 191–261, The 2004 Hicks Lecture.

Cunha, F., J. J. Heckman, and S. M. Schennach (2010): "Estimating the Technology of Cognitive and Noncognitive Skill Formation," *Econometrica*, 78, 883–931.

Currie, J. and E. Moretti (2003): "Mother's Education and the Intergenerational Transmission of Human Capital: Evidence from College Openings," *Quarterly Journal of Economics*, 118, 1495–1532.

Cutler, D. M. and A. Lleras-Muney (2010): "Understanding Differences in Health Behaviors by Education," *Journal of Health Economics*, 29, 1–28.

Diebolt, J. and C. P. Robert (1994): "Estimation of Finite Mixture Distributions Through Bayesian Sampling," *Journal of the Royal Statistical Society Series B: Methodological*, 56, 363–375.

Digman, J. M. (1990): "Personality Structure: Emergence of the Five-Factor Model," *Annual Review of Psychology*, 41, 417–440.

Dohmen, T. J., A. Falk, D. Huffman, and U. Sunde (2007): "Are Risk Aversion and Impatience Related to Cognitive Ability?" Institute for the Study of Labor (IZA) Discussion Paper No. 2735.

Duncan, G. and J. Morgan (1981): "Sense of Efficacy and Subsequent Change in Earnings–A Replication," *Journal of Human Resources*, 16, 649–657.

Duncan, G. J. and R. Dunifon (1998): "Soft-Skills and Long-Run Labor Market Success," *Research in Labor Economics*, 17, 123–149.

Dyk, D. A. v. and X.-L. Meng (2001): "The Art of Data Augmentation," *Journal of Computational and Graphical Statistics*, 10, 1–50.

Escobar, M. D. and M. West (1995): "Bayesian Density Estimation and Inference Using Mixtures," *Journal of the American Statistical Association*, 90, 577–588.

Fahrmeir, L. and A. W. Raach (2007): "A Bayesian Semiparametric Latent Variable Model for Mixed Responses," *Psychometrika*, 72, 327–346.

Falk, A., D. Huffman, and U. Sunde (2006): "Self-Confidence and Search," Institute for the Study of Labor (IZA) Discussion Paper No. 2525.

Feinstein, L. (2000): "The Relative Economic Importance of Academic, Psychological and Behavioural Attributes Developed on Chilhood," Discussion Paper, Centre for Economic Performance, London.

Ferguson, T. S. (1983): "Bayesian Density Estimation by Mixtures of Normal Distributions," in *Recent Advances in Statistics: Papers in Honor of Herman Chernoff on his Sixtieth Birthday*, ed. by H. Chernoff, M. Rizvi, J. Rustagi, and D. Siegmund, New York: Academic Press, 287–302.

Fisher, R. A. (1935): *The Design of Experiments*, London: Oliver and Boyd.

Flossmann, A. L., R. Piatek, and L. Wichert (2007): "Going Beyond Returns to Education: The Effect of Noncognitive Skills on Wages in Germany," Paper presented at the European Meeting of the Econometric Society, Budapest.

Frey, B. S. and S. Meier (2004): "Pro-Social Behavior, Reciprocity or Both?" *Journal of Economic Behavior and Organization*, 54, 65–68.

Frühwirth-Schnatter, S. and H. F. Lopes (2009): "Parsimonious Bayesian Factor Analysis when the Number of Factors is Unknown," Unpublished Technical Report.

Frühwirth-Schnatter, S. and R. Tüchler (2008): "Bayesian Parsimonious Covariance Estimation for Hierarchical Linear Mixed Models," *Statistics and Computing*, 18, 1–13.

Gale, C. R., G. D. Batty, and I. J. Deary (2008): "Locus of Control at Age 10 Years and Health Outcomes and Behaviors at Age 30 Years: the 1970 British Cohort Study," *Psychosomatic Medicine*, 70, 397–403.

Gale, C. R., S. L. Hatch, G. D. Batty, and I. J. Deary (2009): "Intelligence in Childhood and Risk of Psychological Distress in Adulthood: The 1958 National

Child Development Survey and the 1970 British Cohort Study," *Intelligence*, 37, 592–599.

GAMMAGE, P. (1975): "Socialization, Schooling and Locus of Control," Ph.D. thesis, University of Bristol.

GAREN, J. (1984): "The Returns to Schooling: A Selectivity Bias Approach with a Continuous Choice Variable," *Econometrica*, 52, 1199–1218.

GELMAN, A. (2006): "Prior Distributions for Variance Parameters in Hierarchical Models," *Bayesian Analysis*, 1, 515–533.

GEWEKE, J. AND G. ZHOU (1996): "Measuring the Price of the Arbitrage Pricing Theory," *The Review of Financial Studies*, 9, 557–587.

GEWEKE, J. F. AND M. P. KEANE (1999): "Mixture of Normals Probit Models," in *Analysis of Panels and Limited Dependent Variables: A Volume in Honor of G. S. Maddala*, ed. by C. Hsiao, G. S. Maddala, K. Lahiri, and L. F. Lee, Cambridge University Press, 49–78.

GEWEKE, J. F. AND K. J. SINGLETON (1980): "Interpreting the Likelihood Ratio Statistic in Factor Models when Sample Size is Small," *Journal of the American Statistical Association*, 75, 133–137.

GHOSH, J. AND D. B. DUNSON (2009): "Default Prior Distributions and Efficient Posterior Computation in Bayesian Factor Analysis," *Journal of Computational and Graphical Statistics*, 18, 306–320.

GOLDBERG, L. R. (1990): "An Alternative 'Description of Personality': The Big-Five Factor Structure," *Journal of Personality and Social Psychology*, 59, 1216–1229.

GOTTFREDSON, L. S. AND I. J. DEARY (2004): "Intelligence Predicts Health and Longevity, but Why?" *Current Directions in Psychological Science*, 13, 1–4.

GREEN, P. J. (1995): "Reversible Jump Markov Chain Monte Carlo Computation and Bayesian Model Determination," *Biometrika*, 82, 711–732.

GRILICHES, Z. (1977): "Estimating the Returns to Schooling: Some Econometric Problems," *Econometrica*, 45, 1–22.

GROSSMAN, M. (1975): "The Correlation Between Health and Schooling," in *Household Production and Consumption*, ed. by N. E. Terleckyj, New York: Columbia University Press, 147–211.

——— (2006): "Education and Nonmarket Outcomes," in *Handbook of the Economics of Education*, ed. by E. Hanushek and F. Welch, Amsterdam: Elsevier, vol. 1, chap. 10, 577–633.

Hakstian, A. R., W. T. Rogers, and R. B. Cattell (1982): "The Behavior of Number-of-Factors Rules with Simulated Data," *Multivariate Behavioral Research*, 17, 193–219.

Hampson, S. E. and H. S. Friedman (2008): "Personality and Health: A Lifespan Perspective," in *The Handbook of Personality: Theory and Research*, ed. by O. P. John, R. Robins, and L. Pervin, New York: Guilford, 770–794, third ed.

Hansen, K. T., J. J. Heckman, and K. J. Mullen (2004): "The Effect of Schooling and Ability on Achievement Test Scores," *Journal of Econometrics*, 121, 39–98.

Hartog, J. and H. Oosterbeek (1998): "Health, Wealth and Happiness: Why Pursue a Higher Education?" *Economics of Education Review*, 17, 245–256.

Heckman, J. J. (1979): "Sample Selection Bias as a Specification Error," *Econometrica*, 47, 153–161.

——— (1997): "Instrumental Variables: A Study of Implicit Behavioral Assumptions Used in Making Program Evaluations," *Journal of Human Resources*, 32, 441–462, addendum published vol. 33 no. 1 (Winter 1998).

Heckman, J. J., R. J. LaLonde, and J. A. Smith (1999): "The Economics and Econometrics of Active Labor Market Programs," in *Handbook of Labor Economics*, ed. by O. Ashenfelter and D. Card, New York: North-Holland, vol. 3A, chap. 31, 1865–2097.

Heckman, J. J. and J. A. Smith (1998): "Evaluating the Welfare State," in *Econometrics and Economic Theory in the Twentieth Century: The Ragnar Frisch Centennial Symposium*, ed. by S. Strom, New York: Cambridge University Press, 241–318.

Heckman, J. J., J. Stixrud, and S. Urzua (2006): "The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior," *Journal of Labor Economics*, 24, 411–482.

Heckman, J. J. and E. J. Vytlacil (1999): "Local Instrumental Variables and Latent Variable Models for Identifying and Bounding Treatment Effects," *Proceedings of the National Academy of Sciences*, 96, 4730–4734.

——— (2000): "The Relationship Between Treatment Parameters Within a Latent Variable Framework," *Economics Letters*, 66, 33–39.

——— (2001): "Local Instrumental Variables," in *Nonlinear Statistical Modeling: Proceedings of the Thirteenth International Symposium in Economic Theory and Econometrics: Essays in Honor of Takeshi Amemiya*, ed. by C. Hsiao, K. Morimune, and J. L. Powell, New York: Cambridge University Press, 1–46.

——— (2007): "Econometric Evaluation of Social Programs, Part II: Using the Marginal Treatment Effect to Organize Alternative Economic Estimators to Evaluate Social Programs and to Forecast Their Effects in New Environments," in *Handbook of Econometrics*, ed. by J. Heckman and E. Leamer, Amsterdam: Elsevier, vol. 6B, 4875–5144.

Heineck, G. and S. Anger (2010): "The Returns to Cognitive Abilities and Personality Traits in Germany," *Labour Economics*, 17, 535–546.

Hill, M. S. (1979): "The Wage Effects of Marital Status and Children," *Journal of Human Resources*, 14, 579–594.

Jöreskog, K. G. and I. Moustaki (2001): "Factor Analysis of Ordinal Variables: A Comparison of Three Approaches," *Multivariate Behavioral Research*, 36, 347–387.

Judge, T. A. and J. E. Bono (2001): "Relationship of Core Self-Evaluations Traits-Self-Esteem, Generalized Self-Efficacy, Locus of Control, and Emotional Stability-with Job Satisfaction and Job Performance: A Meta-Analysis," *Journal of Applied Psychology*, 86, 80–92.

Kaestner, R. (2009): "Adolescent Cognitive and Non-cognitive Correlates of Adult Health," National Bureau of Economic Research (NBER) Working Paper No. 14924.

Kaiser, H. F. (1960): "The Application of Electronic Computers to Factor Analysis," *Educational and Psychological Measurement*, 20, 141–151.

Kotlarski, I. (1967): "On Characterizing the Gamma and the Normal Distribution," *Pacific Journal of Mathematics*, 20, 69–76.

Lawrence, D. (1973): *Improved Reading through Counselling*, London: Ward Lock Educational.

——— (1978): *Counselling Students with Reading Difficulties: A Handbook for Tutors and Organisors*, Good Reading Limited.

Lee, L. F. (1992): "On Efficiency of Methods of Simulated Moments and Maximum Simulated Likelihood Estimation of Discrete Response Models," *Econometric Theory*, 8, 518–552.

Li, M., D. J. Poirier, and J. L. Tobias (2004): "Do Dropouts Suffer from Dropping out? Estimation and Prediction of Outcome Gains in Generalized Selection Models," *Journal of Applied Econometrics*, 19, 203–225.

Li, M. and J. L. Tobias (2007): "Bayesian Analysis of Treatment Effects in an Ordered Potential Outcomes Model," *Advances in Econometrics*, 21, 57–91.

Lindeboom, M., A. Llena-Nozal, and B. van der Klaauw (2009): "Parental Education and Child Health: Evidence from a Schooling Reform," *Journal of Health Economics*, 28, 109–131.

Linn, R. L. (1968): "A Monte Carlo Approach to the Number of Factors Problem," *Psychometrika*, 33, 37–71.

Liu, C., D. B. Rubin, and Y. N. Wu (1998): "Parameter Expansion to Accelerate EM: the PX-EM Algorithm," *Biometrika*, 85, 755–770.

Liu, J. S. and C. Sabatti (2000): "Generalised Gibbs Sampler and Multigrid Monte Carlo for Bayesian Computation," *Biometrika*, 87, 353–369.

Liu, J. S., W. H. Wong, and A. Kong (1994): "Covariance Structure of the Gibbs Sampler with Applications to the Comparisons of Estimators and Augmentation Schemes," *Biometrika*, 81, 27–40.

Liu, J. S. and Y. N. Wu (1999): "Parameter Expansion for Data Augmentation," *Journal of the American Statistical Association*, 94, 1264–1274.

Liu, X. (2008): "Parameter Expansion for Sampling a Correlation Matrix: An Efficient GPX-RPMH Algorithm," *Journal of Statistical Computation and Simulation*, 78, 1065–1076.

Lopes, H. F. and C. M. Carvalho (2007): "Factor Stochastic Volatility with Time Varying Loadings and Markov Switching Regimes," *Journal of Statistical Planning and Inference*, 137, 3082–3091.

LOPES, H. F., E. SALAZAR, AND D. GAMERMAN (2008): "Spatial Dynamic Factor Models," *Bayesian Analysis*, 3, 759–992.

LOPES, H. F. AND M. WEST (2004): "Bayesian Model Assessment in Factor Analysis," *Statistica Sinica*, 14, 41–67.

MCCLOSKEY, D. N. AND S. T. ZILIAK (1996): "The Standard Error of Regressions," *Journal of Economic Literature*, 34, 97–114.

MCCRAE, R. R. AND O. P. JOHN (1992): "An Introduction to the Five-Factor Model and its Applications." *Journal of personality*, 60, 175–215.

MCCRARY, J. AND H. ROYER (2010): "The Effect of Female Education on Fertility and Infant Health: Evidence from School Entry Policies Using Exact Date of Birth," *American Economic Review*, forthcoming.

MINCER, J. (1958): "Investment in Human Capital and Personal Income Distribution," *Journal of Political Economy*, 66, 281–302.

MUELLER, G. AND E. PLUG (2006): "Estimating the Effect of Personality on Male and Female Earnings," *Industrial and Labor Relations Review*, 60, 3–22.

MURASKO, J. E. (2007): "A Lifecourse Study on Education and Health: The Relationship between Childhood Psychosocial Resources and Outcomes in Adolescence and Young Adulthood," *Social Science Research*, 36, 1348–1370.

NANDRAM, B. AND M.-H. CHEN (1996): "Reparameterizing the Generalized Linear Model to Accelerate Gibbs Sampler Convergence," *Journal of Statistical Computation and Simulation*, 54, 129–144.

NEYMAN, J. (1923): "Statistical Problems in Agricultural Experiments," *Journal of the Royal Statistical Society*, 2 (Supplement), 107–180.

NYHUS, E. K. AND E. PONS (2005): "The Effects of Personality on Earnings," *Journal of Economic Psychology*, 26, 363–384.

O'HAGAN, A. (1995): "Fractional Bayes Factors for Model Comparison," *Journal of the Royal Statistical Society. Series B (Methodological)*, 57, 99–138.

OSBORNE, M. A. (2000): "The Power of Personality: Labor Market Rewards and the Transmission of Earnings," Ph.D. thesis, University of Massachusetts.

RAACH, A. W. (2006): "A Bayesian Semiparametric Latent Variable Model for Binary, Ordinal and Continuous Response," Ph.D. thesis, Ludwig-Maximilians-Universität München.

RICHARDSON, S. AND P. J. GREEN (1997): "On Bayesian Analysis of Mixtures with an Unknown Number of Components," *Journal of the Royal Statistical Society. Series B (Methodological)*, 59, 731–792.

ROBERTS, B. W., P. HARMS, J. L. SMITH, D. WOOD, AND M. WEBB (2006): "Using Multiple Methods in Personality Psychology," in *Handbook of Multimethod Measurement in Psychology*, ed. by M. Eid and E. Diener, Washington, D.C.: American Psychological Association, 321–335.

ROBERTS, B. W., N. R. KUNCEL, R. L. SHINER, A. CASPI, AND L. R. GOLDBERG (2007): "The Power of Personality: The Comparative Validity of Personality Traits, Socioeconomic Status, and Cognitive Ability for Predicting Important Life Outcomes," *Perspectives in Psychological Science*, 2, 313–345.

ROEDER, K. AND L. WASSERMAN (1997): "Practical Bayesian Density Estimation Using Mixtures of Normals," *Journal of the American Statistical Association*, 92, 894–902.

ROSENBERG, M. (1965): *Society and the Adolescent Self-Image*, Princeton, NJ: Princeton University Press.

ROSETT, R. N. AND F. D. NELSON (1975): "Estimation of the Two-limit Probit Regression Model," *Econometrica*, 43, 141–146.

ROTTER, J. B. (1966): "Generalized Expectancies for Internal versus External Control of Reinforcement," *Psychological Monographs: General & Applied*, 80, 1–28.

RUBIN, D. B. (1974): "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies," *Journal of Educational Psychology*, 66, 688–701.

RUTTER, M. (1967): "A Children's Behaviour Questionnaire for Completion by Teachers: Preliminary Findings," *Journal of Child Psychology and Psychiatry*, 8, 1–11.

RUTTER, M., J. TIZARD, AND K. WHITMORE (1970): *Education, health and behaviour*, London, UK: Longmans.

SCOTT, D. W. (1992): *Multivariate Density Estimation: Theory, Practice, and Visualization*, Wiley Interscience.

SHANNON, C. E. (1948): *The Mathematical Theory of Communication*, vol. 27, Bell System Tech. J.

SILVERMAN, B. W. (1986): *Density Estimation for Statistics and Data Analysis*, CRC Press.

SMITH, M. AND R. KOHN (2002): "Parsimonious Covariance Matrix Estimation for Longitudinal Data," *Journal of the American Statistical Association*, 97, 1141–1153.

SPEARMAN, C. (1904): "'General Intelligence,' Objectively Determined and Measured," *American Journal of Psychology*, 15, 201–293.

STUMM, S. V., C. R. GALE, G. D. BATTYA, AND I. J. DEARY (2009): "Childhood Intelligence, Locus of Control and Behaviour Disturbance as Determinants of Intergenerational Social Mobility: British Cohort Study 1970," *Intelligence*, 37, 329–340.

TANNER, M. A. AND W. H. WONG (1987): "The Calculation of Posterior Distributions by Data Augmentation," *Journal of the American Statistical Association*, 82, 528–540.

TRZCINSKI, E. AND E. HOLST (2010): "Interrelationship among Locus of Control and Years in Management and Unemployment: Differences by Gender," Discussion Papers of DIW Berlin 974, DIW Berlin, German Institute for Economic Research.

TÜCHLER, R. (2008): "Bayesian Variable Selection for Logistic Models Using Auxiliary Mixture Sampling," *Journal of Computational and Graphical Statistics*, 17, 76–94.

TUCKER, L. R., R. F. KOOPMAN, AND R. L. LINN (1969): "Evaluation of Factor Analytic Research Procedures by Means of Simulated Correlation Matrices," *Psychometrika*, 34, 421–459.

VAN DYK, D. A. AND X.-L. MENG (2001): "The Art of Data Augmentation," *Journal of Computational & Graphical Statistics*, 10, 1–50.

VELICER, W. F. (1976): "Determining the Number of Components from the Matrix of Partial Correlations," *Psychometrika*, 41, 321–327.

West, M. (2003): "Bayesian Factor Regression Models in the 'Large p, Small n' Paradigm," *Bayesian Statistics*, 7, 723–732.

Whalley, L. J. and I. J. Deary (2001): "Longitudinal Cohort Study of Childhood IQ and Survival up to Age 76," *British Medical Journal*, 322, 819–822.

Wooldridge, J. M. (2003): "Further Results on Instrumental Variables Estimation of Average Treatment Effects in the Correlated Random Coefficient Model," *Economics Letters*, 79, 185–191.

Ziliak, S. T. and D. N. McCloskey (2004): "Size Matters: the Standard Error of Regressions in the American Economic Review," *Journal of Socio-Economics*, 33, 527 – 546.