Personality Traits and Economic Outcomes

–

Models and Measurement, with Two Empirical Applications

Der Wirtschaftswissenschaftlichen Fakultät der
Gottfried Wilhelm Leibniz Universität Hannover
zur Erlangung des akademischen Grades

Doktor der Wirtschaftswissenschaften
– Doctor rerum politicarum –

genehmigte Dissertation
von

Diplom Volkswirt Hendrik Thiel
geboren am 09.05.1982 in Bernburg (Saale)

2015

Referent: Prof. Dr. Stephan L. Thomsen

Korreferent: Prof. Dr. Daniel Schnitzlein

Tag der Promotion: 21. Juli 2015

Acknowledgments

The accomplishment of this dissertation has been one of the more lasting experiences in my life thus far, but one that has also been marked with overwhelming help and support of many persons, to whom I would like to express my deepest gratitude.

First and foremost, I am indebted to my thesis advisor Prof. Dr. Stephan Thomsen. I owe this opportunity to him who offered me a position at his chair of Labor Economics back in due course as a professor at the University of Magdeburg. He also gave me the chance to follow him to the Lower-Saxony Institute for Economic Research (NIW) when he seized the opportunity to take a professorship at the University of Hanover. While having always provided a stimulating and inspiring general research environment, Prof. Thomsen granted me his open-mindedness, his constant enthusiasm, his professional advice, and his unrestricted support in all my undertakings. He always facilitated my personal development by conceding me the freedom to comprehend new scopes and at the same time guided me whenever necessary. Secondly, I would like to thank Prof. Daniel Schnitzlein, who did not hesitate to be my second assessor for this thesis.

Further gratitude goes to my fellows in both institutions, in Magdeburg and Hanover, for worthwhile discussions and advises, as well as for cheery moments that maintained my intrinsic motivation numerous times. Working in such privileged environments is more than a fulfilling professional experience, it has been an everyday pleasure. I also like to enunciate additional gratitude to those postdoctoral fellows at the NIW Hanover who always allowed me to contribute to many interesting research projects as an independent and full-fledged member of the team. I am also indebted to the doctoral programs in Magdeburg and Hanover. Without the opportunities to participate in various conferences, workshops, and meetings, the research papers this thesis builds on would not have been able to prosper as far as they did. But also at various other occasions, I received valuable comments from other researchers. To that effect, special thanks are due to Katja Coneus, Michèle Belot, Philipp Eisenhauer, Marjaana Gunkel, Guido Heineck, Verena Niepel, Friedhelm Pfeiffer, and Pia Pinger, who all took time for distinguished suggestions on preliminary stages of my research.

Finally, I would like to dedicate this work my mother and close friends, who literally had to abide me throughout certain parts of this whole episode, and did so in an unmatched fashion. Their unwavering support and their company gave me the strength to get through many difficult situations.

Summary

This thesis derives from three distinct research papers that all deal with a more recent and interdisciplinary strand of the economic literature. The encompassed field has evolved from the necessity to better account for unobserved abilities and related phenomena within empirical, predominantly labor market settings. For that purpose, concepts from psychology, like measures of cognitive intelligence, have become appropriate means in empirical labor market research. Nonetheless, many issues of individual differences in behavior and selectivity still remained unexplained. Within the last ten year or so, the consideration of personality traits has found its way into the economic literature, and meanwhile substantially contributes to the prediction of various behavioral aspects. In that regard, the economic literature explores the role of personality traits on two broad accounts: as predictors and as causes of various economic entities. In general, personality traits as well as abilities are captured by means of various psychometric constructs, primarily originating from personality psychology.

The first part of the thesis, namely Chapters 1 to 7, provides an overview on the growing and influential literature in this and related fields. The composition and the impact of personality traits with respect to certain outcomes are often less familiar to economists. Therefore, the aim of the literature reviewing part of the thesis is to give a short and introductory guide to a wide audience of readers in economics. This audience includes nonspecialist readers as well as experts in the field. Based on the contemporary literature, central questions and findings regarding measurement, theoretical modeling, and the empirical estimates are summarized within the corresponding chapters. The obtained results shed light on the relation between parental investments, the formation of personality traits and abilities, and later outcomes. The most important result for the explanation of previously unobserved behavioral heterogeneity is that the direct impact of acquired traits on various outcomes is more significant than assumed until the recent past. Not only educational achievements and later earnings, but also important social and health-related outcomes are strongly affected. Moreover, there is some preliminary but relatively clear-cut evidence on the malleability of personality traits over the life course. Early investments are the most crucial inputs into the formation of traits and abilities and should be followed by later ones. As a consequence, early neglect usually cannot be compensated in the aftermath, as the returns to those investments diminish.

In light of these general findings, two empirical applications that have the primary aim of examining the malleability of personality traits at different points in life, constitute the second and third contribution to the thesis. The former study focusses on individuals at the end of adolescence, whereas the latter involves individuals in working age. In the first application, the varying and not conclusively explored malleable of personality traits in late adolescence is addressed. In particular, the impact of factors that represent certain dimensions of the academic environment for a specific secondary-schooling track is examined. Previous findings from the related literature do not reveal whether these environmental factors can be substantial with regard to personality development and whether they are supposed to be affected as the learning intensity at school is increased. The empirical analysis presented here considers exactly this question. In order to give the empirical results a causal meaning, an exogenously induced educational policy reform in the federal state of Saxony-Anhalt is exploited as a natural experiment. At the time of its decree, the reform was intended to reduce the time spent for graduation from higher secondary school by eliminating the final grade. Since the curriculum was roughly maintained the reform also gave rise to an increase in learning intensity.

Along the same arguments of stabilizing personality traits over the life course, the second empirical application seeks to unveil the mediation paths of individual poverty dynamics, in particular as to why poor people are often literally "trapped" in poverty. Among the typically alleged determinants, personality traits that capture individual control beliefs are often considered to be one such mediator. As they are known to stabilize towards adulthood, but also to remain susceptible to environmental influences to a certain degree, the potential reverse causation from past poverty experiences to trait stability is of salient relevance in this empirical setting. With regard to the latter specific causal pathway, the empirical results provide an indication whether permanent environmental changes, such as poverty, can be strong enough to affect certain control related traits at adulthood.

Zusammenfassung

Diese Dissertation beruht auf drei separaten Forschungspapieren, die allesamt einen interdisziplinären Strang der ökonomischen Literatur bedienen. Dieser ist aus der Notwendigkeit heraus entstanden, unbeobachtbare individuelle Heterogenität und Fähigkeit im Rahmen empirischer Fragestellungen besser abbilden zu können. Die Erfassung solcher Fähigkeiten ist sogleich auch die wesentliche Anforderung in der empirischen Untersuchung humankapitaltheoretisch motivierter Zusammenhänge. Hierzu werden seit langem Konzepte aus der Psychologie verwendet, insbesondere zur Messung kognitiver Fähigkeiten wie z.B. dem IQ. Wesentliche Teile individueller Unterschiede bleiben bei einer solchen Approximation aber unerklärt. Der Einbezug von Persönlichkeitsmerkmalen in der jüngeren Forschung hat zu einem erheblichen Erkenntnisgewinn in der Erklärung dieser Unterschiede beigetragen. Entsprechend der Terminologie der Humankapitaltheorie ist in der ökonomischen Literatur die Rolle von Persönlichkeitseigenschaften als erklärendes Merkmal unterschiedlicher Ergebnisgrößen dominierend. Aber auch als Ergebnis verschiedenster Entwicklungseinflüsse werden Persönlichkeitsmerkmale mehr und mehr in die ökonomische Modellbildung integriert. Die der empirischen Erfassung zugrundeliegenden Konzepte stammen jedoch größtenteils aus der psychologischen Literatur.

Aus diesem Grund gibt der erste Teil der Dissertation (Kapitel 1 bis 7) einen Überblick über die umfangreiche relevante Literatur, die sich der Untersuchung von Persönlichkeitseigenschaften und damit verbundenen Fähigkeiten in ökonomischen Problemzusammenhängen widmet. Der Schwerpunkt liegt dabei auf der Messung und Erfassung dieser Fähigkeiten, der theoretischen Erklärung des Entwicklungsprozesses über den Lebenszyklus und der verfügbaren empirischen Evidenz. Die Validität der jeweiligen psychometrischen Konzepte ist jedoch nicht abschließend geklärt. Die Mehrzahl der Maße ist durch Messfehler, "Rückwärtskausalität" oder latente Einflüsse anderer Faktoren verzerrt. Das notwendige methodische Rüstzeug ist deshalb ebenfalls Gegenstand der Ausführungen. Zum besseren Verständnis der Humankapitalentwicklung wird auf ein erweitertes theoretisches Modell Bezug genommen, das explizit kognitive Fähigkeiten sowie Persönlichkeitsmerkmale berücksichtigt. Aufbauend auf diesen Grundlagen wird anschließend die empirische Literatur anhand der zugrunde liegenden Forschungsfragen klassifiziert und die zentralen Resultate werden zusammengefasst. Dabei kann die Tatsache, dass Persönlichkeitseigenschaften einen weitaus nachhaltigeren Einfluss auf viele Größen im Lebensverlauf haben als bislang angenommen, als fundamental und essenziell beurteilt werden. Zu den beeinflussten Größen zählen neben Schulabschluss und Verdienst auch soziale Ergebnisse und die Gesundheit. Als weitere zentrale Ergebnisse aus dieser Über-

sicht lassen sich die folgenden identifizieren: Frühkindliche Umgebungsfaktoren sind die entscheidenden Inputs in die Fähigkeitsentwicklung und somit auch in die Persönlichkeitsentwicklung, sie sollten aber durch spätere Investitionen ergänzt werden. Wichtige Konsequenz hieraus ist, dass Vernachlässigungen in diesem Alter im Nachhinein nur schwer zu kompensieren sind, da Bildungsinvestitionen einem abnehmenden Grenzertrag unterliegen.

Im Kontext dieser allgemeinen Erkenntnisse zur Dynamik von Persönlichkeitsmerkmalen im Lebensverlauf besteht die Arbeit im Weiteren aus zwei empirischen Anwendungen, die deren Formbarkeit in unterschiedlichen Altersspannen zum Gegenstand haben. Die erste Studie bezieht sich auf das späte Jugendalter, während die zweite sich dem Erwerbsalter widmet. Im Jugendalter werden insbesondere Faktoren, die bestimmte Dimensionen des schulischen Umfeldes widerspiegeln, sowie deren Veränderung im Rahmen der Analyse berücksichtigt. Vorangegangene Untersuchungen in diesem Literaturbereich haben diese Frage nur unzureichend thematisiert, insbesondere im Hinblick darauf, ob diese schulischen Dimensionen entscheidenden Einfluss auf die Ausprägung und Entwicklung von Persönlichkeitsmerkmalen ausüben können. Um diese Fragestellung empirisch untersuchen zu können, wird eine exogen induzierte Veränderung der Lernintensität herangezogen, die mit einer entsprechenden Veränderung schulischer Determinanten einhergeht. Die Exogenität der Veränderung gewährleistet dabei die kausale Interpretation der Wirkungsmechanismen. Sie basiert auf eine Bildungsreform in Sachsen-Anhalt, welche die Reduktion der gymnasialen Schulzeit zum Ziel hatte. Die Verringerung der Schulzeit um ein Jahr bei nahezu gleichbleibendem Lehrplan impliziert dabei die beschriebene Veränderung der Lernintensität.

Der selben Logik folgend, ist auch im weiteren Lebensverlauf ein Einfluss bestimmter externer Umstände auf die Persönlichkeitsentwicklung- und stabilität denkbar. Deshalb hat die zweite in der Dissertation enthaltene empirische Anwendung das Ziel, Übertragungs- und Interaktionskanäle eines konkreten Persönlichkeitsmerkmals und individuelle Armutserfahrungen zu untersuchen. Im Vordergrund steht dabei die Frage, warum von Armut betroffene Individuen häufig dauerhaft oder zumindest über einen längeren Zeitraum arm bleiben. Kontrollbezogene Persönlichkeitsmerkmale und deren vermutete Abwertung könnten dabei einen solchen Übertragungsmechanismus darstellen, da sie sich im Erwachsenenalter zwar stabilisieren, aber dennoch eine Restformbarkeit durch Umwelteinflüsse erhalten bleibt. Die empirische Untersuchung gibt somit Aufschluss darüber, ob Armutsepisoden in dieser Hinsicht eine hinreichend starke Umweltveränderung darstellen.

Schlagworte

Englisch:

- Personality Traits
- Personality Formation
- Latent Variabel Models

Deutsch:

- Persönlichkeitseigenschaften
- Persönlichkeitsentwicklung
- Modelle für latente Variablen

Contents

List of Figures

List of Tables

1

Introduction*

There is a long-standing literature in economics that investigates the sources and mechanisms underlying individual differences in labor market related outcomes. Starting with the seminal works of Becker (1964) and Ben-Porath (1967), various approaches that model the relations of innate abilities, acquired skills, educational investment, and economic outcomes to educational achievements or labor market success have been established in the literature. Unfortunately, empirical analysis in this field has always been burdened with a lack of observability in these individual determinants. This has led to a burgeoning diversity of attempts to quantify the skills and abilities involved. In most cases, measures of cognitive achievements, such as IQ tests or similar scores, have been used for empirical assessments.[1] In the psychological field, so-called personality traits that represent another source of behavioral differences have been subject to a longstanding literature already. Psychologists still argue out on the appropriateness of traits in settings economists try to accommodate (see, e.g., Roberts et al., 2007, for an overview), fortunately with an agreeing tendency in the recent past. Thus the constructs originating from this literature become more and more ubiquitous among economists. Of particular relevance are those personality traits that represent relatively persistent dimensions of an individual's behavior, most of which can be measured by means of relatively simple psychometric assessment instruments.[2] For traits specifically related to human capital outcomes, economists predominantly use the term noncognitive skills. The consideration of trait measures in empirical analysis contributes to a better understanding of the genesis and the evolvement of productivity enhancing skills beyond those of formal education and labor market experience. Moreover, a profound understanding of these mechanisms also has important implications for various kinds of policy recommendations.

Despite the compelling appeal of personality traits for many economic questions, one should always be aware of the fact that the objectives in applying such measurement constructs fundamentally differ in economics and psychology. Economists are interested

* The discussions presented in this chapter are published in Thiel and Thomsen (2013).

1 See, for example, Hause (1972), Leibowitz (1974), Bound et al. (1986), and Blackburn and Neumark (1992). See also Griliches (1977) for an overview.

2 Psychometrics is the field of psychology that deals with measurement of psychological constructs, including personality traits.

in establishing traits as productivity enhancing skills for rather specific settings. Personality psychologists, on the contrary, try to explain an individual's complete spectrum of behaviors and thoughts. Using sets of psychometric measures or other tools from psychological field without consideration of its underlying objectives may lead to very ill-advised applications.

To give an account on the issues involved in this topic is what the first chapters of this thesis are devoted to. I will provide an overview on the central problems regarding definitions, measurement, development, theoretical and empirical modeling, and outcomes that are most likely to arise when combining personality traits and micro-founded empirical strategies. The studies reviewed shall offer some practical guidance for economists who are little familiar with the relevant literature and parlance from the psychological field. It focuses on notational and methodological specifics of the psychological field and links it to the concepts that prevail in economics. For this purpose, some critical assumptions that are necessary to establish the mere existence of persistent personality traits in the sense of the human capital theory are addressed. Some of the relevant aspects have already attracted a large attention in the economic literature, and thus are subject to a good deal of review articles already in place. Most of these treatises involve the formation process of personality traits along with overviews on affected outcomes.[3] Though the aim here is not to repeat these findings in full length, a cursory summary of such studies is given as a basis for an ensuing guide to empirical analysis. The focus in the corresponding chapters is on the methodological and definitional challenges inherent to the analysis of personality traits.

From Chapter 7 on, the emphasis of the thesis is slightly shifted towards eliciting the peculiarities of trait formation in different age spans. The already existent empirical studies that deal with the formation process over the life course show a degree of plasticity that generally decreases as age increases. It follows that personality traits are alleged to be set until early adulthood. Consequently, most empirical evidence for this formation pattern focusses on earlier periods of life (see, e.g., Cunha et al., 2010). For this reason, two empirical applications that result from two separate research papers are added to the literature reviewing general discussion. By and large, both applications pick a specific part of the formation process and seek for stability pattern under rather specific environmental settings. The two populations of interest exclusively comprise individuals beyond adolescence and thus differentiate from most of literature discussed in the general part.

[3] First reviews on the topic may be found in Cunha et al. (2006). Later literature overviews are due to Borghans et al. (2008a), Cunha and Heckman (2009), and Almlund et al. (2011).

Regarding the issues raised throughout the previous paragraphs, the corresponding chapters of the thesis at hand are organized as follows. The first chapter will start out with some crucial definitions and will elicit how the notion of personality traits is embedded in the psychologic literature. In Chapter 3, a selection of psychometric measures for personality traits will be presented and evaluated with respect to their virtues and drawbacks. In addition, I give an introductory overview on how to check for validity and reliability of the measures commonly used in psychometrics, and what should be considered when applying both criteria for construct choice. Chapter 4 outlines some intuitive notions and first evidence on how to map personality traits into economic preference parameters. Some practical guidance regarding econometric approaches that properly account for the error-proneness of raw test scores in representing latent personality traits will be provided in Chapter 5. Chapter 6 reviews a number of studies that establish causal inference for personality traits on several outcomes. Chapter 7 embeds the psychologic and sociologic literature on personality development into a formal framework of human capital formation suggested by Cunha and Heckman (2007). The general patterns induced by this framework are augmented by the findings of the two empirical analyses presented in Chapters 8 and 9. The final chapter concludes and puts the empirical findings into perspective with the reviewed literature.

2

Some Terminology[*]

Due to its origins in the intersection of economics and psychology, some concurrent terms have evolved throughout the seminal period of the literature concerned. This chapter clarifies on the most important ones among them.

The term "noncognitive skills" originates from the economic literature and started to emerge in course of the work by Heckman and Rubinstein (2001). It comprises the notion of personality traits that are, besides pure intelligence, particularly relevant for several human capital outcomes, such as educational or labor market achievements. Henceforth, I will make use of the latter term in all following discussions.[1] Personality traits constitute, along with other determinants, an individual's personality. In economic parlance, personality is kind of a response function to various tasks (see Almlund et al., 2011). There are several approaches in the psychologic literature that target at modeling personality in light of environmental entities.

A good point of departure is the model suggested by Roberts et al. (2006). It shall serve as a reference framework for the remainder of the discussion. It designates four core factors of personality: personality traits, (cognitive) abilities, motives, and narratives. Together with social roles and cultural determinants, these core factors produce an individual's identity and reputation. Identity, in that regard, is the consciously available self-image about the four factors, including self-reports about them. Reputation is the entity that includes others' perspectives into the framework. The Roberts framework also accounts for the possibility of feedback processes, i.e., the possibility of environment activating the core factors and vice versa. In its original definition, personality traits are relatively persistent attributes of behavior, feelings and thoughts and hence are largely non-situational (see Allport, 1937). However, the prevalence of consistency (or at least a certain degree of it) across situations is not without controversy in the literature. I will elaborate on this point in the discussion of measurement constructs in Chapter 3.

[*] The discussions presented in this chapter are published in Thiel and Thomsen (2013).

[1] After some years in the making, the Heckman-driven part of the literature made itself an effort to switch to the term "personality traits". Presumably, the study of Borghans et al. (2008a) is the turning point for this terminology.

Prominent examples for personality traits are self-discipline, self-control, agreeableness, self-esteem, or conscientiousness, just to mention a few.[2] As the Roberts model suggests, few issues of personality are devoid of cognition. Sometimes it is even hard to conceptually, not to mention empirically, distinguish cognitive abilities and personality traits. For instance, emotional intelligence (see Salovey et al., 2004), which describes the processing ability to anticipate the consequences of feelings and the resulting behavior, is a marginal case in terms of cognitive factors and personality traits.[3] Hence, the denotation noncognitive is rather imprecise. Notwithstanding this fuzzyness, the terms noncognitive skills and personality traits are often used interchangeably in most parts of the economic literature.

Recall that the primary interest is in working out those trait dimensions that are in some sense productivity enhancing. One can thus relate the notions from psychology to economic terms by construing cognitive abilities and personality traits as a partially acquired and partially inherited stock of human capital. To attain definitional exactness, it is furthermore necessary to clarify on the drawn distinction between skills and abilities that usually prevails in the human capital literature. For instance, Becker (1964) differentiates both terms in that abilities are innate and genetically predetermined, whereas skills are acquired over the life cycle. According to this view, skills and abilities can be seen as two distinct determinants of potential outcomes.[4] On the contrary, the more recent literature that incorporates personality and intelligence constructs as an inventive means of measuring human capital emphasizes that inherited and acquired factors act somewhat jointly in forging stocks of human capital. Along with prenatal environmental factors, genitacal determinants provide the initial inputs in the process of trait and ability formation (see Blomeyer et al., 2009, Cunha et al., 2010). The ensuing gene-environment interactions are highly complex in nature and will be discussed in more detail in Chapter 7. These findings suggest to construe skills and abilities rather as complements in generating outcomes of interest.[5] I will thus use both terms interchangeably throughout the rest of the discussions.

Put together, it is obvious that the human personality is a highly complex construct that goes beyond the concept of personality traits and requires consideration of multiple factors combined in an interactional fashion. As the subsequent chapters will show,

---

2 For example, Allport and Odbert (1936) obtained about 18,000 attributes describing individual differences in the English language.

3 Borghans et al. (2008a) discuss further examples like cognitive style, typical intellectual engagement, and practical intelligence.

4 To that effect, Becker (1964) figures out that acquired skills possess higher explanatory power for future earnings than innate abilities do.

5 Accordingly, Cunha and Heckman (2009) rather use the term capabilities in order to elude notational conflicts.

personality and its impact on various outcomes are of particular interest for the field of economics. Notwithstanding this appeal, for empirical analyses one has to presume a sufficient degree of stability and also has to accept certain simplifications. Somewhat fortunately, the objective in economics is rarely to model and project all facets of personality, but to identify relatively stable and conveniently assessable determinants of the particular outcome of interest. Given this fact and a general tendency to reconcile different views about cross-situational stability of the personality in the psychological field (see Roberts, 2009), empirical operationalizations can be attained with less difficulties than apparent at the first glance. This will be made clear throughout the discussions over the next three chapters.

3

Measures and Constructs for Personality Traits*

There is no uniform opinion about adequate personality models and the resulting assessment of personality in the field of psychology. Hence, a brief overview on the relevant psychologic literature is a sensible first step. The crucial issue in terms of postulating a persistent stock of skills is to ensure a sufficient degree of stability of personality traits across situations. First, bring into mind the difference between a behavioral instability obeying some genuine randomness and behavioral instability due to situations. If there is instability of actions in the sense that someone acts very emotional and this is a persistent phenomenon irrespective of situations, one can quantify this inclination by means of an appropriate personality construct.[1] If inconsistency of actions, however, pertains to more than the individual's emotional stability and is generally driven by situational and contextual determinants, it would be meaningless to impose traits as a stock of human capital for outcome prediction. In such a case, traits would be themselves to much of an outcome of situations and individual circumstances in order to be a reliable predictor. To clarify these points, the prevailing view in the literature is briefly discussed in what follows.

3.1 Personality and Situations

The existence of persistent traits has been subject to vigorous discussion in the psychological literature over the last decades. The common understanding of the influential work by Mischel (1968) is that all patterns of behavior, feelings, and thoughts are manifestations of specific situations, not of stable personality traits. Mischel and his proponents dileniate this as a misinterpretation in the aftermath. Mischel (1973) merely endeavors to incorporate situations into whatever drives stable characteristics (e.g. personality traits), dubbing it the if-then signature of personality. It characterizes an individual's variabil-

* The discussions presented in this chapter are published in Thiel and Thomsen (2013).
[1] As will be addressed below, the Big Five factor Neuroticism addresses exactly these issues.

ity by relatively stable patterns across situations. Orom and Cervone (2009) therefore highlight that Mischel's initial point was that cross-situational consistency in personality assessment is only low when focussing attention to global, nomothetic trait constructs. This implies that cross-situational consistency exists, but only under certain conditions. These conditions generally apply when relaxing the ambitious view of rigid and globally valid traits, and instead allow for some other factors to affect measured personality. The ensuing discourse in the literature led to alternative notions of personality traits that meanwhile also provide some consensus.

The so-called social-cognitive approach established and advocated by Mischel (1973) and Bandura (1986) provides such an alternative notion of personality. It mainly focuses on explaining the cognitive processing which underlies thoughts and behaviors. Accordingly, people differ in terms of cognitive abilities relevant for the implementation of certain behaviors. The awareness of these abilities in conjunction with expectations about self-efficacy, goals, and valuation standards constitute the personality. All four subsystems of personality are interactional in nature and therefore not separately assessable. As such, social-cognitive theorists rather rely on qualitative types of assessments and would not assign a certain score or number to one of these systems. The evaluator has to account for the situation as perceived by the observed individual and thus has to analyze consistent patterns in this situational context. One of Bandura's contributions is the concept of reciprocal determinism. It essentially states that there is no actual source of behavior as asserted by trait theorist or behaviorists. Instead there is a triangular feedback system which consists of personal characteristics, behavior, and environmental factors. A potential explanation for the interior processing underlying this system is the cognitive-affective processing system by Mischel and Shoda (1995). It interrelates the abovementioned subsystems (abilities, expectations, goals, and valuation standards) by means of cognition and affects. Individual differences therefore arise from differences in activation levels of cognitions and affects. The accessability of activation levels differs over various situations. This has very far reaching implications for the proper construction of assessment tools.

A contrary view to the whole situation debate is held by the proponents of the global dispositional approach. It is best exemplified by concepts like the Five Factor Model of Goldberg (1971).[2] Proponents of the Five Factor Model constantly and above all highlight the stability of personality over most of the lifetime and across situations.

The most widely advocated approach in contemporary personality psychology is to combine the assumption of a certain stability in traits with elements of the social-cognitive

2 The most widely applied version is the Big Five inventory of Costa and McCrae (2008).

approach, such as goals expectations, and assign them to different levels of analysis.[3] The Roberts model, which I already mentioned to be a good contemporary reference, accounts for these elements and their interaction with environmental factors. Roberts (2009, p. 138) vividly summarizes this unifying view with the following words: “The trait psychologists can continue to focus on factor structure and test retest stability. The social cognitive psychologists can study goals, motives, beliefs, and affect - things that putatively change.” Such a view maintains the notion of personality traits without imposing any assumptions about the underlying cognitive processing function, and also incorporates other entities in the personality system that are necessary to represent contextual and situational factors.

Given the specific setting, one should be aware of the fact that all observed or otherwise assessed measures of personality traits can also be manifestations of the other entities addressed in the Roberts model. For instance, fulfilling a certain social role at the time an assessment takes place, is a contextualizing variable one has to control for (see Wood, 2007).[4] This approach gives rise to a certain stability of personality traits across situations and therefore paves the way for application of the kind of personality tests empirical economists are most interested in. Some crucial features of such tests are summarized in the next section.

3.2 Assessment Tools

I will now sketch how trait theorists or social-cognitive theorists assess the entities in their respective models and what are the pros and cons of the respective methods with regard to different assessment situations. There are three main dimensions an evaluator has to decide on: (1) the type of assessment, (2) the person to be assessed, and (3) the dimension.

(1) Proponents of the social-cognitive approach usually rely on qualitative assessment methods conducted by experts who passively observe or actively interview a person. These methods involve variations of situational stimuli and substitutions of the assessed person until systematic evidence for the underlying processing is revealed. For applications within large scale data instruments (in the field or in experimental settings), which definitely dominate in empirical economics, this type of assessment is rather cumbersome

[3] Even very early definitions of personality traits implicitly account for situational variance in behavior (Allport, 1961, p. 347)

[4] As will briefly address in context of personality development, permanently fulfilling certain social roles does not solely affect measures of personality, but induces changes of personality traits as well.

and costly. For such kind of investigation, quantitative assessment methods are undeniably more appealing. Generally, the aim of the latter type of methods is to provide scores for respective dimensions of the personality. These scores are directly used for analyses or employed to derive underlying latent constructs.

(2) The evaluator has to choose between self-reports and observer-reports. Self-reported measures are convenient due to their simple implementation, but implicitly assume that respondents are capable to consciously perceive their personality, or at least the actions that are supposed to represent it. This prerequisite does not generally apply. For instance, infants and children are often not capable of doing so, and thus are usually assessed by observers from their social environment (parents or teachers), or by experts. Distortions of self-reports or observer-ratings can also be more generic. For traits related to typical social-environmental settings, like meeting a stranger or having a discussion with another person, observer-ratings tend to better predict behavior than self-reports, particularly since the potential for disorder in self-perception is usually high with regard to such situations. For instance, what the narrator of a joke believes to be funny is not necessarily perceived by others in the same manner. Vice versa, self-reported personality ratings are more strongly related to assessments of emotional issues driven by interior processes and less shared with others. An illustrative example for such pattern is that a person who suffers from depression would usually try to conceal this fact from others. A potential drawback of observer-ratings, at least in some settings, is their generally lower suitability for survey instruments within simple questionnaires. The choice of the person to be assessed therefore strongly depends on the trait of interest and the specifics of the survey setting.

(3) The last aspect is the extent to which the measure captures personality. Besides various low-dimensional constructs for assessing the magnitude of very specific traits, there is a large number of taxonomies mapping human personality as a whole. Proponents of these high-order personality inventories advocate the global dispositional approach discussed above and therefore construe these models as comprehensive representation of the personality, usually without further consideration of situational aspects.

Higher-Order Constructs: Most mappings of personality impute hierarchical structures that are derived from factor analytic approaches. Many of them derive from initial studies that exploit lexical-linguistic patterns prevailing in their respective language of origin.[5] The retrieved taxonomies are then used to set up inventories of closely related questions. The resulting measurement systems bundle a number of questions, which are

[5] This follows the tradition of Allport and Odbert (1936), who were the first to do so for the English language.

also called items, relating to specific trait dimensions. They follow similar designs as those used for the assessment of general IQ expressions.[6] But compared to intelligence tests, the level of abstraction is lower in case of personality traits. Despite early efforts to identify a general factor for personality (see Webb, 1915), personality inventories from the prosperity period of the global dispositional approach (see the previous section) usually assume at least three major factors. Table 3.1 provides an overview on the global constructs most often used in the psychological literature.

A widely accepted taxonomy is the Five Factor Model established by Goldberg (1971) and the related Big Five by Costa and McCrae (1992). The identification of five high-order factors is not uncontested in the literature, though. Some factor analytic results suggest a lower number of dimensions, whereas others claim a higher number. Eysenck (1991), for example, provides a model with just three factors. Digman (1997) curtails the Big Five distinction to only two principal factors.[7] In contrast to that, Hough (1992) proposes a more stratified version of the Big Five taxonomy, the so-called Big Nine. Due to their data reducing genesis, virtually all the aforementioned concepts lack a theoretical foundation, and,therefore are largely inconsistent with the type of personality models discussed above. Only for exceptional cases, neurological support for the constructs is available (see, e.g., Canli, 2006, pertaining to the Big Five).

As a consequence, low predictive power of a high-order factor does not necessarily imply that all of the lower-order factors in Table 3.1 exert no influence on an outcome of interest as well. Using lower-order constructs or even uni-dimensional factors often entails a gain in explanatory power, but at the potential cost of not covering all relevant personality facets.

Lower-Order Constructs: There also exist several lower-order constructs which, in light of the points addressed above, may be more appropriate for settings where contextualization is an issue. This property simply follows from the fact that it is easier to align and substantiate low-order constructs with regard to specific settings. Prominent examples in the context of educational outcomes are self-control (see Wolfe and Johnson, 1995) and the related self-discipline (see Duckworth and Seligman, 2005). The Brief Self-Control Scale (Tangney et al., 2004) is a commonly used means of assessing self-control. It includes 13 items that add up to an overall score increasing with higher degrees of self-control. The Internal-External Locus of Control by Rotter (1966) is often

[6] A version of Cattell (1971) includes fluid intelligence, i.e., the ability to solve novel problems, and crystallized intelligence, comprising knowledge and developed abilities.

[7] The factors are not presented in Table 3.1 since they are simply denoted metatraits without further specification.

Table 3.1: Personality Models and Sub-Factors

Inventory	Factors	Lower-order Factors
Big Five (Costa and McCrae, 1992)[a]	Openness to Experience	Fantasy, Aesthetics, Feelings, Actions, Ideas, Values
	Conscientiousness	Competence, Order, Dutifulness, Achievement Striving, Self-Control/ Self-Discipline, Deliberation
	Extraversion	Warmth, Gregariousness, Assertiveness, Activity, Excitement Seeking, Positive Emotions
	Agreeableness	Trust, Straightforwardness, Altruism, Compliance, Modesty, Tender-Mindedness
	Neuroticism	Anxiety, Vulnerability, Depression, Self-Consciousness, *Impulsiveness*, Hostility
MPQ[b] (Tellegen, 1985)	Negative Emotionality	Stress Reaction, Alienation, Aggression
	Constraint	Control, Traditionalism, *Harm Avoidance*
	Positive Emotionality	*Achievement*, Social Closeness, Well-Being
Big Three (Eysenck, 1991)	Neuroticism	Anxious, Depressed, Guilt-Feeling, Low Self-Esteem, Tease, Irrational, Shy, Moody, Emotional
	Psychoticism	Aggressive, Cold, Egocentric, Impersonal, Anti-Social, Unempathic, Tough-Minded, Impulsive
	Extraversion	Venturesome, Active, Sociable, Carefree, Lively, Assertive, Dominant
JPI[c] (Jackson, 1976)	Anxiety, Breadth of Interest, Complexity, Conformity, Energy Level, Innovation,Interpersonal Warmth, Organization, Responsibility, Risk Taking, Self-Esteem, Social Adroitness, Social , Participation, Tolerance, Value Orthodoxy, Infrequency	
Big Nine (Hough, 1992)	Adjustment, Agreeableness, Rugged Individualism, Dependability, Locus of Control, Achievement, Affiliation, Potency, Intelligence	

italic: Affiliation of facet is still in debate (see Bouchard and Loehlin, 2001).
[a] see also Costa and McCrae (2008)
[b] **M**ultidimensional **P**ersonality **Q**uestionnaire.
[c] **J**ackson **P**ersonality **I**nventory.
Source: Bouchard and Loehlin (2001) and own illustration.

perceived as a related measure, but exhibits a fundamental difference. It merely assesses an individual's attitude on how self-directed or how coincidental attainments in life are, but not how successful one could be in governing this fate. The original Locus of Control (Rotter, 1966) comprises 60 items. Usually, longitudinal datasets apply abbreviated versions.[8] A similar scale for Locus of Control is the Internal Control Index (Duttweiler, 1984), a 28-item scale that scores in the internal direction. Self-esteem provides another important determinant of educational and labor market outcomes (see Heckman et al., 2006). It is often quantified by means of the Rosenberg Self-Esteem Scale (Rosenberg, 1965), a 10-item scale. For the assessment of the personality of a child, measurement constructs usually draw on observations from other persons. A corresponding scale based on observational reports of teachers or parents is the Self-Control Rating Scale by Kendall and Wilcox (1979), a 33-item scale indicating the ability of inhibiting impulsiveness.

The need to assess behavioral structures of children has led to a related field in psychology that deals with a behavioral dimension called temperament. Temperamental research is as sub-discipline of developmental psychology. The latter investigates all kinds of psychological changes over the life course, not only changes of personality. The major focus, however, is on infancy and childhood. Constructs to assess temperament rather refer to behavioral tendencies instead of pure behavioral acts. They are thus similar to selected trait dimensions of adult inventories. An influential model has been suggested by Thomas et al. (1968). It stratifies temperament into nine categories, each of which is further grouped into three types of intensity. There are further established concepts of temperament, for instance those of Buss and Plomin (1975) and Rothbart (1981). But even the more recent literature is still involved in this topic (see, e.g., Rothbart and Bates, 2006).[9] Meanwhile, some interrelations between concepts of personality psychology and developmental psychology have been established. For instance, Caspi (2000) reveals links between the extent of temperamental facets at age 3 and personality at adulthood. Temperament at infancy and early childhood designates later personality but remittently affects behavior as the individual matures. According to Thomas and Chess (1977), purely temperamental expressions at later age are only likely in case of being faced with a new environmental setting, often abrupt or extreme. However, the inferences from studies linking temperament and personality are far from being conclusive (see Rothbart et al., 2000, Shiner and Caspi, 2003, Caspi et al., 2005, for a review of the literature).

[8] The German Socio Economic Panel (SOEP), for instance, comprises a 10 item version, whereas the National Longitudinal Survey of Youth uses a 23 item version.

[9] See Goldsmith et al. (1987) for an overview on temperamental measures.

3.3 Reliability of Items

Reliability refers to the consistency of answers to a psychometric task over time or across observations. The most convenient way to test for reliability is by means of test-retest correlations over time. Generally, each test item i can be expressed as

$$T_i = \alpha_i \tau_i + \varepsilon_i, \tag{3.1}$$

where T_i is the attained score, τ_i is the true score with α_i as the corresponding scaling parameter, and ε_i is an error term. Since test-retest settings are rarely at hand, other coefficients prevail in the literature. A standard measure to quantify reliability across several items is Cronbach's alpha (see Cronbach, 1951) which can be determined as follows.

$$\rho_\alpha = \left(\frac{l}{l-1}\right)\left(1 - \frac{\sum_{i=1}^{l} Var(T_i)}{Var(\sum_{i=1}^{l} T_i)}\right), \tag{3.2}$$

where l is the number of items used to measure the true score. It relates item variance to the variance of the total score and therefore increases with rising inner consistency of the construct. This procedure originates from methods of classical test theory, one of the very first fields analyzing issues of measurement error in psychometric constructs.

Given the nature of the observed test scores T_i in equation (3.1), some important implications for item selection and the corresponding degree of reliability arise. As suggested by the Roberts model introduced above, the consistency of an item with regard to a specific trait may be imperiled if also other entities in the interactional framework are captured inadvertently. Separately assessing these units is difficult as they do not occur in isolation but instead simultaneously influence each other. For instance, when measuring a certain personality trait by means of a questionnaire, it is important to not prompt the respondent to project his thoughts into a particular situation. In this case the score can be a manifestation of the trait of interest, but also of motivation, past experiences, or narratives and abilities helpful for comprehension of the task.

Though proponents of the global dispositional approach claim that most of the entities in the Roberts model can be mapped into at least one of the dimensions of the global personality inventories discussed above (see Costa and McCrae, 1992, for emprical evidence), this result is dissatisfying when it comes to the measurement of more specific traits and their relation to particular economic outcomes. Global personality mappings like the Big Five are derived by exploratory factor analysis tools.[10] In case of low-order constructs or even uni-dimensional factors, exploratory factor analysis is primarily used for verification of the assumed structure. In either instance, neglecting the influences of accompanying determinants can be harmful for the resulting trait scores.[11] By construction, exploratory factor analysis cannot disentangle the effects of immediate pathways via common factors and indirect pathways. Therefore, the identification problem that inheres a lack of contextualization frequently causes some variation to be attributed to measurement error or spurious pattern of the trait under study. The former may occur if the item formulation unsystematically induces the measured scale to include framing effects due to motivation or social roles. The latter is likely to result from more systematic distortions. In order to elude these drawbacks, it is necessary to contextualize the measurement, i.e., to control for situational determinants that potentially affect the expression of abilities, motivation and the like. When using questionnaires as an assessment tool, the item framing should avoid to mentally force the respondent into specific situations to answer an item. Intuitively, low-dimensional or uni-dimensional constructs are less susceptible to these phenomena and are easier to validate by means of other constructs or outcomes. Such cross-validations are discussed in the next section.

Contextualization addresses most interactions between entities of the personality that may distort the measurement of "true" traits. If this distortion, however, is intended by the responder, the phenomenon the researcher has to deal with is called faking. Evidently, the potential for faking is higher for measures of personality traits than for cognitive abilities, as it is generally easier to pretend a fraudulent level of a trait than of intelligence. It might be that the background of an assessment can urge the respondent to understate and/or overstate. As an example consider a test administered for making a hiring decision. The faking behavior in tests is also a projection of other personality traits or cognitive capabilities. Borghans et al. (2008b) provide evidence for an interrelation between personality and incentive responsiveness. Morgeson et al. (2007) conclude that correcting for intentional faking does not improve the validity of measures, presumably since it mostly offsets across observations.

[10] A comprehensive introduction into the methods of exploratory factor analysis is provided by Mulaik (2010).

[11] A vivid impression of construct development is given in Tangney et al. (2004).

3.4 Validity of Item Sets

After a construct has been developed by means of some data reduction like exploratory factor analysis or by means of some theoretical considerations, another point that has to be minded is the validity of the resulting set of items. It is concerned with the question as to whether a chosen item inventory actually measures what it is supposed to measure. It should thus always be tested when developing a scale, but should also be considered whenever an existing construct is applied to a new kind of data. In the psychometric literature, three types of validity are distinguished (see, e.g., Cervone et al., 2005).

Content Validity: Content Validity is a qualitative type of validity and requires sound theoretical foundation of construct to be assessed. It evaluates whether the considered theoretical domain is captured by the data. For instance, if a construct justified by some theory comprises three different dimensions, i.e., three latent factors, one needs measures for all of them. Otherwise, content validity is questionable. A potential lack of theoretical consensus is the major weakness of this kind of validity. Therefore, two further data-based types of validity have to be utilized in general.

Criterion Validity: To test for criterion validity one needs a variable that constitutes a standard measure to which the personality dimension under assessment is related to, called a criterion variable. It can be a concurrent measure from the same or a similar measurement system, or a predictive measure derived from a resulting outcome. The magnitude is usually represented by means of correlations between measurement and criterion variables. It can be shown (see Bollen, 1989, for a detailed discussion) that the magnitude is largely sensitive to unsystematic error variance in both, measurement and criterion variable, and depends on the choice of the criterion variables. To illustrate this point, consider both variables, the measure T and the criterion C, in an additive separable factor representation.

$$
\begin{aligned}
T &= \lambda_1\theta + \varepsilon_1 \\
C &= \lambda_2\theta + \varepsilon_2,
\end{aligned}
$$

where θ is the latent factor constituting both measures with respective factor loadings λ_1 and λ_2, and ε_1 and ε_2 are uncorrelated residual terms. The correlation between T and C, which Lord and Novick (1968) denote as validity coefficient, is

$$
\rho_{T,C} = \frac{\lambda_1\lambda_2\phi}{\sqrt{Var(T)Var(C)}},
$$

with ϕ representing the factor variance. It turns out that even if all measures are standardized, which extends to the latent factor and leads to a vanishing denominator, the validity coefficient still depends on both factor loadings. Hence, not only the quality of T as a proxy for θ quantified by λ_1 is relevant, but also the quality of the criterion variable captured by λ_2. This result should be minded whenever criterion reliability is examined. Moreover, as validity measures are based on factor models, they are not in general capturing causal relationships.

Construct Validity: For many constructs in psychometrics it is difficult to find measures that establish criterion validity. Instead one has to rely on construct validity. It assesses to what extent a construct relates to other constructs in a fashion that is in line with underlying theory. The resulting coefficient is again a correlation. By arguments similar to those invoked for criterion validity, other driving forces apart from the quality of the proxy, like factor correlation and reliability of the measure, can contaminate the validity coefficient. A formalization of this point is less straightforward than in the previous case but can be sketched as follows. Consider two measures T_1 and T_2 for two latent traits θ_1 and θ_2 with different loadings λ_{11} and λ_{22}. Hence, one has

$$\begin{aligned} T_1 &= \lambda_{11}\theta_1 + \varepsilon_1 \\ T_2 &= \lambda_{22}\theta_2 + \varepsilon_2. \end{aligned} \tag{3.3}$$

It can be shown that the construct validity depends on more than the association between the latent factors. The particular relation reads

$$\rho_{T_1T_2} = \sqrt{\rho_{T_1T_1}\rho_{T_2T_2}}\rho_{\theta_1\theta_2},$$

where $\rho_{T_1T_1}$ and $\rho_{T_2T_2}$ represent reliability. Moreover, it should be noted that the choice of comparison constructs is arbitrary. A more systematic approach of establishing construct validity is the multitrait-multimethod design suggested by Campbell and Fiske (1959). It requires that two or more traits are measured by two or more constructs, where each respective construct is one method.[12] It thus is an extension to the setting in equation (3.3). If the correlations for the same trait across different methods are sufficiently large, there is evidence for so-called convergent validity. Discriminant validity arises if convergent validity is higher than the correlation between measures which neither share trait nor method and higher than the correlation between different traits measured with the same method. Again, the magnitude of convergent validity can be sensitive for other reasons than closeness of the measure, like latent factor correlation and reliability.

12 Hence, the terminology multitrait-multimethod.

As the foregoing discussion suggests, there is a twofold circularity to be resolved in order to obtain reasonable validity measures. The first is circularity in a statistical sense, i.e., a simultaneous causality between measures and the latent traits. It particularly arises for concurrent real world outcomes used to establish validity. The second is a circularity in justification of genuine measures and resultant measures of validation. This is what Almlund et al. (2011) denote an intrinsic identification problem rather than a parameter identification problem. Loosely speaking, one should always be aware of the "chicken and egg problem" of choosing a construct and validating it by means of constructs that were established in the same manner. In order to resolve the former problem and to ensure causality one has to rely on structural equation approaches. To deal with the latter, at least one "dedicated" measurement equation per trait would be desirable (following the notation of Carneiro et al., 2003), i.e., a measure that exclusively depends on a particular trait. For illustration, one may consider a psychometric task like responding to a questionnaire item. Even if one controls for situational determinants, identification is restricted to tuples of traits without dedicated measures. In case of low-order constructs and respective real world outcomes the reasoning for a dedicated measure is generally easier to achieve. For more general dimensions it requires a more profound justification in choosing measurement and validation constructs.

4

Personality Traits and Economic Preference Parameters*

It is quite intuitive to assume a relationship between the expressions of traits and economic preference parameters. As has been elaborated in the previous chapters, personality traits are primarily intended to project dimensions of behavior with a focus on generality, situation-invariance, and durability. Behavioral preference parameters, whether self-related or other-regarding, rather refer to mathematical laws that link specific stimuli to behavioral responses, usually within experimental settings. Preference parameters and traits therefore roughly represent the same causes of individual behavior, albeit in different hypothetical frameworks. An integrative framework for preference parameters and personality traits is yet not explicitly established, though some indicative results have been suggested in the literature. To give an example, one may consider the patience of an individual to be related to his or her time preference. As Borghans et al. (2008a) summarize, from an economic point of view, it is meaningful to relate personality concepts to common parameters like time-preference, risk-preference, and leisure-preference, but also to the more recently emerged concepts of other-regarding preferences, like altruism and reciprocity (Fehr and Gächter, 2000).

Relating traits and preference parameters in a causal way requires a notion of the underlying associations. Due to the complexity of human thoughts and behavior, a theoretical foundation of such mechanisms is difficult to establish. Even without the incorporation of traits, specifying explicit value functions that account for multiple forms of other-regarding preferences is almost infeasible (see Fehr and Schmidt, 2006, for a discussion). To illustrate what may be in effect, an implicit representation has to be used instead. Following Almlund et al. (2011) and their various model suggestions, personality traits can be construed as both preferences and constraints.[1] Formally, a utility maximizing agent who faces some uncertainty could be characterized by the following implicit expected utility.[2]

* The discussions presented in this chapter are published in Thiel and Thomsen (2013) as well as in Thiel et al. (2014).

1 In contrast, preferences are rarely seen as constraints in economic theory.

2 As long as the the representation is implicit and no axiomatic foundation is required, one does not

$$E\left[U(\mathbf{x}, \mathbf{P}_{\boldsymbol{\theta},\mathbf{e}}, \mathbf{e}|\boldsymbol{\psi}_{\boldsymbol{\theta}})|\mathcal{I}_{\boldsymbol{\theta}}\right] \text{ s.t. } \mathbf{I} + \mathbf{r}'\mathbf{P}_{\boldsymbol{\theta},\mathbf{e}} \geq \mathbf{x}'\mathbf{w} \text{ and } \mathbf{1}'\mathbf{e} \leq \bar{e} \tag{4.1}$$

All variables with $\boldsymbol{\theta}$ as a subscript constitute a possible pathway of the influence from traits to the economic representation of an agent's response function. Utility $U(\cdot|\cdot)$ depends on preferences $\boldsymbol{\psi}$, which in turn are related to $\boldsymbol{\theta}$. $E\left[U(\cdot|\cdot)|\mathcal{I}_{\boldsymbol{\theta}}\right]$ is the expected utility for the arguments $\mathbf{x}$, $\mathbf{P}(\cdot)$, and $\mathbf{e}$ conditional on the information set $\mathcal{I}$. The latter may also depend on traits $\boldsymbol{\theta}$. All arguments are vectors. $\mathbf{x}$ is a vector of consumption goods and $\mathbf{e}$ is the vector of effort devoted to all possible tasks, where the sum of its elements cannot exceed $\bar{e}$.[3] Since effort can cause kind of a "good feeling" related to agents' endeavors, it also enters the utility function directly. In addition, effort is a complement for the vector of available traits $\boldsymbol{\theta}$ in the vector function for productivity $\mathbf{P}(\boldsymbol{\theta}, \mathbf{e})$, which maps $\boldsymbol{\theta}$ and $\mathbf{e}$ into outcomes for all possible tasks.[4] $\mathbf{P}(\cdot)$ is the "intangible" mediating path of productivity into utility, whereas the "tangible" one is through consumption goods. The goods with price vector $\mathbf{w}$ are settled with income that does not depend on productivity for tasks $\mathbf{I}$ and with the income from performing tasks for task-specific rewards $\mathbf{r}$.

Traits can also be a constraint in another sense than in equation (4.1). Dohmen et al. (2010) discuss the potential for confounding due to the observational equivalence of differences in actual preferences and differences in capabilities required to perform the task that is used to measure the preferences. In terms of the representation given in equation (4.1) this means that it is difficult to disentangle $\boldsymbol{\psi}$ and $\mathcal{I}$. As an example, consider the degree of numeracy that affects the comprehension of an investment decision used to assess time preference. As to that, Dohmen et al. (2010) show that cognitive capabilities are inversely related to risk aversion and impatience. As will be discussed in greater detail in Chapter 7, there is some inheritability, or at least intergenerational stability, of traits, including cognitive factors. Given the associations between some preferences and cognitive factors, these findings open up a new pathway for intergenerational transmission of preferences. In line with that, Kosse and Pfeiffer (2012) show a correspondence between mothers' and children's degree of impatience. In particular, this intergenerational relation appears to hold for short-run impatience (see Kosse and Pfeiffer, 2013), which has its neural correlate in the limbic system (see McClure et al., 2004). Borghans et al. (2008b) examine potential links between personality traits and responsiveness to incentives when completing some cognitive tests. The responsiveness is captured by common economic

have to commit oneself to a value function based on utility theory. Other forms of well-being (see Sen, 1999) are also in line with these general considerations.

[3] Think of effort as a representation of the situational parameters discussed in the psychologic literature above.

[4] Of course, complementarity between $\boldsymbol{\theta}$ and $\mathbf{e}$ only holds to a certain degree as effort can only compensate for a lack of $\boldsymbol{\theta}$ within certain ranges of $\mathbf{P}(\cdot)$.

preference parameters. They find a negative correlation between the Internal Locus of Control and the personal discount-rate, and likewise a negative correlation between emotional stability and risk-preference. Both results appear intuitively plausible. Dohmen et al. (2008) use data from the German Socioeconomic Panel (SOEP) and reveal some connection between Big Five personality traits, measures of reciprocity, and trust. All Big Five factors exert significant a positive relation to positive reciprocity, especially conscientiousness and agreeableness. Moreover, neuroticism seems to be positively related to trust and negative reciprocity.

Given the obvious complexity in making the above general framework explicit, studies that rely on correlations between traits and economic preference parameters provide only vague and sometimes inconclusive evidence on the associations between both concepts. Without further ado, generalizations of the documented relationships are rather inadvisable (see Becker et al., 2012). Unfortunately, preferences are usually more difficult to survey. This becomes apparent as large scale studies with assessments via questionnaires are likely to suffer from a number of possible problems. The observed preferences are mostly stated, i.e., they refer to hypothetical items. If revealed preferences derived from real actions are used, they take place within an isolated non-market setting. Yet, it is ambiguous whether preferences for artificial and real market settings are identical (see Kirby, 1997, and Madden et al., 2003, for two opposing views). If an experimental assessment embodies real rewards, choosing the respective payoffs binds the participant to maintain his or her choice. In a real-life setting, however, the individual also has to withstand other opportunities and there may be a higher degree of uncertainty for future payoffs. It proves difficult to partial out time preference from risk-aversion (see Borghans et al., 2008a, and the literature they refer to). Moreover, measures of time preference may be subject to framing effects. Non-linearities with respect to the payoffs are also likely and limit the external validity of experimental findings. Further caveats, most of which resulting in identification problems, are outlined in Almlund et al. (2011). Though the respective frameworks to model personality and preferences are both ought to describe human behavior and decision making, they originate from different disciplines with very different objectives. To that effect, the status of this research field is still premature and more interdisciplinary research, possibly with neurological foundations (see, e.g., McClure et al., 2004), seems promising in closing this gap. Whenever causation between the two concepts is difficult to establish, it may also be worthwhile to jointly use traits and preferences as predictors for causally determined real world outcomes. This has already been done separately by Dohmen et al. (2011) for risk preference and by Heckman et al. (2006) for personality traits. In such applications, both concepts are projected onto one common and economically interpretable entity, like wage for instance.

5

Methodological Challenges[*]

This chapter intends to give a brief overview on the eligibility of different estimation strategies in order to deal with the specific requirements that arise from personality test scores. The discussion thus far has revealed various sources of simultaneity and measurement error. Therefore, one has to carefully scrutinize the process underlying the data before using psychometric test scores for empirical analysis. It follows that it is generally most convenient to decompose each personality construct into unidimensional chunks that can be analyzed separately. As even unidimensional constructs depend on sets of items, relying on unweighted raw scores or arbitrary selections from the available items does not necessarily lead to the best projections attainable. To find such item combinations, exploratory factor models are usually employed.

5.1 Obtaining Relevant Item Sets

I will sketch a rather simple frequentist approach due to its ease of implementation and its practical relevance. Note that other procedures that are somewhat more flexible, but unfortunately also more involved regarding their implementation, have come up in the recent past (see, e.g., Conti et al., 2014). The starting point of the frequentist approach is the common factor model of Anderson and Rubin (1956). As the strategy is also used within the empirical applications in Chapters 8 and 9, the following discussion is more detailed.

For each individual one obtains a measurement vector $\mathbf{T}$, which refer to a specific group of items. These in turn are deemed to represent a certain personality trait. However, there may be more than one trait dimensions $\boldsymbol{\theta}$ underlying the initial set of items $\mathbf{T}$. As such, the respective mean and covariance patterns are

$$\begin{aligned} \mathbf{T} &= \boldsymbol{\Lambda}\boldsymbol{\theta} + \boldsymbol{\nu} \\ \mathbf{S} &= \boldsymbol{\Lambda}\boldsymbol{\Psi}\boldsymbol{\Lambda}' + \boldsymbol{\Theta}, \end{aligned} \tag{5.1}$$

[*] The discussions presented in this chapter are published in Thiel and Thomsen (2013) and build on Thiel and Thomsen (2015).

where $\boldsymbol{\Theta}$ is the covariance structure of the unique components $\boldsymbol{\nu}$. Furthermore, $\boldsymbol{\Psi} = \mathbf{I}$ implies that the common factors are a priori uncorrelated and have unit variance for the sake of identification.[1] If the elements of $\mathbf{S}$ are additionally normalized it changes to a correlation matrix. The residuals have mean zero and are uncorrelated with all elements of $\boldsymbol{\theta}$ and among each other. Given these presumptions, the factor extraction is exclusively based on the observed matrix $\mathbf{S}$. The mean structure becomes irrelevant if the scores are already centralized around the corresponding observed means. Moreover, in order to reduce potential statistical artifacts resulting from the categorial nature of the responses, it is common to take the underlying nature of $\mathbf{S}$ into account. For categorial item responses so-called polychoric correlations are presumed. Consequently, the responses in $\mathbf{T}$ are based on a latent continuum $\tilde{\mathbf{T}}$, the correlations of which are the "true" ones. The diagonal elements of the transformed $\tilde{\mathbf{S}}$ are again unity, where only half of the off-diagonal elements have to be evaluated due to the symmetry of $\tilde{\mathbf{S}}$. For these $\tilde{S}_{ij}$, suppose the underlying continuous $\tilde{T}_i$ and $\tilde{T}_j$ are

$$\begin{pmatrix} \tilde{T}_i \\ \tilde{T}_j \end{pmatrix} \sim \mathcal{BN}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho_{ij} \\ \rho_{ij} & 1 \end{pmatrix} \right),$$

where both, $\tilde{T}_i$ and $\tilde{T}_j$, have mean zero and unit variance. This implies that

$$\Phi(\tilde{T}_i, \tilde{T}_j, \rho_{ij}) = \frac{1}{2\pi\sqrt{1-\rho_{ij}^2}} \int_{-\infty}^{\tilde{T}_i} \int_{-\infty}^{\tilde{T}_j} e^{\frac{1}{2(1-\rho_{ij}^2)}(\tilde{T}_i^2 - 2\rho_{ij}\tilde{T}_i\tilde{T}_j + \tilde{T}_j^2)} d\tilde{T}_i d\tilde{T}_j. \tag{5.2}$$

Given that for every two items (i, j) one has $k = 1 \ldots K$ and $l = 1 \ldots K$ response categories, one would obtain the (i, j)-specific log-likelihood contribution

$$\ell_{ij} = \ln C + \sum_{k=1}^{K} \sum_{l=1}^{K} n_{kl} \ln \eta_{kl}, \tag{5.3}$$

where η_{kl} is the cell probability of a response combination (k, l) which is observed n_{kl} times in the data, and C comprises the constant elements of the likelihood. The contribution for η_{kl} is obtained from the double-difference of the cumulated density function defined in equation (5.2), where the intervals of the differences depend on unknown cutoff-points

[1] Such normalization on either the factor loading or the corresponding factor variance are always required in factor models, as the overall scale is otherwise unidentifiable (see Anderson and Rubin, 1956). In exploratory factor models, it is common to normalize the factor variance. The orthogonality assumption that prescribes $\boldsymbol{\Psi}$ to be diagonal can be replaced by other restrictions on $\boldsymbol{\Lambda}$, which are hard to reason in exploratory settings, however. As diagonalization leads to parsimony in terms of underlying factors and can be relaxed later on, it is common to proceed in this fashion. Moreover, if only few items are available, non-orthogonality between factor may prevent identification.

$\gamma_{i,k}$ and $\gamma_{j,l}$ $(k = 1 \ldots K,\, l = 1 \ldots K)$. More specifically, one obtains

$$\eta_{kl} = \left[\Phi(\gamma_{i,k}, \gamma_{j,l}, \rho_{ij}) - \Phi(\gamma_{i,k-1}, \gamma_{j,l}, \rho_{ij})\right] - \left[\Phi(\gamma_{i,k}, \gamma_{j,l-1}\rho_{ij}) - \Phi(\gamma_{i,k-1}, \gamma_{j,l-1}, \rho_{ij})\right],$$

which, when substituted into the above likelihood, provides full-information estimates of the respective ρ_{ij}.[2] This procedure is repeated for all triangular item combinations in $\tilde{\mathbf{S}}$.

As a next step, the common factors that produce a relatively high share of the common variance and, at the same time, a high number of retained items, can be extracted from the estimated $\hat{\tilde{\mathbf{S}}}$ by common methods for factor extraction. A convenient choice is the principal factor analysis with iterated communalities, for which the covariance structure in equation (5.1) can be rewritten as follows.

$$\tilde{\mathbf{S}} - \boldsymbol{\Theta} = \boldsymbol{\Lambda}\mathbf{I}\boldsymbol{\Lambda}'$$

By definition, this step only affects the diagonal elements of $\tilde{\mathbf{S}}$, where the reduced values are called communalities. An initial estimate for the i-th communality $\hat{h}_i^2$ can be obtained from $1 - 1/r_{ii}$, where r_{ii} is the i-th diagonal element of $\tilde{\mathbf{S}}^{-1}$ (see Mulaik, 2009, for a derivation). An estimate $\hat{\boldsymbol{\Lambda}}$ of the factor loadings results from the factorization

$$\hat{\boldsymbol{\Lambda}}\hat{\boldsymbol{\Lambda}}' = \hat{\tilde{\mathbf{S}}} - \hat{\boldsymbol{\Theta}} = \mathbf{CDC}' = \mathbf{CD}^{1/2}\mathbf{D}^{1/2}\mathbf{C}',$$

which is the so-called spectral decomposition of the symmetric matrix $\hat{\tilde{\mathbf{S}}} - \hat{\boldsymbol{\Theta}}$ with $\mathbf{D}$ being the diagonal matrix of eigenvalues and $\mathbf{C}$ being the matrix of the corresponding characteristic vectors.[3] Furthermore, as $\hat{\tilde{\mathbf{S}}} - \hat{\boldsymbol{\Theta}}$ is positive semi-definite, so is $\mathbf{D}$. This allows for the factorization $\mathbf{D} = \mathbf{D}^{1/2}\mathbf{D}^{1/2}$ such that $\hat{\boldsymbol{\Lambda}} = \mathbf{CD}^{1/2}$. Since the factorized matrix is standardized, all factor loadings represent correlations and their item-specific sums are new guesses of the communalities. Hence, the communalities in $\hat{\tilde{\mathbf{S}}} - \hat{\boldsymbol{\Theta}}$ can be updated iteration-wise until they converge (see, e.g., Rencher, 2004). Sometimes the iterative nature leads to corner solutions. Such so-called Heywood cases (see Thompson, 2004) are usually discarded and the respective second-best combinations are used instead. Following Costello and Osborne (2005), it is expedient to end up with a "clean" factor structure where the factor loadings associate as much items as possible with one major common factor explaining most of the variance.

---

2 There are also three-step procedures based on conditional likelihood estimates available, but combination-specific cutoff-estimates ought to perform better than row-specific first-stage estimates in most cases.

3 Each column of $\mathbf{C}$ forms a characteristic vector with orthonormalization such that $\mathbf{c}'_i\mathbf{c}_j = 0\ \forall i \neq j$ and $\mathbf{c}'_i\mathbf{c}_j = 1\ \forall i = j$.

5.2 Accounting for Measurement Error

When unidimensional item sets are identified, the next step is to find an appropriate method to account for measurement error. There are occasions when measurement error apparently is of minor importance, usually when measured traits are employed as an outcome variable in program evaluation settings. The common aim of declaring a personality trait as a dependent variable is to examine environmental influences on its formation process. The implications of measurement errors in such settings depend on whether a regression based approach (see Wansbeek and Meijer, 2000) or a program evaluation approach is uses (see Imbens and Wooldridge, 2009). Apart from the case of randomized program evaluations, such environmental influences are not exogenous to the individual's personality.[4] This follows from the points made in the first three chapters. As discussed by Cunha and Heckman (2008), the multiplicity and self-selectivity of investments and environments that foster the development of personality traits causes a general endogeneity problem. To overcome the resulting consistency issues, one always needs some structural assumptions that, in the simplest case, comprise no more than an exclusion restriction. Todd and Wolpin (2003) provide some estimation strategies that follow similar notions and are relatively simple to use. Without such data features, however, more of the presumed underlying processes have to be modeled. This usually leads to more involved estimation approaches, such as those discussed by Cunha et al. (2010). I will not discuss both types of models here any further as similar applications are less frequent. However, Chapter 8 will discuss a specific program evaluation setting in greater depth.

Most research questions that deal with personality traits or their relation to human capital outcomes, incorporate them as explanatory variables. In this case, the parameter consistency in standard regression approaches is jeopardized as well. Comparable to the case above, instruments are almost impossible to be found for determinants like personality.[5] As an alternative, one can try to correct standard estimates for the inherent measurement error and avoid settings with obvious simultaneity, or one can use latent variable approaches imposing some additional structure.[6] I will briefly discuss both approaches in what follows.

4 It should be noted however, that the precisions of estimates obtained in a program evaluation context still can be impaired by measurement error.

5 See Card (1999) for a discussion on these points in context of wage determination.

6 I follow the definition of Aigner et al. (1984) for latent variables. According to their definition, as opposed to unobserved variables, latent variables cannot be represented as a linear combination of observed variables.

5.2.1 Adjusted Regression

The virtue of measurement error correction, as opposed to factor analytic or more structural approaches, is its simplicity. The relatively simple estimation, however, comes at the cost of requiring relatively precise information about the magnitude of measurement error. Such a source of information may be one of the reliability measures addressed above, which however impose very strong assumptions on the relation between true and measured scores (see the above discussion). For instance, Cronbach's alpha requires the scaling parameters between measured and true score to be equal across items in order to yield a consistent reliability estimate (see Bollen, 1989, for discussion). For most measures this assumption does not hold and reliability is therefore underestimated. Given that a consistent estimate of the share of measurement is available, it is straightforward to adjust least square estimates by weighting the variation in the erroneous explanatory variables. In the univariate case one would simply use

$$\hat{\beta}_A = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2 - nVar(\text{error})}.$$

Using an arbitrary coefficient of reliability ρ, this expression can also be written as

$$\hat{\beta}_A = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\rho \sum_{i=1}^{n}(x_i - \bar{x})^2}.$$

The multivariate case is derived by Schneeweiss (1976), among others. Besides the fact that accounting for measurement error requires consistent coefficients of reliability, no solution for the often accompanying simultaneity is provided when choosing this approach. To resolve this problem structural approaches with latent variables have to be used instead.

5.2.2 Methods based on Factor Analysis

Latent variable or factor models are a generalization of error-in-measurement (EIV) models. In either case the observed personality score is a manifestation of the latent true score (see Aigner et al., 1984). However, the aim of EIV and factor models is fundamentally different. The former primarily intends to obtain consistent estimates when some explanatory variables are erroneous. In contrast, latent factor approaches do not only aim at removing some nuisance, but also at the estimation of structural parameters that represent the relationship between latent factors and observed response variables, as well as the estimation of individual scores of the latent factors.[7] The fundamental equations for

[7] To illustrate this close but differing relation, consider the following representations for an erroneous explanatory variable and a simple factor model with T representing the observed manifestation of

the analysis of factor models are the same as those in equation (5.1), possibly with an additional intercept. Following the terminology from the econometric literature, parameters that are related to the model structure (such as the loadings) are referred to as structural parameters, whereas those parameters that vary across observations, like latent trait scores, are denoted incidental parameters.[8] The most common estimation approaches are different kinds of maximum likelihood methods (see Skrondal and Rabe-Hesketh, 2004, for an overview).

In EIV models it is common to maximize the conditional likelihoods iteratively (see Aigner et al., 1984) or to integrate out the latent factor numerically in order to obtain a closed form expression (see Skrondal and Rabe-Hesketh, 2004). In case of the former approach, stepwise conditional maximization is required as simultaneous estimation of structural and incidental parameters can cause severe consistency problems (see Neyman and Scott, 1948).[9] In case of integrating out the latent factors and maximizing the marginal likelihood, one has to impose quite restrictive distributional assumptions on the latent factors. When assuming the latent factors to represent personality traits, such distributional predefinitions are at least questionable (see Heckman et al., 2006). Yet, both approaches provide no inference about the individual specific latent factor scores whatsoever.

For that reason, factor analysis or factor structure models additionally seek to provide estimates for trait scores.[10] This generally requires some further identification restrictions on the latent factors and their factor loadings (see, e.g, Jöreskog, 1977, Aigner et al., 1984, for general discussions). The identification problems follow the same notions as those for the exploratory factor model addressed above. Due to the somewhat different practical aim, other identification restrictions are commonly imposed, however. The first identification problem is not related to any consideration of the multiplicity of equations that represent the latent traits. It results from the fact that, even in the single equation

a latent factor θ and an unexplained residual ε.

$$T = \lambda\theta + \varepsilon$$
$$T = \theta + \varepsilon.$$

The obvious difference is that factor analysis is interested in identification of both, factor loadings λ and latent factors scores θ. In the EIV case, the aim merely is to obtain a consistent estimate for the intercept $\theta + \varepsilon$, or for the slope parameters if covariates would have been included.

8 The distinction between structural and reduced-form parameters that is used in the econometric literature on simultaneous equations (see Intriligator, 1983) can be maintained for factor structure models as well.

9 Baker and Kim (2004) discuss assumptions for iterative estimation to resolve this problem.

10 For the discussion provided here, the choice of factor analysis or factor structure models does not imply any differences. Factor structure models simply extend factor analytic measurement systems in that they allow for structural dependencies between latent traits and also for the incorporation of observable covariates.

case, multiplying a factor loading with an arbitrary scalar is observationally equivalent to the corresponding factor being divided by the same scalar. More formally $\tilde{\lambda} = \lambda \cdot c \Leftrightarrow \tilde{\theta} = \frac{1}{c} \cdot \theta$. There is an infinite number of such observationally equivalent combinations of factors and loadings.

An approach that is often pursued to warrant identification is to fix the variances of the latent traits to unity and to impose independence among latent factors, at least if more than one trait is considered within the framework. Additional sign restrictions on the factors are needed in this case, as the association between variances and the signs of the corresponding factors is twofold. Moreover, across all employed measurement equations, some lower-triangular form for the matrix of factor loadings has to be prespecified (see Geweke and Zhou, 1996). In general, restrictions on factor loadings can be relaxed to some extent when the mean structure of the equation system is incorporated in addition to its covariance structure.[11] If theoretically justifiable, another form of restriction is preferable, though. Depending on the number of latent factors and the number of measurement equations, some known associations between a certain trait and a certain observable measure can be exploited for this purpose. If one sets a particular factor loading in the measurement system to unity, Carneiro et al. (2003) show that this normalization is an alternative to the variance and sign restrictions explained above. Given a subtle choice of this normalization one can anchor the estimated parameters into appropriate real world outcomes (see, e.g., Cunha and Heckman, 2008) and thus assign an interpretable metric to them. It should be noted, however, that the scale for the remaining loadings can still be somewhat arbitrary, which implies that the estimates are only interpretable to a limited extent. One may overcome this limitation in simulating the estimated model with different conditional data points (see Piatek and Pinger, 2010, for an example).

As with the estimation of errors-in-variables models, obtaining a closed form of the objective function is a major issue. The LISREL approach by Jöreskog (1977) estimates the parameters of the complete model by minimizing the discrepancy between the sample correlation matrix and the correlation matrix imputed by the model. This can be conducted by different estimation techniques such as maximum likelihood or least squares (see Bollen, 1989, for a comparison of the different approaches with regard to their efficiency).[12] Under normality assumptions for the vector of observable measures and for the latent traits, a particular convenient form for likelihood estimation results. It is the only approach that directly leads to a closed-form likelihood. Given normally distributed measurement variables as well as normally distributed latent traits, it can be derived

[11] Usually whenever the measured scores are not centralized around zero.

[12] The major virtue of the LISREL approach is its still steadily maintained implementation as a ready-to-use software package.

that the covariance matrix of the measurement system is Wishart distributed (see, e.g., Anderson, 2003). The resulting discrepancy function (see Jöreskog, 1967) then is

$$F = \ln|\hat{\mathbf{\Sigma}}| + tr(\mathbf{S}\hat{\mathbf{\Sigma}}^{-1}) - \ln \mathbf{S} - (2j) + (\mathbf{m} - \hat{\boldsymbol{\mu}})'\hat{\mathbf{\Sigma}}^{-1}(\mathbf{m} - \hat{\boldsymbol{\mu}}), \tag{5.4}$$

with $\mathbf{S}$ and $\mathbf{m}$ representing the empirical covariances and means, and $\hat{\mathbf{\Sigma}}$ and $\hat{\boldsymbol{\mu}}$ being their estimated counterparts implied by the model structure. If the measurement scores are standardized and thus the estimation only builds on the covariance structure, the latter term of the sum can be discarded. As noted above, it is difficult to vindicate the very strong normality assumption in practice (in this case, even a dual one). It is relatively easy to relax it at least for the observed measurement variables. However, algebraic simplifications as in equation (5.4) do not exist any longer in this case. Closed-form estimation of the structural parameters then again has to draw on integrating out the latent factors by numerical or Monte Carlo methods (see Skrondal and Rabe-Hesketh, 2004).

Having chosen one of the estimation approaches for the structural parameters discussed thus far, a logical next step for any practical application would be to obtain estimates of the factor scores for each individual. Given the structural estimates and the individual specific realizations of the measurement equations, one can use a simple linear projection onto the factor continuum (see, e.g., Heckman et al., 2013). It can be shown that for an individual specific scalar trait θ_i and the corresponding response vector $\mathbf{T}_i$

$$\theta_i = \mathbf{L}'\mathbf{T}_i = (\tilde{\boldsymbol{\lambda}}'\mathbf{\Psi}^{-1}\tilde{\boldsymbol{\lambda}})^{-1}\tilde{\boldsymbol{\lambda}}'\mathbf{\Psi}^{-1}\mathbf{T}_i, \tag{5.5}$$

where $\tilde{\boldsymbol{\lambda}}$ is the vector of loadings of θ_i across all measurement equations.[13] Evidently, this term simplifies when the assumption $\mathbf{\Psi} = \mathbf{I}$ is maintained. The projection in equation (5.5) will almost never have an exact solution. Most conveniently, some L^2-approximation like least squares in case of $\mathbf{\Psi} = \mathbf{I}$ or generalized least squares for arbitrary $\mathbf{\Psi}$ can be used to obtain a consistent estimate of θ_i.

A quite critical assumption in the classical factor structure models discussed thus far, is the strong distributional dependence of the procedures with respect to the latent factors. As already mentioned, especially for latent personality traits, this presumption is often inappropriate (see Heckman et al., 2006). Carneiro et al. (2003) discuss identification assumptions for more general types of factor structure models that allow for correlated and non-normal factors. Another drawback of classical factor models is that they com-

[13] The tilde indicates this, so as to not confuse $\tilde{\boldsymbol{\lambda}}$ with $\boldsymbol{\lambda}$, which usually stands for the vector of loadings within one measurement equation but across latent factors.

monly presume linear responses in the measurement or outcome equations. In particular when not only traits specific items, but also real outcome variables are included into the structure, linearity in factor seems too restrictive in general. Though generalizations to different types of link functions are known and existent (see Skrondal and Rabe-Hesketh, 2004, for an overview), the promptly increasing computational complexity of such model is often regarded as daunting in the applied literature.

Once again, the identification of such relaxed models can still be based on exploiting covariance structures, depending on the nature of the measures employed.[14] Moreover, one can relax the independence among latent factors when so-called dedicated measures are a priori defined. This means that one or more measurement variables exclusively represent one latent factor, the same set of measures and some additional ones represent the first and a further factor, and so on.[15] The theoretical fundamentals to vindicate such proceeding depends on the specific setting (see Hansen et al., 2004, for such an application).

Given covariance structures, only the first two moments of the distributions of latent factors can be identified, however.[16] This is a major problem when relaxing the normality assumption toward more general distributions with higher moments. Carneiro et al. (2003) show conditions under which the complete distribution of a latent factor is nonparametrically identified. Summarizing their relatively involved discussion on this point, the identification requires combinations of continuous and discrete response variables. The latent traits can then be represented by finite mixtures of normals, which provide enough flexibility to approximate most surmisable distributions of θ, given that the number of mixture components is sufficiently high (see Diebolt and Robert, 1984).[17]

For estimation of these kinds of models, extensions of maximum likelihood and least square methods described above for ordered discrete response variables are available (see Jöreskog and Moustaki, 2001, for an overview). Likewise, finite mixtures of normals can be accommodated by common estimation methods such as maximum likelihood. Unfortunately, this requires numerical approximations of various multidimensional integrals.[18] As such, these methods become more and more intricate as the number of factors, mixture components, or the number of discrete response equations increases. The same holds for

[14] Recall that this has also been the case for factor models with normally distributed latent traits.

[15] Without this restriction, it can be shown that factor models with dependent latent factors are observationally equivalent to factor models with independent latent traits.

[16] More precisely, only the first two moments of every random variable involved into estimation procedure can be identified in that way.

[17] The implementation is also discussed by Piatek (2010).

[18] As techniques for numerical integration are used quite extensively within the applications that are presented in Chapters 8 and 9, a thorough discussion about the underlying theory is provided in the first part of Appendix A.

most other suitable frequentist approaches like expectation maximization (see Dempster et al., 1977) or maximum simulated likelihood (see Gouriéroux and Monfort, 1991). More advantageous approaches drawing on Bayesian techniques, particularly Bayesian Markov Chain Monte Carlo methods or MCMC (see Gelman et al., 1995), have evolved in the last years due to a considerable progress in computational speed. The basic principle of Bayesian estimation is to enhance the imposed assumptions on the data generation process, which is the likelihood in frequentist parlance, by prior beliefs about the parameter distributions. Applying Bayes' Theorem yields a posterior distribution that unifies the assumptions made on the data generating process and on the unconditional beliefs about parameter distributions. Somewhat more formally, this implies that

$$p(\boldsymbol{\Theta}|data) = \frac{f(data|\boldsymbol{\Theta})f(\boldsymbol{\Theta})}{f(data)} \propto L(data|\boldsymbol{\Theta})f(\boldsymbol{\Theta}), \tag{5.6}$$

where $\boldsymbol{\Theta}$ is the relevant parameter set, $p(\boldsymbol{\Theta}|data)$ is the posterior distribution, $L(data|\boldsymbol{\Theta})$ is the data generating process or the likelihood, and $f(\boldsymbol{\Theta})$ is the prior distribution of the parameters. A comprehensive introduction to the topic including discussion of consistency and asymptotic behavior is provided by Geweke (2005). Estimates from the posterior distribution can be obtained in different ways. One possibility is to compute marginal distributions for the parameters of interest from the joint posterior $p(\boldsymbol{\Theta}|data)$ by means of numerical or Monte Carlo integration. Respective moments of the marginals can be easily obtained then. Another way is to directly simulate draws from the posterior. However, both these methods basically suffer from the same problems in evaluating high-dimensional integrals as the above frequentist approaches do. The problem is simply postponed to another step.

This is where MCMC algorithms come into play (see, e.g., Gilks et al., 1996,). A highly convenient property of Bayesian MCMC approaches is their ability to deal with high-dimensional integrals, which, apart from any general controversy about Bayesian versus frequentist approaches, make them a particularly attractive method for the estimation of factor models with more complex structures. This is done by sampling from chains of distributions that are more simple in nature than the joint posterior. The simulation finally converges to draws from the joint posterior. The resulting chains fulfil the Markov condition in that transition probabilities between states only depend on the current state. After a sufficient number of iterations, the probability of being in a particular state of the parameter space is supposed to be independent of the initial state. Such chains are called stationary. The most appropriate feature for estimation of factor structure models has the Gibbs sampler (Geman and Geman, 1984), as it samples from chains of rather simple univariate conditional parameter distributions. Following the Markov property, these conditional draws are supposed to converge to draws from the joint posterior after a sufficient number of repetitions.

What actually completes the fit for factor models, is its extension due to Tanner and Wong (1987), the so-called Data Augmentation (see also van Dyk and Meng, 2001). It uses an astonishing contrivance for the computation of all the unobserved components that arise in non-linear factor settings with non-normal latent traits. These components usually comprise latent index variables, cutoff points, as well as latent factors and their corresponding mixture parameters. Instead of specifying the posterior of the parameters conditional on observed data and on latent components, one can factorize the posterior into distinct posteriors for parameters, latent responses, and latent factors. The Gibbs sampling is then sequentially conducted over the elements in these sub-groups, namely exactly in this order (see Diebolt and Robert, 1984).[19] The procedure thus is a multi-step version of the Gibbs sampler. The sampling of all latent components is also called the imputation step, whereas sampling of parameters conditional on the imputation step is called the posterior step. The algorithm cycles between imputation and posterior step until convergence.[20] A precise summary for practitioners along with some refinements to improve convergence is provided by Piatek (2010).

The described algorithm makes the use of follow-up steps for estimation of the trait scores (as in equation 5.5) obsolete. After a so-called burn-in phase, the sampled results for the latent factors conditional on individual specific data can be stored. After convergence, the respective last value can be used apart, or together with the previous draws for the estimation of additional precision statistics or other moments.

5.2.3 Item Response Theory

Another strand of the psychometric literature which arose from classical test theory is item response theory (IRT, see Sijtsma and Junker, 2006, for a historical classification). It traces back to the work of Lazarsfeld (1950) and Lord (1952). The breakthrough contributions in the psychometric field are due to Lord and Novick (1968) and Samejima (1952). The basic notion that IRT exploits is to establish a probabilistic relation between latent traits θ and categorical responses. A straightforward way to illustrate an item response model is to consider a dichotomous item. If a respective trait is positively related to the item responses, the response probability is usually modeled by means of a sigmoid response function like the normal or logistic distribution function. This yields the so-called

[19] See also Piatek (2010) for a specific discussion in context of factor models.
[20] In that sense, Data Augmentation is the Bayesian equivalent to the Expectation Maximization algorithm of Dempster et al. (1977).

item characteristic curve (ICC).[21] When using a normal distribution, the mean provides the scale location, which for some kinds of traits represents the difficulty of the item, whereas the variance determines the discriminatory power of the item. An item characteristic curve can be estimated for each item by using all available sample observations. The extension to polytomous responses is analogous. The ICC for the lowest response on the scale has an inverted shape, i.e., it decreases with increasing θ, whereas the ICC for the highest response has the usual sigmoid shape of some normal or logistic distribution. The response realizations in between these two cases have bell shaped probabilities with their locations shifting from left to the right as the response category, and thus the latent trait θ, increases. A more convenient functional form for maximum likelihood estimation of the relevant parameters is obtained when the characteristic curves are cumulated. For K response categories this yields $K-1$ non-intersecting and monotonic cumulated characteristic curves that all have the intended sigmoid shape. Non-intersection is guaranteed when the discrimination parameter (or variance) is restricted to be the same across all cumulated ICCs. It follows that the ICCs that set up the likelihood can be simply stated as differences of adjacent cumulated curves.[22] Identification of such parametric item response models is not subject to general indeterminacies like factor rotation and follows quite simple rules of thumb (see Skrondal and Rabe-Hesketh, 2004). Most scales realizations are commonly identified with as little as three items per trait.

The aim is then to estimate the parameters determining the shape of the ICC, i.e., the structural parameters, and the incidental parameters θ_i that represent the respective location of the assessed individuals on the latent trait scale. Given some independence assumptions, traditional estimation approaches commonly use an iterative maximum likelihood procedure that cycles between conditional likelihoods for incidental and structural parameters (see Birnbaum, 1968). As with the above discussion of factor models, joint estimation of structural and incidental again lead to inconsistency due to the result by Neyman and Scott (1948). Joint estimation is only feasible for so-called Rasch Models (see Andersen, 1972), which are models with a discrimination parameter generally fixed at unity. Enhancements in computational speed have led to more robust strategies like the Bock-Aitken solution (see Bock and Aitken, 1981) that uses Expectation Maximization over numerically integrated marginal likelihoods.[23]

[21] For the opposite relation the corresponding survival function can be used.

[22] Though the order of the differences can be inverted, this formulation of the likelihood is the same as for an ordered response model without covariates.

[23] See Baker and Kim (2004) for a discussion of further methods. Commonly, item response models can also be reformulated as Generalized Linear Latent and Mixed Models (GLLAMM) and estimated accordingly (see, Rijmen et al., 2003, Rabe-Hesketh et al., 2007, for a detailed treatise). GLLAMM procedures are readily available in statistical packages.

A potential drawback of the techniques discussed thus far is that they impose several parametric assumptions to enforce properties like monotonicity and non-intersection. There are less parametric approaches available, though, all of which use approaches from the econometric literature that rely on semiparametric estimation approaches.[24] Some general conditions for the cumulated ICCs in these types of models are established by Spady (2006). These include monotonicity, stochastic dominance and local independence assumptions, where the latter means independence of the item responses conditional on the latent traits.[25] By construction, monotonicity is necessary to give the cumulated ICCs the intended interpretation. Stochastic dominance is one possible way to imply non-intersection.

Suggestions for the implementation of potential estimation procedures based on maximum likelihood are provided by Spady (2007). It largely depends on the specifics of the data to what extent parametric assumptions can be relaxed. One possible realization is to use a Sieve maximum likelihood strategy established by Grenander (1981), with order of approximation increasing as the sample size increases.[26] It should be noted, however, that the convergence of such flexible likelihood functions can be poor for quite low numbers of item-curve specific parameters, already.[27] As such, I will focus on a variant with only three curve specific parameters that still accommodates the most common curve shapes. The identification of this specific model under fairly general premises is established in Appendix B.[28]

Again, consider there is a positive relation between latent traits and the corresponding responses on a K-point Likert-scale. For each of J potential items representing θ, the

[24] To be more specific, following the notation of Chen (2007), these methods are semi-nonparametric as both structural and incidental parameters can defined as being infinite-dimensional. See also Härdle et al. (2004) and Horowitz (2009) for general overviews on semiparametric and semi-nonparametric estimation approaches.

[25] This conditional independence also implies that background characteristics affect the response probabilities only through latent abilities.

[26] Sieve methods can be extended to other estimators than maximum likelihood. Without distributional assumptions the parameter space for the response functions is infinite and maximization of the criterion function is therefore infeasible. The method of Sieves defines a series of approximation spaces in order to reduce dimensionality of the previously infinite dimensional parameter space. For concave optimization problems with finite dimensional linear sieve spaces, this technique is also denoted series estimation (see Geman and Hwang, 1982, Barron and Sheu, 1991, and Chen, 2007 for technical overviews). Appropriate base functions are orthogonal polynomials, trigonometric polynomials and shape-preserving splines, just to mention a few (see Härdle, 1995, and Chen, 2007).

[27] For practioners, it may be worthwhile to implement this model in association with a parameter-bounded numerical optimization procedure, such as the one by Zhu et al. (1997). Otherwise, machine precision problems are bound to arise, especially at the boundaries of the support for θ, where the cumulated item curves get very close to each other.

[28] The discussion of this method is more detailed as the two later empirical applications I am going to discuss build on it.

probability of response K approaches zero for low θ and one for very high θ. Conversely, the probability of giving response $k = 1$ is one for very low levels of θ and zero for high levels. As in the parametric case, responses $k = 2$ to $k = K - 1$ have bell-shaped probabilities with their location shifting from left to right as k and θ increase. For K responses, in general $K - 1$ cumulated response curves $P(r \leq k|\theta)$ that are monotonically decreasing in θ and non-intersecting have to be estimated. Non-intersection is bound to arise by setting up the cumulated response probabilities as follows.

$$\begin{aligned}
&P(r \leq K-1|\theta) = 1 - G(K-1|u)\\
&P(r \leq K-2|\theta) = [1 - G(K-2|u)]\ P(r \leq K-1|\theta)\\
&\qquad\vdots\\
&P(r \leq 1|\theta) = [1 - G(1|u)]\ P(r \leq 2|\theta)
\end{aligned} \tag{5.7}$$

Since $P(r \leq K-1|\theta)$ is decreasing in θ, $G(K-1|u)$ is increasing in it via some monotone mapping u. Spady (2007) establishes general conditions under which polynomial series are a flexible way to approximate the respective $G(k|u)$ for $k = 1, \ldots, K$. More explicitly, one may use shifted Legendre polynomials of the third degree (see Judd, 1998) to approximate $G(k|u)$ using an exponential tilting factor as in Barron and Sheu (1991). Then one has

$$G(u) = \frac{\int_0^\theta e^{t_1\gamma_1(u)+t_2\gamma_2(u)+t_3\gamma_3(u)}du}{\int_0^1 e^{t_1\gamma_1(u)+t_2\gamma_2(u)+t_3\gamma_3(u)}du}, \tag{5.8}$$

where the transform $u = \Phi(\theta)$ is used to match the support of θ with the domain $[0, 1]$ of the polynomial basis, and t_1 to t_3 are the parameters to be estimated for each θ and its corresponding items. As such, there are $3 \times K - 1 \times J$ structural parameters to be estimated in the given setting. The bound θ is a placeholder for the respective $u = \Phi(\theta)$. Hence, every $P(r = k|\theta)$ can simply be expressed as differences of the respective cumulated response curves as modeled in equation (5.7). The resulting likelihood contribution for the ith individual is

$$\begin{aligned}
p_i(r_{i1}, r_{i2}, \ldots, r_{iJ}) &= \int p_{ip}(r_{i1}, r_{i2}, \ldots, r_{iJ}|\theta_i) f(\theta_i) d\theta_i\\
&= \int p_{i1}(r_{i1}|\theta_i) p_{i2}(r_{i2}|\theta_i) \ldots p_{iJ}(r_{iJ}|\theta_i) f(\theta_i) d\theta_i.
\end{aligned} \tag{5.9}$$

The second expression requires local independence, which states that all individual characteristics that may influence the response probabilities are conveyed in θ. Hence, response probabilities across items $j = 1, \ldots, J$ are independent conditional on θ. In order to obtain a likelihood expression unconditional on unobserved traits, θ_i can easily be integrated out by assuming $f(\theta)$ to be $\mathcal{N}(0,1)$ distributed. A consideration of the approximation errors of this expression can be found in the second part of Appendix A.

The implementation of the objective function that has to be set up from the components in equations (5.7) to (5.9) involves a high number of recurrences and loops. Moreover, the polynomial bases change from step to step, making the resulting transformations hard to "vectorize" by means of built-in functions common to most statistical packages. As such, it is often annoyingly inefficient to code up and estimate the model in interpreted statistical languages such as R or Stata. For that reason, it is recommendable to swap these routines to a compiled lower-level language such as C or Fortran.[29] Appendix C provides one possible implementation using Fortran.

Given the estimated polynomial coefficients obtained from equation (5.9), it is possible to predict an individual's θ_i by means of a so-called Empirical Modal Bayes approach (see Skrondal and Rabe-Hesketh, 2004). Each response pattern uniquely determines θ_i by finding the mode of the implied empirical posterior.

$$f(\theta_i|r) = \frac{f(\theta, r)}{p(r)} = \frac{p_1(r_1|\theta)\dots p_J(r_m|\theta)f(\theta)}{\int p_1(r_1|\theta)p_2(r_2|\theta)\dots p_J(r_J|\theta)f(\theta)d\theta} \tag{5.10}$$

The denominator is approximated in the same way as the integral in equation (5.9) (see also Appendix A for a more detailed discussion).

[29] Just-in-time compiled interfaces such as Stata's Mata language may provide similar performance, but the compiled general-purpose solution may be preferable due to its unlimited extensibility to existent open source libraries.

6

Direct and Indirect Outcomes of Personality Traits*

Until recently, personality traits have not played an important role in explaining labor market outcomes. Bowles et al. (2001) review that quite early contributions to the literature on the explanation of wage differentials came up with concepts like disequilibrium rents (Schumpeter, 1934) and incentive-related effects (Coase, 1937). Disequilibrium rents are producer rents and surpluses that are induced by technical change, product innovation, changes in business organization, and by other forms of external shocks. Traits determine in how far employees differ in their ability to gain from these rents. Incentive effects arise as certain traits act incentive-enhancing. Both these explanations have an intuitive association with personality traits. In the sense of these concepts, the respective traits can also be construed as not necessarily being productivity enhancing. Bowles et al. (2001) also stress that it is important to distinguish different segments of the labor market. Two illustrative examples demonstrate this: in a working environment where monitoring is difficult, behavioral traits like truth telling may be higher rewarded than in other cases. Considering a low-skill labor market, docility, dependability, and persistence may be highly rewarded, whereas self-direction may generate higher earnings for someone who is more of a white-collar worker. Besides different rewards in different occupation segments of the labor market, people also opt in these occupations owing to their personality traits (see Antecol and Cobb-Clark, 2010). By similar arguments, John and Thomsen (2014) show that returns to specific traits within occupational groups provide sort of a mixed signal due to group-specific returns and self-selection.

To that effect, it is difficult to determine if certain traits increase wages by affecting occupational choice, productivity, or if market mechanisms additionally induce wage premiums along the lines of (Schumpeter, 1934) or (Coase, 1937) for certain traits. On a more general level Borghans et al. (2008c) show that supply and demand regarding workers who are more or less endowed with "directness" relative to "caring", create a wage premium for directness. Another explanation is that the society solidifies certain expec-

* The discussions presented in this chapter are published in Thiel and Thomsen (2013).

tations about appropriate traits and behavior, and rewards or punishes individuals who deviate from them in either direction. This interpretation is fostered by the results of Mueller and Plug (2006) for the gender wage gap in US data. They show that particularly men obtain a wage penalty for Big Five agreeableness, a trait stronger associated with women.

The aim in what follows is to give a very short review of empirical studies that deal with predictive power of various personality traits. Tables 6.1 and 6.2 provide some key facts for a small selection of studies. A more widespread overview on studies that relate traits to various outcomes is given in Borghans et al. (2008a) and Almlund et al. (2011), including a larger focus on literature from other disciplines. In summary, the results presented in Tables 6.1 and 6.2 indicate that personality traits have a substantial impact on labor market remuneration, making the promotion of certain abilities and traits a worthwhile policy objective. However, many mediating variables, like educational achievements, seem to operate on the facilitation of the later labor market success.

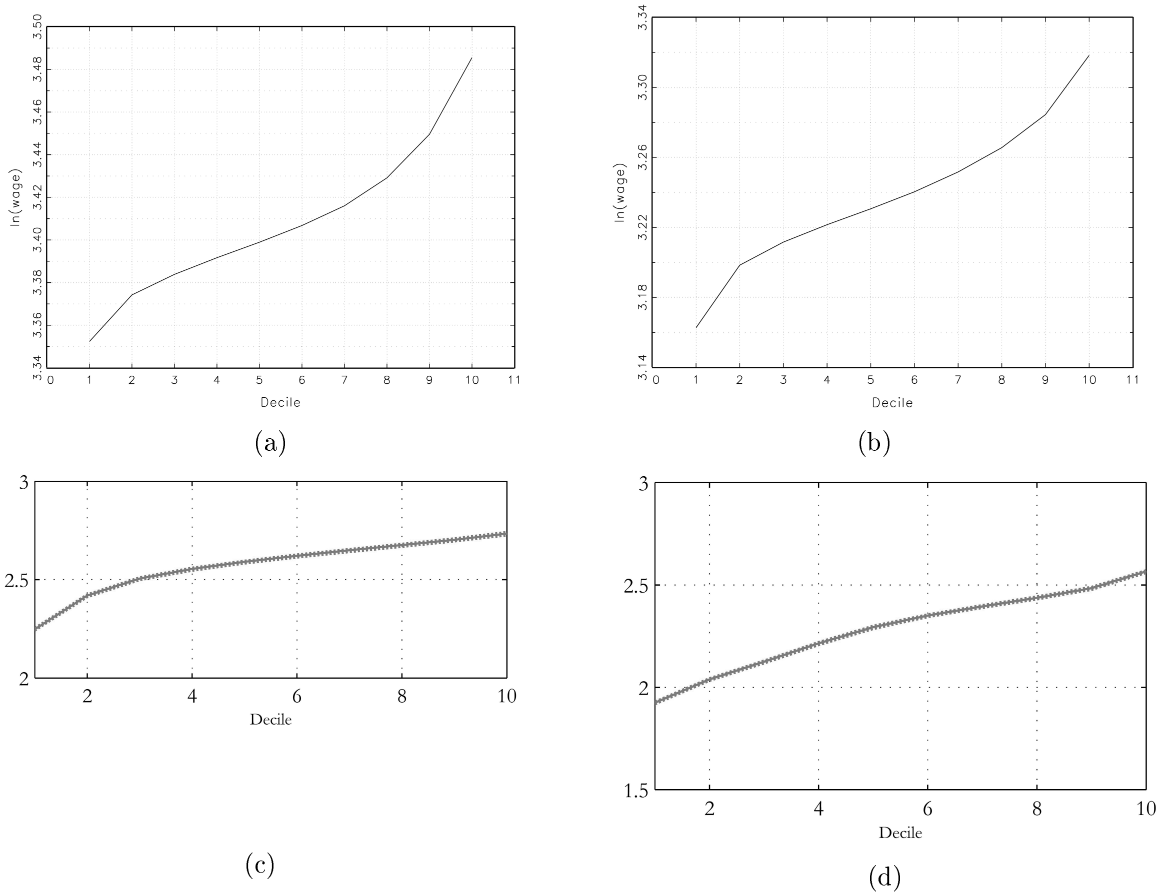

Figure 6.1: Net effect of control related traits on log wages for 30-year old males and females in Germany and the United States. Upper panel: males (a) and females (b) in Germany. Lower panel: males (c) and females (d) in the United States. Sources: Flossmann et al. (2007) and Heckman et al. (2006).

Irrespective of how exactly traits are valued in the labor market, they explain differences in the earnings structure relatively well. Picking out some of the represented studies clarifies this point. Heckman et al. (2006) provide empirical evidence on the effects of traits related to self-control and self-esteem on log hourly wages. Especially for the lower deciles of the distribution of latent control traits, a strong influence is revealed. Flossmann et al. (2007) reproduce these results for German data with quite similar findings. Figure 6.1 compares these net effects on log wages obtained in the two studies. Particularly for the upper and lower deciles of the distribution of internal control, the marginal effects are higher. Both results provide an important point on how personality traits affect earnings.

Abilities and traits do not solely affect wages, but educational outcomes as well. Presumably, the major effects of abilities on wages are mediated through the endogenous schooling choice (see Piatek and Pinger, 2010). The structural approach pursued by Heckman et al. (2006) and Flossmann et al. (2007) accounts for this issue. Besides wages in general, Heckman et al. (2006) also assess the effects of cognitive abilities and personality traits on wages given certain levels of schooling, as well as on the probability of obtaining certain educational degrees. For instance, for males, control related traits hardly affect the probability of being a regular high school dropout, but rather promote the probabilities of being a GED participant, of graduating from high school, of graduating from a two-year, and from a four-year college.[1]

Hence, it is of particular interest to identify which traits affect educational performance and along with it, schooling choices. Duckworth and Seligman (2005) show that self-discipline even exceeds the explanatory power of IQ in predicting performance at school. They define self-discipline as a hybrid of impulsiveness and self-control. Highly self-disciplined adolescents outperform their peers on all inquired outcomes including average grades, achievement-test scores, and school attendance.

The choice of self-discipline as a personality trait of particular interest with regard to educational achievement is related to the findings by Wolfe and Johnson (1995). They assess which trait is most eligible for predicting grade point averages (GPA) in a sample of 201 psychology students. The outstanding GPA predictors are measures displaying the level of control and items closely related, like self-discipline. Thus, besides cognitive skills, some personality traits play an equally important role in affecting schooling choices or years of schooling, respectively.

Since personality is malleable throughout adolescence and IQ is fairly set earlier in life (the next chapters will elaborate on this point), the inverse causation also applies. This

[1] GED stands for General Educational Development and is a test that certifies college eligibility of US high school graduates.

induces the aforementioned simultaneity. Hansen et al. (2004) determine causal effects of schooling on achievement tests. They reveal that an additional year of schooling increases the Armed Forces Qualification Test (AFQT) score by 3 to 4 points. Achievement tests provide a mixed signal constituted of IQ and personality traits (see Borghans et al., 2011), where IQ is relatively stable from school age on.

Personality traits also exhibit a substantial influence on social outcomes. Closely related to the previously discussed wage achievements are employment status and mean work experience, which are likewise affected by the personality.[2] Further outcomes, like the probabilities of daily smoking, of incarceration, and of drug abuse, are also significantly determined by control related traits, albeit to different extents.

[2] See also Heckman et al. (2006) for a detailed discussion and the corresponding effect sizes.

Table 6.1: Overview of Studies Analyzing the Effects of Personality Traits on Various Outcomes

Study	Data	Sample Size	Time Horizon	Research Question	Method	Results
Coneus and Laucht (2011)	Mannheim Study of Children at Risk (MARS)	$N = 384$	Personality traits at ages 3 months and 2 yrs., outcomes between ages 8 and 19	Effects of early temperamental measures (expert ratings) on grades and social outcomes	Children fixed-effects	Particularly low attention span (benchmark: high attention span) shrinks math grades ($\beta = .55^{***}$), grades in German ($\beta = .35^{***}$), number of delinquencies ($\beta = 3.11^{***}$), probability of smoking ($\beta = .15^{**}$), and number of alcoholic drinks per month ($\beta = 22.77^{***}$)
Duckworth and Seligman (2005)	Two consecutive cohorts of US eighth-grade high school students	$N_1 = 140$, $N_2 = 164$	2002 and 2003	Effect of self-discipline (composed of standardized self-control and impulsiveness ratings) on grade point average	OLS	One standard deviation increase in self-discipline increases GPA by $.1^{***}$ for the 2002 and by $.08^{**}$ for the 2003 cohort
Coneus, Gernandt, and Saam (2011)	German Socio Economic Panel (SOEP) youth questionnaire, waves 2000-2005	$N = 3,650$	Personality traits, academic skills, and background variables were assessed from 2000 on, information on school tracks up to (including) 2005	Effect of internal control at age 17 on later educational dropout (up to age 21)	Correlated random effects	One standard deviation increase in internal control decreases dropout probability at age 17 by 2.5% and by 6% at age 19
Wolfe and Johnson (1995)	Students from State University of New York	$N = 201$	Cross section	Influence of various personality inventories and scales on college grade point average (GPA)	OLS	Particular influence of traits related to control, i.e., organization (from JPI) $\beta = .27^{***}$, control (from BIG3) $\beta = .32^{***}$ and conscientiousness (from Big Five) $\beta = .31^{***}$
Carneiro, Crawford, and Goodman (2007)	National Child Development Survey 1958 (NCDS58)	Initial cohort size $\approx 17,000$	Background variables at 1958 and 1965, skill measures at 1965 and 1969, schooling outcomes at 1974 and labor market outcomes at 2000	Effect of cognitive skills and traits (social adjustment) at childhood on various outcomes	OLS Probit	Standardized social adjustment score effects, e.g., probability of staying on at school until age 16 ($.038^{***}$), employment status at 42 ($.026^{***}$) and log hourly wages at 42 ($.033^{***}$)

continued on next page

Study	Data	Sample Size	Time Horizon	Research Question	Method	Results
Murnane, Willett, Braatz, and Duhaldeborde (2000)	National Longitudinal Survey of Youth 1979 (NLSY79) subsample for males	$N = 1,448$	Measures assessed in 1980, wages (age 27/28) from 1990 to 1993	Effect of self-esteem (controlling for cognitive speed, scholastic achievement, ethnic group and calendar year) of 15-18 year old males on log hourly wages at age 27/28	OLS	One point increase in Rosenberg self-esteem scale increases log hourly wage by $3,7\%$ ($\beta = .037^{***}$)
Heckman, Stixrud, and Urzua (2006)	National Longitudinal Survey of Youth (NLSY79)	$N = 6,111$	Annual assessment beginning 1979 on background variables, test scores only in 1979	Influence of cognitive skills and personality traits (internal control and self-esteem) assessed in adolescence (ages 14-21) on various social and labor market outcomes at age 30, controlling for schooling and background	Factor structure models (Bayesian Markov Chain Monte Carlo)	- Personality traits predict log wages even stronger for most educational degrees, especially at the tails of the distribution[b] - Further influence on probability of unemployment, of being a white collar worker, of graduating from college, of smoking and marijuana use and of incarceration
Flossmann, Piatek, and Wichert (2007)	German Socio Economic Panel (SOEP) wave from 1999	$N_m = 1,549$ $N_f = 695$ Males/ females	Cross section	Effect of internal control on wages	Factor structure models (Bayesian Markov Chain Monte Carlo)	Particularly strong effect at the tails of the estimated distribution of control attitudes[b]
Osborne-Groves (2005)	-National Longitudinal Survey of Young Women 1968 (NLSYW68) for the US -National Child Development Study 1958 (NCDS58) for the UK	$N_{US} = 380$ $N_{UK} = 1,123$	- NLSYW68: earnings in 1991/1993, background variables in 1968/1969 and self-control in 1970 and 1988 - NCDS58: aggression and withdrawal in 1965 and 1969, different controls 1969-1974 and log hourly wages in 1991	Effect of the respective personality measures on women's later log hourly wages (accounting for simultaneity, measurement error and sample selection)	IV	- NLSYW68: one standard deviation increase in internal control directly lower hourly wages by 6.7%*** (when controlling for schooling)[a] - NCDS58: a one SD increase in aggression and withdrawal lowers hourly wages by 7.6%*** and 3.3%***

continued on next page

Study	Data	Sample Size	Time Horizon	Research Question	Method	Results
Heineck and Anger (2010)	German Socio Economic Panel (SOEP) waves 1991-2006	$N = 1,580$ (ability measures available in 2005/2006, respectively, matched on previous waves)	Background variables from all panel waves, personality measures from 2005 and IQ from 2006	Effect of Big Five factors, internal control and positive/ negative reciprocity on log hourly wages	- OLS (controlling for simultaneity, measurement error and sample selection) - Random Effects - Fixed Effects Vector Decomposition (FEVD)	- Internal control is the strongest predictor across methods for women ($-.080^{***}$) and men ($-.067^{***}$)[a]

*** $p \leq .01$, ** $p \leq .05$ and *** $p \leq .1$

[a] The self-control scale points towards the internal direction.

[b] The MCMC results are presented graphically.

Table 6.2: Overview of Studies Analyzing the Effects of Personality Traits on Wage Gaps

Study	Data	Sample Size	Time Horizon	Research Question	Method	Results
Fortin (2008)	- National Longitudinal Survey of High School Class of 1972 (NLS72) - 8th grade students from the National Educational Longitudinal Study (NELS88)	Not reported (subsamples are used for comparability)	- NLS72: background and tests from 1973/74/76 and 1979, wages from 1979 - NELS88: background etc. surveyed in 1990/92/94 and 2000, wages from 2000	Effect of internal control, self-esteem, importance of money/work and importance of people/family on log hourly wages at ages 24 (NELS88) and 25 (NLS72), controlling for cognitive skills, experience and other variables	OLS (Oaxaca-Ransom type decomposition)	NLS72: 6.4% of the .24 log wage gap due to personality traits, particularly by importance of money/work NELS88: 5.3% of .18 gap, again mainly due to importance of money/work
Urzua (2008)	National Longitudinal Survey of Youth (NLSY79) representative sample and subsample for oversampling blacks	$N = 3,423$	Annual assessment beginning 1979 on background variables, test scores only in 1979	Relationship between cognitive (Armed Service Vocational Aptitude Battery sub-tests, ASVAB) and traits (internal control, self-esteem) abilities, schooling choices, and black-white labor market differentials	Factor structure models (Bayesian MCMC)	Personality traits are strong predictors of schooling choices and wages for both groups, but explain little of the racial wage gap [a]
Braakmann (2009)	German Socio Economic Panel (SOEP), 2005 wave for 25-55 years old full time employed respondents	$N = 4,123$	Cross-section [b]	Effects of Big Five, internal control and positive/negative reciprocity on log hourly wages	OLS (Blinder-Oaxaca decomposition)	Using males as reference, personality traits explain up to 17.7%[c] of the .18 log wage gap, mainly due to women's higher degree of agreeableness and neuroticism

continued on next page

Study	Data	Sample Size	Time Horizon	Research Question	Method	Results
Mueller and Plug (2006)	Wisconsin Longitudinal Study (WLS), high school graduates in 1957	$N = 5,025$	Background variables from 1964, 1975 and 1992; IQ from 1957; trait measures from 1992	Effect of standardized Big Five measures on gender wage gap (controlling for background, IQ, education and tenure)	OLS (Oaxaca-Ransom type decomposition)	- Differences in personality traits explain 7.3% and differences in rewards 4.5% of the .58 log wage gap in favor of men (driven by women's higher degree of agreeableness and neuroticism) - Penalty for agreeableness and neuroticism is 2.9% higher in case of women

*** $p \leq .01$, ** $p \leq .05$ and *** $p \leq .1$

[a] The MCMC results are only presented graphically.

[b] A measure for risk-aversion was surveyed in the 2004 wave.

[c] Using the male coefficients as reference.

7

The Determinants and Dynamics of Personality Traits*

Arguably, only those components of the personality that are sufficiently stable across situations, i.e., personality traits and cognitive abilities, can be construed as skills in the sense of the human capital literature. Chapter 3 has discussed assumptions and conditions that are necessary for their existence. The previous chapter has also highlighted the relevance of these traits within the labor market and other parts of the society.

Given this subtle notion about a permanent presence of personality traits, the logical follow-up questions with a particular relevance for their governance are (1) what drives individuals to differ in terms of their personality traits and (2) whether there is scope for interventions if their development is unsatisfactory. In the following I will illustrate both issues by means of a theoretical approach known as the Technology of Skill Formation (see Cunha and Heckman, 2007), along with a brief overview on the underlying literature.

7.1 Empirical and Neurobiological Facts

Similar to the Roberts model of thoughts and behavior in an environmental context (see Chapter 2), the interactional pattern between personality traits and IQ has to be considered for the formation process.[1] As a very first framework to unify the underlying processes, Cunha et al. (2006) refer to a range of intervention studies that capture different periods of childhood and adolescence. The respective results are summarized in Table 7.1. Most of the data used in the empirical studies cover childhood and adolescence retrospectively and only provide measures of cognitive abilities and scholastic achievement. Fortunately, there is a strong consensus in the literature that IQ largely stabilizes before schooling age. If scholastic achievements are an outcome of intelligence and some other

* The discussions presented in this chapter are published in Thiel and Thomsen (2013).

[1] As previously mentioned, cognitive capabilities can have an impact on faking behavior in responding to a personality test. Vice versa, IQ tests never exactly measure pure cognitive intelligence. The results also can reflect motivational and thus aspects of personality traits.

abilities, and if a certain treatment results in a permanent shift in achievements but not in IQ, this yields an indication for other (presumably personality-related) skills to be affected by the specific treatment or intervention (see Borghans et al., 2011).

Although the evaluation of interventions that provide such kind of treatment provides only implicit evidence for the formation process of personality traits, Cunha et al. (2006) reveal a clear formation pattern which is characterized by two important features: self-productivity and dynamic complementarity. Self-productivity postulates that traits and abilities acquired at one stage enhance the formation of traits and abilities at later stages. Dynamic complementarity denotes the observation that a higher stocks of such entities at an earlier stage of life enhance the productivity of investments in abilities and traits in the ensuing stages. Furthermore, early investments should be followed by later ones in order to be retained. As a result of both features, the early periods of life are ought to constitute a bottleneck period for investments in the formation process, that is, early differences in investments clearly have an impact on the stock of abilities and traits and this impact decreases as individuals get older. Depending on whether external factors exclusively or predominantly operate within certain age spans, these periods are called critical or sensitive, respectively.

Evidence for the existence of such bottleneck periods is provided by research from various disciplines. For instance, in clinical psychology O'Connor et al. (2000) assess cognitive abilities among a group of Romanian orphans who were adopted into UK families between 1990 and 1992 and compare them to within-UK-adoptions at age four and six. As opposed to the Romanian orphans, the UK orphans were all placed into their new families before the age of six months. The study suggest that early deprived children never catch up. However, in case of personality traits the time period for malleability is longer than for cognitive abilities. Intervention studies that aim at children in school age usually report gains in behavioral measures. As the findings in Table 7.1 illustrate, even interventions at primary school age boost scholastic performance in a lasting manner without permanently raising IQ. By the above arguments, these findings provide implicit evidence on the susceptibility of personality beyond early childhood. This is in line with the literature in pediatric psychiatry (see, e.g., Dahl, 2004), that highlights the role of the prefrontal cortex in governing emotion and self-regulation and its malleability up into adulthood in specific cases. The evidence on the stabilization of personality traits from the psychological field results in somewhat different age spans and therefore is sometimes misleading at the first glance. For instance, Roberts and Delvecchio (2000) show that the rank-order of the Big Five factors stabilizes in adolescence, but there are still moderate changes until age 50.

Table 7.1: Overview of Selected Empirical Studies on the Formation of Traits

Feature	Study	Data	Sample Size	Duration	Research Question	Method	Results
Sensitive /Critical Periods	Hopkins and Bracht (1975)	Parts of 2 high school graduating classes ($\approx$ 20,000 enrollments) from district Boulder County (Colorado)	Varies between $N = 236$ and $N = 1709$	Assessment (unbalanced): $\approx$ 10 yrs.	stability of group verbal and group nonverbal IQ scores throughout grade 1,2,4,7,9 and 11	Correlation analysis	- Stability becomes evident between grades 4 and 7, i.e., correlation of adjacent measurements exceeds $r = .7$ - Nonverbal IQ is less stable
	Johnson and Newport (1989), Newport (1990)	Native Chinese or Korean speaking students (University of Illinois) with English as a second language and immigration ages between 3 and 39	$N = 46$	Cross-section	Immigration-age dependent differences in second language proficiency in terms of syntax and morphology	Correlation analysis	- Correlation between age of arrival and test performance: $r = -.77^{***}$ (controlling for various background variables) - For early arrivals (age 3-15, $N = 23$): $r = -.87^{***}$; for late arrivals (age 17-39, $N = 23$): no significant correlation
	O'Connor, Rutter, Beckett, Keaveney, Kreppner, and the English and Romanian Adoptees Study Team (2000)	Romanian adoptees from deprived environments placed at ages 0 to 42 months (between 1990 and 1992) and early adoptees from within the UK placed between ages of 0 and 6 months as control group	Treatment group (Romanians): $N = 165$; control group: $N = 52$	Assessment at ages 4 and 6 (except for 48 Romanian orphans placed after age of 24 months who were only assessed at age 6)	Effects of early deprived environments on cognitive competence (Global Cognitive Index, GCI) and possible remediation	Correlation analysis; repeated ANOVA	- At age six (whole sample): correlation between duration of deprivation and GCI, $r = -.77^{***}$ - Repeated ANOVA (data from age 4 and 6): group-factor, $F(2, 150) = 14.89^{***}$; age-factor, $F(1, 150) = 54.90^{***}$; no significant interaction term, i.e., gains over time are equal across groups and early deficits are maintained
Self-Productivity /Complementarity	Campbell and Ramey (1994), Carolina Abecedarian	- Children from low-income families randomly assigned to treatment and control group at infancy and again before entering kindergarten (TT, TC, CT, CC) - Four cohorts from 1972-1977	$N = 111$ ($T = 57$; $C = 53$)	-Treatment: infancy - age 8 for TT, infancy - age 5 for TC, age 5 - 8 for CT and no treatment for CC group - Assessment: at least annually up to age 8 and additionally at age 12	Effects of different timing of 1. Preschool program, including full day care with additional involvement and advisory for parents 2. School age program, providing home school teachers ...on Wechsler Intelligence Scale for children (WISC, longitudinal) and Woodcock Test of Academic Achievement (WTAA, at age 12)	Repeated ANOVA	- Longitudinal data: T-group constantly shows a significant advantage in IQ up to age of 8 - Age 12 data: $TT > TC > CT > CC$ concerning overall score (WISC and WTAA), $F(6, 170) = 2.63^{**}$

continued on next page

Technology Feature	Study/ Program	Data	Sample Size	Duration	Research Question	Method	Results
	Johnson and Walker (1991), Houston Parent-Child Development Center (PCDC)	Children from low-income Mexican-American families randomly assigned to treatment and control group	N=216 (initially), $T = 97$, $C = 119$ (follow-up data only partially available)	- Treatment: 2 yrs. (starting 1970) - Follow-up assessment, cross-sectional	Effect of home visits for parent support in the first year (about age 1 to 2) and center-based programs for parents and children of aggregated 400 hrs. in year two (age 2 to 3), on grades, Iowa Test of Basic Skills (ITBS) and Classroom Behavioral Inventory (CBI) at ages 8-11	ANOVA	- At time of program completion, program children had superior IQ scores (Andrews et al., 1982) - At ages 8-11 $T = C$ for grades - ITBS: $T > C$ for vocabulary score ($F(1,107) = 7.40^{***}$), reading ($F(1,107) = 6.72^{**}$) and language ($F(1,107) = 5.70^{**}$) - CBI: $T < C$ for hostility ($F(1,134) = 7.62^{***}$)
Self-Productivity /Complementarity	Schweinhart, Barnes, and Weikart (1993), High/Scope Perry Preschool Project	Black children from adverse socioeconomic backgrounds from Ypsilanti (MI) randomly assigned to T and C group	$N = 123$ ($T = 58$, $C = 65$)	- Treatment: 2 yrs. (in five waves from 1962-1965) - Assessments: scattered between age 3 to 27	Effect of daily $2\frac{1}{2}$ hr. classroom and weekly $1\frac{1}{2}$ hr. home visit on Stanford-Binet Intelligence Scale (SBI) at ages 3-9, Wechsler Intelligence Scale for Children (WISC) at age 14, California Achievement Test (CAT) and grades at ages 7-11 and 14, among various others	Group-mean comparison	- SBI: $\mu_T = 91.3 > \mu_C = 86.3^{**}$ (at age 6) fades out to $\mu_T = 85$ and $\mu_C = 84.6$ (age 10) - No differences for WISC, $\mu_T = 81$, $\mu_C = 80.7$ at age 14 - CAT score significantly higher at age 14, $\mu_T = 122.2$, $\mu_C = 94.5^{***}$
	Fuerst and Fuerst (1993), Chicago Child Parent Center Program (CPC)	Children from impoverished neighborhoods in Chicago assigned on a first-come-first-served basis to 6 CPCs affiliated to public schools (1965-1977)	- Treatment: $N^T = 683$ - 2 controls: $C_1 = 372$, $C_2 = 304$	- Program entry at 3 to 4 yrs. of age - Exit at age 8/3rd grade (in one CPC up to 6th grade) - Last assessment at 8th grade	Effects of daily 6 hrs. center care with special curricula, parents-training and support from social workers, nurses and nutritionists on California Achievement Test (CAT, up to 3rd grade) and Iowa Test of Basic Skills (ITBS, afterwards)	Group-mean comparison	- During treatment (up to third grade) T>C; at 8th grade: $T = C$ - Graduation-rate: $T = 62\%$, $C = 49\%$

continued on next page

Technology Feature	Study/ Program	Data	Sample Size	Duration	Research Question	Method	Results
	Hill, Brooks-Gunn, and Waldfogel (2002), Infant Health and Development Project (IHDP)	- Low-birth-weight (lbw, $< 2500g$) premature infants from 8 US sites randomly assigned to treatment (1985)	$N = 1,082$ ($T = 416$, $C = 666$)	- Treatment: 3 yrs., from 4 weeks of age on - Assessment: at age 3, 5 and 8 yrs.	- Effects of weekly home visits by trained staff in the 1st year (biweekly afterwards) and daily (weekdays) center-based care (2nd and 3rd yr.) on Peabody Picture Vocabulary Test - Revised (PPVT-R) for achievement at age 3, 5 and 8, Stanfort-Binet Intelligence Scale (SBI) at age 3 and Wechsler Preschool Primary Scale of Intelligence - Revised (WPPSI-R) at age 5, Wechsler Intelligence Scale for children (WISC) at age 8 and Woodcock-Johnson-Psycho-Educational Battery (WJ, broad math and reading) at age 8	Propensity score matching	-Age 3 PPVT-R[a]: $TE_{MM} = 14.7$ ($ITT = 7.0$, $\mu_C = 82.1$)[b]; SBI: $TE_{MM} = 17.5$ ($ITT = 8.9$, $\mu_C = 84.2$) - Age 5 PPVT-R: $TE_{MM} = 9.8$ ($ITT = 1.8$, $\mu_C = 79.3$); WPPSI-R: $TE_{MM} = 7.0$ ($ITT = -.5$, $\mu_C = 91.1$) - Age 8 PPVT-R: $TE_{MM} = 9.8$ ($ITT = 1.8$, $\mu_C = 84.3$); WJ Broad Reading: $TE_{MM} = 7.4$ ($ITT = 1.0$, $\mu_C = 96.6$); WJ Broad Math $TE_{MM} = 11.1$ ($ITT = -.1$, $\mu_C = 95.1$)
	Currie and Thomas (1995), Head Start	Children from National Longitudinal Survey's Child-Mother File (NLSCM) and mothers' data from the NLSY79 (only households with 2 or more at least 3 year old children)	$N \approx 5000$ children	- Treatment: 2-4 yrs. - Survey: 1986-1990 biennially	Effects of Head Start participation of disadvantaged white and black children on Peabody Picture Vocabulary Test (PPTV) percentile-points and probability of no-grade-repetition compared to no-preschool siblings	Mother/ household fixed-effects	- White: 6-percentile increase in PPTV score ($\beta = 5.875$); significant increase compared to other pre-schools (F(Head Start = other preschool) = 7.45***); 47% increase in probability of never repeating a grade - Afro-American: no significant effects for both outcomes - Inclusion of Head Start × Age interaction shows that insignificant effect for blacks is due to rapid fade out of PPVT score gains[c]

continued on next page

Technology Feature	Study/ Program	Data	Sample Size	Duration	Research Question	Method	Results
Self-Productivity	Coneus and Pfeiffer (2007)	SOEP Mother-Child Questionnaire for children 0-18 months at date of assessment and Mother-Child Questionnaire 2 for 26-42 months old children	- Age 0: $N =$ 730 - Age 3-18 months: $N =$ 580 - Age 26-42 months: $N = 192$	- birth cohorts 2002-2005 - Birth cohort 2002 again in 2005 when they were 26-42 months old	Influence of various skill-indicators of the previous period on indicators of current stock of skills when controlling for investment and further background variables	OLS, 2SLS, 3SLS	- partial elasticities of period 1 (ln birth-weight) on period 2 (3-18 months) skill indicators: e.g. on satisfaction (.28***), cry (.43***), console (.25**), health (.64***) - Elasticities for period 3 skill indicators: e.g. $t = 2$ meta skill [d] on social skill (.63***), $t = 1$ ln birth weight on $t = 3$ everyday skill (1.04***), activity in $t = 2$ and $t = 3$ (.81**), meta skill in in $t = 2$ and $t = 3$ (.32*)
	Blomeyer, Coneus, Laucht, and Pfeiffer (2009)	First-born children from the Mannheim Study of Children at Risk (MARS) born between February 1986 and February 1988	$N = 384$	Assessment waves at 3 months, 2 yrs., 4.5 yrs., 8 yrs. and 11 yrs. of age	Explanatory power of IQ and persistence measures of the previous assessment wave on current ones when controlling for nine different organic-psychosocial risk-combinations and indicators of investment	OLS	Partial elasticities among IQ measures ($t-1$ on t): - .23** (at t =2 yrs.), .53** (t =4.5 yrs.), .84** (at t =8 yrs.) and .89** (at t =11 yrs.) personality trait (perseverance at $t-1$ on t): - $-.008$ (at t =2 yrs.), .18** (t =4.5 yrs.), .29** (at t =8 yrs.) and .31** (at t =11 yrs.)

*** $p \leq .01$, ** $p \leq .05$ and *** $p \leq .1$
[a] Only results for high-intensity treatment group with more than 400 days in center-care are displayed.
[b] Benchmark: Intention-to-Treat Effect(ITT) and control group mean (μ^C) for no-match.
[c] By age 10 African-Americans have lost any gains in terms of PPVT, whereas white children maintain an advantage of 5 percentile-points at that age.
[d] Meta skill is the arithmetic mean of the Likert-Scale transformations of health, satisfaction, cry, console and activity.

This confirms the hypothesis of decreasing returns to remediation efforts. However, for the investigation of stability patterns, mean-level consistency is also important. Roberts et al. (2006) show the highest mean-level change to be concentrated on young adulthood. The authors suggest that these changes are induced by persistent shifts in social roles and role expectations common to most individuals. This age pattern is moderated when intra-individual measures are employed (see Cobb-Clark and Schurer, 2013). Though intra-individual measures suggest a higher degree of stability for the working age population, no complete time-invariance can be established from the findings in either case. Given the above discussion on the appropriateness of the Big Five and related global inventories over varying contexts, the findings on the adaption to changing social roles seem quite reasonable. A social role is also a situational factor that determines measured traits. As long as there are changes in social roles over the life course, it is tempting to interpret them as instability in actual traits. According to Almlund et al. (2011), there can be kind of a feedback between traits and situations since many situations are a consequence of trait endowment earlier in life. Maybe, the only useful distinction between changes of situations, social roles, or permanent traits is due to their different levels of sustainability. The findings from Table 7.1 somewhat confirm this notion. It shows that early interventions which involve a long-term treatment are most successful, implying that only a sufficiently enduring environmental change warrants actual improvements. Conversely, most of the gains fade out if no follow-up efforts are made.

On the other hand, sole remediation attempts in adolescence exhibit only weak effects, implying that the general efficiency of interventions in adolescence is definitely lower compared to earlier ones. This overall pattern seems to hold for all kinds of environmental changes (see, e.g., Almlund et al., 2011, and the literature they cite). There are a number of studies that document this pattern by evaluations of adolescent mentoring programs, like the Big Brothers/Big Sisters (BB/BS) and the Philadelphia Futures Sponsor-A-Scholar (SAS) program. The BB/BS assigns educated volunteers to youths from single parent households for the purpose of providing surrogate parenthood or at least an adult friend. Grossman and Tierney (1998) stress that meeting with mentors decreases the probability of initial drug and alcohol abuse, exertion of violence, and absence from school. Moreover, the participants had higher grade points and felt more competent in their school activities. SAS targets at public high school students and supports them in making it to college by academic and financial support. Johnson (1996) reveals a significant increase in grade point average and college attendance.

Summarizing the empirical picture from the listed intervention studies, there is broad evidence for the existence of crucial early life periods in terms of self-productivity, and further indication for the same efforts having substantially lower returns in later periods. The latter observations thus indicate dynamic complementarity. In neurobiology the existence of such patterns is attributed to a superior susceptibility of neural circuits and brain architecture in early lifetime (see Knudsen, 2004, Knudsen et al., 2006). One reason for this plasticity is that without already established neuronal connections, it takes less stimuli to form new ones. The second reason is, that the underlying molecular mechanisms are more active in early years of life, which induces a higher rate of changes in brain chemistry and gene expression (see Knudsen et al., 2006, for a detailed discussion of both arguments).

7.2 A Formal Representation

A theoretical representation of the formation process should account for all of the aforementioned facts. A canonical version of such a framework has been established by Cunha and Heckman (2007), and has been refined by Almlund et al. (2011), among others. The development of personality traits and cognitive factors follows a pattern that is best exemplified by means of a production function, where available resources, general environments, and other investments are the inputs. The basic relation defines an individual's traits and abilities in period t as a state variable that results from previous investments and previous stocks of personality traits. The implicit relation would read

$$\boldsymbol{\theta}_t = f(\boldsymbol{\theta}_{t-1}, \mathbf{I}_{t-1}, \mathbf{h}_{t-1}), \tag{7.1}$$

where $\boldsymbol{\theta}_t$ is the vector of personality traits and $\boldsymbol{\theta}_{t-1}$ represents the same vector in the previous period. $\mathbf{I}_{t-1}$ is a vector of investments that directly promote the formation of $\boldsymbol{\theta}_t$. $\mathbf{h}_{t-1}$ is a less directed, more intangible and situation specific input vector that represents general environment. For better analytical tractability, f is assumed to be increasing, concave, and twice differentiable in each of its arguments.

In order to make the functional form in equation (7.1) explicit, Cunha and Heckman (2007) show that a CES production function provides enough flexibility to account for the above features observed from the data. It allows for different elasticities of substitution between investments and capabilities/traits at different stages and therefore gives rise to dynamic complementarity and self-productivity. For notational simplicity consider a scalar cognitive ability and a scalar personality trait. The relation for successive periods

$t \in \{1, \ldots, T\}$ then may be expressed as follows.

$$\theta_t^j = \left[\gamma_{1,t-1}^j (I_{t-1}^j)^{\rho_{t-1}^j} + \gamma_{2,t-1}^j (\theta_{t-1}^C)^{\rho_{t-1}^j} + \gamma_{3,t-1}^j (h_{t-1})^{\rho_{t-1}^j} + \gamma_{4,t-1}^j (\theta_{t-1}^P)^{\rho_{t-1}^j}\right]^{\frac{1}{\rho_{t-1}^j}}, \quad (7.2)$$

where $\gamma_{4,t-1}^j = 1 - \gamma_{1,t-1}^j - \gamma_{2,t-1}^j - \gamma_{3,t-1}^j$ and θ with $j \in \{C, P\}$ denotes a specific cognitive ability (C) and a specific personality trait (P). Moreover, I_t denotes one respective scalar investment. The representation in equation (7.2) therefore accounts for cross-productivity between cognitive abilities and personality traits, which follows an intuitive reasoning but may also be relevant by the arguments presented in Chapter 4. Moreover, ρ_t^j, $\gamma_{1,t}^j$, $\gamma_{2,t}^j$, and $\gamma_{3,t}^j$ are the respective complementarity and multiplier parameters. An extension to the more realistic vector case of equation (7.1) is straightforward but provides no additional insights. To give the above notion of productivity and complementarity for some personality trait θ^P a formal meaning, one has

$$\text{Self-productivity: } \frac{\partial \theta_t^P}{\partial \theta_{t-1}^P} > 0,$$

$$\text{Cross-productivity: } \frac{\partial \theta_t^P}{\partial \theta_{t-1}^C} > 0,$$

$$\text{Static Complementarity: } \frac{\partial^2 \theta_t^P}{\partial \theta_{t-1}^P \partial I_{t-1}^P} > 0 \Leftrightarrow \frac{\partial^2 \theta_t^P}{\partial I_{t-1}^P \partial \theta_{t-1}^P} > 0,$$

$$\text{Dynamic Complementarity: } \frac{\partial^2 \theta_t^P}{\partial I_{t-1}^P \partial I_{t-2}^P} = \frac{\partial^2 \theta_t^P}{\partial I_{t-1}^P \partial \theta_{t-1}^P} \frac{\partial \theta_{t-1}^P}{\partial I_{t-2}^P} > 0 \Leftrightarrow \frac{\partial^2 \theta_t^P}{\partial I_{t-2}^P \partial I_{t-1}^P} > 0,$$

$$\forall\, t \in \{1, \ldots, T\},$$

where the latter two equivalence relations follow from Young's Theorem. For estimation purposes relatively sparse parameterization as in equation (7.2) are quite useful (see, e.g., Cunha et al., 2010). For theoretical exposition, however, further levels of details can be easily accommodated due to the CES specification. This may include parental stocks of traits and abilities, as well as health features and many other factors. Examples of such more nuanced version of the equation (7.2) can be found in Cunha et al. (2006) or in the supplemental material of Cunha et al. (2010).[2] The empirical feasibility of such input refinements largely depends on data availability, which is usually not given for all of these dimensions. Given the formalization of the formation process of θ^P and θ^C it is

[2] Further suggestions in terms of the implicit form in equation (7.1) are made in Almlund et al. (2011).

also meaningful to account for some entities that may arise from the stocks of θ^P and θ^C in more detail. For one thing, this may imply to augment the implicit function 7.1 by an environmental variable

$$\mathbf{h}_t = g(\mathbf{h}_{t-1}, \mathbf{I}_{t-1})$$

that depends on previous environmental states and investment efforts as a simultaneous but interdependent process (see Almlund et al., 2011). When adulthood is attained, the disposable stock of human capital can be regarded as the outcome of the acquired cognitive abilities and personality traits developed up to this point in the formation process. Cunha and Heckman (2006) present approaches to obtain estimates of the parameters of equation (7.2) and thereby quantify the degrees of self-productivity and complementarity. The data they use comprise measures of cognitive ability, temperament, motor and social development, behavioral problems, and information on the home environment. The results yield strong evidence for self-productivity within the production of the respective skill and trait types.[3] The cross-effects are weaker. Complementarity is evident for both, cognitive and noncognitive stocks, but somewhat higher in case of the former. The average parameter estimate is just below zero and thus indicates that the production technology is well approximated by a Cobb-Douglas function. Slightly altered estimation strategies that, however, yield quite similar results are provided by Cunha and Heckman (2008) and Cunha et al. (2010).

The estimation approaches used to quantify the parameter values in equation (7.2) yield factor loadings that represent the roles played by different environmental resources in the skill formation process. According to these results, indicators that relate to cultural and educational involvement, like having special lessons or going to the theater, are of particular importance. Family income however is less important when controlling for the aforementioned factors.[4] As Currie (2009) suggests, parents obtaining higher labor market returns may invest less time in children and are only partially able to compensate this neglect by provision of substituting goods. The properties of the skill formation discussed above suggest that schooling, in particular post-primary schooling, is a minor determinant compared to investments outside school. The major foundation is already set in preschool age. Adding to this view, Todd and Wolpin (2007) argue that in context of education production functions it is generally difficult to find data that combine rich information on schooling and home resources. If so, there is always less variation

[3] The identification strategy is in spirit of the factor structure models discussed in section 5. It allows for endogenous choice variables and measurement error in indicators.

[4] It nonetheless is a good indicator for the provision of home resources (see, Almlund et al., 2011).

in more aggregated indicators for schooling resources, which could lead to additional attenuation of the estimated effects. Notwithstanding the predominance of environments in earlier years of life as opposed to later, particularly school related factors, Chapters 8 and 9 provide empirical treatises that examine lifespans beyond childhood and early adolescence.

7.3 Initial Environmental Conditions

Given the formal production identity, one may allow for the formation to start within a prenatal stage already, since the time before birth can be crucial as well (see, e.g., Coneus and Pfeiffer, 2007; Cunha and Heckman, 2009). To be in line with the production process in equation (7.1), genetical endowments have to enter the process as an underlying of the initial capability and trait states (see, e.g., Blomeyer et al. , 2009), and have to be subject to interactions with investments and environments from then on. When genetical endowments are incorporated in this way, one automatically accounts for the fact that the impact of investments and environments on genes are decreasing in age (see, e.g. Cunha and Heckman, 2009, for a review of the related literature).

One should be aware of the fact that most of the summarized findings on the interaction of genes and environments have been established of late. Particularly disciplines like economics, that simply draw on such relations in order to vindicate certain settings, often refer to norms that have been overhauled in the originary disciplines in the very recent past. Ontogeny clearly is such a case. Though genetic endowments and environments are meanwhile known to follow the pattern induced by equation (7.1), they have been construed to be simply additive for a long time. However, twin and adoption studies from various fields of social science, for example in Turkheimer et al. (2003), show that a simple additive structure is inappropriate to capture the complexity of the interactions that are in place for abilities and environments. Instead, there is evidence for substantial nonlinear interactions between genes and environment in IQ generation. The fact that most empirical results from adoption studies promote genetical factors as the main driving force of formation is due to the low share of adoptive families from adverse environments in these samples, that is, the differences in environments are mostly too minor in order to explain much. The relative importance of such environmental differences apparently varies with the overall level in socioeconomic status. Studies from neurobiology and behavioral genetics draw a similar picture. By the arguments already addressed above, Knudsen et al. (2006) summarize a number of studies that show early life conditions and experiences to be particularly critical for formation processes, as the underlying molecular mechanisms take place at a higher rate in early years. This enables neural circuits to be subject to substantial changes by external factors, among others via so-called DNA methylation. DNA methylation represents a form of transcription-related genome mark that induces gene repression throughout replications (see Cedar and Bergman, 2012). A large literature from behavioral genetics deals with this phenomenon. For instance, Fraga et al. (2005) reveal that monozygotic twins who are exerted to different environmental stimuli throughout early childhood can exhibit significantly different gene expressions due to differences in DNA methylation. Personality and behavioral patterns arise out of the same neurobiological principles and therefore the same reasoning applies. Regarding neurobiological findings more related to personality, Caspi et al. (2002) reveal this relationship for psycho-pathologic phenomena like antisocial behavior.[5]

[5] A further discussion including additional empirical evidence is given in Heckman (2008) and Cunha and Heckman (2009).

8

Application I: Stability of Traits in Late Adolescents*

The empirical studies discussed in the previous chapter have shown a major plasticity of personality traits until early adulthood. Moreover, environments outside of school seem to have a stronger impact on trait formation. Whether schooling is an integral determinant in this context has not yet been analyzed coherently. This question is subject to the following chapter, though within a rather specific setting. While having left the curriculum largely unchanged, most German states have abolished the final year of higher secondary schooling to enable earlier graduation. The empirical application presented in what follows uses this exogenous policy shift to evaluate the effects of an increase in the amount of curriculum per unit of time on different personality traits.

8.1 Motivation

During the past decade, almost all German federal states with a 13-year school system have implemented policies designed to reduce the time spent for higher secondary education by eliminating the final grade. In the federal state of Saxony-Anhalt, such a reform was announced in 2003. For students in grade nine at that time, this meant a reduction of overall time for graduation. The academic schedule remained largely unchanged, though. The learning intensity, defined as the curricular workload per unit of instructional time, increased substantially. In light of the findings presented in the previous chapter, the increased workload that went along with the reform may have affected the development of the students' personality traits. The potential mechanisms through which these impacts could have operated are diverse, including persistent shifts in so-called inputs into the formation process, more general environmental changes, as well as simple changes in constraints or other factors relevant for students' decision-making. As students who concurrently attended the tenth grade continued to graduate after 13 years, the policy

* The results presented in this chapter are published in Thiel et al. (2014).

change provides a natural experiment with a double cohort of graduates in 2007. This graduation cohort is used to identify potential effects of the increased learning intensity on the development of different personality traits in late adolescence. The employed measures of personality traits are derived from a short version of the Big-Five inventory of Goldberg (1971), a short version of the Locus of Control scale established by Rotter (1966), and the Brief Self-Control Scale by Tangney et al. (2004).

The insight of investigating the potential impacts is twofold. The first contribution adds to the role that schooling plays in the formation process of personality traits in late adolescence and thus to the discussion in the previous chapter. Recall the studies that analyze the process of capability formation and personality development in earlier periods of life. Some further studies that have not been subject to the general discussion provided in Chapter 7 have a particular relevance for the question at hand. For example, Heckman et al. (2010) consider the long-term effects of the High/Scope Perry Preschool Program. Cunha et al. (2010) formulate and estimate a multistage model, where cognitive skills and personality traits are determined by home investments in different periods of childhood. Both studies find evidence for plasticity of personality traits across the complete age spectrum investigated (although decreasing with time), whereas cognitive skills are shown to be exclusively malleable at preschool age (see also Borghans et al., 2008b). In line with these findings, Cobb-Clark and Schurer (2013) show a modest variation for the intra-personal stability of a specific personality trait that also partially prevails in adulthood. These investigations notwithstanding, less research exists on the impact of schooling factors on traits, in particular for stages of later secondary education. Secondly, a less general, but still relevant contribution from a policy perspective is implied by the research question at hand. The countries of the European Union have converging designs of their education systems, in particular with regard to secondary and tertiary education. Evaluating the direct effects of an educational reform, such as the one implemented in Saxony-Anhalt, provides important information for future decisions in education policy.

8.2 Institutional Background

8.2.1 Schooling in Germany

Compared to other industrial countries, university graduates in Germany are significantly older. This fact gave rise to a debate on reforming higher education. The longer time spent obtaining an academic degree was primarily due to the very comprehensive curricula in higher secondary and tertiary education. The university curriculum has therefore been revised in the course of the Bologna Process, which has led to a replacement of the former German academic degrees by the bachelor and master degrees. For the same reason, the secondary schooling system has also been altered. It is important to know that in Germany, the responsibility for education policy, including the funding of public schools, is entrusted to the federal states. Despite this decentralized nature, the differences in the education systems between federal states are rather marginal. Children are commonly enrolled in elementary school at the age of six. After four years of elementary schooling they are assigned to one of three secondary schooling tracks. The two tracks for students with lower previous grade achievements are the "Hauptschule" and the "Realschule". The only track that directly entitles a student to university entrance is the "Gymnasium", which (prior to the reform) required nine years of attendance (except for the states of Saxony and Thuringia). The German federal states have jointly decided to reduce the overall time for graduation from the "Gymnasium" to eight years.

8.2.2 Implementation in Saxony-Anhalt – a Quasi-Experiment

In Saxony-Anhalt, the decision was passed into legislation in 2003 and came into force few months later at the beginning of the 2003/2004 academic year. The first students to be affected were in the ninth grade at that time. Accordingly, they were the first to receive their degree (Abitur) after 12 years of overall schooling. The academic requirements, however, remained almost unaltered. Graduation after 13 years was maintained for students attending the tenth grade at that time. In spring 2007, a double cohort of students simultaneously passed their final exams. For the 12-year graduates, the curriculum of the former eleventh grade was partially shifted to lower grades. In the subjects German literature and foreign languages, this applied to the whole curriculum, whereas only minor reductions took place in mathematics and chemistry. In some other subjects, for example biology and history, parts of the eleventh grade curriculum were transformed into elective courses. Moreover, three extra class hours per week were added to the syllabus in the ninth and tenth grade. Schools were allowed to decide which subjects would receive the additional weekly hours. For most subjects, however, the only change was a reduction in the net time for graduation without a compensating reduction in the graduation requirements.

The modalities just described provide some good arguments in favor of the education reform in Saxony-Anhalt complying with the requirements for a natural experiment. Consecutive graduating classes are not supposed to differ in any substantial manner other than their cohort affiliation. If so, these differences should be captured by the students' observable characteristics. Moreover, as described above, the reform was announced and implemented in rather quick succession. As the students in the sample had been attending their respective academic track for several years already, accommodations in terms of a different track choice were very unlikely. Generally, all possible avoidance actions seem to have involved disproportionate costs relative to the extent of the reform and therefore seem negligible.

8.3 Potential Mechanisms

The curricular change that has been induced by the reform may have affected the students involved in a variety of ways. With regard to scholastic achievement, Büttner and Thomsen (forthcoming) find significant reform effects on the students' performance in mathematics. 12-year students score significantly worse in mathematics but not in German language proficiency, indicating that subject-specific routines have most likely been affected by the reform. Since changes are found for mathematics only, the underlying mechanisms may be related to students' knowledge-based skills rather than their more "fluid" verbal skills. Borghans et al. (2011) or Heckman and Kautz (2012) show that scholastic achievements and achievement test scores are a mixed signal of underlying personality traits (such as self-discipline, perceived control, agreeableness, etc.) and cognitive capabilities (such as fluid intelligence, numeracy, and so forth). Scholastic achievements are therefore supposed to be mediating factors instead of genuine outcomes in the sense of human capital theory (see Heckman and Pinto, 2013a, for a related discussion). This view is promoted by the fact that scholastic achievements do not necessarily remain at their immediate post-treatment levels (see, e.g., Heckman et al., 2013) in experimental studies. As such, it is straightforward to assume that observed effects in achievements and other outcomes are partially driven by effects on underlying cognitive abilities and personality traits. In the case of the former, such an effect at the age of graduation from secondary schooling can be ruled out, as the plasticity of cognitive abilities is known to end approximately at the age of school enrollment (see Almlund et al., 2011, or Thiel and Thomsen, 2013, for overviews of the corresponding literature).

For personality traits, in contrast, the related literature posits malleability in response to environmental factors beyond preschool age, although it clearly decreases from there on. Recall some of the facts from the previous chapter on general trait formation. Findings in pediatric psychiatry (see, e.g., Dahl, 2004) emphasize that the brain region that governs emotion and self-regulation is malleable up to the age of 20 or 25. Evidence from psychology shows mean-changes for different age cohorts in cross-sectional data until age 30 (see, e.g., Roberts et al., 2006). For other traits, this age pattern is weakened when longitudinal data are used instead (see Cobb-Clark and Schurer, 2013). Although such intra-individual measures suggest a lower degree of plasticity, no complete invariance can be established. Given this ambiguity about the susceptibility of personality traits in adolescence and early adulthood, abolishing a complete year of schooling may have led to an environmental change strong enough to affect the involved students' personality development. For instance, one can imagine that having coerced students to prepare for graduation from higher secondary school in less time could have increased their self-discipline. It is also possible, however, that learning requirements have become too demanding, with students thus having lost self-confidence.

8.3.1 The Basic Development of Traits

In order to sketch the potential mechanisms through which the shift in instructional intensity may have affected personality, one may use the general framework for trait formation contained in equation (7.1) of the previous chapter. Recall that the basic relation defines an individual's traits and abilities in period t as a state variable that results from previous investments and previous stocks of personality traits,

$$\boldsymbol{\theta}_t = f(\boldsymbol{\theta}_{t-1}, \mathbf{I}_{t-1}, \mathbf{h}_{t-1}), \tag{8.1}$$

where $\boldsymbol{\theta}_t$ is a p-vector of personality traits and $\boldsymbol{\theta}_{t-1}$ represents the same vector in the previous period. $\mathbf{I}_{t-1}$ is a vector of parental and self-investments that directly promote the formation of $\boldsymbol{\theta}_t$. $\mathbf{h}_{t-1}$ is a less directed, more intangible and situation specific input vector that represents general environment. To obtain a better understanding of the constituent factors of I and h, consider the following possible vectors that may have played a role in the present setting.

$$\mathbf{I}_t = \begin{pmatrix} \textit{books} \\ \textit{hobbies} \\ \textit{instrument} \\ \textit{sports} \\ \dots \\ \textit{teacher quality} \\ \textit{teacher effort} \\ \textit{class size} \\ \textit{general resources} \end{pmatrix}, \ \mathbf{h}_t = \begin{pmatrix} \textit{friendship} \\ \textit{relationship} \\ \textit{intact family} \\ \dots \\ \textit{rel. to classmates} \end{pmatrix},$$

where the upper rows of $\mathbf{I}_t$ and $\mathbf{h}_t$ refer to elements attributed to environments outside the school and the lower ones to those inside the school (see Hanushek and Woessmann, 2011, for a review of the underlying literature).

From the self-productivity and the dynamic complementarity that result from equation (8.1) it is clear that early investments are more effective than later ones, but they should be maintained throughout all later stages of life. It is also important to note that at each t neither $\mathbf{I}_t$ nor $\mathbf{h}_t$ are exogenous to $\boldsymbol{\theta}_t$. Self-investments contained in $\mathbf{I}_t$ (such as playing team sports) are clearly endogenous to $\boldsymbol{\theta}_t$ as they are intrinsic. In the case of investments that are conducted by the students' parents, endogeneity also takes effect, since parents probably have better insights into their child's $\boldsymbol{\theta}_t$ and usually act according to them. Of course, the same reasoning holds for the more general environmental vector $\mathbf{h}_t$.

To operationalize the decisions that a student or her/his parents make in awareness of individual characteristics and situations, first consider a variation of the concept that Almlund et al. (2011) refer to as an action. Following their original definition, actions capture the style of behavior, such as simply being kind to others. Here, a slightly different notion of actions is used, in that they shall also comprise activities for which the consideration of productivity in performing these activities is largely useless. Spending time with friends is probably a good example of such an activity. Each action in the contemporaneous index set of all feasible actions M depends on current traits $\boldsymbol{\theta}_t$ and effort e_j devoted to it. As actions can be very unspecific, it seems sensible to distinguish between more nuanced forms of behavior where productivity indeed plays a role. Following Almlund et al. (2011) such activities are tasks.[1] A task can be a particular piece of homework or a test at school, which is productivity related in that it is more efficiently accomplished by individuals possessing the corresponding capacities and traits. As the relevance of capacities varies across tasks, consider a finite index set J of contemporaneous tasks an individual can choose from. In performing these tasks, each student possesses a J-vector of task-specific productivities $\mathbf{P} = g(\boldsymbol{\theta}_t, \mathbf{I}_t, \mathbf{h}_t, \mathbf{e})$. The sum of effort devoted to different tasks and actions $\sum_{j \in J \bigcup M} e_j$ is constrained to $\bar{e}$. Following notions presented in Chapter 4, the decision problem that arises from facing different tasks and actions in varying situations can be thought of as an individual maximization problem over the resulting expected utilities. Then,

$$E\left[U(\mathbf{P}, \mathbf{e}, \mathbf{a}, Y|\boldsymbol{\psi})|\mathcal{I}\right], \tag{8.2}$$

[1] In contrast to their definition, however, a more appropriate definition for actions and tasks in the given case is to construe them as rather complementary notions.

where parental characteristics Y (mostly resources provided) may play a role as well. Furthermore, $\boldsymbol{\psi}$ is a vector of preference parameters and $\mathcal{I}$ denotes the individual's information set. Both, $\boldsymbol{\psi}$ and $\mathcal{I}$, can be mapped into the period specific vector $\boldsymbol{\theta}_t$ of traits, though in an undefined fashion. The actions and task choices that result from equation (8.2) at each t partially drive $\mathbf{h}_t$ and the self-related part of $\mathbf{I}_t$. Both evolve over time in a similar manner as the traits in equation (7.1), leading to

$$\begin{aligned} \mathbf{I}_t &= b(\mathbf{I}_{t-1}, \mathbf{P}_t, \mathbf{a}_t, \mathbf{h}_{t-1}), \\ \mathbf{h}_t &= q(\mathbf{h}_{t-1}, \mathbf{P}_t, \mathbf{a}_t, \mathbf{I}_{t-1}). \end{aligned} \tag{8.3}$$

Equation (8.3) completes the circular relation between traits, tasks, actions, and environments.

8.3.2 Potential Effects of the Reform on Personality Traits

Recall from Section 8.2.2 that there were two stages in which potential effects of the reform may have occurred. The first one started with the commencement of the reform back in 2003 and ended at graduation. The second relevant period began with graduation in 2007 and consists of the time from then onwards.

Given that D denotes a student's cohort membership, $D = 1$ indicates graduation after 12 years, while $D = 0$ signifies the counterfactual cohort of 13-year students. The increase in learning intensity for $D = 1$ and the resulting extra burden could have affected the entities defined in equations (8.1) to (8.3) in six different ways. All possible pathways are related to the above stated concepts of actions, tasks, the decisions rules, and the resulting environmental and investment factors.

The first possible pathway (i) captures treatment-induced changes in school-related elements of $\mathbf{I}_t$, most obviously teacher effort. These may have affected the productivity $p_j(\cdot)$ of all schooling specific tasks. Changes of this type were not related to individual decisions of the students at all. All remaining pathways, however, have more or less resulted from individual decisions, related endowments, and constraints. One possibility is via a shift in the number of tasks and actions that have been allocated to the time at school or via a shift in their intensity. Both, the number of actions and tasks, and the effort devoted to it, are supposed to have changed the outcomes of the decision rule in equation (8.2) and the respective development of $\mathbf{I}_t$ and $\mathbf{h}_t$ in equation (8.3). Consider $J_s \subseteq J$ and $M_s \subseteq M$ to be the subsets of tasks and actions that happened inside the school environment, with corresponding counting measures $\mu(J_s)$ and $\mu(M_s)$. Analogously, $\mu(J_l)$ and $\mu(M_l)$ count tasks and actions that took place within the students' leisure time. The potential mechanism (ii) addresses the fact that the ratios $\mu(J_s)/\mu(J_l)$ and $\mu(M_s)/\mu(M_l)$ could have depended on individuals being in treatment or counterfactual state, i.e. on $\mu(J_s)/\mu(J_l) \not\perp D$ and $\mu(M_s)/\mu(M_l) \not\perp D$. The third potential path (iii) supposes that the constraints for effort devoted to tasks and actions inside the school $\bar{e}_s$ or outside the school $\bar{e}_l$ (where $\bar{e}_s + \bar{e}_l = \bar{e}$) could have depended on D (i.e. $\bar{e}_s \not\perp D$ and $\bar{e}_l \not\perp D$). Effects (ii) and (iii) may have operated through changes in productivity $\mathbf{P}(\cdot)$, through changes in actions, or directly as an argument in equation (8.2).

If one considers preferences $\boldsymbol{\psi}$ to be just a different, utility-related representation of the overall $\boldsymbol{\theta}_t$, there is no need to assume that $\boldsymbol{\psi}$ has been affected by D other than via $\boldsymbol{\theta}_t$. Another pathway therefore may have been D affecting $\boldsymbol{\theta}_t$ via the choice-relevant information set $\mathcal{I}$. Put differently, path (iv) means $\mathcal{I} \not\perp D$.[2]

The second set of pathways may not have resulted from the direct impact of the reform prior to graduation, but from the period afterwards. Such perpetuating effects could have arisen from the dynamic nature of the formation process. As D has possibly altered $\boldsymbol{\theta}_t$ before graduation, which in turn has been relevant for the determination of $\boldsymbol{\theta}_{t+1}$, the potential effects (i) to (iv) could have perpetuated due to self-productivity and dynamic complementarity. Whereas this mechanism is straightforward, there may have been a medium-run pathway that is less obvious. As stated at the outset, changes induced by (ii) to (iv) could have become stuck in the formation process if personality traits were no longer malleable at the age of graduation. Consider in contrast that D has marginally increased the expected utilities of some individuals just enough for them to become engaged in a major action or task. If actions/tasks have been sufficient in magnitude to improve the next periods $\mathbf{I}_t$ and/or $\mathbf{h}_t$, this improvement could have been large enough to alter the resulting $\boldsymbol{\theta}_{t+1}$. To make things more explicit, consider an individual who has just slightly been against enrolling abroad directly after graduation. Suppose that the changes due to D have not been strong enough to alter the corresponding $\boldsymbol{\theta}_t$, but strong enough to result in a decision in favor of studying abroad. This would have been a major change of the corresponding $\mathbf{I}_t$ or $\mathbf{h}_t$, one that was possibly large enough to affect the subsequent $\boldsymbol{\theta}_{t+1}$. If anything, such an effect (v) has probably pertained to minor fractions of the relevant population, but nevertheless should be considered. In addition to the suggested pathways (i) to (v), there could have been a sixth, more trivial pathway (vi), which is a pure age-effect. If so, it results from the fact that 13-year students could have simply "self-(re)produced" their existing trait endowments due to being one year older on average.

[2] This is one possible interpretation of the results presented in Büttner and Thomsen (forthcoming) since $\mathcal{I}$ can be thought of as representing formal knowledge or experience that evolves in a similar fashion to $\mathbf{I}$ and $\mathbf{h}$ in equation (8.3).

8.4 Data

8.4.1 The Sample

The empirical analysis is based on primary data obtained from a pen-and-paper questionnaire that was sent to members of the double cohort of graduates in Saxony-Anhalt. The survey was conducted from February to April 2009, i.e. almost two years after graduation. The resulting data comprise 101 responses relating to various aspects of the students' personality traits, social background, and educational experiences. In total, the sample consists of students from 12 schools. The schools were selected in order to fully cover two prototypical commuting areas for the school type Gymnasium in the federal state of Saxony-Anhalt. The first survey area comprises all ten such schools in the city of Magdeburg (the state capital). Magdeburg (pop. 230,000) is located near the center of Saxony-Anhalt and is an exemplary urban area (there is only one further urban area of this size in the federal state). The two other schools are the only Gymnasia in the county of Halberstadt. Regarding its size (pop. 41,000), Halberstadt properly represents typical county town areas in the whole federal state. All 12 schools are public schools that can be attended without any tuition fees. Although each school is a primary sampling unit, the contact to the respondents was not established via the school administration of the federal state. As there is no central register providing the addresses of students, other sources had to be drawn on. Two main channels were used: firstly, address information from published yearbooks of the schools (with approval of the official authorities) were used. Secondly, the principals of the schools were contacted and asked for their support. Some of the schools agreed to dispense the questionnaires via the principal's office using their own registry, while others had to be dispatched by mail.

From a total of 1,628 graduates in all 12 schools, 164 were not contactable. 1,464 questionnaires were successfully delivered, of which 805 were completed and returned. The response rates are the same for students from Magdeburg and Halberstadt. More generally, the lowest response rate only deviates from the highest one by some four percentage points. The number of returned questionnaires is almost equally split between 12 and 13-year students. At the time of the survey, a total of 81 respondents had already spent a year abroad. These observations are discarded from the sample due to the resulting age difference. The final sample size used for the analysis amounts to 724 observations. According to the Federal Statistical Office, the number of observations corresponds to about 5% of the 2007 population of graduates in the state of Saxony-Anhalt. The ratio of female graduates in the sample is slightly higher than in that population (63% as opposed to 59%).

8.4.2 Measures of Personality Traits

The set of personality measures employed in the empirical analysis comprises three inventories (see Appendix D). The first is a short version of the Big Five Inventory (see Dehne and Schupp, 2007). It incorporates the factors Openness to Experience, Conscientiousness, Extraversion, Agreeableness, and Neuroticism. Conscientiousness describes the degree to which a person is willing to comply with conventional rules, norms, and standards. Extraversion refers to the individual's need for attention and social interaction, warmth, and gregariousness. Openness to Experience is related to an individual's need for intellectual stimulation, change, and variety. Agreeableness broadly reflects the degree to which a person needs pleasant and harmonious relations with others. The final dimension, Neuroticism, describes the degree to which a person experiences the world as threatening and as something beyond their control. It covers a range of factors, such as anxiety, depression, self-consciousness, and suffering from stress. Two further measures that are more narrowly related to "real world" tasks are Locus of Control and Self-Control. Locus of Control is based on Rotter (1966) and assesses an individual's attitude to how self-directed (internal) or how coincidental achievements in his or her life are. As opposed to concepts from motivational research (such as Self-Efficacy), Locus of Control does not capture the beliefs as to how successful one could be in governing one's fate. Here, a 10-item version of the original Rotter scale is used. Moreover, the Self-Control scale by Tangney et al. (2004) is applied. Self-Control refers to the capability of adapting to one's environment by controlling thoughts, emotions, impulses, and performance.

8.4.3 Descriptive Statistics

Pre-reform characteristics are a major indicator of whether 12 and 13-year students differ in respects other than graduation time, as these characteristics are likely to have once been pivotal to individual trait formation. The related empirical literature indicates that family background and the accompanying parental investments are important ingredients of the formation process of cognitive skills and personality traits. Home items, such as the availability of newspapers or the number of books, predict the developments of cognitive skills as well as personality formation (see Todd and Wolpin, 2006, Cunha et al., 2010). Similarly, participation in cultural activities, such as theater visits, is a major indicator for parental investments into the development of personality traits (see Cunha and Heckman, 2008). The data provide three related indicators, namely "number of books" (measured by ordinal dummies), whether the parents possess artifacts at home (dummy), and the availability of an internet connection (dummy). A comparable item is the "own TV" dummy. The respective numbers indicate that the parental households are well endowed with such items and do not differ between graduation cohorts. Besides resource items, Table 8.1 reports some general background variables which characterize the individual's situation over the entire period of schooling. Mathematics and German grades at grade 7 were simply averaged in order to obtain a very general indicator of previous skill endowment.

Table 8.1: Means of Background Characteristics by Gender for Treatment and Control Group

	Male			Female		
	Gr. 13	Gr. 12	p-value[a]	Gr. 13	Gr. 12	p-value[a]
Schooltime family background						
Age at enrollment	6.227	6.223	0.951	6.189	6.115	0.050
No. of siblings	0.922	1.014	0.402	0.940	0.904	0.649
Mother's age at birth	25.429	25.927	0.290	25.616	26.055	0.316
Mother unemployed $<$ 1 year (D)	0.141	0.201	0.190	0.127	0.124	0.944
Father unemployed $<$ 1 year (D)	0.129	0.081	0.205	0.136	0.086	0.096
Gr. 7 Math/German avrg.[b]	2.291	2.219	0.324	2.172	2.157	0.763
No. of moves	1.656	1.604	0.771	1.638	1.627	0.939
Family disruption (D)[c]	0.307	0.194	0.033	0.322	0.250	0.091
Mother religious (D)	0.203	0.209	0.912	0.133	0.168	0.296
Father religious (D)	0.195	0.194	0.983	0.129	0.123	0.847
Mother leading position (D)	0.270	0.206	0.225	0.231	0.207	0.541
Father leading postiton (D)	0.309	0.348	0.504	0.350	0.338	0.802
Preschool background						
Mother unemp. preschl. age (D)	0.031	0.079	0.090	0.074	0.060	0.547
Father unemp. preschl. age (D)	0.016	0.015	0.926	0.031	0.024	0.660
Day nursery (D)	0.828	0.805	0.624	0.863	0.877	0.658
Home resources during schooltime						
Own TV (D)	0.703	0.734	0.579	0.760	0.691	0.101
Internet access (D)	0.922	0.906	0.656	0.918	0.886	0.250
Artifacts at home (D)	0.109	0.165	0.186	0.150	0.182	0.367
50-250 books (D)[d]	0.378	0.423		0.349	0.393	
250+ books (D)	0.496	0.460	0.913	0.496	0.507	0.276
N	128	139		233	220	

Dummy variables are indicated by (D).
[a] p-value from t-test on equality of means. For the book dummies the p-values refer to the χ^2-test on the complete contingency table.
[b] Mean of both grades. The best grade is 1.0 (very good), the worst is 6.0 (fail).
[c] At least one parent lives outside the household for longer than one year.
[d] Group 0-50 books is the baseline category.

8.5 Methodological Considerations

8.5.1 Further Identification Issues

Given the discussion thus far, it seems straightforward to assume the policy change to represent a substantial change in some of the model entities discussed in Section 8.3. Given the potential mechanisms involved in the framework, distinct identification of the imposed pathways is infeasible, as only cross-sectional data for the graduates are observed. Moreover, even with panel data for the entire time span available, it should be noted that most entities of the hypothetical model outlined in Section 8.3 are not or at best imperfectly observed. As a consequence of the natural experiment assumption, I use a treatment evaluation approach that imposes far fewer assumptions than an evaluation of the hypothetical model would require. This comes, however, at the cost of not being able to disentangle all "structural" parameters addressed in Section 8.3. Moreover, implied by the special nature of the outcomes, some additional threats to the randomization assumption have to be addressed in the following (see Heckman and Vytlacil, 2006).

Referring to the points in time declared in Section 8.3, one only observes individuals of both cohorts approximately two years after graduation. Consider θ_p to be a specific scalar personality trait. Conditional on a vector of observables $\mathbf{x}_i$, the individual change in the outcome induced by the reform amounts to $\Delta_{ip} = \theta^1_{ip} - \theta^0_{ip}$, where $\theta^1_{ip} = \mathbf{x'}_i\boldsymbol{\beta}_{ip} + \Delta_{ip} + u^1_{ip}$ and $\theta^0_{ip} = \mathbf{x'}_i\boldsymbol{\beta}_{ip} + u^0_{ip}$ are the stocks of personality trait p in counterfactual states. The differences in potential outcomes can be expected to depend on mechanisms (i) to (vi), i.e. $\Delta_{ip} = \Delta^{(i)}_{ip} + \Delta^{(ii)}_{ip} + \Delta^{(iii)}_{ip} + \Delta^{(iv)}_{ip} + \Delta^{(v)}_{ip} + \Delta^{(vi)}_{ip}$, where even the respective signs are unknown.

Following common representation from the treatment literature, the observed outcome of an individual is

$$\theta_{ip} = \theta^0_{ip} + (\theta^1_{ip} - \theta^0_{ip}) \cdot D, \tag{8.4}$$

depending on the hypothetical state D in which the individual is observed. The aim is to model switchings in D in a fashion such that concurrent changes in the unobserved part of the potential outcome in equation (8.4) are precluded. D is then said to be fixed at that state, i.e., is allowed to vary freely without systematic co-movements in other parts of the relation (see Heckman and Pinto, 2013b).[3] What impedes this assumption for now is the fact that one does not observe θ^0_{ip} and θ^1_{ip}, but two imperfect measures

[3] The idea of fixing an element in a system stems from the literature on causal inference (see Pearl, 2009). Heckman and Vytlacil (2006) were the first to associate this notion to matters of program evaluation.

$T_{ip}^0 = \theta_{ip}^0 - \zeta_{ip}^0$ and $T_{ip}^1 = \theta_{ip}^1 - \zeta_{ip}^1$ that are subject to measurement errors ζ_{ip}^0 and ζ_{ip}^1. This is the measurement model of classical test theory (see Lord and Novick, 1968) discussed in Chapter 3. It is the most simplified representation of a measurement error model, but is sufficient to illustrate the potentially resulting problems here. In fact, the following results extend to the case of T_{ip} not being an observed test score, but any erroneous factor score of the true latent trait. Using T_{ip} when the outcome of interest is θ_{ip}, the (sample) average effect of the reform on the treated outcome $E(\Delta_{ip}) = E\left(\theta_{ip}^1 - \theta_{ip}^0 | D, \mathbf{x}\right)$ would be the conditional expectation

$$E\left(\theta_{ip}|D, \mathbf{x}\right) = \mu(\mathbf{x}) + E(\Delta_{ip})D + E(\zeta_{ip}^0) + \left[E(\zeta_{ip}^1) - E(\zeta_{ip}^0)\right] D. \tag{8.5}$$

For $E(\Delta_{ip})$ to be identified, one requires Rosenbaum and Rubin's (1983) strong ignorability assumption to hold. On the one hand, it means that there is sufficient overlap in the covariates between both groups and that the functional $\mu(\mathbf{x})$ is the same for treated and non-treated. On the other hand, it means that conditional on $\mathbf{x}$, participation in the reform is independent of heterogeneity in individual gains, and also independent of differences in the non-treated or unexplained parts of the potential outcomes. Though neither assumption is directly testable, they are highly promoted by the quasi-experiment outlined in Section 8.2 and by the results presented in Table 8.1. As to that, non-observable selection on gains and on non-treated outcomes are not likely to be an issue in the underlying setting.

The measurement error term in equation (8.5) of course is not a selection mechanism in that individuals consciously act on it. Technically, however, it can have the same confounding impact as a selection on non-treated outcomes, namely whenever $E(\zeta_{ip})$ differs for the treatment and control group. If $E(\zeta_{ip}) \neq 0$ but is equal for $D = 0$ and $D = 1$ it would increase the error variance by $Var(\zeta_{ip}^0)$ and thereby decrease the precision of the estimates (see, e.g., Wansbeek and Meijer, 2000). The resulting problems can be resolved by the methods discussed in Chapter 5.

Under the assumption that measurement error is properly accounted for, the average effect of the reform $E(\Delta_{ip})$ is assumed to be identified for the complete target population of graduates from higher secondary schools in Saxony-Anhalt. Given some additional smoothness conditions for the unconditional distributions of θ_p^1 and θ_p^0, treatment effects $\Delta_{\tau p}$ at arbitrary unconditional quantiles $\tau \in [0, 1]$ of the distribution of θ_p^0 are also generally identified (see Firpo, 2007).

8.5.2 Building Unidimensional Measurement Aggregates

The first necessity for identification given imperfectly measured personality traits is unidimensionality, as scalar outcomes are required for the framework in equation (8.5). Since the personality traits under study comprise three to ten items that are stated on 7-point Likert scales, relying on unweighted raw scores or arbitrary selections from all available items does not necessarily lead to unidimensional outcomes. To find the best item combinations in terms of dimensionality, I apply the exploratory approach introduced in Chapter 5. For each individual, $p = 7$ measurement vectors $\mathbf{T}_{ip}$ are obtained, where each refers to a specific group of items that represent one particular personality trait θ_{ip}. However, there may potentially be more than one personality trait $\boldsymbol{\theta}_i$ underlying $\mathbf{T}_{ip}$. The Principal Factor Analysis with iterated communalities uses the fitted factor loadings to determine communalities of the correlation matrix and updates the communalities at every iteration until they converge (see, e.g., Rencher, 2004). Sometimes an iterative approach leads to corner solutions. I discard these so-called Heywood cases (see, e.g., Thompson, 2004) and use the respective second-best combinations instead. Following Costello and Osborne (2005), the next step is to check the factor loadings of the single items for a "clean" factor structure, i.e., for high common variances with high corresponding loadings. The factor loadings as well as the Eigenvalues and the shares of common variances for the respective item combinations are presented in Table 8.2.

Since the scarce empirical findings about the formation process of personality traits are based on samples that are homogeneous with respect to gender (see Cunha et al., 2010) and since there is evidence that program effects in childhood differ according to gender (see Heckman et al., 2013), a pooled and a gender-specific version are examined. A high proportion of variance explained by one common factor in Table 8.2 indicates that the corresponding item combination is likely to be unidimensional. For Locus of Control, the common share of variance for the principal factor falls short of 90% only for the male sample. The same holds true for Big Five Agreeableness in the female and the pooled sample. The differences across samples suggest that a gender-specific extraction is preferable. The factor structures for the first two rotated factors are shown in the lower panel of Table 8.2. Costello and Osborne's (2005) rule of thumb states that loadings on the principle factors should be above .30 and that there should be no substantial cross-loadings on subordinate factors. This holds for almost all cases. As with the common variances, only Big Five Agreeableness shows some sign of cross-loading on the minor factor. Performing Horn's Parallel Analysis (see Horn, 1965) indicates that a five-factor structure for the Big Five inventory is nonetheless appropriate. There is a co-movement of the Eigenvalues of the matrix of actual test scores and a random matrix of the same

Table 8.2: Iterated Principle Factor Analysis

	Common Variance					
	(male)		(female)		(pooled)	
	Eigenvalue	Proportion	Eigenvalue	Proportion	Eigenvalue	Proportion
Locus of Control	0.980	0.879	1.567	0.902	1.185	0.912
Self-Control	1.612	0.939	1.644	0.912	1.630	0.910
Openness to Experience	1.225	0.994	1.162	0.982	1.186	0.989
Conscientiousness	1.293	0.990	1.302	0.984	1.333	0.985
Extraversion	1.816	0.990	2.163	0.997	2.022	0.995
Agreeableness	0.992	0.911	1.173	0.870	1.114	0.880
Neuroticism	1.096	0.953	1.034	0.986	1.249	0.983
	Factor Structure					
	(male)		(female)		(pooled)	
	Factor 1	Factor 2	Factor 1	Factor 2	Factor 1	Factor 2
Locus of Control:						
Item 1	0.658	0.062	0.422	-0.154	0.682	-0.062
Item 2	0.389	-0.119	0.479	0.003	0.475	0.136
Item 3	0.372	0.062	0.389	0.131	0.339	0.106
Item 4	0.433	-0.151	0.686	0.094	0.655	-0.117
Item 5	0.364	0.181	0.732	-0.124	0.079	0.069
Self-Control:						
Item 1	0.448	0.187	0.442	0.133	0.417	0.237
Item 2	0.479	-0.166	0.589	-0.075	0.538	-0.021
Item 3	0.647	0.069	0.551	-0.181	0.608	-0.216
Item 4	0.609	0.042	0.496	0.190	0.513	0.183
Item 5	0.624	-0.119	0.741	-0.071	0.728	-0.069
Openness to Experience:						
Item 1	0.666	-0.036	0.593	-0.100	0.618	-0.081
Item 2	0.570	0.071	0.574	0.105	0.573	0.092
Item 3	0.675	-0.023	0.692	-0.001	0.688	-0.003
Conscientiousness:						
Item 1	0.723	-0.027	0.757	-0.005	0.754	-0.010
Item 2	0.713	-0.046	0.728	-0.072	0.723	-0.070
Item 3	0.512	0.103	0.444	0.127	0.490	0.121
Extraversion:						
Item 1	0.822	-0.079	0.900	-0.029	0.867	-0.051
Item 2	0.864	-0.007	0.900	-0.029	0.886	-0.017
Item 3	0.625	0.114	0.735	0.072	0.694	0.086
Agreeableness:						
Item 1	0.642	-0.167	0.600	-0.269	0.622	-0.237
Item 2	0.348	0.154	0.412	0.218	0.354	0.205
Item 3	0.719	0.061	0.801	0.038	0.775	0.051
Neuroticism:						
Item 1	0.489	0.174	0.631	-0.057	0.615	-0.090
Item 2	0.715	0.006	0.653	-0.018	0.687	-0.003
Item 3	0.585	-0.153	0.456	0.105	0.545	0.106

Presented are the principal common factors for the item combination that provides the best picture in terms of unidimensionality.

rank that vanishes from the fifth factor onwards (see Figure 8.1). Finally, I check the plausibility of the extracted factors as related to what the initial item inventories intend to measure. Figure 8.2 shows the correlation matrix for Quartimin rotated factors. As the Big Five factors can be thought of as representing all dimensions of personality on the highest achievable level of abstraction, they should be rather orthogonal. This even holds for the employed three-item versions, however, with a minor exception for Openness and Extraversion. By construction (see, e.g., Almlund et al., 2011), Self-Control is related to Big Five Conscientiousness, and (External) Locus of Control (negatively) to Neuroticism. These patterns known from the literature are retained by the used item selection. In addition, there are moderate correlations between Locus of Control and Extraversion, and Locus of Control and Self-Control. It is important to note that for the unconfoundedness assumption in equation (8.5) to hold, it is merely required that the way one latent trait is interrelated to another to be independent of D.[4] However, there seems to be no rationale for such a dependence. Orthogonality of outcomes therefore is not necessary per se.

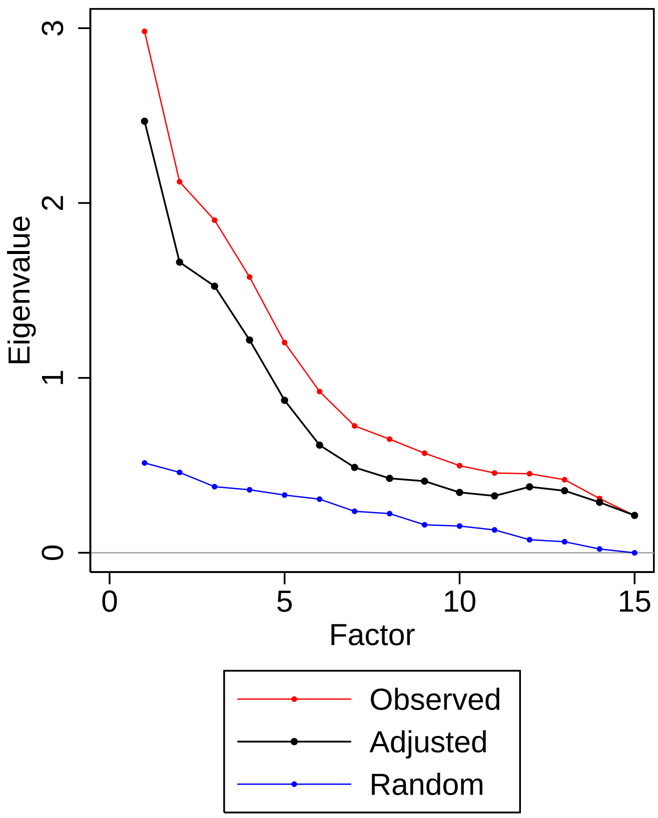

Figure 8.1: Horn's Parallel Analysis (displayed for pooled sample).

[4] Recall that this again is a mechanism that is technically equivalent to selection on non-treated outcomes discussed above.

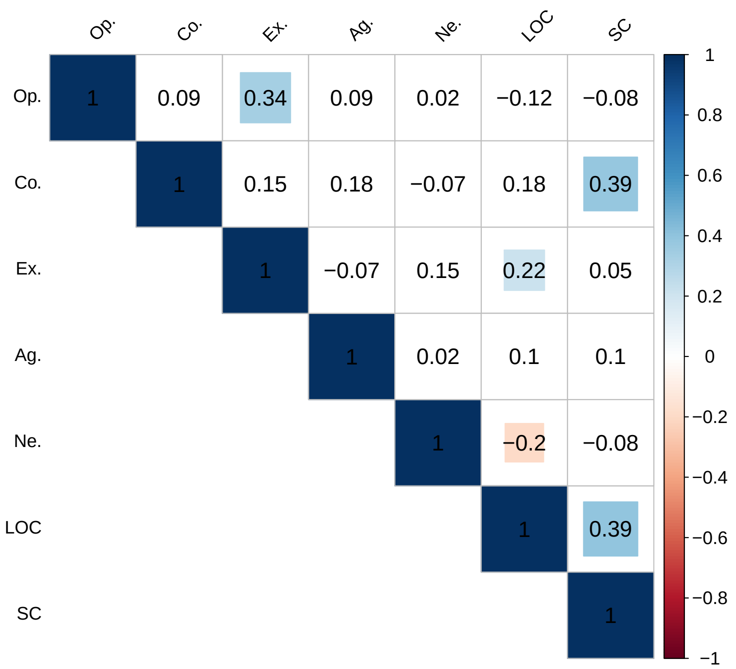

Figure 8.2: Factor Correlation Matrix (after Oblique Quartimin Rotation)

8.5.3 Estimating Latent Personality Traits

Given the unidimensionality of the factors, the next step is to estimate the latent traits for each individual. For this purpose I use the semiparametric item response model of Spady (2007) that was broadly discussed in Chapter 5. It is implemented by gender, since, apart from the fact that programs may differently affect traits, there is considerable evidence that measurement systems for personality traits also differ for males and females (see Heckman et al., 2006).

8.5.4 Identifying Treatment Effects under Measurement Variance

Having addressed the estimation approach for obtaining the factor scores of the p personality traits, one remaining issue has to be clarified. The identification of the treatment effects according to equation (8.5) depends on the assumption that treatment solely affects the latent traits, not the parameters of the measurement system used to identify them. If this were the case, the factor scores estimated by the item response framework would be a compound of true parameters for the treatment group and true parameters for the control group. To make this more explicit, consider the case where the impact of $D = 1$ on the parameters of the item system implies a true latent score for the treatment group $\tilde{\theta}^1$ that lies above the estimated one. Correspondingly, the true scores for the control group $\tilde{\theta}^0$ would be below those estimated. In this case, one would underestimate the true treatment effect, as the final term in equation (8.5) would not add up for the treatment and control group, except for the case of equal error components in both groups.

The item response model explicitly accounts for each categorial threshold. In order to test for measurement invariance between the treatment and control group, I use a linear expansion around the middle axis of all response curves. This simplification is necessary, as the polynomial coefficients have no meaningful interpretation in terms of testing for group invariance of the locations and scales of the response curves.[5] A linearization of categorial responses has previously been undertaken in comparable evaluation settings (see Heckman et al., 2013) and can be thought of as a trait-specific factor model with an intercept. Hence,

$$\mathbf{T}_p = \boldsymbol{\alpha}_p + \boldsymbol{\lambda}_p \theta_p + \boldsymbol{\nu}_p. \tag{8.6}$$

Instead of multiple threshold curves, the relationship for each item involves one overall intercept α_{pj} determining the location, and one item-specific slope parameter λ_{pj} that sets the scale (see Forero and Maydeu-Olivares, 2009, for a discussion with respect to parametric Item Response Models). The approach used here to test for measurement invariance treats the presumably identical item sets for the treatment and control group as if they actually differ. If personality trait p is measured by J items, one therefore obtains an overall system of $2J$ equations staggered in $\mathbf{T}_p$. The vector of intercepts $\boldsymbol{\alpha}_p$, the factor loadings $\boldsymbol{\lambda}_p$, and the vector of unique factors $\boldsymbol{\nu}_p$ are of the same length. Furthermore, $Cov(\nu_j, \nu_l) = 0$ for all $j \neq l$. Given the information contained in the first two moments, there are $2J^2 + 3J$ moment structures to identify $6J + 2$ free parameters, i.e. one needs a minimum of two items per personality trait in order to identify all relevant parameters. Even then, there still is an indeterminacy in the factor model in equation (8.6), since adding arbitrary scalar constants c_1 and c_2 to the model produces the same observed data structure by the identity $\tilde{\boldsymbol{\lambda}}_p = \boldsymbol{\lambda} \cdot \frac{1}{c_1} \Leftrightarrow \tilde{\theta}_p = c_1 \cdot \theta_p$ and $\tilde{\boldsymbol{\alpha}}_p = \boldsymbol{\alpha}_p - \boldsymbol{\lambda}_p c_2 \Leftrightarrow \tilde{\theta}_p = \theta_p + c_2$ (see, e.g., Anderson and Rubin, 1956).

To overcome the first indeterminacy, I choose to designate some marker item j where is set to $\lambda_{pj} = 1$. To resolve the second one, an arbitrary intercept α_{pj} can be fixed to zero (preferably for the same item $\lambda_{pj} = 1$ has been applied to). Hence, there are $6J$ unidentified parameters left. To render the reparameterized model identified (due to the mean structure, identification follows slightly different notions than in common factor models), a theorem established by Rothenberg (1971) is applied. It requires the Jacobian of the vector of all first and second moment equations $J\left[\mathbf{m}(\boldsymbol{\alpha_p}, \boldsymbol{\lambda_p}, \boldsymbol{\theta_p})\right]$ to have full rank for local identification. A sufficient condition for global identification is based

5 This results from the fact that, although being uniquely identified, quite resemblant polynomial fits can be achieved by very different combinations of polynomial coefficients.

on results from calculus and requires a positive determinant of some $(6J) \times (6J)$ submatrix of $J\left[\mathbf{m}(\boldsymbol{\alpha_p}, \boldsymbol{\lambda_p}, \boldsymbol{\theta_p})\right]$ (see Gale and Nikaido, 1965, Rothenberg, 1971). Given the reparameterization of a marker item, identification is generally established for $J \geq 2$.

Recall from Chapter 5 that under normality assumptions for the vector of observable test scores $\mathbf{T}_p$ and the latent variable θ_p, the covariance matrix of the test scores is Wishart distributed (see, e.g., Anderson, 2003). The normality assumption for the test scores is only critical if the distribution cannot be validly summarized by first and second moments (see Bollen, 1989). As noted in Chapter 5 it is the only factor analytic approach leading to a closed-form likelihood that, for computational convenience, can be reexpressed as a discrepancy function to be minimized (see Jöreskog, 1967).

Further parameter restrictions are imposed in order to test for model invariance between the treatment and control group (see, e.g., Meredith, 1993). The first set of parameter restrictions refers to scale invariance, i.e. one has to set the latter j elements of λ to be equal to the corresponding first j elements. The second set of restrictions tests for location invariance by equating the intercept parameters for both groups. If the parameters of the measurement system are not affected by participation D, imposing parameter restrictions in the described order is not supposed to change the model fit. Table 8.3 displays the respective absolute and relative χ^2-statistics. The absolute value for the baseline model has $(2J^2 + 3J) - 6J$ degrees of freedom.

Table 8.3: Parameter Restrictions and implied Changes in Model Fit (χ^2-Statistics)

	Baseline	Scale	Location
	$\chi^2 = 21.666$	$\Delta\chi^2 = 3.059$	$\Delta\chi^2 = 2.964$
Openness	16.031 (9*df*)	2.412 (+2*df*)	1.892 (+2*df*)
Conscientiousness	15.867 (9*df*)	3.164 (+2*df*)	1.786 (+2*df*)
Extraversion	12.894 (9*df*)	2.175 (+2*df*)	2.569 (+2*df*)
Agreeableness	17.315 (9*df*)	1.568 (+2*df*)	2.606 (+2*df*)
Neuroticism	11.411 (9*df*)	1.928 (+2*df*)	1.265 (+2*df*)
	$\chi^2 = 57.342$	$\Delta\chi^2 = 5.104$	$\Delta\chi^2 = 5.031$
Locus of Control	23.653 (35*df*)	3.689 (+4*df*)	3.178 (+4*df*)
Self Control	35.451(35*df*)	4.112 (+4*df*)	1.384 (+4*df*)

Level of significance: $\alpha = 0.01$. *Baseline* is the measurement system with no parameter restriction except those for just-identification. *Scale* restricts the factor loadings to be equal for the respective items in both groups. *Location* imposes the same restrictions on the intercept parameters.
Increases in degrees of freedom due to the respective restrictions are reported in parentheses. Critical χ^2- values/changes are on top of the respective panels and should not be exceeded for group invariance to be valid.

The χ^2-values in Table 8.3 refer to $(N-1)$ times the F-value of the discrepancy function at the minimum. The absolute χ^2-value refers to a model with (unrealistic) perfect fit. This drawback, however, is irrelevant for the present case, in that only relative changes in χ^2-values with respect to the same baseline model are to be considered. For changes below the critical one, one fails to reject the null of an unchanged model. Table 8.3 indicates that neither of the imposed restrictions leads to a significant change in model fit. This provides evidence that the treatment has no overall impact on location and scale of the response pattern. One can therefore interpret the reform effects on the estimated latent personality scores as the true ones. It should be noted, however, that one cannot separately test for a potential age effect that may distinctly act on the measurement systems of both cohorts (Section 8.7 provides some indicative evidence against such an effect).

8.5.5 Estimation of the Reform Effects

Given the factor scores that account for the issues addressed in Section 8.5.1, one can use conventional regression frameworks to estimate the reform effects. As the sample is from 12 different schools, school-specific influences may play a role. Examples for potential school-specific effects include differences in the quality of teachers, differences in infrastructure, or differences in the number (and background) of peers. Therefore school-fixed effects are included. To account for the potential gender differences already mentioned, I estimate gender specific models as well as a pooled one. The general specification of the model is given as follows.

$$\theta_{pg} = \alpha_{pg} + \Delta_{pg} D + \mathbf{s}'\boldsymbol{\delta}_{pg} + \mathbf{x}'\boldsymbol{\beta}_{pg} + \varepsilon_{pg}, \tag{8.7}$$

where $\alpha_{pg} + \mathbf{s}'\boldsymbol{\delta}_{pg} + \mathbf{x}'\boldsymbol{\beta}_{pg} = \mu(\mathbf{x})$ and $g \in \{\text{male, female, pooled}\}$. Accordingly, $\mathbf{s}$ contains the school dummies, $\mathbf{x}$ includes observed pre-treatment characteristics, and θ_{pg} is the standardized personality factor score obtained from equation (5.10) as introduced in Chapter 5. Finally, $p = 1, \ldots, 7$ indexes the personality dimensions under study and D is the treatment dummy with $\Delta_{pg} = \Delta_{pg}^{(i)} + \Delta_{pg}^{(ii)} + \Delta_{pg}^{(iii)} + \Delta_{pg}^{(iv)} + \Delta_{pg}^{(v)} + \Delta_{pg}^{(vi)}$ being the parameter of interest.

Based on the general specification in equation (8.7) two estimands are considered: the Average Treatment Effect (ATE) and the Unconditional Quantile Treatment Effect (UQTE). The former provides information about the average effect of the policy change and is easily obtained by estimating equation (8.7) by OLS. Obtaining additional insights into potential distributional impacts of the reform is, however, also meaningful. There is substantial evidence that sign switches of the labor market remuneration for personality traits can occur along their distributions (see Heckman et al., 2006, John and Thomsen, 2014, among others). As, for instance, very low and very high expressions of Conscientiousness are punished with regard to females' wages (see Heineck and Anger, 2010), schemes that compress the distribution of such traits at the tails may be more desirable from a policy perspective than schemes that benefit the average individual. In order to account for those effects with regard to their desirability, treatment effects for different points of the support of the (personality) factor score density are estimated in addition. The choice of the estimator that is used for this purpose is motivated by the fact that, compared to conditional mean regression, there is no equivalent for the law of iterated expectation for conditional quantiles (see Firpo et al., 2009, for a formal exposition). Without this property, one cannot infer about the impact of the policy change at the τth quantile of the unconditional distribution of the pth trait. Conditional and unconditional quantile effects are only equivalent for location shift models (see Doksum, 1974). As with the average effects, however, further covariates $\mathbf{x}$ are used to raise the efficiency of the estimates. An approach that adapts conditional quantile regression of Koenker and Basset (1978) in the respective way is due to Firpo (2007). It augments the objective function for the conditional quantile estimator by an inverse probability weighting factor, implying

$$\hat{\Delta}_p^\tau = \arg\min \sum_i \left(\frac{D_i}{\hat{p}(D=1|\mathbf{x}_i)} + \frac{1-D_i}{1-\hat{p}(D_i=1|\mathbf{x}_i)} \right) \rho^\tau(\theta_{ip} - \alpha_p - D_i\Delta_p), \tag{8.8}$$

where the sum is over all individuals i of $g \in \{\text{male}, \text{female}, \text{pooled}\}$. The covariate vector $\mathbf{x}$ only occurs in the probability weights. The propensity score $\hat{p}(D=1|\mathbf{x}_i)$ is estimated by means of a non-parametric *local logit* approach (see Frölich, 2006).

8.6 Estimation Results

The estimation results for the gender-specific average treatment effects are provided in Table 8.4. As standardized factor scores are used in all regressions, the slope parameters have to be interpreted in terms of standard deviations. All estimates of treatment effects control for four different specifications of the background variables presented in Table 8.1. The model specifications vary across estimates presented in Table 8.4 and are based on the respective AIC values. As the presumption of a natural experiment suggests, the variation of the effect sizes is very low across specifications (see Figure 8.3 for male respondents). Details about the specifications of covariates can be found below Table 8.4.

Table 8.4: Regression Estimates: ATE (Gender Specific)

	Male				Female				Pooled (male)			
	Δ_p		SE	Spec.	Δ_p		SE	Spec.	Δ_p		$\Delta_p + Int.$	
Openn. to Exp.	0.086	**	0.036	2	-0.007		0.006	2	0.004	**	0.078	**
Conscientiousness	0.038	**	0.017	2	-0.097	***	0.032	1	-0.113	**	0.027	***
Extraversion	-0.032	**	0.015	1	0.095	***	0.035	3	0.073	***	-0.055	**
Agreeableness	0.042		0.044	1	0.069		0.132	2	0.039		0.036	
Neuroticism	-0.185	***	0.016	1	-0.131		0.083	1	-0.134	***	-0.181	***
Locus of Control	-0.098	***	0.015	3	-0.066	***	0.017	2	-0.116	***	- 0.060	**
Self-Control	-0.099	**	0.040	3	0.054		0.045	4	0.114	**	- 0.068	***

$\Delta_p + Int.$ indicates pooled Δ_p plus male treatment interaction. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

[1] Spec. 1 (schooltime background): age at enrollment, No. of siblings, mother's age at birth, mother's unemployment spell (months), father's unemployment spell(months), math/german composite grade (at gr.7), No. of moves, family disruption (D),m other religious (D), father religious (D), mother leading position (D), father leading position (D).

[2] Spec. 2 (*Spec. 1.* + preschool): ..., mother unemployed, preschool age (D), father unemployed, preschool age (D, day nursery (D).

[3] Spec. 3 (*Spec. 1.* + home items): ..., own TV (D), internet access (D), artifacts at home (D), 50-250 books (D), 250+ books (D).

[4] Spec. 4: all covariates included.

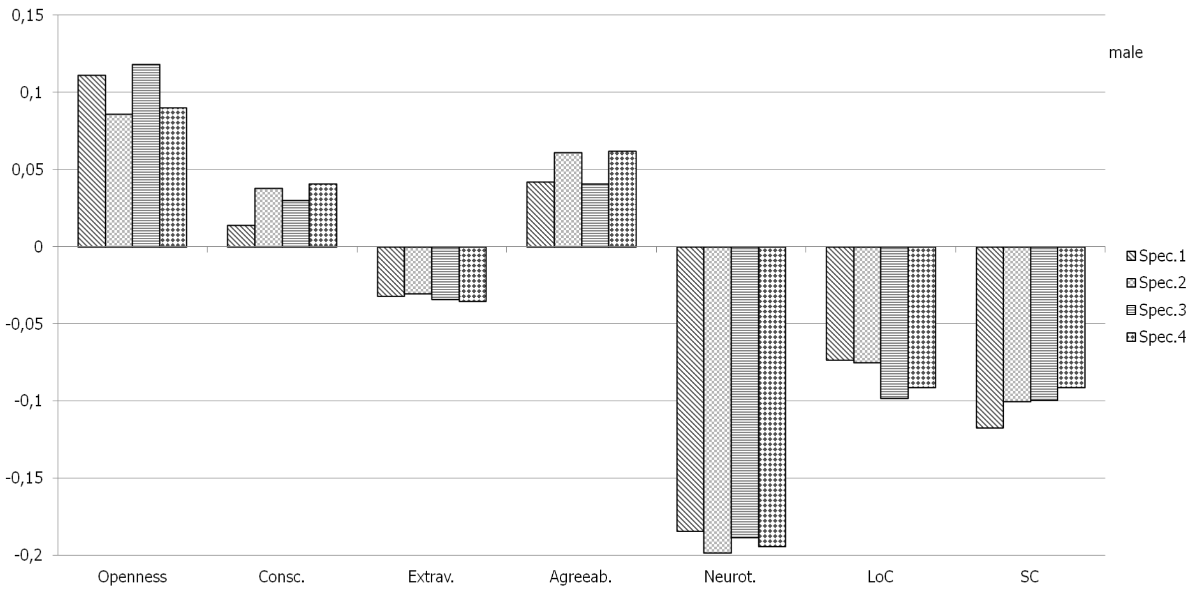

Figure 8.3: Robustness of effect-sizes to covariate-specifications (males).

The estimates show that the curricular changes induced by the reform do not affect any of the personality dimensions assessed in a sizable magnitude. Economically, the mean effects are only moderately significant for the estimated models (between -0.185 and 0.095 standard deviations). In terms of statistical significance, Openness, Conscientiousness, Extraversion, Neuroticism, Locus of Control, and Self-Control allow clear inference for the effects in the male sample. Openness increases by 0.086 standard deviations. Conscientiousness is moderately improved by 0.038 standard deviations. A small negative average effect of the same magnitude can be observed for Extraversion. Locus of Control is slightly more affected (-0.098 standard deviations), as is Self-Control. The most salient average effect is the 0.185 standard deviation decrease for Neuroticism. In the case of the female respondents, Conscientiousness, Extraversion, and Locus of Control are significantly affected by the reform. As opposed to the male sample, the effect for Conscientiousness is negative (-0.097), but comparatively moderate. On the other hand, female Extraversion is increased, but likewise by less than 10% of a standard deviation. Again, contrary to the effects for males, Locus of Control decreases by -0.066 standard deviations. Comparing the estimates to those obtained from the models using the corresponding raw scores reveals some interesting findings. Notwithstanding the persistently low magnitude of the effect sizes after measurement error correction, most of the estimates notably improve due to the procedure (see Figure 8.4) as some of the effects sizes on the raw test scores are quite different (e.g., for female Self-Control). In line with these spreads, the precision of the estimates in the models with measurement error correction is increased, and with it the statistical significance. In the case of the raw score, however, one thus cannot be too sure about the correctness of the effect signs (for instance, the effect for female Self-Control can easily become positive given its bounds of significance). By and large, these findings indicate that the issues of precision and consistency addressed in Section 8.5.1 seem to apply. Another concern may be the relatively small sample size. With regard to that, I estimate a pooled model with an interaction effect for males (see the right column of Table 8.4). The composite effect is the effect for male participants compared to female non-participants. Although some differences occur compared to the gender specific models, there are no substantial changes. This indicates that the control groups do not differ much. The gains in statistical significance are rather moderate (except for Self-Control) given the fact that the sample size has more than doubled. Therefore, the sample size is obviously not an issue in terms of statistical significance. It still might be one in terms of the power of the tests though. Table 8.5 reports (absolute) minimum effect sizes required to reject a false H_0 under a true H_A. It shows that most of the gender-specific effect sizes are apparently too small under power-consideration. The decreases in standard errors that are induced by the increased sample size in the pooled model, however, leads to admissible minimum effect sizes. As the inference of the pooled model is otherwise in line with the gender-specific results, however, it is likely that the test results for the latter ones are credible.

Table 8.5: Minimum Effect Sizes (Gender Specific)

Male	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$
Openness to Experience	0.094	0.071	0.059
Conscientiousness	0.044	0.033	0.028
Extraversion	-0.114	- 0.087	- 0.073
Agreeableness	0.039	0.030	0.025
Neuroticism	-0.042	- 0.032	- 0.026
Locus of Control	-0.039	- 0.030	- 0.025
Self-Control	-0.104	- 0.079	- 0.066
Female	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$
Openness to Experience	-0.016	-0.012	- 0.010
Conscientiousness	-0.083	- 0.063	- 0.053
Extraversion	0.091	0.069	0.058
Agreeableness	0.343	0.260	0.218
Neuroticism	-0.216	- 0.163	- 0.137
Locus of Control	-0.044	- 0.033	- 0.028
Self-Control	0.117	0.089	0.074
Pooled (male interaction)	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$
Openness to Experience	0.012	0.009	0.007
Conscientiousness	0.062	0.047	0.040
Extraversion	-0.068	-0.052	-0.043
Agreeableness	0.257	0.195	0.163
Neuroticism	-0.162	-0.123	-0.103
Locus of Control	-0.033	-0.025	-0.021
Self-Control	-0.088	-0.066	-0.056

Effect sizes for two-sided t-tests are computed under the estimated standard errors.

With regard to the potential transmission paths (i) to (vi), there can be several explanations for the empirical findings. First and foremost, it is likely that the induced changes in inputs have been too minor to affect personality development in the age span considered (see Cunha and Heckman, 2006); this interpretation is in line with previous empirical results that focus on home investments (see, e.g., Cunha et al., 2010). Alternatively, one can suppose potential cross-compensation of the paths (i) to (vi), i.e. the distinct effects of the increased curricular burden have somehow added up. One may refute this presumption by checking two additional outcomes in Section 8.7 below.

The unconditional quantile treatment results are presented in Table 8.6. The reported quantiles for the male and female sample are chosen according to the properties of the estimator discussed above. Pairs of antithetic quantiles are used as they provide enough support for consistent coefficient estimation and for bootstrapping the standard errors. The estimated magnitudes are in the same range as for the mean. More importantly, there

Table 8.6: Regression Estimates: UQTE (Gender Specific)

Male	Δ_p^τ			
	$\tau = 0.15$	$\tau = 0.40$	$\tau = 0.60$	$\tau = 0.85$
Openness to Experience	0.033	0.017 *	0.014	0.013 *
Conscientiousness	0.012 *	0.039 **	0.042 **	0.007
Extraversion	-0.039	-0.011 *	-0.019 *	-0.016
Agreeableness	0.031 *	0.056 *	0.009	-0.017
Neuroticism	-0.008	-0.029 *	-0.084 **	-0.012 *
Locus of Control	-0.044 *	-0.023	-0.061 *	0.003
Self-Control	-0.037 *	0.002	-0.021 **	-0.055
Female	Δ_p^τ			
	$\tau = 0.20$	$\tau = 0.40$	$\tau = 0.60$	$\tau = 0.80$
Openness to Experience	-0.009 *	-0.009	-0.021 **	0.014
Conscientiousness	-0.128	0.089 *	-0.131 **	-0.013 *
Extraversion	-0.014	-0.061	-0.046 *	-0.017 *
Agreeableness	0.051 **	0.023 *	0.015	-0.007
Neuroticism	0.019	0.097 **	0.147 **	0.121 *
Locus of Control	-0.025 *	-0.093 *	-0.077 *	-0.011
Self-Control	-0.013	0.009	-0.062 *	-0.047 *

* $p < 0.1$, ** $p < 0.05$

Standard errors bootstrapped for clusters on school level (150 replications).

Specification 3 was used for inverse probability weights: age at enrollment, No. of siblings, mother's age at birth, mother's unemployment spell (months), father's unemployment spell(months), math/german composite grade (at gr.7), No. of moves, family disruption (D),m other religious (D), father religious (D), mother leading position (D), father leading position (D), own TV (D), internet access (D), artifacts at home (D), 50-250 books (D), 250+ books (D).

Dummy variables are indicated by (D).

is no evidence for a compression at one or both ends of the distribution. Additionally, the modest effects for the upper and lower quantiles are largely insignificant.

Summarizing the results, the estimates for the impact of the analyzed educational policy reform on personality traits differ from the effect on grade achievements. Whereas the higher learning intensity has negatively affected students' academic achievements (see Büttner and Thomsen, forthcoming), personality apparently remains unaffected. Hence, schooling at the considered age seems to promote more specific forms of human capital. The consensus that higher academic requirements at school come at the expense of an impeded personality development cannot be supported. The presented findings indicate that the development of students' personality is not at odds with the increased learning intensity. As mentioned above, all estimates refer to standard deviations of the latent factor scores, which is a somewhat artificial metric when it comes to pinning down the results. Taking the entire support of the factor score distributions as a benchmark (see

Figure 8.5) reveals that even 19% of one standard deviation (for male Neuroticism) is still a minor effect size. For early childhood interventions (see Heckman et al., 2013), the effect sizes for some of the personality traits under investigation amount to almost thrice that magnitude. To give another benchmark, I use wage regression estimates for a representative German working age population that are provided in the study by Heineck and Anger (2010). Apart from Self-Control, they consider the same personality traits that are assessed here. Given their estimates, the male effect size for Locus of Control, which is one of the more highly rewarded personality traits in the labor market, would decrease the average hourly earnings for a German male in working age by 0.7%. For females, earnings would be lowered by 0.5%. The stronger effects for Neuroticism would hardly transmit into hourly wages, since the wage gradient for this trait is almost zero in the labor market. This shows that the effect sizes induced by the reform are also minor in terms of later achievements.

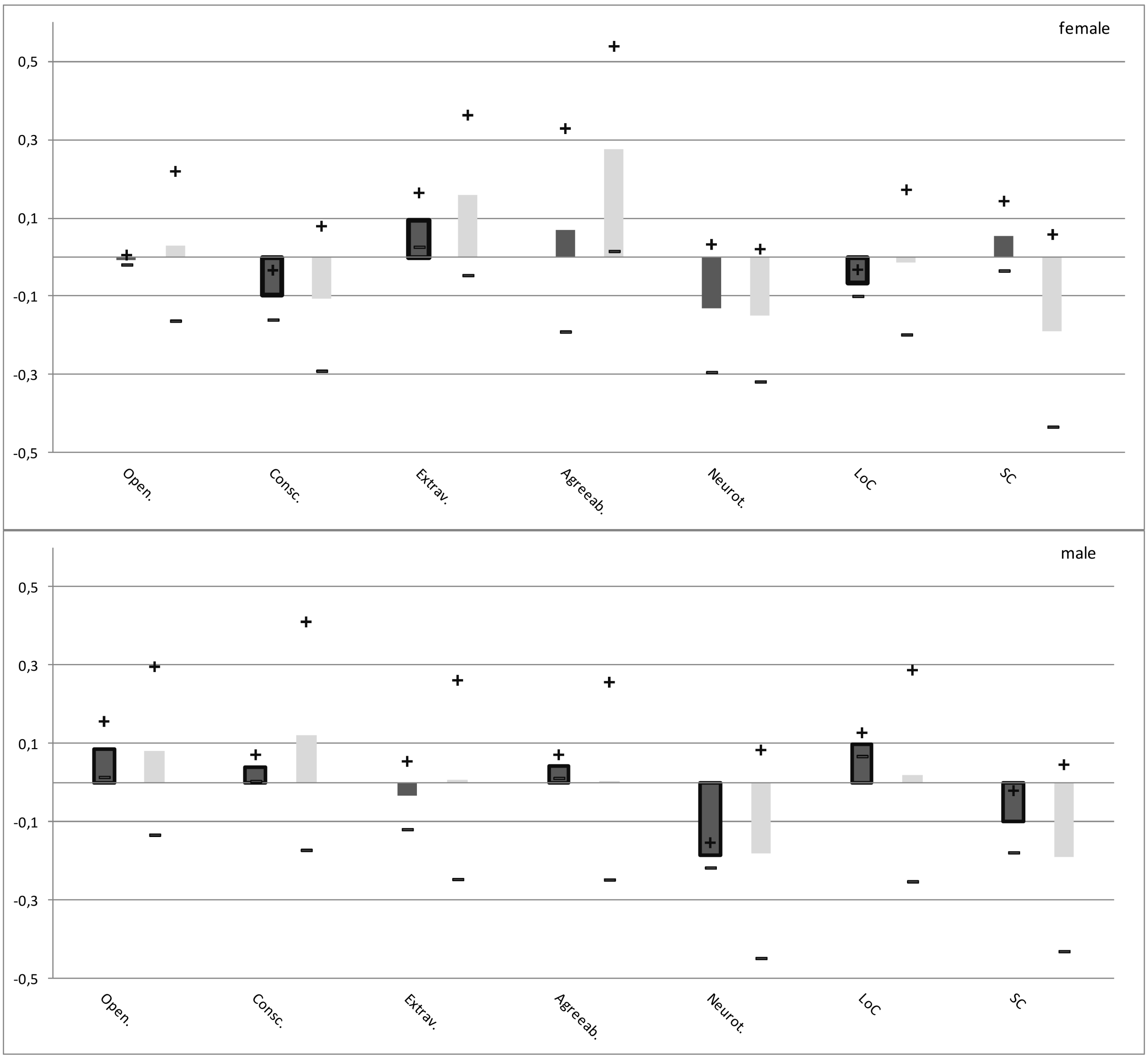

Figure 8.4: Variation of effect-sizes due to measurement error correction (dark-grey columns are error-corrected). Statistically significant effects are bold-framed. Upper and lower significance bounds are indicated by + and −.

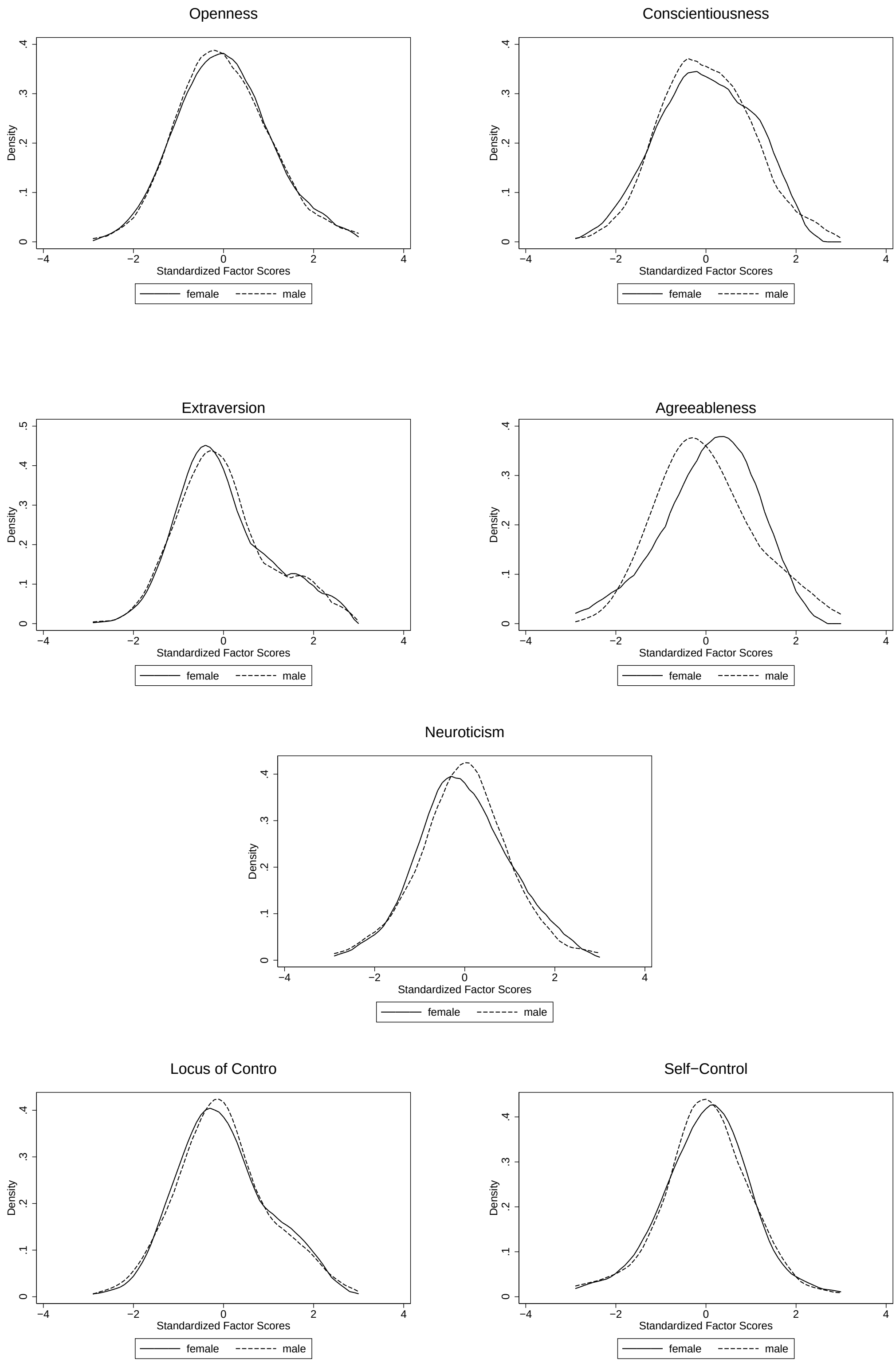

Figure 8.5: Kernel smoothed (Epanechnikov) latent factor scores (100 grid points) for male and female students

8.7 Robustness of the Results

Although the most likely explanation for the small effect sizes observed from the data is that none of the six pathways suggested in Section 8.3 were still active for the students in their late adolescence, a mixing of the potential pathways to a total of zero should be ruled out wherever possible. Recall that one of these potential influences could have been $\Delta_{pg}^{(vi)}$, as students differ in age by a full year on average. To check for a net age effect in factor scores (tests with raw scores provided comparable findings), I estimate separate models within both cohorts (as treatment and age are almost perfectly collinear and the resulting estimates are very imprecise). As before, the models are defined separately for each gender, that is, four subsamples are used: females/$D = 1$, females/$D = 0$, males/$D = 1$ and males/$D = 0$. These subsamples are confined to students born in the first three months and in the last three months of the respective age spans. Relying on quarter years is sensible in order to obtain sufficiently high numbers of observations and clusters. I then regress all personality scores on the baseline specification and include an additional dummy variable taking the value one if the student was born during the first three months and zero if the student was born during the last three months of the stretch. The corresponding effects are presented in Table 8.7.

Table 8.7: Age Effects

	(male)		(female)		(pooled)	
	$D = 0$	$D = 1$	$D = 0$	$D = 1$	$D = 0$	$D = 1$
Openness to Experience	0.009*	0.011*	0.017	0.006**	0.003*	-0.002**
Conscientiousness	-0.001*	-0.003*	0.004	-0.013	0.005**	0.007*
Extraversion	0.034	0.018*	0.002*	-0.017	0.006*	-0.000
Agreeableness	-0.002	-0.007	0.061*	0.072**	0.007**	-0.009*
Neuroticism	-0.011*	0.014	0.017*	-0.004	-0.003	0.014***
Locus of Control	0.003**	-0.006*	-0.019	-0.007**	0.001*	-0.005**
Self-Control	-0.015*	0.021*	-0.001	0.013	0.008**	0.002**
	N=63	N=60	N=95	N=97	N=160	N=155

$D = 1$: 12 year graduates; $D = 0$: 13 year graduates. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

The estimates show no clear-cut pattern for pure age effects. Most of the coefficients are weakly significant and additionally weak in magnitude. This picture is likely to result from the relatively low number of observations in the subsamples. In order to weaken this problem I also estimate a pooled version, which increases the significance of parameters and leads to a similar picture. Though not jointly estimable with the outcome equation (8.5), these findings suggest ruling out major age effects between the two cohorts. In order to test for potential added-up effects in the remaining pathways (i) to (v), one has to rely on two indicators obtained from the questionnaire, as it is not possible to directly observe the relevant entities. These indicators comprise leisure information and perceived curricular workload. Due to the simultaneity concerns with the main outcome variables, separate models with the available indicators as a dependent variable are estimated. The leisure variable is constructed using information on the weekly mandatory curriculum and statements regarding additional elective courses and hours spent on tasks such as homework, learning, chores, taking care of siblings etc. The respective results are presented in Table 8.8. Neither for males nor for females a significant and conclusive effect of the dummy on disposable leisure in hours per week can be found. The conclusion therefore rather reads that there has been no relevant trade-off between leisure and schooling investments that may veil the impact of the policy change. Moreover, no significant cohort differences can be found for two further questionnaire items that indicate the perceived scholastic workload in the final year of schooling (results not presented).

Table 8.8: Regression of leisure (hours per week) on specifications 1 to 4

	(male)	(female)
Specification 1	2.482	-0.053
	(1.764)	(1.336)
Specification 2	2.014	0.075
	(1.732)	(1.356)
Specification 3	2.341	0.177
	(1.835)	(1.329)
Specification 4	1.922	0.263
	(1.798)	(1.349)

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$ (Standard errors in parentheses)
Spec. 1 (schooltime background): age at enrollment, No. of siblings, mother's age at birth, mother's unemployment spell (months), father's unemployment spell(months), math/german composite grade (at gr.7), No. of moves, family disruption (D),m other religious (D), father religious (D), mother leading position (D), father leading position (D).
Spec. 2 (*Spec. 1.* + preschool): ..., mother unemployed, preschool age (D), father unemployed, preschool age (D, day nursery (D).
Spec. 3 (*Spec. 1.* + home items): ..., own TV (D), internet access (D), artifacts at home (D), 50-250 books (D), 250+ books (D).
Spec. 4: all covariates included.

Given this admittedly indicative evidence, it seams reasonable to presume that no pathways related to individual effort (ii) and (iii) could have affected the outcomes. If there still has been a mixed impact not observable from the data, it has likely arisen from various combinations of, $\Delta_{pg}^{(i)}$, $\Delta_{pg}^{(iv)}$ and $\Delta_{pg}^{(v)}$. Findings from a study analyzing tertiary education decisions for students from the same cohort suggest slight effects of the reform for female students (see Meyer and Thomsen, 2013). They are more likely to delay university entrance for the purpose of prior vocational education. This kind of behavior is what is addressed by means of pathway (v). Whether this decision-based change has been strong enough to (perhaps later) produce a feedback effect on personality formation cannot be observed from the data, but is at least questionable. As such, the remaining pathways that may have added up refer to external environmental changes (i) and information sets (iv). However, since the picture is very homogeneous for all outcomes, one may consider this eventuality to be unrealistic.

8.8 External Validity

Since primary data is used for the investigation, it seems expedient to provide some evidence for representativity of the sample in order to promote the external validity of the estimates. For this purpose, the information provided by the German Socio-Economic Panel Study (GSOEP, see Wagner et al., 2007) can be used. More precisely, as an approximation for the target population the comparison draws on the cohort of 18 to 24-year-old GSOEP participants, who are currently attending a *Gymnasium* or have already graduated from one. In order to obtain a sufficiently large sample for comparison, one has to pool the waves 2000 to 2008. Table 8.9 provides mean and frequency comparisons for parental and leisure characteristics.

The differences in most of the displayed benchmark values are largely minor in magnitude. The only notable exception is the gap in the employment status of the respondents' mothers. At a 10% significance level, the null of mean equivalence between samples for the sociodemographic variables, the educational degree of mothers, professional qualification of both, and occupational status of mothers is rejected. In interpreting these results, however, one should bear in mind that the underlying variance estimates are likely to be inconsistent for their respective population counterparts, mainly due to the very complex survey design of the GSOEP and accompanying panel attrition. Moreover, despite some statistically significant differences, all compared variables point to small differences in magnitude only. One may therefore co the sample as being representative for the respective German students overall. With respect to the external validity of the results, some further points remain to be discussed.

Table 8.9: Means of Selected Characteristics from Own Sample Compared to Means from selected GSOEP subsamples

	Student Survey	GSOEP Data Germany	p-value[a]
Sociodemographic variables	N=722	N=2994	
Age	20.70	22.07	0.00
Country of birth[c]	0.98	0.97	0.02
Number of siblings[d]	0.94	1.13	0.00
Educational degree (father)	N=685	N=2691	
Dropout	0.00	0.00	
Secondary school degree	0.58	0.60	
Higher secondary school degree	0.42	0.39	0.24
Educational degree (mother)	N=713	N=2715	
Dropout	0.00	0.00	
Secondary school degree	0.62	0.69	
Higher secondary school degree	0.38	0.30	0.00
Professional qualification (father)	N=693	N=2874	
No occupational training	0.00	0.02	
Apprenticeship training	0.58	0.55	
University / university of applied sciences	0.41	0.42	0.01
Professional qualification (mother)	N=716	N=2885	
No occupational training	0.01	0.06	
Apprenticeship training	0.55	0.61	
University / university of applied sciences	0.44	0.33	0.00
Occupational status (father)	N=682	N=2858	
Not employed	0.07	0.07	
Blue-/white-collar worker, civil servant	0.77	0.77	
Self-employed	0.15	0.16	0.69
Occupational status (mother)	N=709	N=1744	
Not employed	0.08	0.15	
Blue-/white-collar worker, civil servant	0.84	0.74	
Self-employed	0.08	0.11	0.00
Occupational position of parents[e]	N=695	N=2999	
Leading position of father	0.34	0.33	0.54
Leading position of mother	0.23	0.13	0.00
Number of books at home	N=719	N=690	
0 to 100	0.28	0.34	
101 to 500	0.47	0.48	
More than 500	0.25	0.18	0.00
Leisure activities during childhood[e]	N=723	N=1030	
Sport	0.76	0.73	0.24
Music	0.51	0.54	0.17

[a] p-value from t-test on equality of means; for categorial variables: p-value from χ^2-test.
[b] Dummy variable: 0 (foreign countries), 1 (Germany)
[c] Number of observations: 1019 (Germany), 368 (East Germany)
[d] Dummy variable: 0 (no), 1 (yes)

The effects of the reform on students' personality traits may have varied with previous levels of learning intensity. Since the intensity in the university preparatory track (Gymnasium) prior to the reform had already been relatively high, a remission in other personality-shaping activities in leisure is more unlikely than in lower-tier tracks. For those, however, it is more apt to assume that other mechanisms could have operated. On the other hand, the role personality traits potentially play regarding track choices may have produced a selected sample in terms of trait endowments prior to the reform. In this case, the picture for lower-tier secondary schooling tracks could have been different as well. Likewise, the increase in learning intensity has only affected the end of schooling time. Given the nature of personality formation discussed in Section 8.3, an implementation of the same changes in earlier grades may have induced different effects.

8.9 Discussion

The empirical analysis conducted in this chapter is a first step towards evaluating the effects of a substantial educational reform in late adolescence on students' personality traits. The loss of an entire school year without a compensating reduction in the graduation requirements has reduced the time available for instruction, homework, and leisure activities. As Büttner and Thomsen (forthcoming) point out, the lack of compensation has resulted in significant negative effects of the policy change on grade achievements in mathematics. In contrast to their findings, the empirical results at hand cannot discover any significant effects of an increased learning intensity on the personality dimensions under investigation. Referring to the personality formation literature, the most salient explanation for these findings is that personality was fairly set at the age of higher secondary schooling, and (scholastic) environmental changes did not have the same impacts as those known from the literature for earlier periods of life. Moreover, the change in curricular intensity may have been too minor in magnitude to play a pivotal role for personality development in the present case. Besides this most likely interpretation, it is also possible that various partial effects have added up to a total of zero. However, neither a systematic age effect nor clear-cut indications for an adding-up of curricular and leisure effects can be found. Therefore, gains in knowledge, exogenous environmental effects, and self-determined environmental effects remain as a possible explanation for such a "hidden" impact. An implicit argument against this possibility is that the picture for the assessed personality dimensions would probably be less unequivocal, as the considered age period is not supposed to have affected all personality dimensions in the same way and magnitude. The analysis of the background characteristics reveals that both cohorts are, on average, equally endowed with relevant inputs prior to the reform. Therefore, the consensus view that higher academic requirements at school come at the expense of personality development cannot be supported. Despite the relevance for academic discussions, this result is also of political importance. Protests in the German federal states which have recently implemented the reform are aiming for its reversal. One of the main arguments involved is that an abbreviated school time impedes the development of the students' personality. The results show that this claim is arguably not justified: there is no evidence that the education reform itself significantly affects students' personality traits.

9

Application II: Stability of Control Related Traits in Working Age*

Along the arguments on the general malleability of personality traits that were introduced in Chapter 7, the empirical investigation of the last chapter has established some consensus view on the stability of various traits in late adolescence, at least with respect to the quite specific setting considered there. The following chapter shifts the focus towards an even later stage of the life course, namely individuals in working age. More specifically, it investigates whether control-related attitudes affect the probability of getting trapped in poverty, and whether the reverse association from poverty to control-related attitudes is also detectable. The latter association would coincide with an instability of control related traits induced by poverty experiences. To hypothesize this point, one has to consider that individual poverty is highly state-dependent, such that the poor are often literally "trapped". As the underlying process is a black box in large parts, it is difficult to unveil the true mediation paths. Those may range from imposed budget or time constraints to physical and psychological reactions. Regarding the latter, a consensus view across the results discussed throughout the previous chapters suggests control-related attitudes to be one such mediator. Though the findings presented thus far indicate that personality traits generally stabilize towards adulthood, it has also been shown that they remain susceptible to environmental influences to a certain degree. This may particularly hold for more vigorous changes in life circumstances, like poverty experiences.

9.1 Motivation

Developed countries are typically beyond the appropriate scope of poverty definitions that relate to levels of physical subsistence or nutrition (see Lambert, 2001). Instead, purely income-based concepts, where people are presumed to be poor whenever their income falls short relative to certain reference regions of the income distribution, are preferable in such

* The results presented in this chapter build on Thiel and Thomsen (2015).

cases.[1] Albeit a large literature deals with the cross-sectional aggregation and comparison of this conception of poverty (see, e.g., Zheng, 1997), only few studies assess the important dynamic implications of poverty experiences on the individual level.[2] Among the few ones, intertemporal associations that give rise to individual poverty paths are implicitly modeled as state or duration dependence. To clarify more on the relevant aspects, the study at hand examines potential interrelations between poverty paths and the dynamics of other potentially involved determinants. These may include typical choice variables that relate to the individual level as well as to the household level, but also entities that capture the individuals' psychological conditions. Regarding the former, it is assumed that decision variables like childbearing and household formation, as well as employment are likely to interact with poverty in that way. With respect to the latter, I suppose perceived control to represent a major dimension of the psychological conditions in such an income-based setup. Using longitudinal data from the German Socio-Economic Panel (GSOEP) a dynamic structural model that relaxes strict exogeneity assumptions between the model components is considered. I examine whether control perception has some direct impact on the development of poverty in terms of different income-based metrics, as well as some indirect effects via other entities involved. Conversely, I also account for a feedback of previous poverty experiences on control attitudes and on the mediating variables. From a methodological point of view, these relaxed exogeneity assumptions follow Wooldridge (2000).

Obtaining a deeper understanding of causal dependencies for individual poverty seems worthwhile for a number of reasons. The main line of argumentation invoked in most debates on poverty builds on the use of occasionally imprecise indicators and often premature inference based on them. This point is best exemplified by annual aggregates of headcount ratios that are the ubiquitous instrument for reportings on poverty (see, e.g., Zheng, 1997), but which are also subject to major interpretational and conceptual pitfalls (see, e.g., Foster et al., 2013). Furthermore, headcount aggregates do not take into account how poor the persons concerned are.[3] Another drawback, especially for the evaluation of causal interrelations, is that the cross-sectional perspective of poverty aggregates is uninformative in terms of the inter-temporal dimension that poverty evidently

[1] As opposed to an absolute, somatic and nutritional notion of poverty, an income-related concept of poverty is relative in nature. As income is just one means to achieve well-being, another view on poverty pioneered by Sen (1982) relates to well-being arising from the freedom of choice among potential achievements that income enables. This generalized poverty concept thus extends beyond matters of income, but is hardly implementable in empirical terms.

[2] See Aassve et al. (2006a) for an outline of the respective literature, that has not overly increased ever since.

[3] In the literature on axiomatic approaches to poverty, this feature is called distribution-sensitiveness (see, e.g., Zheng, 1997). Further axioms classifying the properties an aggregate poverty measures should comply with are also given in Sen (1976), as well as in Foster and Sen (1997).

possesses and furthermore provides no reasoning on the individual level.[4] If poverty were a transitory phenomenon that bears on different parts of the population over time, a cross-sectional perspective may be adequate. However, as has been shown in various studies (see Stevens, 1999, among others), poor people are often trapped in poverty. Beyond this well documented pattern, the underlying individual causes should be disentangled more explicitly in order to deduce potential counter-measures.

Following the dedicated strands of the literature, two main mechanisms causing such persistence may be in order. On the one hand, individuals can differ in terms of characteristics that are relevant for the propensity to slip into poverty. Especially when it is assumed that poverty is rooted in income only, the understanding of the relevant causes is well developed and subject to a long-standing literature (see, e.g., Heckman et al., 2006). As of late, the incorporation of cognitive and affective factors stemming from the psychological field (like the control attitudes considered here) adds to this literature (see, e.g., Almlund et al., 2011), also in explaining other outcomes related to labor market success. In economics, such cognitive and affective factors are better known as traits or preferences. On the other hand, affective components and other individual characteristics may be further deteriorated by past poverty experiences, thus locking-in the persons concerned. Such mechanisms have been hypothesized in the sociological literature on poverty for a long time already (see, e.g., Sher, 1977).[5] In empirical economics a reasoning based on changes in attitudes or deprecation in human capital is usually alleged as an implicit explanation for the observed state dependence (see, e.g., Aassve et al., 2006a). As with the causal relation between traits and economic outcomes, a perspective that draws on disciplines other than economics extends the set of potential mediating pathways. A meta-analysis conducted by Haushofer and Fehr (2014) shows that apart from plain economic explanations, like credit constraints, psychological factors (cognitive and affective ones) and even neurobiological factors are evident predictors of poverty traps.[6]

By now, frameworks that allow for a circular causality between poverty and individual characteristics are bound to a theoretical literature on life cycle saving and wealth accumulation. This particular branch uses concepts from behavioral economics, like hyperbolic discounting, to explain individual heterogeneity in accumulation paths and feedback that trap individuals within respective trajectories (see, e.g., Bernheim et al., 2013, for a

[4] On the aggregate level, endeavors to incorporate dynamic aspects into measures of poverty have been made (see Hojman and Kast, 2009, and the literature they cite). By construction, however, even these dynamic metrics cannot account for individual determinants, as no conditioning sets are accounted for.

[5] Sher (1977) invokes disinvestments as people become poorer and less self-confident, though he does not only consider investments at the individual level, but also at the community level.

[6] For instance, poor living conditions may impede achievements in subsequent tasks via decreased self-regulating capabilities (see Muraven and Baumeister, 2000).

recent example). Hyperbolic discounting has behavioral implications that are often paraphrased as self-control or self-regulation (see, e.g., Ainslie, 1991). Preference parameters and traits therefore roughly represent the same causes of individual behavior, albeit in different hypothetical frameworks.[7] On that account, the various empirical assessment tools that exist in the field of trait psychology (see, e.g., Rotter, 1966, or Tangney et al., 2004) capture different aspects of control attitudes, at least to a decent extent.[8]

The following empirical analysis combines psychometric measures of control-attitudes and poverty formation in an interacting fashion within a panel framework. It provides an end-to-end treatise along the whole line of argumentation hypothesized by the respective parts of the literature. It is a matter of course that it does so with a necessary abstraction from more detailed model entities that prevail within the self-contained scopes of the different subdisciplines. Using trait measures to explain individual poverty status adds to the literature of poverty constitution, primarily by providing an additional source for typically unobserved individual heterogeneity. As a spinoff, allowing for interdependencies between both entities contributes to the literature on general malleability of traits throughout adulthood (see, e.g., Roberts et al., 2006, Cobb-Clark and Schurer, 2013).[9]

9.2 The Measurement of Poverty

An initial point to be clarified is as to why an understanding of poverty based on individual valuation or well-being does not always has to coincide with a single-dimensioned lack of income.[10] For someone to be declared poor or not poor, it may not be sufficient to know that person's current income status, as the well-being derived from monetary endowments is likely to vary across individuals. As such, some preliminary assumptions are needed in order to make income a meaningful stand-alone objective.

[7] Recall from Chapter 4 that preference parameters are utility-related representations of behavioral differences, whereas a trait is seen as more of a task-specific skill or ability in the sense of human capital literature (see, e.g., Almlund et al., 2011).

[8] Though the associations between psychometric constructs and preferences in behavioral economics are far from perfect (see, e.g., Almlund et al., 2011).

[9] Large scale cross-sectional analyses show peaks of mean-changes for highly nuanced age cohorts until age 30 (see, e.g., Roberts et al., 2006). For other traits, this age pattern is moderated when intra-individual measures are employed (see Cobb-Clark and Schurer, 2013). Though intra-individual measures suggest a higher degree of stability for the working age population, in general no complete time-invariance can be established.

[10] I do not employ the term "utility" in this context as some general utilitarian axioms are unduly strict for the evaluation of income inequality and poverty. Foster and Sen (1997) and much of the related literature elaborate on this criticism. To make this distinction more apparent, alternative terms like "well-being" or "valuation" are used instead.

9.2.1 Poverty Based on Income

The understanding that underlies an association of income and well-being is that income results from rational behavior that seeks to maximize well-being. A common approach to concatenate both concepts is to use some additional data to approximate differences in needs, prices, and household composition. Unfortunately, even for individuals that are observationally homogeneous in that sense, preferences, motives and enjoyment abilities are diverse, making it still problematic to compare individual levels of income and infer different well-being from such variation. If one allows for a comparison of individual differences in ratios of well-being derived from income (see Foster and Sen, 1997), it is possible to relate income and well-being via an expenditure function.[11] This setup would require multiple income realizations in a very close time interval (or some stated equivalents, see Dagsvik et al., 2006). Without closeness in time one runs the risk that constraints and preferences change in between. In most settings, including the one used here, such information is not available. As a consequence, it is inevitable to impose some normative assumptions on the individual well-being derived from income. One possible approach is to make normative presumptions on the complete functional form of individual well-being and thus allow for interpersonal level-comparisons. This understanding of the potential use made from income may be too strict and can be relaxed to some extent. A second possible approach is less narrow and follows from the relativeness of income poverty. In this context, relativeness means that preferences are not claimed to be completely self-interested, but can depend on some distributional reference point.[12] A threshold income that discerns poor and non-poor individuals complies with this requirement. What remains to be assumed is that the interpersonal difference in well-being that is induced by a certain deficit of the realized income with respect to the reference point monotonically increases as the distance between both income levels grows. Conversely, a change in well-being arising from a shift towards that reference point has to follow the same rules for all individuals. These assumptions follow the notion of Atkinson's (1970) "ethical observer" in that it is merely assumed that certain hypothetical differences are based on comparable valuations.[13] A rather critical point in this assumption

[11] In utility theory, the comparison of difference ratios is referred to as a "cardinal" measure.

[12] As discussed by Lambert (2001), imputing pure self-interest in individual income valuations is necessary when all incomes in a given population have to be assessed, as no objectively superior (or inferior) reference income can be defined in these circumstances. This is usually the case for inequality measures. If the mappings from incomes to valuations are to be defined for subsets of the population only (e.g., for all poor individuals), a reference value is can be meaningfully defined, however.

[13] It should be noted, that the presumptions on the functional form for individual distance-comparisons are somewhat stricter than those originally required by Atkinson for the aggregate level. This follows from the fact that on the aggregate level, exactly equivalent gains and losses from marginal redistributions of incomes have to be considered, whereas on the individual level with a reference

is that the awareness of where this reference point is located also has to coincide across individuals to a very large extent. Otherwise, no judgements about derived well-being can be achieved. Following these presumptions provides a "working definition" that gives individual poverty levels some projection into well-being, however.

9.2.2 Generalized Poverty

For completeness, it should be noted that multi-dimensional poverty concepts that go beyond the connection of income and well-being have also gained considerable attention. In the most prominent extension pioneered by Sen (1982), poverty is not characterized by a stand-alone entity like available income and the corresponding level of well-being (see, e.g., Foster and Sen, 1997). Accordingly, commodities that are achievable as a matter of income, are only a means to satisfy needs. Individuals use commodities according to some common transformations the purpose of which is to comply with bundles of characteristics the individual seeks to be fulfilled in one way or another. For instance, transportation is one particular characteristic a car fulfills, but one that is also available from buying a bus ticket.[14] Furthermore, there exist individual-specific "patterns of use" for the transformed commodities, the so-called functionings. Well-being derived from these functionings does not result from their actual realization, but from their realization given the possibility to choose from various other functionings. Sen calls this freedom of choice a capability set.[15] The capability approach implies that individuals with the same observed functionings may have different well-being because their choice sets, i.e., their capabilities, are different. For instance, a paraplegic person does not have the freedom to choose driving by car as a transportation mode at all. Unfortunately, empirical implementations of these concepts (see Schokkaert, 2007, for an overview) exhibit a high degree of complexity in static settings already. I thus refrain from further considerations of generalized poverty in what follows and use the terms income-poverty and poverty interchangeably henceforth.

9.3 Potential Mediators of Poverty

In what follows, some pertinent mediation processes that emerge from previous findings in various fields of the literature are sketched, albeit only as an excerpt of the most recent ones. Their origins evolve from family economics and more recent strands of the human capital literature, from behavioral economics, and from interdisciplinary research

point, varying differences in well-being and varying margins occur at the same time.

[14] In the literature dealing with formalizations of Sen's approach, such transformations are frequently compared with consumer transformations in spirit of Gorman (1980) and Lancaster (1966).

[15] For instance, people may have the capability to elude hunger, but may choose to diet anyway.

on psychological and neurobiological factors. A consideration of these findings when setting up the empirical framework seems fruitful as they help to identify those driving forces where strict exogeneity seems implausible.

9.3.1 Socio-Demographic Factors

As already addressed above, the relevance of income differs as the needs of people differ. Many of those needs are objective ones, in that they can be defined by fairly general individual characteristics. Following the family-economic literature, such characteristics may evolve successively or parallel and comprise decisions like household formation, childbearing, and labor market participation (see, e.g., Aassve et al., 2006b). They are assumed to take place on an individual basis, but with some collective aims underlying them (see Browning et al., 2011).

The determination of household income is intrinsically rooted in these factors. However, predicting dynamic cross-effects by means of established theoretical frameworks is difficult, as the directions and magnitudes are largely unforeseeable.[16] For instance, gains arising from household formation may include the ability to exploit economies of scale or comparative advantages in transforming market commodities to household goods (see Becker, 1993). Moreover, the publicness of household goods among household members usually leads to budget increases for further affordable goods.

Individuals draw their decisions on these factors to a more or less extent, but owing to their unobserved preferences. Therefore, these features are what the concept of "equivalent incomes" seeks to mimic in empirical investigations of income data. But there are also household characteristics involved in income generation that are not captured by equivalent incomes at all. Several unobserved household patterns may impinge on time constraints or credit constraints, but at the same time may be outcomes of decisions that depend on these constraints. Labor market participation and childbearing are two examples that follow this logic (see Aassve, 2006b). The potential to share risk may be another important point in explaining household constitution (see Browning et al., 2011), one that may be particularly relevant in the present framework as it may manifest in changed attitudes or income paths. As such, the likely occurrence of factors not captured by equivalent incomes urges for their additional consideration for a proper representation of poverty dynamics.

[16] Becker (1993) and Browning et al. (2011) give a comprehensive account on these and related topics.

9.3.2 Preferences and Traits

Though depending on joint decisions, the incomes of households eventually arise from their member's incomes, implying that individual characteristics are still crucial. Apart from socioeconomic characteristics and other observables that affect incomes and other achievements, the discussion in the previous chapters has shown that traits and preferences capture more and more attention in the related fields of economics and psychology (see Almlund et al., 2011). Recall the fact that there is a compliance in that literature that certain types of preferences and traits have akin behavioral implications, though the angle of assessment is somewhat different. As has been made clear above, psychological traits are primarily intended to project various dimensions of behavior into a lower-dimensional continuum, focussing on generality, situation-invariance, and durability. In economics such traits are usually seen as a productivity enhancing human capital stock, where productivity refers to tasks in a wider sense, not only those envisaged on the labor market.[17] Behavioral preference parameters, on the contrary, refer to mathematical laws that link specific stimuli to behavioral responses. In economics, the interest in such parameters is mostly limited to decision and optimization frameworks. As already pointed out, an integrative framework for preference parameters and personality traits is yet not explicitly established. Almlund et al. (2011) provide an overview on some first correlation studies that reveal largely intuitive relationships between both concepts.[18] As such, most of the following findings on the role of preferences and traits in poverty constitution suggest similar mediation paths, though they stem from largely unrelated fields of the economic literature.

An impact of productivity enhancing traits on incomes and related entities is shown in various empirical studies (see, e.g., Heckman et al., 2006) that have been discussed in Chapter 6. An explicit consideration of poverty constitution, however, is limited to preference-related studies that deal with life-cycle savings. These (mostly theoretical) models attribute interpersonal variation in saving behavior to differences in time preferences, risk tolerance, exposure to uncertainty, and relative tastes for work and leisure, with a particular focus on non-standard types of preferences.[19] They establish that differ-

[17] Indeed, human capital of this kind has been addressed in the literature all along (see Becker, 1964), but has not been made explicit due to a lack of measurability.

[18] Becker et al. (2012), for instance, show that parameter measures for time preference are predominantly correlated to traits like openness and neuroticism, the latter of which in turn is moderately associated with perceived control. Moreover, a verbatim compliance holds for perceived control and aspiration-striving, as people with external control perception, i.e. people who believe achievements in life are due to luck or fate, are likely to exhibit low levels of aspiration in their future life.

[19] Non-standard preference parameters arise from findings that agents are not generally capable of solving complex multi-stage optimization problems (see Thaler, 1994, for a discussion) as assumed in the traditional literature on life cycle savings (see Modigliani and Brumberg, 1954). The doubts

ent endowment conditions can lead to individual saving paths that can be understood as a poverty trap. Hyperbolic time preferences as defined by Ainslie (1991), together with borrowing constraints, can lead to occasional exuberance in consumption that in turn leads to low wealth-accumulation in which individuals get trapped (see, Laibson, 1997, Bernheim et al., 2013). Such local deviations from individually rational accumulation plans are a form of time inconsistency in preferences, a behavior that Ainslie (1975) has introduced as self-control.[20]

Somewhat related to this notion of executive control or self-control is a person's so-called "capacity to aspire" (see Appadurai, 2004).[21] In an economic context (see Genicot and Ray, 2012, Dalton et al., 2013), a lack of aspiration can be construed as a factor that endogenously lowers reference points in valuation (relative to agents with higher levels of aspiration) that lead to lower accumulation pathes of wealth. There is a circular relation between lower aspirations, wealth levels, and valuations drawn from both. Thus, poverty self-perpetuates in a downward circle, as individuals may lose their aspirations when low income is persistently experienced.[22] There also is a growing empirical support for these mostly model-based mechanisms addressed to this point. Haushofer and Fehr (2014) provide an intriguing argumentation by summarizing experimental and empirical findings from various fields. For one thing, poverty and other unpleasant life events are shown to be causally related to well-being, affect, and stress, where stress levels are gathered through self-information and measured hormone levels. These in turn, are known to have a significant impact on time and risk preferences, building on a substantial literature of behavioral lab-experiments. For another thing, the authors also emphasize that poor people are more liquidity-constrained, making changes in their saving behavior often a matter of external factors rather than of intrinsic preferences and traits. As such, non-normative changes in life circumstances, like a major income drop, impinge on several behavioral parameters, and thus possibly on related traits like control perception.

casted are as to whether *(i)* the complexity of the problem is to high, *(ii)* the chance to learn is low (as the consequences of saving decisions are not immediately revealed), *(iii)*, no easy rule of thumb is available.

20 More explicitly, poor people with low assets are more prone to consumption sprees as the "severity of punishment" is lower for these individuals. There are some empirical facts underpinning this notion, in that poor people frequently engage in all kinds of commitments in order to stick with their initial saving plans (see, Bernheim et al., 2013, and the literature they cite). For instance, Thaler and Benartzi (2004) show that employee commitments on savings from future wage gains, significantly increased saving rates.

21 Aspirations in Appadurai's anthropologic sense reflect wants, preferences, choices, and calculations.

22 In case of increasing aspiration levels, Dalton et al. (2013) paraphrases it as "every ceiling, when reached, becomes a floor...". As opposed to the control-related approaches, where individual poverty traps result from initial conditions only, the "capacity to aspire" approach is characterized by a distinctive feature. It explicitly accounts for a parameter which mimics that experienced poverty may further deteriorate the respective individual path.

As has been addressed throughout the previous two chapters, dynamics over the life course that allow for a comparable reasoning about experience of poverty are long established in the psychological field, though predominantly for normative environmental changes that are supposed to happen to every person within a certain age span. The corresponding literature shows the highest degree of susceptibility for personality traits in early childhood. From there on, it steadily decreases throughout later childhood, adolescence, and adulthood. For age spans beyond adolescence, large scale cross-sectional analyses show peaks of mean-changes until age 30 (see, e.g., Roberts et al., 2006). These results, however, are moderated when intra-individual measures and very specific or non-normative life events, like death of a spouse, are used. Following Cobb-Clark and Schurer (2013) among others, the effects become even weaker for the working age population, though no complete time-invariance can be established.

Summarizing the above studies, there are several surmisable associations between income and control attitudes, not all of them in a coherent way regarding low incomes and perceived control, though. Some persistent changes seem to have an impact on traits, but are normative in nature and thus also happen to people with higher income. Evidence on non-normative life events, as those that happen to poor or deprived people, are usually based on those events that are onetime occurrences. They may, however, permanently affect the social roles of the people concerned. Thus, the consequences for more persistent but non-normative events, like poverty, are less foreseeable.

9.4 Data

9.4.1 The Sample

For the empirical analysis, I use data from the German Social Economic Panel (GSOEP). The GSOEP is a longitudinal survey conducted since 1984 by the German Institute for Economic Research (see Wagner et al., 2007). It provides comprehensive information on a representative sample of German households, including annual information on household income, decision variables related to household composition and employment, as well as other characteristics that are of particular interest in the analysis. Further information provide aspects about labor market history, health, biography, well-being, family background, and living-conditions. In the waves 1994, 1995, 1996, 1999, 2005, and 2010 the survey contains inventories that measure control attitudes (see Appendix D). In the latter four waves the inventories are a version of the Locus of Control established in Chapter 3. As these attitudes are among the outcomes of primary interest, the empirical analysis is predominantly based on the corresponding waves. But also the sampling periods in between are used exploit additional information on some of the mediating factors. Based on the register of the 2010 wave, a total of 28,776 individual observations are available. As income determination plays a crucial role in analyzing poverty dynamics, the focus is on sample members in working age (18 to 65). Considering the timespan from 1994 to 2010, and given panel attrition and unit non-response, one ends up with about 13,000 (gross) observations in each wave, where the exact cross-sectional sample sizes feature substantial further drops due to item non-response.

9.4.2 Measuring Perceived Control

In order to involve control-related attitudes into the empirical model of poverty formation, two specific trait inventories that are part of the GSOEP can be used. Both comprise questions related to certain dimensions of behavior and/or attitudes. As with the general case discussed in Chapter 3, the respective responses are stated on Likert scales that cover manifestations from “completely disagree ” to “completely agree”. The fact that several items are used in order to obtain the individual scores increases the reliability of the constructs. The item inventories used as of the 1999 wave are based on the seminal “Locus of Control” scale of Rotter (1966). As previously mentioned, it assesses an individual’s attitude on how self-directed (internal) or how coincidental attainments in her or his life are. It thus fundamentally relates to the notion of self-control addressed in Section 9.3.2, but does not capture exactly the same facets. Locus of Control merely captures individual beliefs in whether self-determination exists, not in how successful one could be in governing it.[23] Self-control, on the contrary, also encompasses such motivational concepts, often denoted Self-Efficacy. The GSOEP uses a 10-item version of the original Rotter scale. It has to be coded such that high internal (low external) attitudes represent a high degree of control-perception. Exactly similar versions of this scale are available for the waves 1999, 2005, and 2010. A slightly different prequel version can be found in waves 1994 to 1996.

Since the trait inventories build on multiple items, relying on unweighted raw scores or arbitrary selections from all available items does not necessarily lead to unidimensional and errorless measures of individual control attitudes. To solve the former problem, one has to obtain a favorable item selection from an explorative factor model as suggested in Chapter 5. Table 9.1 presents the results for the finally selected item combinations. For waves 1999, 2005, and 2010, the same items are used and the resulting pattern is quite stable across waves. In order to account for potential gender differences, the common

[23] These beliefs are a major driving force with respect to educational attainments and wages (see, e.g., Heckman et al., 2006, Mueller and Plug, 2006, Heineck and Anger, 2010).

Table 9.1: Iterated Principle Factor Analysis for Polychoric Item Correlations

		Common Variances			
		(female)		(male)	
		Eigenvalue	Proportion	Eigenvalue	Proportion
2010	1st Principal Factor	1.671	0.896	1.805	0.876
	2nd Principal Factor	0.143	0.077	0.131	0.064
2005	1st Principal Factor	1.588	0.886	1.721	0.940
	2nd Principal Factor	0.127	0.071	0.069	0.038
1999	1st Principal Factor	1.699	0.939	1.761	0.959
	2nd Principal Factor	0.074	0.041	0.052	0.029
1995	1st Principal Factor	2.394	0.884	2.459	0.898
	2nd Principal Factor	0.201	0.074	0.202	0.074
		Factor Structure			
		Factor 1	Factor 2	Factor 1	Factor 2
2010	Item 1	0.576	0.136	0.628	0.251
	Item 2	0.508	0.062	0.531	0.083
	Item 3	0.673	-0.106	0.669	-0.243
	Item 4	0.368	0.268	0.405	-0.014
	Item 5	0.702	-0.195	0.720	-0.047
2005	Item 1	0.526	0.101	0.591	0.058
	Item 2	0.527	0.079	0.523	0.143
	Item 3	0.647	-0.121	0.653	-0.113
	Item 4	0.363	0.264	0.398	0.131
	Item 5	0.696	-0.161	0.716	-0.123
1999	Item 1	0.616	0.056	0.649	-0.080
	Item 2	0.512	0.031	0.537	0.087
	Item 3	0.667	-0.027	0.658	0.061
	Item 4	0.357	0.212	0.366	0.140
	Item 5	0.697	-0.155	0.696	-0.123
1995	Item 1	0.626	0.017	0.657	-0.058
	Item 2	0.624	0.177	0.631	0.215
	Item 3	0.548	0.308	0.565	0.293
	Item 4	0.818	-0.252	0.824	-0.246
	Item 5	0.802	-0.104	0.795	-0.076

factor model is separately estimated for female and male sample members. There are combinations with less than five retained items that exhibit slightly higher shares of common variance. Nonetheless, I opt for the five-item alternative as a higher number of measurement equations generally increases the quality of the factor scores derived later on. Apart from the share of common variance, the second objective of the item selection was to obtain homogeneous loadings on the first principal factor. Surprisingly, the item selection that complies most with both aims is the same for females and males. As mentioned in the data section, the 1995 variant of control perception is a slightly different prequel of the later one. It is thus meaningful to jointly examine the factor structure and factor pattern of the 1995 and the 1999 item inventory. The corresponding results are provided in Table 9.2.

Table 9.2: Factor Structure and Factor Pattern for the Association of the 1995 and 1999 Perceived-Control Inventories

	Common Variances			
	(female)		(male)	
	Eigenvalue	Proportion	Eigenvalue	Proportion
1^{st} Principal Factor	2.791	0.693	2.982	0.713
2^{nd} Principal Factor	1.236	0.307	1.199	0.287
3^{rd} Principal Factor	0.110	0.027	0.142	0.034
	Factor Structure			
	Factor 1	Factor 2	Factor 1	Factor 2
Item 1 (1995)	0.584	-0.193	0.619	-0.221
Item 2 (1995)	0.538	-0.319	0.535	-0.324
Item 3 (1995)	0.509	-0.160	0.542	-0.121
Item 4 (1995)	0.721	-0.343	0.737	-0.340
Item 5 (1995)	0.741	-0.288	0.730	-0.305
Item 1 (1999)	0.412	0.451	0.471	0.447
Item 2 (1999)	0.425	0.318	0.435	0.316
Item 3 (1999)	0.396	0.578	0.399	0.562
Item 4 (1999)	0.268	0.202	0.326	0.200
Item 5 (1999)	0.504	0.440	0.516	0.408
	Factor Pattern (after Rotation)			
Item 1 (1995)	0.553	-0.270	0.589	-0.292
Item 2 (1995)	0.490	-0.387	0.494	-0.384
Item 3 (1995)	0.483	-0.226	0.524	-0.183
Item 4 (1995)	0.669	-0.436	0.692	-0.423
Item 5 (1995)	0.696	-0.384	0.689	-0.388
Item 1 (1999)	0.469	0.391	0.520	0.389
Item 2 (1999)	0.464	0.259	0.469	0.263
Item 3 (1999)	0.470	0.520	0.462	0.512
Item 4 (1999)	0.292	0.165	0.347	0.161
Item 5 (1999)	0.558	0.368	0.561	0.345

Factor structure refers to the factor loadings under factor orthogonality that can be seen as correlation coefficients. After rotation, the reported coefficients are only interpretable as *factor pattern/weights* and do not represent correlations any longer.

Given that the item combinations for 1995 and 1999 are jointly evaluated, the share of the common variance is reduced by some 15 to 20 percentage points. Instead, a second principal factor, which accounts for about 30 percent of the overall variance, occurs. The next common factor is again negligible. The loadings suggest that the association of all selected 1995 and 1999 items with respect to the principal projection axis is as intended. However, the second axis obviously implies a full reversal for the projection of both item blocks. Fortunately, this second dimension is almost orthogonal to the first factor, making the prequel version of the perceived control scale a still descent approximation to the later one.

To reduce the error proneness, the semiparametric item response model suggested in Chapter 5 is fitted to the extracted item sets. The resulting response parameters are in turn used to obtain latent factor scores for each individual in the sample. The procedure is applied to all waves that contain control related measures.

9.5 Descriptive Results

In this section I present some first descriptive results that suggest some cursory patterns of the dynamics in income related poverty. Furthermore, individual characteristics that are supposed to be important determinants or endogenous mediators of poverty are presented.

9.5.1 Poverty and Equivalence Incomes

In Section 9.2 some of the problems that arise when inferring from incomes to individual well-being were addressed. A first step in order to make income an indication of individual well-being is to adjust the former for observable interpersonal differences in needs. Most commonly, so-called equivalence weights (see, e.g., Cowell, 2011) are used for this purpose. I apply a modified OECD scale (see Atkinson et al., 1995). It assigns a weight of 1 to the adult head of a household, a weight of 0.5 to each additional adult member, and a weight of 0.3 to each child being below age 15. The weights are summed for each household in order to obtain the total of equivalent adults that have to share a respective net household income, where household income comprises earned income and capital income.[24] The rationale for doing so is that individuals who live together in one household experience gains in terms of the usability of collective goods. They have economies of scales in the transformation from market goods to household goods. Though

[24] More specifically, the GSOEP also allows for the consideration of home ownership (i.e. saved rent), social transfers, other transfers, as well annual extra payments. Subsequently, tax payments are computed based on these and other relevant magnitudes.

equivalence scales only adjust incomes and not affordable commodities arising from them, the underlying notion is somewhat similar to those addressed in Section 9.2. It should be noted, however, that equivalence incomes are more of an "empirical crutch" to somehow accommodate the theoretical requirements necessary for inference about well-being. As already pointed out, it remains to be the only feasible solution without the availability of multiple income realizations per individual (see, e.g., Dagsvik et al., 2006) or without additional assumptions on the connection between income and well-being (see Layard et al., 2008, for such assumptions).

Using the scale by Atkinson et al. (1995) and designating the cutoff value, which separates the poor from the non-poor, to be six tenth of the median equivalence income, some first descriptive results are obtained. As illustrated by Figure 9.1, the poverty line in Germany has increased in nominal terms (on a monthly income-basis) throughout the period from 1995 to 2010. The increase amounts to almost 50 percent, which is only partially on account of an increased price level, as the CPI increase in the same time span is about 25 percent (according to the Federal Statistical Office). Another reason is that some of the skewed frequency mass, especially at the lower tail, shifts to the right between 1995 and 2010, and with it, the reference for the cutoff point.

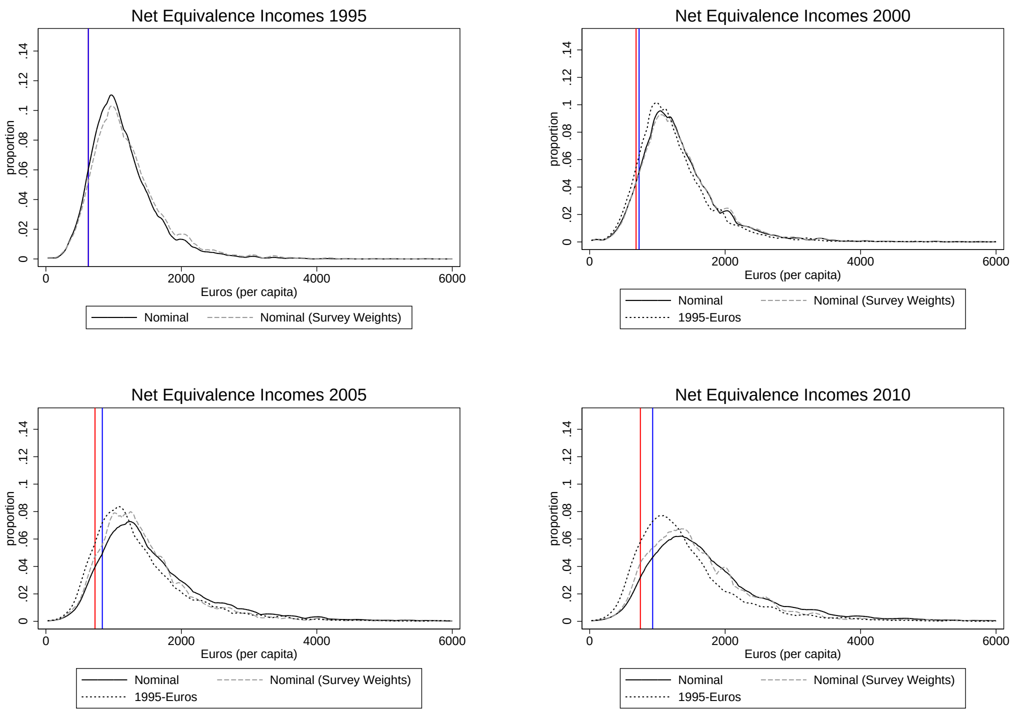

Figure 9.1: Development of Monthly Net Equivalence Incomes and the Poverty Line (Nominal = blue)

Considering the five-year increments displayed in Figure 9.1, the increase in the cutoff value has been steady. The corresponding changes in the shares of poor people in the sample are not so, however. They amount to slightly more than 13 percent in 2005 and 2010, and to 9.8 and 9.6 percent in 1995 and 2000. This finding is often invoked as one of the major weaknesses of headcount ratios, as the concentration (or distribution) of individuals around the cutoff affects the sensitivity of the headcount ratio in response to small but erratic income changes (see Foster et al., 2013). The headcount ratios also do not differ substantially between female and male GSOEP respondents. In case of the 2005 wave, it amounts to 12.7 percent as opposed to 11.1, which is largely in line with ratios provided by census data.

The development of the density plots over time do not provide information whatsoever about the time people remain in the lower tails of Figure 9.1. To illustrate the fluctuations among the poor and the non-poor over different timespans, mobility plots as presented in

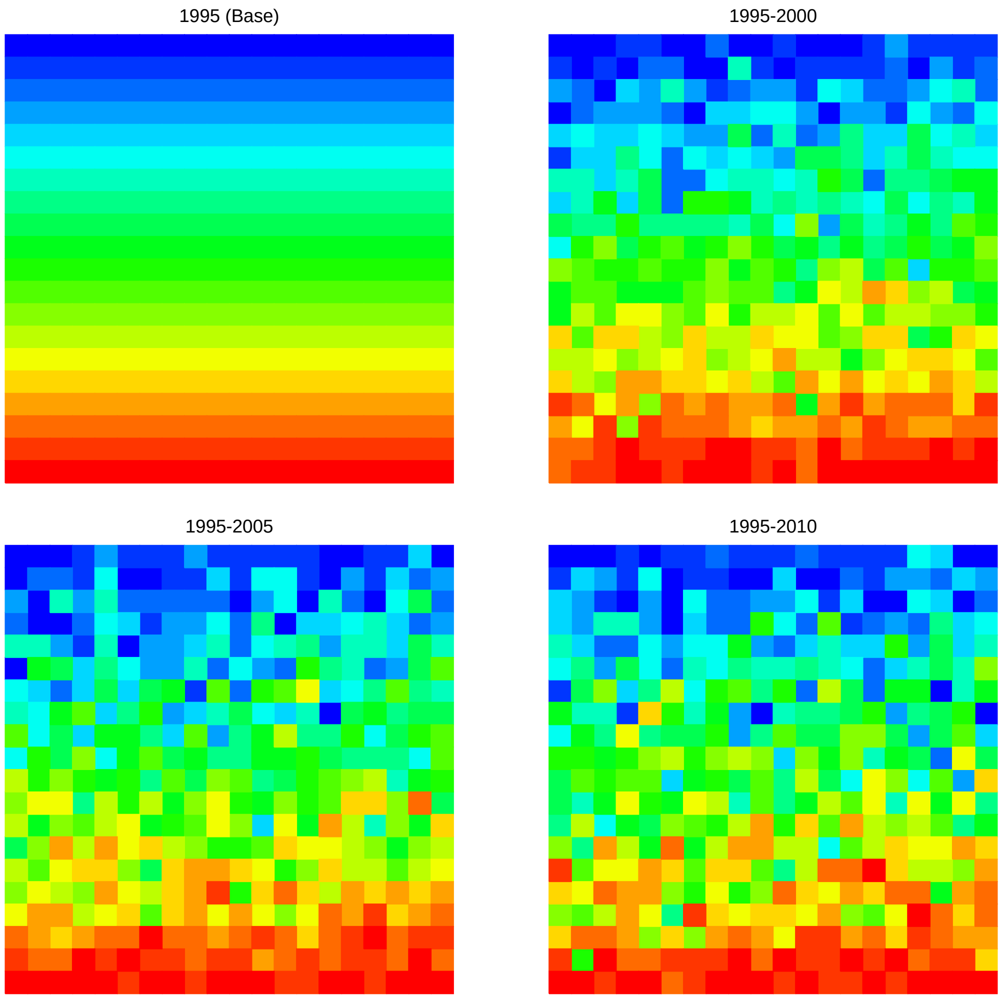

Figure 9.2: Mobility plots for net equivalence incomes with 1995 as a base period. The reference locations of the 1995-members for the 400 quantile increments in 2000, 2005, and 2010 are sorted in row-major order from left to right.

Figure 9.2 are more suitable. The patterns suggest a substantial degree of persistence for the net equivalence incomes in the different rank-groups over the time intervals 5, 10, and 15 years. Using samples of those who are respectively observed at both points in time, the chances of leaving the lower regions of the distribution seem to slightly increase along with the considered timespans, but the dependence on the initial state is yet tremendous. Even after 15 years, most poor rank-groups are still poor with regard to their equivalence incomes.

In summary, the descriptive results reveal that path dependence obviously is a major factor, due to reasons whatsoever, and therefore should be analyzed on the individual level. Moreover, the described ambiguities with regard to the shares of the poor urge for some refined measures in order to better capture the extent of income poverty.

9.5.2 Poverty and Background Characteristics

Selected descriptives on sample characteristics are presented in Table 9.3 for the 2005 wave. The results for other waves do not differ substantially. In line with the literature on demographic transitions (see Aassve et al., 2006b), characteristics that are tied to decision variables underlying household constitution, labor market participation, and the like, are to be considered.

The share of full-time employees is more than twice as large in the male sample, for poor as well as for non-poor individuals. Another substantial divergence holds for full-time job experience and the share of persons that hold a university degree. The remaining means and shares of the variables are relatively equal. Irrespective of their poverty state, about 30 percent of the female and male respondents have a higher secondary schooling degree (overall shares are not displayed). Roughly 65 percent have an eight or ten-year schooling degree. The average age for the working age sample is roughly above 40 years. The share of east Germans in the sample largely coincides with the fraction in the overall population.[25] About 68 percent of the respondents live together with at least one child below age 18. About 60 percent live together with a partner, fiancé(e) or spouse. A higher extent of mean differences occur when poverty states are considered as well. Only for secondary school degrees, one fails to reject the null of equal mean shares, though only at the 1-percent level and only for higher secondary schooling in the male sample. For the remaining characteristics contained in Table 9.3 some substantial differences between poor and non-poor individuals are apparent, most of them with quite similar patterns for female and male respondents. Most remarkably, the share of full-time employed among the non-poor is more than three times higher. In line with this, the share of university

[25] According to the Federal Statistical Office.

Table 9.3: Sample Descriptives (Wave 2005)

	Non-Poverty		Poverty		Mean-Diff.
	Mean	Stand. Dev.	Mean	Stand. Dev.	p-value[b]
Female					
Employed (D)	0.330	0.470	0.099	0.299	0.000
Perceived Control[a]	0.050	0.984	-0.366	1.054	0.000
Some School (D)	0.667	0.471	0.671	0.470	0.779
Higher Secondary (D)	0.297	0.457	0.235	0.424	0.000
University (D)	0.125	0.331	0.037	0.190	0.000
Job Experience (Full Time)	11.373	10.760	7.731	9.370	0.000
Age	42.137	13.170	38.319	13.658	0.000
East German (D)	0.205	0.403	0.319	0.466	0.000
Child(ren) in HH (D)	0.411	0.492	0.524	0.500	0.000
Living with Partner (D)	0.635	0.481	0.378	0.485	0.000
Sample Size (N)	7,692		1,120		
Male					
Employed (D)	0.716	0.451	0.222	0.416	0.000
Perceived Control[a]	0.048	0.970	-0.481	1.106	0.000
Some School (D)	0.659	0.474	0.641	0.480	0.303
Higher Secondary (D)	0.301	0.459	0.258	0.438	0.012
University (D)	0.152	0.359	0.035	0.185	0.000
Job Experience (Full Time)	19.091	12.768	13.918	12.416	0.000
Age	42.256	13.335	38.647	13.865	0.000
East German (D)	0.206	0.404	0.368	0.483	0.000
Child(ren) in HH (D)	0.395	0.489	0.465	0.499	0.000
Living with Partner (D)	0.608	0.488	0.414	0.493	0.000
Sample Size (N)	7,545		939		

Dummy variables are indicated by (D).
[a] Standardized raw scores.
[b] Two-sample equality of mean t-test.

graduates among the non-poor exceeds that among the poor by almost the same order, but even more for men. Corresponding to the above hypothesis, one finds that poor sample members lack a notable level of control perception. Females in poverty fall behind by an average of more than 0.4 standard deviations, males even by more than 0.5 standard deviation. For all other characteristics displayed in Table 9.3, the differences are also sizeable, but to a less extent.

9.6 Measuring Income Poverty

The first prerequisite for an empirical assessement of poverty was in defining a poverty line L that separates the poor from the non-poor. The OECD scale that is used for this purpose was exposed in the previous section. Up next is to find an appropriate measure that properly mirrors the extent of poverty.

The implementations of individual poverty measures can be derived from measures of poverty on the aggregate level, as these build on underlying axioms with well understood implications (see Zheng, 1997). Moreover, much of the usefulness implied by these axioms readily translates to the individual level. Robust inference can only be established if the findings are coherent across all poverty measures. For this purpose, I consider three poverty measures that originate from different classes with varying degrees of axiomatic foundation, namely the headcount ratio, the poverty deficit, and the Watts measure. The selected measures have to comply with the focus axiom (see Zheng, 1997), i.e., they are non-zero only for those individuals who have equivalent incomes below the poverty line L. On an aggregated level, this property has let to the use of right censored income distributions in order to parametrically approximate empirical distributions of poverty (see Zheng, 1997, for an overview). In case of modeling individual magnitudes of poverty, this censoring basically reverses, as measures are zero for non-poor observations and strictly positive otherwise (however, not necessarily continuous).

Let $\mathscr{Y}_i$ be a placeholder for the three poverty metrics defined in what follows. Each $\mathscr{Y}_i$ in the sample depends on the corresponding equivalent income y_i and the poverty line L, both assumed to be random variables. Basically, the domain for individual equivalence incomes is the positive real line $\mathbb{R}_{\geq 0}$, but given that individuals may face different feasible income ranges, the support S^i may vary considerably across individuals. For the whole sample, the hypothetical support $\mathrm{S}= \bigcup_{i=1}^{N} S^i$ therefore does not cover the complete positive real line. The individual poverty metric is a mapping $\mathscr{Y}_i(y_i, L) : S^i \times \mathrm{S} \rightarrow \mathbb{R}_{\geq 0}$, where the possible realizations of L depend on the exact way in which the mapping is defined.[26] Given that L is determined outside the data generating process that finally results in the empirical distribution of equivalence incomes F_Y, it may take on any value in $\mathbb{R}_{\geq 0}$. If, however, L directly results from a fraction of a distributional statistic of F_Y (here, six tenth of the median), L is bound to be somewhere in $\{L \in \mathbb{R}_{\geq 0} : L \leq F_Y^{-1}(0.5)\}$.[27] For empirical evaluations on the individual level, it is meaningful to preassign exactly one

[26] For some cases, e.g. the binary individual contribution used for the headcount measure, $\mathbb{Q}_{\geq 0}$ (when adjusted for the sample size) or even $\mathbb{N}_{\geq 0}$ would suffice.

[27] If L were a quantile and not a fraction of a quantile, it would be restricted to be within the support $\mathrm{S}= \bigcup_{i=1}^{N} S^i$ of F_Y.

$L = y$ for all N (as done in the previous section). The individual magnitude of poverty $\mathscr{Y}_i(y_i, L)$ would then change to a conditional measure $\mathscr{Y}_i(y_i|L)$. However, the fact that L depends on F_Y, which in turn depends on other y_j $\forall j \neq i$, introduces a problem common to all empirical strategies that model outcomes derived from a distributional statistic of F_Y under *iid* assumption. L is not absolutely independent with respect to the other random variables $\mathscr{Y}_j$ $\forall j \neq i$, as all the considered entities are derived from the same empirical distribution of y_i.[28] By similar reasoning, each y_i additionally depends on those of potential household members. The necessary change from the joint $\mathscr{Y}_i(y_i, L)$ to the conditional $\mathscr{Y}_i(y_i|L)$ thus only holds as an approximation. It follows that $\mathscr{Y}_i(y_i|L)$ is not exactly *iid*, but gets close to it as N grows.[29] This mild violation of the *iid* assumption has to be tolerated in general.[30] The employed measures $\mathscr{Y}_i(y_i|L)$ are derived from aggregated poverty measures that are simple (weighted) sums over individual contributions in the sample, hence its decomposability. The first one derives from the headcount ratio (see Sen, 1976) and simply is defined as

$$H_i(y_i|L) = \mathbb{1}(y_i \leq L). \tag{9.1}$$

The second one derives from the poverty deficit (see Lambert, 2001) and has the virtue to account for the magnitude of poverty as well. It reads

$$PD_i(y_i|L) = (L - y_i)\mathbb{1}(y_i \leq L). \tag{9.2}$$

The third alternative is, when aggregated over observations, the only measure considered here which is completely distribution sensitive. It has been established by Watts (1968) and is related to the entropy concept from information theory (see Theil, 1967). The individual-specific contribution reads as follows.

$$W_i(y_i|L) = (\log L - \log y_i)\mathbb{1}(y_i \leq L)^{31} \tag{9.3}$$

Apart from the decomposability, the latter two measures also quantify the distance that was established as a necessity for an interpretation in terms of well-being in Section 9.2.

[28] To illustrate this point, recall that $\mathscr{Y}_i$ can change from zero to some positive value just because another person $j \neq i$ has changed its position in F_Y and thereby affects L.

[29] At least if the sample on y_i is well-behaved.

[30] A possible account for this mild interdependence would be to condition each $\mathscr{Y}_i(y_i|L)$ on a control function term that is made up from sensitivity measures of L w.r.t. all other $y_{j\neq i}$, such as the Influence Function due to Hampel (1974). Though unlikely, even this concept does not capture the possibility of erratic jumps in y across points in time, but only the influence at the current position and magnitude.

[31] In its aggregated form, the Theil entropy measure for all N with $y_i \leq L$, $T(y|L)$, enters the Watts measure by $W(y|L) = H(y|L)\left[T(y|L) - \log(1 - \frac{PD(y|L)}{H(y|L)L})\right]$.

In addition to the mentioned focus axiom, the headcount and the Watts measure also share the property of scale invariance (see Zheng, 1992).[32] Scale invariance implies that a common factor applied to the y_i of all poor individuals, does not change the aggregate measure. It translates into the individual specific contributions as well. However, complying with scale invariance does usually not suffice to account for price level changes over time, except when exactly the same share of income is affected by the price level change for all poor individuals. As even the most basic commodity bundles represent different relative shares of the respective overall incomes, this assumption is unreasonable though. As such, price level changes should be considered for the computations of the poverty measures on the individual level.

9.7 Empirical Approach

9.7.1 Identification and Consistency

Keeping dependencies on y_i implicit, consider $D(\mathscr{Y}_{i|L})$ to be a parametric distribution that properly represents the individual contribution to one of the respective poverty measures addressed in the previous section, where $\mathscr{Y}_{i|L}$ is a placeholder for the measure-specific scalar random variable.[33] For instance, in case of the binary headcount contribution, the distribution for $\mathscr{Y}_{i|L}$ would be Bernoulli with respective conditional expectation and link function (see McCullagh and Nelder, 1989).

As the considered mediating pathways suggest, it is important to account for three features that impinge on the model structure in a dynamic perspective: (*i*) the path/state dependence of individual poverty formation, (*ii*) potential feedbacks from the current states to at least some determinants of poverty in the future, (*iii*) the initial conditions of the poverty paths at the beginning of the sampling period.

State Dependence (i) and Lagged Feedback (ii)

Firstly, in order to properly account for a poverty-trap, some kind of state dependence for the poverty measure under study has to be introduced into the empirical model. A first order autoregressive process for the outcome variable is mostly sufficient, as the interest

32 The headcount ratio is additionally characterized by location invariance, a property that no distribution sensitive poverty measure fulfills in general (see Zheng, 1994).

33 Depending on the random variable that represents poverty, the measures defined on its support can be Lebesque, counting, or combinations of both (see Davidson, 1994). All of the identification results extend to more general parameterizations of $D(\cdot|\cdot)$, i.e., to other measures of poverty not considered here.

is generally not in an outright representation of the individual time paths over large T. Secondly, one has to take into consideration that at least some individual determinants that drive poverty are not independent of previous poverty experiences, as it is likely that past poverty experiences further deprecate those individual characteristics. Such behavior, which Wooldridge (2000) terms a feedback, implies that the development of some explanatory variables $\mathbf{Z} = (\mathbf{z}'_1, \ldots, \mathbf{z}'_T)$ can be considered to take place outside the model throughout the whole sampling period, whereas for variables that are subject to feedback this only holds for some sampling periods. For every period t, the latter are contained in the vector $\mathbf{w}_t$. Moreover, the panel structure of the data allows for the incorporation of some otherwise unobserved individual heterogeneity c_i that is assumed to be time invariant. Given this distinction, the respective distribution of individual poverty measures $\mathscr{Y}_{it|L}$ conditional on covariates $\mathbf{z}_{it}$ and $\mathbf{w}_{it}$, as well as on unobserved heterogeneity c_i, reads

$$D_t(\mathscr{Y}_{it|L}|\mathbf{w}_{it}, \mathbf{z}_{it}, \mathbf{x}_{it-1}, c_i), \text{ with } t = 1, 2, \ldots, T \text{ and } \mathbf{x}_{it} = (\mathscr{Y}_{it|L}, \mathbf{w}'_{it}).^{34} \tag{9.4}$$

Treating c_i as an incidental parameter to be estimated causes severe consistency problems (see Neyman and Scott, 1948). Giving an explicit account on c_i has some clear advantages over this. Following the approaches of Mundlak (1978) and Chamberlain (1982b), one can parameterize c_i conditional on covariates.[35] Modeling c_i in that way eludes arbitrary dependence among the error terms of $\mathscr{Y}_{it|L}$ and does not restrict observed and unobserved factors to be independent, i.e., $\mathbf{w}_{it}, \mathbf{z}_{it} \not\perp c_i$ is allowed for. However, the lagged dependent part of $\mathbf{x}_{it-1}$ in equation (9.4) depends on c_i by construction. Putting aside this dependence for the moment, one can formally restate the above arguments on $\mathbf{Z}_i$ as a requirement that each $\mathbf{z}_{it}$ is strictly exogenous, implying that $D_t(\mathscr{Y}_{it|L}|\mathbf{z}_{iT}, \mathbf{z}_{iT-1}, \ldots, \mathbf{z}_{i1}, c_i) = D_t(\mathscr{Y}_{it|L}|\mathbf{z}_{it}, c_i)$, or in terms of conditional expectations that $E(\mathscr{Y}_{it|L}|\mathbf{z}_{iT}, \mathbf{z}_{iT-1}, \ldots, \mathbf{z}_{i1}, c_i) = E(\mathscr{Y}_{it|L}|\mathbf{z}_{it}, c_i)$.[36] According to the definition of Engle et al. (1983), $\mathbf{z}_{it}$ is also weakly exogenous such that its data generating process takes place outside of the conditional model in equation (9.4), without any overlap in the pa-

[34] Note that without addtional requirements, higher order lags of $\mathscr{Y}_{it|L}$ and $\mathbf{w}_t$ could be included. Then the conditional distribution changes to $D_t(\mathscr{Y}_{it|L}|\mathbf{w}_{it}, \mathbf{z}_{it}, \mathbf{X}_{it-1}, c_i)$, with $\mathbf{X}_{it-1} = (\mathbf{x}_{it-1}, \ldots, \mathbf{x}_{i1})$ and $\mathbf{x}_{it} = (\mathscr{Y}_{it|L}, \mathbf{w}'_{it})$.

[35] The explicit realization is not relevant for the identification and consistency considerations here. It will be discussed below.

[36] Following Arellano and Honoré (2001), this is the (projection based) statistical definition of strict exogeneity. It results from the yet implicit representation in equation (9.4), but is equivalent to strict exogeneity relative to the error terms of an explicitly stated econometric model. The corresponding formalizations in the panel literature are usually of the form $E(u_{it}|\mathbf{z}_{iT}, \mathbf{z}_{iT-1}, \ldots, z_{i1}, c_i) = 0 \ \forall \ t = 1, 2, \ldots, T$. In practice, panel models are rarely specified with dynamics that require independence of $\mathbf{z}$ and u over the full time path. In such cases it is sufficient to assume $E(u_{it}|\mathbf{z}_{it}, \mathbf{z}_{it-1}, \ldots, \mathbf{z}_{i1}, c_i) = 0$ or $E(\mathscr{Y}_{it|L}|\mathbf{z}_{it}, \mathbf{z}_{it-1}, \ldots, \mathbf{z}_{i1}, c_i) = E(\mathscr{Y}_{it|L}|\mathbf{z}_{it}, c_i)$ (see Wooldridge, 1997).

rameter vectors. It is thus possible to refrain from any further discussion on the marginal distributions of $\mathbf{z}_{it} \forall\, t$. Recall that the vector $\mathbf{w}_{it}$ ($\forall\, t = 1, 2, \ldots, T$) contains the mediating processes of poverty along with perceived-control. Much like $\mathbf{z}_{it}$, the elements of $\mathbf{w}_{it}$ are driving forces of poverty, but they are deemed to be affected by past poverty states. Besides perceived control, outcomes like childbearing, household formation, and employment are assumed to be affected by a similar reversion. As initially stated, such feedbacks urge a partial relaxation of the strict exogeneity assumption. In the terminology of Engle et al. (1983), $\mathbf{w}_{it}$ is predetermined with respect to $\mathscr{Y}_{it|L}$ for $t-1, \ldots, 0$, implying that for each t, $\mathbf{w}_{it}$ is independent of the current and future error terms $s \geq t$ of $\mathscr{Y}_{it|L}$.[37]

This relaxation complicates the modeling of the joint distribution $\prod_{t=1}^{T} D_t(\mathscr{Y}_{it|L}|\mathbf{w}_{it}, \mathbf{z}_{it}, \mathbf{x}_{it-1}, c_i)$, as one cannot apply the same simplification as in case of $\mathbf{Z}_i$, or $\mathbf{z}_{it}$ respectively. Without $\mathbf{w}_{it}$, it would suffice to properly account for the initial poverty state in $t = 0$ to make the joint distribution a product of the T conditionally independent distributions $D_t(\mathscr{Y}_{it|L}|\mathbf{w}_{it}, \mathbf{z}_{it}, c_i)$. In presence of the feedback effect on $\mathbf{w}_{it}$, this property no longer holds (see, e.g., Arellano and Honoré, 2001). Given the set of properties discussed for the time paths of $\mathscr{Y}_{it|L}$, $\mathbf{z}_{it}$, and $\mathbf{w}_{it}$ thus far, two frameworks that can consistently estimate the parameters of interest may be considered.

Partial Likelihood Approach: One possible solution is to refrain from any independence assumption discussed within the last paragraph, and thus from any assumption on the joint distribution of the individual paths over T. Instead, one merely has to settle for the correct specification of the period-specific distributions $D_t(\mathscr{Y}_{it|L}|\cdot)$ for all $t = 1, \ldots, T$. If these period specific distributions are correctly specified and treated like distributional contributions in a pooled sampling context, strict exogeneity is not a necessary condition for consistency any longer. This finding builds on a special case of general consistency results in presence of partial misspecification for maximum likelihood and extremum estimators (see White, 1982).[38] Following the Kullback-Leibler identity, it can be shown that averages over single factors of a joint distributions suffice in order to establish consistent estimates. In the case of averaging over the joint distribution along the time dimension, Wooldridge (2002) calls this a partial likelihood approach.[39] However, it cannot jointly quantify the dynamic interactions between $\mathscr{Y}_{it|L}$ and $\mathbf{w}_{it}$, as would be the case given more structure along the time dimension. Moreover, it should be noted that contemporaneous exclusion restrictions among some of the possible combinations of the variables in $\mathbf{w}_{it}$ have to be imposed. As opposed to the case where the equations for $\mathscr{Y}_{it|L}$ and $\mathbf{w}_{it}$ are to

37 See also Arellano and Honoré (2001) for a discussion of predetermindness in panel data context.

38 See also Amemiya (1985).

39 In case of averaging over more general types of dimensions, such as multinomial choices, it is called a quasi-maximum likelihood approach.

be considered simultaneously, unrestricted contemporaneous cross-effects are not a matter of identification. Instead, they would result in a form of self-imposed simultaneity bias. One thus still has to make sensible choices about which elements of $\mathbf{w}_{it}$ contemporaneously enter the partial likelihood models for other elements of $\mathbf{w}_{it}$. As the order cannot be empirically inferred, one has to base the restrictions on economic theory.

Structural Approach: The second empirical approach pursued in the present setting is more structural, but likewise relaxes the strict exogeneity assumptions for $\mathbf{w}_{it}$. The difference to the partial likelihood approach is that it jointly models the contemporaneous effects and the lagged feedbacks. It builds on the results discussed in Wooldridge (2000), who suggests to factorizes the individual processes for $\mathscr{Y}_{it|L}$ and the set of predetermined covariates $\mathbf{w}_{it}$, $\mathbf{x}_{it} = (\mathscr{Y}_{it|L}, \mathbf{w}'_{it})$. If one assumes that, in addition to strict exogeneity with respect to $\mathscr{Y}_{it|L}$, $\mathbf{z}_{it}$ is also strictly exogenous with regard to $\mathbf{w}_{it}$, one can write

$$\begin{aligned} D(\mathbf{x}_{it}, \ldots, \mathbf{x}_{i1}|\mathbf{z}_{iT}, \ldots, \mathbf{z}_{i1}, c_i) &= \prod_{t=1}^{T} D_t(\mathbf{x}_{it}|\mathbf{z}_{it}, \mathbf{x}_{it-1}, c_i) \text{ with factorization} \\ D_t(\mathbf{x}_{it}|\mathbf{z}_{it}, \mathbf{x}_{it-1}, c_i) &= D_t(\mathscr{Y}_{it}|\mathbf{w}_{it}, \mathbf{z}_{it}, \mathbf{x}_{it-1}, c_i) D_t(\mathbf{w}_{it}|\mathbf{z}_{it}, \mathbf{x}_{it-1}, c_i). \end{aligned} \tag{9.5}$$

Assuming that all conditioning variables in equation (9.5) enter the distributions of $\mathscr{Y}_{it|L}$ and $\mathbf{w}_{it}$ in a linear-additive fashion given some link function, standard identification theory based on cross-equation restrictions, exclusion restrictions, and covariance restrictions can be applied in order to render the model identified.[40] However, given the particular mixture of linear, binary, and corner solution link functions that arise from the variable-types in $\mathscr{Y}_{it|L}$ and $\mathbf{w}_{it}$, some peculiarities compared to the linear case are in order. These requirement kind of predesignate the first identification restriction. As shown by Maddalla (1983), all systems of binary or censored endogenous variables (or mixtures of them) should be recursive with respect to contemporaneous cross-effects.[41] Omission of this recursive design leads to the case where at least some of the equations involved are logically inconsistent, i.e., the sum over all joint probabilities do not generally sum to one in this case. Recursiveness implies logical consistency, but is not a necessary condition in all possible realizations.[42] If one imposes no restrictions on the equations for $\mathscr{Y}_{it|L}$, the re-

[40] If a model is in single-index form, the conditional mean $\mu \equiv E(y|\upsilon) = g(\upsilon)$, where $\upsilon = \mathbf{x}'\boldsymbol{\beta}$ usually holds. A link function is defined to be $g^{-1}(\mu)$ (see McCullagh and Nelder, 1989). Henceforth, I use this term for the inverse, i.e. for $g(\upsilon)$.

[41] The general multi-variate case is discussed in Schmidt (1981).

[42] For corner solution equations, logical consistency depends on specific parameter realization and restrictions may be weaker than recursiveness. The necessary and sufficient conditions on the parameter space of the contemporaneous endogenous variables would not be feasible as a reparameterization, but only as an inequality-constraint optimization. This is relatively impractical and, furthermore, the resulting model has no meaningful economic interpretation. For binary links involved, however, the recursiveness assumption is strictly necessary.

cursiveness assumption in the adjacent equation in $\mathbf{w}_{it}$ is mathematically equivalent to the requirement for predeterminedness of this mediating variables with respect to $\mathscr{Y}_{it|L}$. Off course, for logical consistency, recursiveness and thus predeterminedness have to extend to the contemporaneous cross-relations among all further variables in $\mathbf{w}_{it}$ as well. It follows that the contemporaneous cross-effects have to decrease row-wise.

For complete identification of the simultaneous structure in equation (9.5), one has to introduce a second type of restriction. Since the unobserved effects are explicitly modeled, cross-equation covariance restrictions among the residuals are a tenable option. As c_i is properly accounted for and is allowed to vary by equation, it does not seem too restrictive to do so. Alternatively, exclusion restrictions on the respective $\mathbf{z}_{it}$-vectors could be imposed, but justifying the required instrument is a more difficult task in the current setting.

Initial Conditions (iii)

Irrespective of using the partial likelihood or the structural approach to allow for predeterminedness, a final requirement is that the initial poverty status for the start of the sampling period in $t = 0$ has to be addressed. For dynamic panel data models with rather small T, misspecified initial conditions $\mathscr{Y}_{i0|L}$ and $\mathbf{w}_{i0}$ are a serious confounder for parameter consistency, as opposed to time series frameworks with large T. Treating the initial conditions as a non-stochastic component would also mean that they are not allowed to depend on heterogeneity c_i, which is not very plausible. If the initial conditions are assumed to be stochastic, Hsiao (2003) discusses cases of equilibrium initial conditions that allow to retrieve their distribution functions and to consider them as part of the joint distribution in equation (9.5), rather than as a conditioning variable. However, such presumptions are not testable in practice and it is unlikely that the starts of the processes $\mathscr{Y}_{it|L}$ and $\mathbf{w}_{it}$ always coincide with the start of the sampling period. I use an approach introduced by Wooldridge (2005), instead. It models c_i as a function of $\mathscr{Y}_{i0|L}$, the elements of $\mathbf{w}_{i0}$, the individual specific time averages $\bar{\mathbf{z}}_i$, and a remainder of unobserved heterogeneity a_i, implying

$$D(c_i|\mathscr{Y}_{i0|L}, \mathbf{w}_{i0}, \bar{\mathbf{z}}_i, a_i), \tag{9.6}$$

where the components $\mathscr{Y}_{i0|L}$, $\mathbf{w}_{i0}$, $\bar{\mathbf{z}}_i$, and a_i are linear and additive. Given this specification, one does not make the initial conditions part of the joint distribution. Instead, by solely conditioning on $\mathscr{Y}_{i0|L}$ and $\mathbf{w}_{i0}$, one can remain unconcerned about the distributions of the initial conditions. The distribution $D(\cdot|\cdot)$ is chosen to coincide with that of the respective outcome $\mathscr{Y}_{i|L}$ or $\mathbf{w}_i$, where for normal-based distribution types both terms conflate to one linear-additive condition set. In case of the partial likelihood approach, the explicit consideration of a time invariant remainder term a_i is meaningless as no time paths are modeled. Thus, a_i can be absorbed into the time-specific innovation term. Furthermore, the consideration of all initial states $\mathbf{x}_{i0}$ as in equation (9.6) is not necessary in this case.

Sample Spacing

One additional problem in the current setting is imposed by the fact that perceived control is not sampled in even intervals. Without formal derivation, it is immediately obvious that the models considered thus far cannot consistently estimate the state dependencies within the paths of poverty experiences and predetermined variables when sampling periods t are unequally spaced.[43] That being the case, the reference period for the underlying data generating process, usually termed the unit period (see Fuleky, 2012), does not coincide with the observational interval. Some approaches that account for these issues are existent (see Baltagi and Song, 2006, for an overview), but are not applicable to non-linear dynamic settings. As such, it is necessary to set up different subsets of the data with varying but equally spaced sampling gaps and to cross-validate the results derived from them.

It should be noted that by the above definitions equal observational intervals also represent an irregular spacing regarding the unit period and the data generating process.[44] It can be shown that the state dependence parameter of the true process mixes with the error term of the observed model in this case (see Millimet and McDonough, 2013). The resulting estimates are consistent, but formally with respect to the "wrong" model parameters. Given equal spacing, the misspecification can be regarded as being constant, though. This may allow for meaningful inference but restricts comparison to estimates derived from other, differently spaced settings. Refraining from this point and setting the observational unit equal to the unit period is common practice in discrete longitudinal (see Baltagi and Song, 2006) and time series settings (see Hamilton, 1994). I follow this premise here.

9.7.2 Parameter Estimation

For the structural approach, the aforementioned focus on the labor force, i.e. on individuals aged 18 to 65, implies to retain only those individuals in the sample that are in working age for the complete time path to be considered. Given time paths T plus the

[43] A formal representation is given in Millimet and McDonough (2013).

[44] This follows from the fact that the unit period at which the individual is supposed to make consecutive decision almost never complies with the rate at which the sampling occurs (e.g., annually).

initial periods, all observations in $t = 0$ are aged between 18 and $65 - (T + 1)$, whereas in $t = T$ the age varies between $19 + T$ and 65. On the one hand, this proceeding has the virtue of decreasing the relative weight of probably aberrated transition periods out of the labor market, since only the last sample waves get close to the legal retirement age. On the other hand, rather practical contemplations underly this step, as the structural approach requires contiguous individual time paths to set up the likelihood contribution and only few waves provide information on perceived control. The partial likelihood approach is less "data hungry" as only two adjacent intra-individual observations are needed in order to obtain a consistent partial likelihood contribution. Thus, the number of observations is generally higher for the pooled models. Moreover, I consider gender specific subsamples for the analysis. This has the intuitive reasoning that human capital pricing and thus income, as well as labor market participation and other factors, may differ by gender. It also greatly simplifies the underlying structures for estimation and computation of standard errors, since there is relatively little need to account for intra-household correlation. The samples for female and male respondents are very homogeneous in that regard. Almost 89 percent of both gross samples do not live together with another sample member who is in working age and of the same gender in 2010. If only those observations without non-responses in the variables of interest are retained, this share increases to above 99 percent in either case. As such, the dependence structures within the individual time paths seem to be the only ones of actual importance. The gender subscripts are kept implicit in the following formal representations.

Partial Likelihood Approach

Recall the vector $\mathbf{x}_{it} = (\mathscr{Y}_{it|L}, \mathbf{w}'_{it})$ combining the respective poverty measure with the predetermined mediating factors and perceived control from equation (9.4). The partial likelihood approach discussed in the previous section separately estimates the respective equations for all K variables in x_i^k $(k = 1, \ldots, K)$. Each variable x_i^k can be associated with a respective link function that characterizes its conditional expectation, and hence, its probability distribution. The link functions corresponding to the variables x_i^k are summarized in Table 9.4.

Table 9.4: Variable Types and Corresponding Link Functions $g(\upsilon)$

Variables	Link Type	Range of $g(\upsilon)$	$g(\upsilon)$
Poverty Metrics			
Headcount	binary	$\{0,1\}$	$\Phi(\upsilon)$
Poverty Deficit	corner solution	$(0,+\infty)$	$\Phi(\upsilon/\sigma)(\upsilon+\sigma\left[\frac{\phi(\upsilon/\sigma)}{\Phi(\upsilon/\sigma)}\right]$
Watts	corner solution	$(0,+\infty)$	$\Phi(\upsilon/\sigma)(\upsilon+\sigma\left[\frac{\phi(\upsilon/\sigma)}{\Phi(\upsilon/\sigma)}\right]$
(Potentially) Predetermined Variables			
≥ 1 child in HH	binary	$\{0,1\}$	$\Phi(\upsilon)$
Living with partner	binary	$\{0,1\}$	$\Phi(\upsilon)$
Employment (full time)	binary	$\{0,1\}$	$\Phi(\upsilon)$
Perceived Control	identity	$(-\infty,+\infty)$	υ

The dependent variables are conditioned on lagged values $\mathbf{x}_{it-1}$, on strictly exogenous variables $\mathbf{z}_{it}$, and on the unobserved heterogeneity term $c_{i|\mathscr{Y}_{i0|L},\bar{\mathbf{z}}_i}$, or $c_{i|\mathbf{w}_{i0},\bar{\mathbf{z}}_i}$ respectively. As stated above, contemporaneous cross-effects among the elements of $\mathbf{x}_{it}$ cannot be arbitrarily specified, as the estimates are otherwise inconsistent due to a self-defined simultaneity. Given the hypothesis that the feedback effects disseminate from past poverty to perceived control with all other elements of $\mathbf{w}_{it}$ being mediating factors, it is self-evident to allow $\mathscr{Y}_{it|L}$ to be contemporaneously affected by all $\mathbf{w}_{it}$. By the same token, perceived control is the Kth element of $\mathbf{w}_{it}$ with no contemporaneous cross-effects. For the remaining variables in $\mathbf{w}_{it}$, the order of the contemporaneous cross-effects are ad hoc choices that cannot be based on the data at hand. Instead, economic theory suggests that household formation with a partner usually takes place before childbearing decisions are made. I follow this convention here. The positioning of employment is more complex from a theoretical perspective. For women, childbearing is known to negatively affect labor force participation and thus employment (see Aassve et al., 2006b). For men, on the other hand, labor market participation and employment may be more of an preliminary decision, as employment is a promoting factor in mating and search frameworks (see Burdett and Coles, 1999, Aassve et al., 2002). I will test whether differences occur under both presumptions.

To give a more ostensive representation of the partial likelihood specification, consider the case of the binary headcount $\mathscr{Y}_{it|L} = H_{it}$ as a left-hand side example for x^1_{it}. Then, the explicit representation of $D_t(\cdot|\cdot)$ is

$$\Phi\left[(2H_{it}-1)(\boldsymbol{\beta}_1'\mathbf{z}_{it}+\boldsymbol{\beta}_2'\tilde{\mathbf{w}}_{it}+\beta_3 H_{it-1}+\boldsymbol{\beta}_4'\tilde{\mathbf{w}}_{it-1}+\alpha_1 H_{i0}+\boldsymbol{\alpha}_2'\bar{\mathbf{z}}_i)\right],$$

with $\mathbf{z}_{it}$ and $\bar{\mathbf{z}}_i$ having ones as their respective uppermost element. The implementation for the other outcome equations follows the same logic. The resulting log-likelihood contribution for each x_i^k-specific pooled model is

$$\ell_i(\mathbf{\Gamma}_k) = \sum_{t=1}^{T} \ln D_t(x_{it}^k | \mathbf{z}_{it}, \bar{\mathbf{z}}_i, \tilde{\mathbf{w}}_{it}, x_{i0}^k, \mathbf{\Gamma}_k),$$

where $\tilde{\mathbf{w}}_{it}$ is always a $(K-1)$-subset of $\mathbf{w}_{it}$, in this case except for $x_i^k = \mathscr{Y}_{i|L}$, due to the otherwise arising simultaneity problems. Again note that the partial likelihood approach does not explicitly involve the unobserved component a_i. Instead, it is absorbed into the respective error term. This affects the scale normalization for binary models or the variance estimate in the censored and linear case. In all three cases, however, the implied serial error-correlation on the individual level has to be accounted for when standard errors are to be computed.

Structural Approach

In the previous section it has been argued that the use of time-invariant random effects makes the assumption of zero covariances across equations relatively plausible. By the same token, the individual-specific joint distribution over the time dimension can be assumed to require no further free form correlation in the idiosyncratic error terms.[45] Without such correlations, the likelihood derived for the estimation of the structural model can be evaluated without any multidimensional integrals. This presumption is not necessary for identification, but greatly alleviates the estimation procedure. Following this simplifying assumption and given the identification results established in the previous section, one may write the joint distribution of $\mathscr{Y}_{it|L}$ and the K row elements of $\mathbf{w}_{it}$ over the sampling period as a simple product.

$$\begin{aligned} &D(\mathscr{Y}_{i1|L}, \ldots, \mathscr{Y}_{iT|L}, \mathbf{w}_{i1}, \ldots, \mathbf{w}_{iT} | \mathbf{z}_{i1}, \ldots, \mathbf{z}_{iT}, \mathscr{Y}_{i0|L}, \mathbf{w}_{i0}, c_i, \mathbf{\Gamma}) = \\ \prod_{t=1}^{T} \; &D_t(\mathscr{Y}_{it|L} | \mathbf{w}_{it}, \mathscr{Y}_{it-1|L}, \mathbf{w}_{it-1}, \mathbf{z}_{it}, \mathscr{Y}_{i0|L}, \mathbf{w}_{i0}, c_i, \mathbf{\Gamma}_1) \\ &\vdots \\ \cdot \; &D_t(w_{Kit} | \mathscr{Y}_{it-1|L}, \mathbf{w}_{it-1}, \mathbf{z}_{it}, \mathscr{Y}_{i0|L}, \mathbf{w}_{i0}, c_i, \mathbf{\Gamma}_K), \end{aligned} \tag{9.7}$$

where the partitions $\mathbf{\Gamma}_k$ are generally not the same as in case of the partial likelihood approach above. In order to maintain a comparably sparse parameterization, the pa-

[45] The argumentation for this assumption is similar to the one by Butler and Moffitt (1982), though in a slightly different context.

rameters in $D(c_i|\mathscr{Y}_{i0|L}, \mathbf{w}_{i0}, \bar{\mathbf{z}}_i, a_i)$ are not allowed to freely vary across equations, but by an overall scaling factor for c_i in each equation.[46] As such, the parameter blocks in $\mathbf{\Gamma} = (\mathbf{\Gamma}'_1, \mathbf{\Gamma}'_2, \ldots, \mathbf{\Gamma}'_K)$ have the parameters for c_i in common. The link functions for the respective $\mathscr{Y}_{it|L}$ and $\mathbf{w}_{it}$ are the same as those defined for the partial likelihood approach above (see Table 9.4).

For a better illustration of the specifications resulting from equation (9.7), consider again the case of the binary headcount $\mathscr{Y}_{it|L} = H_{it}$, for simplicity only along with perceived control as a scalar predetermined variable $w_{it} = \theta_{it}$. Then one obtains individual time paths

$$\prod_{t=1}^{T} \Phi\left[(2H_{it}-1)(\boldsymbol{\beta}'_1\mathbf{z}_{it} + \beta_2\theta_{it} + \beta_3 H_{it-1} + \beta_4\theta_{it-1} + \psi + \alpha_1 H_{i0} + \alpha_2\theta_{i0} + \boldsymbol{\alpha}'_3\bar{\mathbf{z}}_i + a_i)\right]$$
$$\frac{1}{\sigma}\phi\left[(\theta_{it} - \boldsymbol{\delta}'_1\mathbf{z}_{it} - \delta_2 H_{it-1} - \delta_3\theta_{it-1} - \delta_4(\psi + \alpha_1 H_{i0} + \alpha_2\theta_{i0} + \boldsymbol{\alpha}'_3\bar{\mathbf{z}}_i + a_i))(1/\sigma)\right].$$

Returning to the general case again, it is implied by the above hypothesis that perceived control is always the lowermost equation in the equation (9.7), i.e., it is the variable that is always predetermined with respect to all other dependent variables at each t. Likewise, $\mathscr{Y}_{i1|L}$ is always the variable that is allowed to be contemporaneously affected by all $\mathbf{w}_{it}$, and thus is always the uppermost equation in the system. The remaining endogenous variables in $\mathbf{w}_{it}$ may follow an order of predeterminedness established by the same economic reasoning as in the case of the partial likelihood approach discussed above. The simultaneous estimation pursued here provides another opportunity, though. Instead of ad hoc choices for the order of predeterminedness in $\mathbf{w}_{it}$, one may nest statistical testing procedures in order learn from the data. Unfortunately, such procedures are rare and largely limit to time series applications with large T (see, e.g., Kilian and Vega, 2011).[47] One therefore has to use a more general specification test for simultaneous equation systems suggested by Anderson and Kunitomo (1992). The family of tests derived there test for predeterminedness against the alternative of unrestricted cross-effects among the elements of $\mathscr{Y}_{i1|L}$ and $\mathbf{w}_{it}$. This choice is rooted in one particular limitation imposed by the setting at hand. Following the identification and consistency considerations addressed above, predeterminedness has to be imposed for logical consistency. As such, it is

[46] For the first equation that generally models $\mathscr{Y}_{it|L}$, the scaling factor is always one.

[47] Larger parts of the econometric and statistical literature deal with the detection of Granger non-causality (see, e.g., Engle et al., 1983) and its implications for strict exogeneity. Given the asymptotic importance of the time dimension in such settings, under suitable sampling horizons, vector-autoregressive approaches (see Holtz-Eakin et al., 1988) can be used to derive such properties in panel data context. However, neither Granger's causal interpretation nor strict exogeneity translate into generally valid necessary or sufficient conditions for predeterminedness (see Chamberlain, 1982a; Arellano and Honoré, 2001).

only possible to derive test statistics from (sub-)models under this assumption, since the unrestricted model is logically inconsistent given the above arguments. The Anderson and Kunitomo (1992) framework provides a convenient solution to this problem, as it also suggests Lagrange multiplier criteria that allow for inference based on the restricted models only.

Having solved the issues of predeterminedness and logical consistency, what remains to be addressed is how to treat the time invariant unobserved component a_i. By assuming that $a_i \sim \mathcal{N}(0, \sigma_{a_i})$, the following log-likelihood contribution for individual i over the sampling periods T is obtained.

$$\ell_i(\mathbf{\Gamma}_1, \ldots, \mathbf{\Gamma}_K) = \ln \int D(\cdot|\mathbf{Z}_i, \bar{\mathbf{z}}_i, \mathscr{Y}_{i0|L}, \mathbf{w}_{i0}, \ldots, \mathbf{w}_{iT}, a_i, \mathbf{\Gamma}_1, \ldots, \mathbf{\Gamma}_K) \left(\frac{1}{\sigma_a}\right) \phi\left(\frac{a}{\sigma_a}\right) da,$$

where $D(\cdot|\cdot)$ is the right-hand side product from equation (9.7) and $\mathbf{Z}_i = (\mathbf{z}'_{i1}, \ldots, \mathbf{z}'_{iT})$. The integral over the unobserved a_i can be solved numerically by means of a Gauss-Hermite quadrature. The number of interpolation nodes required for obtaining a relatively accurate approximation result is relatively low in cases where $D(\cdot|\cdot)$ involves a link function based on a normal distribution (see Butler and Moffitt, 1982), which applies to all the link functions in Table 9.4.

9.8 Results

The main results presented in the following section are based on the five-year sampling interval.[48] That being the case, the 1995 wave represents the initial period, whereas the waves 1999 to 2010 model the actual individual specific time paths. Other observational intervals and specifications are presented in Section 9.9. For all results to be considered, note that most of the average partial effects possess a self-explaining magnitude. In case of perceived control and the deficit measures that account for the gap between equivalent incomes and the poverty line, some preliminary explanations may be in order. For perceived control, all effect sizes refer to a change on its standard deviation or from its standard deviation.[49] For the poverty deficit, the average effects can be interpreted

[48] Recall that the second wave (1999) actually has a four-year gap towards the the first wave and a six-year gap towards the third one. Moreover, the interview dates may vary in course of the respective years. Hence, the observational intervals are only approximations.

[49] For estimation of the structural model, however, it is convenient to chose a higher dispersion of the perceived control scores in order to scale the corresponding sub-model such that it provides tantamount contributions to the overall likelihood value. Otherwise, many precision exceptions occur throughout the likelihood optimization. Due to the scale-invariance of the likelihood approach, the estimates can be re-scaled in the aftermath.

in terms of the absolute distance of equivalence incomes to the poverty line. Similarly, for the Watts measure the average change refers to the logs of both entities. Hence, one should always put into perspective that average partial effects for the poverty deficit measures tend to be rather large in magnitude, as they relate to induced changes in equivalent euros. On the other hand, model parts that include the poverty deficit as a right-hand side variable produce comparably small average effects as opposed to those with binary indicators of poverty involved, though their absolute meaning may be quite substantial. For the Watts measure a meaningful interpretation is somewhat more difficult to establish. If the deficit in logs is a left-hand side variable, one may reformulate the average partial effects of continuous variables by means of the exponential function. The resulting average partial effects then refer to the implied average change on the ratio of the poverty line and the net equivalent income. By the same transformation, the average effect sizes of the Watts measure as a right-hand side variable are in terms of a one percent increase in the ratio of poverty line and equivalence income with respect to the corresponding dependent variable. For discrete explanatory variables, the exponential transformation for the log ratio does not apply, as it also depends on its level when the change is discrete rather than marginal.

9.8.1 Partial Likelihood Approach

The empirical results for the female sample using the partial likelihood approach are presented in Tables 9.5 to 9.7. For males, the corresponding results will be presented in Tables 9.8 to 9.10. At the first glance, it becomes obvious, that state dependence plays a dominant role in the model parts for poverty status, full-time employment, living with partner, and childbearing.[50] Being poor in the previous period vastly increases the probability of living in poverty in the ensuing one. The same holds for full-time employment status and the other considered predetermined variables. Being currently full-time employed also significantly reduces the probability of contemporaneously living in poverty.

Regarding the exact magnitudes of the presented estimates, one finds that, probably owing to the large sample size, a lot of statistically significant effects are at hand. The effect sizes for the strictly exogenous variables are largely in line with what could have been expected based on economic rationales. In addition to the contemporaneous exogenous effects, the significant coefficients of the time-averaged indicators suggest the prevalence of some characteristics that are highly correlated with average (unobserved) behavioral driving forces that go beyond perceived control, such as intelligence, further unobserved

[50] I will use “childbearing” as a synonym for “having at least one child” throughout the following discussion.

abilities, and motivational factors. Adding up the time-invariant and time-varying components of the strictly exogenous variables, one finds that having some secondary school degree reduces the average probability of living in poverty by roughly 7 percentage points, obtaining a higher secondary degree (Gymnasium) does so by even 8 percentage points. Similarly sizeable is the 6 percentage point reduction in probability when holding a university degree. Discarding the impact of those time-invariant variables that act as an indicator for unobserved heterogeneity, having some vocational qualification and holding German citizenship also significantly contribute to the explanation of individual poverty states. Those individuals who possess a vocational degree are almost 4 percentage points less likely to live in poverty than those who do not. Holding the German citizenship lowers the probability of living in poverty by roughly the same magnitude, whereas living in the eastern part of Germany has an opposing effect that is twice that size. Moreover, though being jointly significant, the coefficients of age included as contemporaneous regressors in the poverty equation do not show any clear pattern.

Considering the contemporaneous predetermined variables, employment status and living with a partner decrease the probability of the binary poverty status by 6.9 and 6.5 percentage points, respectively. This seems quite intuitive. The effects of the lagged characteristics are not completely in line with what could have been expected. One the one hand, having been gainfully employed in the previous period significantly reduces the poverty risk in a given period. On the other hand, living together with a partner at $t-1$ increases the risk of being poor, though by a comparably small margin of one percentage point. The same holds for having had at least one child in the previous period. The likely explanation for these somewhat contradictory effects might be that, after controlling for potential economies of scale by using equivalence incomes, working-age individuals who live together with others may not solely benefit from living with each other. Quite often, such individuals may be sole earners or at least have to keep one additional household member. This finding is somewhat at odds with those results derived from traditional random effects models under strict exogeneity assumptions (see, e.g., Biewen, 2004), implying that its relaxation is a quite reasonable step in the setting at hand.

Nonetheless, the estimate for the state dependence effect is the strongest one. Having been poor at $t-1$ raises the probability of being poor in the subsequent observational period by about 19 percentage points. This effect highlights that even after controlling for differences in observed and unobserved characteristics, past poverty experience is connected to a higher future poverty risk. The revealed state dependence corresponds to the previous empirical findings that have been discussed throughout the review of the related literature. The fact that the incorporation of perceived control into the model does not change this pattern indicates that this particular trait does not add much to the explanation of this implicit association. Furthermore, it should be noted that given the current setting those who are poor at $t-1$ and again at t may consist of two rather different groups. There are those individuals for whom the two points in time are part of a continuing poverty spell. Additionally, there may be those individuals who have an interrupted spell of poverty, or potentially even more than one. In the setting at hand, a potential mixing of these groups is even more likely as the observational points in time are quite distant. Given that the partial likelihood approach does not distinguish between likelihood contributions across individuals and within individuals along the time axis, this issue is not properly accounted for by the result presented here. The implications of continuing spells and repeated poverty unemployment may be somewhat different. One may learn more about this phenomenon from the data when the observational interval as well as the modeling of the individual time paths are changed. This will be subject to Section 9.8.2 and Section 9.9.

The estimation results presented thus far remain valid when the two remaining poverty measures, namely the poverty deficit and the Watts measure, are considered. Recall that, as opposed to the binary indicator, both measures also capture the extent of poverty, where the Watts measure puts more weight on equivalent incomes in further distance to the poverty line. The corresponding estimation results are presented in Tables 9.6 and 9.7. The order of effect sizes presented for the binary poverty indicator thus far do not change for either measure. Bear in mind that the average effects refer to the conditional expectation for the complete sample, not only to those observations for whom the deficit measures are not censored. Being employed reduces the average deficit by roughly 200 equivalent euros, or given the Watts specification, decreases the log-ratio between the poverty line and the equivalent income by 0.4 percentage points. Analogously to the binary case, living with a partner and the degree of perceived control also exert a substantial effect in terms of poverty reduction. For the strictly exogenous variables, the picture is slightly different compared to the binary case. Educational achievements, which have been a strong predictor for the headcount measure, are less significant for the poverty deficit and the Watts measure. Holding a vocational degree, on the contrary, seems to decrease the poverty deficit by 60 equivalent euros on average, or in terms of the log ratios by 0.12 percentage points. Apart from a rather small impact of age, the time-averaged covariates in the correlated part of the model have lost most of their statistical significance given the two specifications that involve the poverty deficit and Watts measure.

Quite naturally, the results for the predetermined left-hand side variables do not depend on the specific poverty measure being employed.[51] Some of them are somewhat remarkable, though. Having at least one child negatively affects the probability of employment, as does living with a partner. Perceived control has some positive effect on employment, but the magnitude is negligible. Again, the state dependence in the respective dependent variable is the most influential predictor. Having been employed in the previous observational period raises the probability of employment in the current period by 31 percentage points. Similar findings occur in case of having one or more children, and in case of living with a partner. These amount to 50 and 21 percentage points, respectively. It should be noted, however, that as with the state dependence in poverty, two rather different groups of observations are likely to mix up in the constitution of the state dependence in the employment equation. Again, there may be those observations with the two points in time being part of a continuing spell without employment, and those who have one or more intervening spell(s) of employment. As argued above, a model that considers complete individual time paths may contribute to a better understanding in this case. With regard to the exogenous variables, having a university degree is a particularly strong exogenous predictor for employment. The same holds for possessing a degree from a technical college. Considering the time-invariant part of the model, these effects are mitigated to some extent. For all further strictly exogenous variables, the effects are rather moderate in magnitude across all columns. The results for perceived control are quite plausible and in large parts in line with the respective literature on stability (see the summarized findings in Chapter 7). As already stated, they exhibit the anticipated effects on current employment status. The effects are positive in the sense that a higher degree of perceived control increases the probability for those outcomes typically associated with labor market success (see Almlund et al., 2011). As such, it is also possible that an additional impact on poverty status transmits via the effect on employment status. The estimates for the lagged effects of perceived control on the considered dependent variables are lower in magnitude. Background characteristics like formal educational attainments explain some of the differences in perceived control, but most covariates are not statistically significant. Holding some school degree increases perceived control by 4.6 percent of a standard deviation. Moreover, a relatively small combined age effect seem to prevail. Apart from that, the pattern is similar to previous findings from the literature on trait determinants, where indicative individual characteristics usually have low explanatory power.

[51] The possible changes in the estimated coefficients, if anything, affect the rightmost reported decimal digit.

Turning to the estimates that are most important in light of the hypothesized feedback effect, past poverty experiences apparently exert some lagged influence on perceived control. In case of the binary headcount, the feedback effect amounts to 5.2 percent of a standard deviation. Though this is not a major feedback effect, it is still remarkable given the often alleged stability of personality traits in adulthood. One should bear in mind that the binary poverty status may mix the effects of those being "slightly" poor with those who have available an even lower amount of equivalent income. Thus, a more nuanced view may be obtained when additionally considering the extent of poverty by means of the deficit or Watts measure. When poverty is linearly scaled as equivalent income, the corresponding negative effect on perceived control is close to zero. When a higher emphasis is put on those individuals who suffer from a higher degree of poverty, a one percent increase in the ratio of the poverty line and an individual's equivalence income significantly decreases perceived control by almost 13 percent of a standard deviation. This is a fairly large effect that particularly seems to be driven by those individuals who are exerted to a comparably high degree of poverty. To give this effect size a more intuitive meaning, consider the case of a poverty line being located at 700 equivalent euros and an individual who has 350 equivalent euros at her/his disposal. For such an admittedly high degree of poverty, a one percent increase would be equivalent to a 2.40 euro decrease in disposable equivalent income. This finding indicates that the negative perception of small income decreases for poor individuals seems to be sizeable.

Though the overall picture does not substantially differ from what has been estimated for the female sample, some slight variations are apparent for male sample members, though. The results for the set of strictly exogenous covariates again is little surprising. However, human capital achievements and labor market assets, like job experience, imply some differing partial effects. Considering the compound contemporaneous and time-averaged variables, holding some secondary school degree results in almost the same average effect as for the male sample, whereas the poverty reduction due to a higher secondary degree is weaker. The latter amounts to just -1.8 and -2.1 percentage points, as opposed to -3.3 and -5.7 for females. Similarly, the average poverty reduction induced by having some vocational degree is only half that size. On the other hand, degrees from university and technical college greatly reduce the probability of living in poverty, but the latter effect is somewhat weaker than for females. Only the indicator for residence in the eastern parts of Germany increases the likelihood of living in poverty. Compared to the female respondents, a similar pattern holds for the various impacts of the contemporaneous and lagged predetermined variables. Having a full-time employment reduces the poverty risk by almost 7 percentage points and thus exhibits almost the same magnitude. The contemporaneous effect of living with a partner is somewhat weaker, whereas having at least one child increases the probability of poverty by 4.4 percentage points. This impact was not at hand in case of females in working-age. As expected, having been poor in the previous observational period increases the current risk of poverty by 15.3 percentage points. As such, this state dependence is somewhat weaker than in case of the male sample. The estimated coefficients on the remaining lagged dependencies do not differ substantially and are relatively weak in magnitude.

Columns II to IV of Table 9.8 present the estimates for the predetermined variables. As with the above discussion for the female sample, all the regressors apart from poverty status produce almost the same estimates for the three considered poverty measures (see Tables 9.8 to 9.10). Hence, it is again sufficient to discuss only the estimates presented in Table 9.8. Once more, a statistically significant effect of past employment on future employment is found. It seems that the relative magnitudes of the strictly exogenous variables are similar for the gender specific employment models. The compound contemporaneous and time-averaged effects for the secondary schooling degrees are slightly lower in case of the male sample. The impact of holding a university degree is somewhat higher, on the contrary. Likewise, living in the eastern part of Germany seems to have a marginally higher impact on the average risk of being unemployed than in case of females. Such minimal differences in the estimated effect sizes in the exogenous variables continue to hold for the remaining predetermined variables. They do not, however, add anything substantial to the general formation patterns of the considered entities.

The contemporaneous and lagged cross-effects provide more interesting information, however. Having children does not negatively impinge on the probability of employment, but even slightly increases the chance by 3.5 percentage points. The small negative effect of living with a partner is reversed for males. A substantial difference occurs for the association of cohabitation with a partner and the probability of having at least one child. As opposed to just 8 percentage points for females, living with a partner increases the probability of the latter by 34 percentage points. As with the discussion thus far, the state dependence effects are the most substantial determinants for all the predetermined variables involved. In case of employment and child bearing, the state dependence is slightly weaker than for the female sample. In case of the probability of living with a partner, it is 0.5 percentage points higher. For the rightmost model that addresses perceived control, the explanatory contributions of the strictly exogenous variables are comparable to the results for females. Only the indicator for holding a secondary schooling degree provides a seizable explanation for the expression of perceived control, as it increases the score by an average of 16.5 percent of a standard deviation. The effects of the remaining estimates are again quite imprecise. There is also a relatively large state dependence in the model for perceived control, along with lagged feedbacks due to living with a partner and having at least one child. As opposed to the estimates obtained for the female sample, the lagged effect of cohabitation is quite substantial and increases the current period perceived control by 13.2 percent of a standard deviation. This effect is even stronger than the negative feedback exerted from past poverty experiences, which amounts to 4.3 percent of a standard deviation. Also in contrast to females, for whom the magnitude was comparable, this effect is statistically significant. Its interpretation in terms of monetary changes is similar to the example given above.

Table 9.5: Average Partial Effects for Headcount (Female) – Partial Likelihood Approach (1995 – 2010, Sampling Interval = 5 yrs.)

Dependent Variables	Pov. St.	Emp. St.	Childn.	Partn.	PC
Endogenous Variables					
Employment St.	-0.069***	–	–	–	–
Children	0.003**	-0.220**	–	–	–
Partner in HH	-0.065***	-0.029***	0.081***	–	–
Perceived Control (PC)	-0.012***	0.011***	-0.001	0.010***	–
Lagged Variables					
Poverty St.	0.189***	0.009**	0.059***	-0.042***	-0.052
Employment St.	0.017***	0.309***	0.024**	0.001	-0.054**
Children	0.016***	0.127***	0.502***	-0.011***	-0.024
Partner in HH	0.011***	-0.015***	0.176***	0.207***	-0.013
Perceived Control (PC)	-0.007**	0.003**	0.007**	0.006**	0.369***
Strictly Exogenous Variables					
Some School (D)	-0.020***	0.028**	0.113***	-0.056***	0.046*
Higher Secondary (D)	-0.033***	-0.062***	0.082***	-0.080***	-0.130
Some Voc. Train. (D)	-0.038**	0.070**	0.019**	0.008**	-0.040*
University (D)	0.002	0.333***	0.156***	0.084***	0.068*
Tech-Coll. (D)	-0.010	0.264***	0.080***	0.022**	0.093
Job Exp. (Full T.)	0.001*	-0.006***	-0.001**	0.004***	0.002
Age	-0.004***	-0.008**	-0.011***	-0.004***	-0.026***
East German (D)	0.067***	0.050**	-0.041***	-0.138***	-0.050
German (D)	-0.035**	-0.024*	0.194***	0.071***	-0.076
Time Averages					
Some School (D)	-0.055***	0.021**	-0.105**	0.064***	0.087*
Higher Secondary (D)	-0.057***	0.012***	-0.084***	0.077***	0.190
Some Voc. Train. (D)	0.001	0.004**	0.013**	0.014**	0.047
University (D)	-0.060***	-0.158***	-0.114***	-0.076***	0.156*
Tech-Coll. (D)	-0.041***	-0.155***	-0.053***	-0.003	-0.025
Job Exp. (Full T.)	-0.002***	0.019***	-0.003***	-0.005***	0.006*
Age	-0.005***	-0.007**	-0.002***	0.003***	0.018***
East German (D)	0.001	-0.073***	0.004	-0.135***	-0.064
German (D)	-0.001	0.022	-0.245**	-0.121***	-0.025*
N	8,954	9,079	12,069	9,067	8,489

Dummy variables are indicated by (D).
APEs for initial conditions from the Wooldridge-term are not reported. Standard errors underlying the reported significance are cluster robust on the individual level. The p-values for the APEs are approximated by the Delta-Method.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 9.6: Average Partial Effects for Poverty Deficit (Female) – Partial Likelihood Approach (1995 – 2010, Sampling Period = 5 yrs.)

Dependent Variables	Pov. St.	Emp. St.	Childn.	Partn.	PC
Endogenous Variables					
Employment St.	-199.414***	–	–	–	–
Children	-2.240	-0.219***	–	–	–
Partner in HH	-160.355***	-0.028***	0.080***	–	–
Perceived Control (PC)	-18.239***	0.011***	-0.002*	0.010***	–
Lagged Variables					
Poverty St.	0.979***	0.000***	0.000***	-0.000***	-0.000
Employment St.	20.015	0.309***	0.022***	0.002	-0.052**
Children	34.303**	0.127***	0.502***	-0.010***	-0.024
Partner in HH	3.953	-0.015***	0.174***	0.207***	-0.011
Perceived Control (PC)	-17.853***	0.003***	0.006***	0.006***	0.369***
Strictly Exogenous Variables					
Some School (D)	-39.425	0.028**	0.115***	-0.059***	0.041
Higher Secondary (D)	-76.730	-0.061***	0.083***	-0.083***	-0.019
Some Voc. Train. (D)	-59.488**	0.070**	0.021***	0.008**	0.039*
University (D)	14.693	0.333***	0.155***	0.084***	-0.067
Tech-Coll. (D)	18.197	0.264***	0.081***	0.021**	0.092
Job Exp. (Full T.)	2.855	-0.005***	-0.001**	0.004***	0.002
Age	6.456***	-0.008**	-0.010***	-0.004***	-0.026***
East German (D)	86.869**	0.050**	-0.039***	-0.139***	-0.052
German (D)	-45.067	-0.025*	0.197***	0.067***	-0.077
Time Averages					
Some School (D)	-73.328	0.021*	-0.108***	0.067***	0.093
Higher Secondary (D)	-62.061	0.013***	-0.086***	0.079***	0.196*
Some Voc. Train. (D)	-21.336	0.004**	0.011**	0.015**	0.048*
University (D)	-126.505*	-0.158***	-0.115***	-0.076***	0.156*
Tech-Coll. (D)	-124.519*	-0.155***	-0.056***	-0.001	-0.023
Job Exp. (Full T.)	-4.704	0.019***	-0.003***	-0.005***	0.005*
Age	-8.786***	-0.007**	-0.002***	0.003***	0.017***
East German (D)	18.799	-0.073***	0.004	0.135***	-0.064
German (D)	-16.633	0.022	-0.249***	-0.116***	0.251*
N	8,954	9,079	12,069	9,067	8,489

Dummy variables are indicated by (D).
APEs for initial conditions from the Wooldridge-term are not reported. Standard errors underlying the reported significance are cluster robust on the individual level. The p-values for the APEs are approximated by the Delta-Method.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 9.7: Average Partial Effects for Watts Measure (Female) – Partial Likelihood Approach (1995 – 2010, Sampling Period = 5 yrs.)

Dependent Variables	Pov. St.	Emp. St.	Childn.	Partn.	PC
Endogenous Variables					
Employment St.	-0.408***	–	–	–	–
Children	0.018	-.219***	–	–	–
Partner in HH	-0.327***	-.029***	0.080***	–	–
Perceived Control (PC)	-0.034***	.011***	-0.002*	0.010***	–
Lagged Variables					
Poverty St.	0.718***	0.036***	0.109***	-0.059***	-0.129*
Employment St.	0.025	0.309***	0.021***	0.004**	-0.053**
Children	0.055**	0.127***	0.503***	-0.011***	-0.024
Partner in HH	-0.007	-0.015***	0.173***	0.208***	-0.011
Perceived Control (PC)	-0.042***	0.003***	0.006***	0.007***	0.370***
Strictly Exogenous Variables					
Some School (D)	-0.095	0.028**	0.115***	-0.058***	0.039
Higher Secondary (D)	-0.172*	-0.061***	0.083***	-0.082***	-0.020
Some Voc. Train. (D)	-0.125**	0.070***	0.021***	0.008**	0.039*
University (D)	0.023	0.333***	0.156***	0.084***	-0.067
Tech-Coll. (D)	0.020	0.264***	0.081***	0.021**	0.092
Job Exp. (Full T.)	0.008	-0.006***	-0.001**	0.003***	0.002
Age	0.013***	-0.008***	-0.010***	-0.004***	-0.026***
East German (D)	0.168*	0.049***	-0.039***	-0.139***	-0.051
German (D)	-0.118	-0.025*	0.197***	0.066***	-0.077
Time Averages					
Some School (D)	-0.135	0.021*	-0.109***	0.068***	0.093
Higher Secondary (D)	-0.121	0.128***	-0.087***	0.080***	0.196*
Some Voc. Train. (D)	-0.049	0.009**	0.010**	0.015**	0.048*
University (D)	-0.247*	-0.158***	-0.115***	-0.075***	0.156*
Tech-Coll. (D)	-0.247*	-0.156***	-0.056***	-0.001	-0.023
Job Exp. (Full T.)	-0.011	0.019***	-0.003***	-0.005***	0.005*
Age	-0.018***	-0.007***	-0.002***	0.003***	0.017***
East German (D)	0.059	-0.073***	0.005	0.134***	-0.064
German (D)	-0.015	0.023*	-0.249***	-0.115***	0.251*
N	8,954	9,079	12,069	9,067	8,489

Dummy variables are indicated by (D).
APEs for initial conditions from the Wooldridge-term are not reported. Standard errors underlying the reported significance are cluster robust on the individual level. The p-values for the APEs are approximated by the Delta-Method.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 9.8: Average Partial Effects for Headcount (Male) – Partial Likelihood Approach (1995 – 2010, Sampling Interval = 5 yrs.)

Dependent Variables	Pov. St.	Emp. St.	Childn.	Partn.	PC
Endogenous Variables					
Employment St.	-0.068***	–	–	–	–
Children	0.044***	0.035***	–	–	–
Partner in HH	-0.032***	0.045***	0.340***	–	–
Perceived Control (PC)	-0.007***	0.022***	-0.002***	0.008***	–
Lagged Variables					
Poverty St.	0.153***	0.001	0.041***	0.002	-0.043*
Employment St.	0.005***	0.170***	0.111***	0.026***	0.017
Children	0.014***	0.036***	0.426***	-0.014***	-0.079***
Partner in HH	0.003	0.011***	0.088***	0.212***	0.132***
Perceived Control (PC)	-0.002***	0.004***	0.004***	0.009***	0.367***
Strictly Exogenous Variables					
Some School (D)	0.020***	0.069***	0.004	0.018**	0.165*
Higher Secondary (D)	-0.018***	0.120***	0.069***	0.056***	0.079
Some Voc. Train. (D)	-0.009***	0.053***	0.047**	0.029**	0.080*
University (D)	-0.042***	0.193***	0.058***	0.093***	0.035
Tech-Coll. (D)	-0.068***	0.186***	0.031***	0.094**	-0.007
Job Exp. (Full T.)	0.000	0.015***	-0.009**	-0.002***	0.004
Age	0.005***	-0.022***	-0.006***	0.000	-0.036***
East German (D)	0.046***	-0.059***	0.029***	-0.097***	-0.075
German (D)	-0.026**	-0.124***	0.077***	0.062***	-0.157
Time Averages					
Some School (D)	-0.057***	-0.029**	0.001	-0.021**	-0.045
Higher Secondary (D)	-0.021***	-0.108***	-0.081***	-0.046***	0.076
Some Voc. Train. (D)	-0.011***	0.040***	-0.021***	-0.014***	-0.006
University (D)	-0.038***	-0.013	-0.039***	-0.083***	0.044
Tech-Coll. (D)	0.041***	-0.131***	-0.010	-0.069***	0.030
Job Exp. (Full T.)	-0.003***	0.002***	0.002***	0.005***	0.002
Age	-0.003***	-0.005***	-0.003***	-0.003***	0.025***
East German (D)	-0.004	0.000	-0.082***	0.096***	-0.015
German (D)	0.006	0.125***	-0.136**	-0.120***	0.296
N	8,378	8,491	11,217	8,479	7,916

Dummy variables are indicated by (D).
APEs for initial conditions from the Wooldridge-term are not reported. Standard errors underlying the reported significance are cluster robust on the individual level. The p-values for the APEs are approximated by the Delta-Method.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 9.9: Average Partial Effects for Poverty Deficit (Male) – Partial Likelihood Approach (1995 – 2010, Sampling Interval = 5 yrs.)

Dependent Variables	Pov. St.	Emp. St.	Childn.	Partn.	PC
Endogenous Variables					
Employment St.	-223.944***	–	–	–	–
Children	77.177***	0.035***	–	–	–
Partner in HH	-80.747***	0.045***	0.341***	–	–
Perceived Control (PC)	-16.007***	0.022***	-0.002***	0.008***	–
Lagged Variables					
Poverty St.	0.839***	0.000***	0.000***	0.000**	-0.000
Employment St.	1.677	0.171***	0.109***	0.026***	0.021
Children	35.359**	0.036***	0.427***	-0.014***	-0.079***
Partner in HH	-1.653	0.011***	0.089***	0.213***	0.132***
Perceived Control (PC)	-4.629	0.005***	0.004***	0.009***	0.367***
Strictly Exogenous Variables					
Some School (D)	11.1236	0.069***	0.003	0.018**	0.164*
Higher Secondary (D)	-51.2933	0.120***	0.069***	0.056***	0.077
Some Voc. Train. (D)	-11.2031	0.053***	0.047***	0.029**	0.079*
University (D)	-80.1962	0.193***	0.059***	0.093***	0.032
Tech-Coll. (D)	-205.191***	0.186***	0.033***	0.094**	-0.009
Job Exp. (Full T.)	.542645	0.015***	-0.009**	-0.002***	0.004
Age	9.30198***	-0.022***	-0.005***	0.000	-0.036***
East German (D)	41.6851	-0.059***	0.029***	-0.097***	-0.075
German (D)	-83.344	-0.119***	0.077***	0.062***	-0.157
Time Averages					
Some School (D)	-69.969*	-0.029**	0.001	-0.021**	-0.043
Higher Secondary (D)	-7.720	-0.108***	-0.083***	-0.046***	0.079
Some Voc. Train. (D)	-29.148	0.040***	-0.022***	-0.014***	-0.005
University (D)	-112.633*	-0.013	-0.041***	-0.083***	0.048
Tech-Coll. (D)	77.097	-0.130***	-0.011	-0.069***	0.033
Job Exp. (Full T.)	-5.816*	0.002***	0.002***	0.005***	0.002
Age	-5.386**	-0.005***	-0.003***	-0.003***	0.025***
East German (D)	31.783	0.000	-0.082***	0.096***	-0.015
German (D)	42.788	0.122***	-0.136***	-0.120***	0.296
N	8,378	8,491	11,217	8,479	7,916

Dummy variables are indicated by (D).
APEs for initial conditions from the Wooldridge-term are not reported. Standard errors underlying the reported significance are cluster robust on the individual level. The p-values for the APEs are approximated by the Delta-Method.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 9.10: Average Partial Effects for Watts Measure (Male) – Partial Likelihood Approach (1995 – 2010, Sampling Interval = 5 yrs.)

Dependent Variables	Pov. St.	Emp. St.	Childn.	Partn.	PC
Endogenous Variables					
Employment St.	-0.418***	–	–	–	–
Children	0.156***	0.035***	–	–	–
Partner in HH	-0.157***	0.045***	0.341***	–	–
Perceived Control (PC)	-0.031***	0.022***	-0.002***	0.008***	–
Lagged Variables					
Poverty St.	0.608***	0.044***	0.080***	0.022**	-0.069
Employment St.	-0.021	0.171***	0.109***	0.026***	0.020
Children	0.062**	0.036***	0.427***	-0.014***	-0.080***
Partner in HH	-0.002	0.011***	0.089***	0.213***	0.132***
Perceived Control (PC)	-0.011	0.005***	0.004***	0.009***	0.367***
Strictly Exogenous Variables					
Some School (D)	0.002	0.069***	0.003	0.018**	0.164*
Higher Secondary (D)	-0.094	0.120***	0.069***	0.056***	0.077
Some Voc. Train. (D)	-0.023	0.053***	0.047***	0.029**	0.079*
University (D)	-0.147	0.193***	0.059***	0.093***	0.032
Tech-Coll. (D)	-0.377***	0.186***	0.033***	0.094**	-0.009
Job Exp. (Full T.)	0.001	0.015***	-0.009**	-0.002***	0.004
Age	0.017***	-0.022***	-0.005***	0.000	-0.036***
East German (D)	0.061	-0.061***	0.029***	-0.098***	-0.074
German (D)	-0.193	-0.119***	0.076***	0.063***	-0.149
Time Averages					
Some School (D)	-0.123*	-0.029**	0.001	-0.020**	-0.043
Higher Secondary (D)	-0.029	-0.108***	-0.083***	-0.045***	0.079
Some Voc. Train. (D)	-0.053	0.040***	-0.022***	-0.014***	-0.005
University (D)	-0.222*	-0.013	-0.041***	-0.083***	0.048
Tech-Coll. (D)	0.127	-0.129***	-0.011	-0.068***	0.033
Job Exp. (Full T.)	-0.012*	0.002***	0.002***	0.005***	0.002
Age	-0.009**	-0.005***	-0.003***	-0.003***	0.025***
East German (D)	0.083	0.001	-0.082***	0.096***	-0.016
German (D)	0.109	0.121***	-0.135***	-0.122***	0.289
N	8,378	8,491	11,217	8,479	7,916

Dummy variables are indicated by (D).
APEs for initial conditions from the Wooldridge-term are not reported. Standard errors underlying the reported significance are cluster robust on the individual level. The p-values for the APEs are approximated by the Delta-Method.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

9.8.2 Structural Approach

The results for the structural model that explicitly accounts for potential feedback in the predetermined variables are shown in Table 9.11 for females and Table 9.12 for males. Though other possible orders of predeterminedness for the dependent variables in columns II to IV have not been tested yet, the Lagrange multiplier test of Anderson and Kunitomo (1992) fails to reject the null hypothesis of the predeterminedness order as given by Tables 9.11 and 9.12 against overidentified (unrestricted) alternatives in case of females and males. This result shows that the structural model as suggested is at least in line with data. It remains to be seen, whether other model constellations provide LM-statistics that are more distant to the respective critical χ^2-values and thus provide a better fit under the null hypothesis.

Moreover, recall from Section 9.7 that the structural approach considers only complete time paths over the whole observational timespan, leading to a substantially lower number of observations compared to the models discussed in the previous section. For the female sample, some of the revealed effects are different, but not in a way that is inconsistent with the previous findings. Looking at the impact of the strictly exogenous variables, apart from the general secondary schooling degree, education significantly decreases the risk of living in poverty. The indicators for holding a higher secondary schooling degree, a university degree, or a technical college degree are the only education variables the time-means of which have a substantial poverty reducing effect, probably due to their role as projections of unobserved abilities. As for the partial likelihood estimates, living in the eastern part of Germany can be associated with an increase in poverty risk. The poverty reducing effect of full time job experience is slightly higher then suggested by the previous models. The results for the impact of holding a German citizenship are comparable under both models considered thus far. All further exogenous partial effects are rather negligible within the poverty equation. Regarding the contemporaneous cross-effects, the mediating role of employment seems to be more of a factor within the structural setup. As such, it may also be possible that some of the exogenous covariates additionally operate on poverty status via the employment equation. When jointly considering the contemporaneous exogenous variables and the correlated part of the employment equation, the compound effects have not changed much. The overall impacts of the human capital related characteristics have a positive influence on employment probability in large parts, whereas the impacts of graduation from university and technical college have changed their signs. By and large, there also are no dramatic changes in the exogenous and correlated model parts for the equations representing partnership and having children, though some effects are even reversed. For instance, the impact of a technical college degree on the probability of having at least one child changes from 3 percentage point to -7 percentage points, an effect that is relatively weak in magnitude though. Other effects remain almost unchanged as is exemplified by the average probability change exerted from holding a university degree to living in a partnership. Likewise maintained is the almost 30 percentage point reduction in the probability of living in a partnership for those individuals who live in eastern Germany. Returning the attention to the contemporaneous interrelations between the model equations, the effects of having children, living with a partner, and of the individual degree of perceived control are still comparable to those found within the partial likelihood approach. The impacts of partnership and childbearing are still negative with regard to employment but have decreased, whereas the magnitude of perceived control in the partnership equation has increased to almost 7 percentage points. Furthermore, the average effect from cohabitation to having a child is up by some 3.5 percentage points compared to the previous framework.

With regard to lagged effects of the predetermined variables and poverty, the strong state dependence within the respective model parts remains for the structural model as well. In case of poverty and employment, it has decreased, whereas for the children and cohabitation sub-models, there is a slight increase in the state dependencies. The previously positive lagged effects from poverty and partnership on the probability of having a child, reverse into quite small negative effects. The explanatory associations for the model part on perceived control is still rather diffuse and at best allows to infer some significant relations with regard to age. What turns out to be the most important finding of the structural estimates, however, is that the sizeable feedback effect from previous poverty experiences to control expression seems to be confirmed. As opposed to the partial likelihood model, it has even increased to -7.6 percent of a standard deviation.

Table 9.12 displays the results for the structural model given the male sample. Concerning the coefficients for the poverty equation, the strictly exogenous and time-invariant effects are again comparable in magnitude. Most of the human capital related predictors lower the probability of living in poverty, as does having the German citizenship. Living in eastern Germany, on the other hand, is again negatively associated with poverty reduction. As with the female sample, the contemporaneous impact of employment is slightly higher in case of the structural model, which again may be an argument in favor of contemplating employment as a mediator of poverty. The other effects are remarkably similar, though the direct impact of perceived control on poverty is substantially lower. The effects of the variables on employment are also in line with prior expectations. Higher educational qualifications are generally associated with higher employment probabilities. In the structural model, the effect of age on the employment probability has inverted, but still is relatively weak in magnitude. The role of the exogenous variables in the remaining model parts also follow the previous discussions in large parts. What should be noted, however, is that the structural model again suggests sizable and significant state dependence effects across the entities involved in the five equations. Those for poverty, living with a partner, and perceived control are even stronger than in the previous models, whereas the state dependence for employment and having a child have decreased. Likewise, the effects for the other lagged predetermined variables show very similar patterns to those in the previous models. The results for the employment equation are rather weak. The fact that lagged poverty slightly increases the employment probability of a given period is somewhat at odds with what one could have expected. All other lagged cross-effects are rather low with regard to their magnitudes and their interpretations. Again, the effect of primary interest has increased (in absolute terms) by more than 3 percentage points compared to the partial likelihood model. This finding provides further evidence that, on average, past poverty experiences seem to negatively impinge on control perception.

Given the previous findings on the feedback effects when the poverty relations are set up from the Watts measure, the structural estimates presented in this section also suggest that the feedback from poverty to future control perception may be even stronger when the more nuanced Watts measures is employed instead of the binary headcount. This model is yet not implemented, however. A final note on the estimates for the variance component σ_a, which can be directly quantified in the structural model, may be in order. As opposed to the true state dependence, the unobserved component is relatively low. This may be owed to the fact, that controlling for perceived control is expected to significantly reduce unobserved heterogeneity that usually prevails in the compound error.

Table 9.11: Average Partial Effects for Headcount (Female) – Structural Approach (1995 – 2010, Sampling Interval = 5 yrs.)

Dependent Variables	Pov. St.	Emp. St.	Childn.	Partn.	PC
Endogenous Variables					
Employment St.	-0.098***	–	–	–	–
Children	0.046**	-0.054***	–	–	–
Partner in HH	-0.011**	-0.033**	0.115***	–	–
Perceived Control (PC)	-0.079**	0.046*	0.059**	0.069***	–
Lagged Variables					
Poverty St.	0.143**	0.018**	-0.022**	0.006*	-0.076**
Employment St.	0.003**	0.101*	0.106**	0.010**	-0.015**
Children	0.014***	-0.098*	0.786**	-0.014***	-0.008*
Partner in HH	0.005*	-0.011**	-0.013*	0.212***	0.018
Perceived Control (PC)	-0.007**	0.004*	-0.059*	0.010*	0.167**
Strictly Exogenous Variables					
Some School (D)	0.020**	0.039***	0.020**	0.085	0.054*
Higher Secondary (D)	-0.021*	0.043*	0.021***	-0.058***	-0.213
Some Voc. Train. (D)	-0.009***	0.126***	0.121*	0.014*	-0.025*
University (D)	-0.052**	0.148***	0.109**	0.058***	0.071*
Tech-Coll. (D)	-0.068*	0.137*	-0.013**	0.009**	0.019
Job Exp. (Full T.)	0.005**	0.177**	0.030*	0.033**	0.002*
Age	-0.005**	-0.086***	-0.049*	0.009**	-0.021**
East German (D)	0.055***	0.059**	-0.025*	-0.171***	-0.055
German (D)	-0.026*	0.061*	0.163**	0.026*	-0.038*
Time Averages					
Some School (D)	0.013**	0.011**	-0.105**	0.016**	0.124*
Higher Secondary (D)	-0.036**	0.001*	-0.084***	0.067**	0.212
Some Voc. Train. (D)	0.007	0.013**	0.013**	0.012**	0.066
University (D)	-0.046***	-0.158***	-0.114***	-0.043***	0.135*
Tech-Coll. (D)	-0.024***	-0.155***	-0.053***	-0.007*	-0.023
Job Exp. (Full T.)	-0.012**	0.019***	-0.003***	-0.003*	-0.001*
Age	-0.011***	-0.007**	-0.002***	0.001**	0.065***
East German (D)	0.003*	-0.073***	0.004	-0.117**	-0.061
German (D)	-0.016**	0.022	-0.245**	-0.093***	-0.021*
σ_a	0.134***	0.221**	0.245**	0.121***	0.251**
N					1,489

Dummy variables are indicated by (D).
APEs for initial conditions from the Wooldridge-term are not reported. Standard errors underlying the reported significance are cluster robust on the individual level. The p-values for the APEs are approximated by the Delta-Method.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 9.12: Average Partial Effects for Headcount (Male) – Structural Approach (1995 – 2010, Sampling Interval = 5 yrs.)

Dependent Variables	Pov. St.	Emp. St.	Childn.	Partn.	PC
Endogenous Variables					
Employment St.	-0.081**	–	–	–	–
Children	0.065***	0.048**	–	–	–
Partner in HH	-0.019*	0.019**	0.567**	–	–
Perceived Control (PC)	-0.093***	0.031***	0.008**	0.006***	–
Lagged Variables					
Poverty St.	0.251***	0.008**	0.037**	0.007***	-0.081***
Employment St.	0.012**	0.127***	0.109**	0.031**	0.025**
Children	0.021*	0.089*	0.368***	-0.021***	-0.001*
Partner in HH	0.009**	-0.043**	-0.001*	0.332**	0.005*
Perceived Control (PC)	-0.003***	0.013**	-0.005*	0.007*	0.587**
Strictly Exogenous Variables					
Some School (D)	0.015***	-0.190**	0.060	0.004***	-0.562*
Higher Secondary (D)	-0.033**	-0.076**	0.025**	0.009**	-0.002*
Some Voc. Train. (D)	-0.003***	0.184**	-0.015*	0.018**	0.067
University (D)	-0.054**	0.268***	-0.088**	0.049***	-0.012
Tech-Coll. (D)	-0.078*	0.156**	-0.187**	0.079***	0.537*
Job Exp. (Full T.)	0.003**	-0.143***	-0.016***	-0.005*	0.041
Age	-0.001**	0.064***	0.014***	-0.009***	-0.004
East German (D)	0.067***	-0.125**	-0.146***	-0.399***	0.248*
German (D)	-0.015**	0.170**	0.072**	0.013*	-0.368
Time Averages					
Some School (D)	-0.005***	0.265***	-0.009	0.006***	0.434*
Higher Secondary (D)	-0.036*	0.116***	-0.007*	0.015*	0.005
Some Voc. Train. (D)	0.007	-0.177***	0.025**	-0.007**	-0.005
University (D)	-0.061**	-0.153***	0.084***	-0.033***	0.261*
Tech-Coll. (D)	-0.012**	-0.118***	0.115*	-0.047**	-0.351*
Job Exp. (Full T.)	-0.005**	0.198*	0.012*	0.004*	-0.047*
Age	-0.001**	-0.072***	-0.026**	0.008***	-0.002
East German (D)	0.005***	0.054**	0.098***	0.168**	-0.319**
German (D)	-0.041***	-0.168*	-0.047***	-0.005	0.226
σ_a	0.346**	0.198***	0.451**	0.219*	0.571***
N					1,351

Dummy variables are indicated by (D).
APEs for initial conditions from the Wooldridge-term are not reported. Standard errors underlying the reported significance are cluster robust on the individual level. The p-values for the APEs are approximated by the Delta-Method.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

9.9 Robustness Checks

As noted above, several caveats should be considered given the estimates discussed thus far. The first issue that has been addressed was that the comparably long observational interval is likely to affect the dynamic cross-effects and state-dependencies in the setting at hand. The other feature of the data that should be taken into account is that the time-averaged model parts fluctuate somewhat more than the other derived effects. This may be an artifact of the relatively low number of waves that can be used for the considered models. This section addresses both points in successive order.

Regarding the potential problems arising from the quite distant observational points, Tables 9.13 and 9.14 provide partial likelihood estimates for the annually available waves from 1994 to 1996, where the 1994 wave acts as initial period. Due to the nature of the partial likelihood approach, only the time-averaged effects are likely to suffer from this even shorter timespan. All other effects may provide a viable comparison to the main results. As such, I will not focus on the differences in the strictly exogenous variables and the correlated model parts here. Regarding the cross-dependencies for the female sample, one finds that relatively minor changes occur for the contemporaneous impacts of the predetermined variables. The risk-reducing effect of employment on poverty has decreased by an absolute margin of 2 percentage points, whereas the other contemporaneous effects in the poverty model remain remarkably stable compared to the original model. For the employment and childbearing model, only the effects of having a child and living with a partner have become weaker. Substantial changes occur for the state dependencies in all five models. This result is likely to arise due to the large discrepancy between the unit period and the observational interval in the original models. Apart from few exceptions, the lagged cross effects seem to be less influenced by this issue. In particular, the feedback effect from poverty to perceived control is only 0.7 percentage points weaker in absolute terms. This indicates a sufficient degree of robustness for the main results on this association.

For the male sample, the picture is quite similar. Some of the contemporaneous cross-effects have changed, but not in a substantial way. Again, the employment effect on poverty is somewhat weaker. Moreover, the effect of living with a partner has reversed, but only amounts to a 4 percentage point change regarding the probability of living in poverty. For the remaining cross-effects, the changes are negligible. As opposed to the female sample, the vast increases in the state dependencies pertain to that of having at least one child, only. All other path dependencies are only moderately increased. Many of the lagged cross-dependencies that are subject to some changes are below 5 percentage points in both models and thus only provide quite unsystematic findings. The negative feedback from past poverty to perceived control has more than doubled, on the other hand. If one sees the former results on this effect as a lower bound estimate, this result at least does not jeopardize the hypothesis of a non-zero feedback on perceived control. In order to establish a sufficient degree of robustness, however, further model estimations should be conducted.

Looking at the results in Tables 9.15 and 9.16, some robustness checks on the potential effect of the rather low number of waves in the models thus far are provided. The employed waves comprise those from 1994 to 2010, with the 1994 wave again representing the initial period. As perceived control is not available on an annual basis for this timespan, the checks only comprise the first four equations of the original model. Regarding the female respondents, the differences in the correlated model parts are only minor when comparing the long panel with original model. The low number of within-individual observations seems to be more problematic for the contemporaneous exogenous variables, as the differences are most substantial in these model parts. The contemporaneous cross-effects for the predetermined variables are roughly in line with the original ones and those for the first robustness check. Fortunately, the same seems to hold for the lagged model parts. As with the model for the waves from 1994 to 1996, the state dependencies have substantially increased as a result of the annual observational interval. However, the lagged cross-effects seem to remain comparable in large parts. It may be cautiously concluded that this would also translate to the lagged feedback on perceived control, given it would have been available. For the male sample the same pattern of changes seems to apply. Again, there are quite substantial differences within the exogenous and correlated model parts, as well as increases in the magnitudes of the state dependencies within the four equations. On the other hand, the contemporaneous and lagged crosse-effects are rather in line with the previously examined model specifications.

Naturally, such separate treatises of the potential problems arising from the small panel length and the large observational intervals are not completely conclusive. Most likely, some data imputation scheme that allows for consideration of perceived control on an annual basis should be included in order to obtain more definite robustness results. Moreover, the strong changes in the state dependencies provide evidence in favor of the points previously made on the mixing of different spell types. Especially for poverty and employment in the five-year observational interval, those individuals for whom the two points in time are part of a continuing spell and those individuals who have one or more interrupted spell(s) obviously mix up in the original estimate.

Table 9.13: Average Partial Effects for Headcount (Female) – Partial Likelihood Approach (1994 – 1996, Sampling Interval = 1 yr.)

Dependent Variables	Pov. St.	Emp. St.	Childn.	Partn.	PC
Endogenous Variables					
Employment St.	-0.048***	–	–	–	–
Children	0.011***	-0.159***	–	–	–
Partner in HH	-0.054***	-0.008**	0.001	–	–
Perceived Control (PC)	-0.010***	0.009***	-0.003***	0.009***	–
Lagged Variables					
Poverty St.	0.179***	0.017**	0.023***	-0.007***	-0.045*
Employment St.	0.019***	0.364***	-0.001	0.002	0.002
Children	0.009***	0.096***	0.878***	0.004***	-0.004
Partner in HH	0.064***	-0.012***	0.082***	0.219***	0.026*
Perceived Control (PC)	-0.006***	-0.005**	0.005***	-0.003**	0.793***
Strictly Exogenous Variables					
Some School (D)	-0.068***	-0.028**	0.191***	-0.035***	0.121*
Higher Secondary (D)	-0.061***	-0.016**	0.153***	-0.044***	0.196
Some Voc. Train. (D)	0.034***	0.543***	0.040***	0.058***	-0.208*
University (D)	0.016	0.495***	0.005**	0.008	0.192*
Tech-Coll. (D)	-0.068***	0.613***	0.137***	0.198***	-0.102
Job Exp. (Full T.)	-0.002	-0.105***	-0.002	-0.006**	0.001
Age	-0.012***	0.031***	0.004***	0.001	0.019*
East German (D)	-0.063***	0.179***	0.259***	0.035*	-0.087
German (D)	0.159***	0.119***	-0.216***	0.054***	-0.054
Time Averages					
Some School (D)	0.192***	0.098***	-0.103***	0.062***	-0.139
Higher Secondary (D)	0.136***	0.056***	-0.078***	0.069***	-0.135
Some Voc. Train. (D)	-0.044***	-0.389***	-0.010	-0.057***	0.271**
University (D)	-0.041***	-0.282***	0.028***	-0.010	-0.037
Tech-Coll. (D)	0.142***	-0.487***	-0.087***	-0.199***	0.250
Job Exp. (Full T.)	0.002	0.109***	0.001	0.005**	-0.001
Age	0.011***	-0.037***	-0.007***	-0.001	-0.022*
East German (D)	0.148***	-0.152***	-0.189***	-0.030	0.053
German (D)	-0.109***	-0.146***	0.119***	-0.040***	0.183
N	8,218	8,303	8,660	8,301	8,663

Dummy variables are indicated by (D).
APEs for initial conditions from the Wooldridge-term are not reported. Standard errors underlying the reported significance are cluster robust on the individual level. The p-values for the APEs are approximated by the Delta-Method.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 9.14: Average Partial Effects for Headcount (Male) – Partial Likelihood Approach (1994 –1996, Sampling Interval = 1 yr.)

Dependent Variables	Pov. St.	Emp. St.	Childn.	Partn.	PC
Endogenous Variables					
Employment St.	-0.042***	–	–	–	–
Children	0.054***	0.048***	–	–	–
Partner in HH	0.016***	0.006	0.243***	–	–
Perceived Control (PC)	-0.002***	0.006***	0.005***	0.003***	–
Lagged Variables					
Poverty St.	0.189***	0.019**	0.006*	-0.008***	-0.105**
Employment St.	0.040***	0.203***	0.024***	0.009***	0.028
Children	-0.027***	-0.012***	0.798***	0.004***	0.006
Partner in HH	-0.013***	0.032***	-0.045***	0.229***	0.068**
Perceived Control (PC)	-0.007***	0.003**	0.002***	-0.001**	0.378***
Strictly Exogenous Variables					
Some School (D)	-0.029***	-0.191***	0.016	0.002***	-0.507*
Higher Secondary (D)	0.002	-0.077***	0.021***	0.008***	-0.003
Some Voc. Train. (D)	-0.027***	0.181***	-0.012	0.017**	0.063
University (D)	0.029***	0.207***	-0.081***	0.045***	-0.013
Tech-Coll. (D)	-0.021	0.162***	-0.182***	0.085***	0.533**
Job Exp. (Full T.)	-0.014***	-0.143***	-0.019***	-0.003	0.047
Age	0.006***	0.069***	0.018***	-0.009***	-0.003
East German (D)	-0.041***	-0.101***	-0.152***	-0.352***	0.239
German (D)	-0.036***	0.151***	0.065***	0.011	-0.327
Time Averages					
Some School (D)	0.032***	0.207***	-0.008	0.009***	0.442*
Higher Secondary (D)	-0.008**	0.115***	-0.004	0.011***	0.012
Some Voc. Train. (D)	0.039***	-0.173***	0.026**	-0.010	-0.016
University (D)	-0.056***	-0.164***	0.083***	-0.030***	0.232
Tech-Coll. (D)	0.010	-0.125***	0.119***	-0.049***	-0.342*
Job Exp. (Full T.)	0.015***	0.147***	0.019***	0.002	-0.049*
Age	-0.006***	-0.075***	-0.022***	0.010***	-0.000
East German (D)	0.089***	0.055***	0.099***	0.143***	-0.303
German (D)	0.023**	-0.175***	-0.043***	-0.004	0.208
N	7,953	8,027	8,379	8,027	7,992

Dummy variables are indicated by (D).
APEs for initial conditions from the Wooldridge-term are not reported. Standard errors underlying the reported significance are cluster robust on the individual level. The p-values for the APEs are approximated by the Delta-Method.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 9.15: Average Partial Effects for Headcount (Female) – Partial Likelihood Approach (1994 – 2010, Sampling Interval = 1 yr.)

Dependent Variables	Pov. St.	Emp. St.	Childn.	Partn.
Endogenous Variables				
Employment St.	-0.062***	–	–	–
Children	-0.002*	-0.111***	–	–
Partner in HH	-0.056***	-0.015***	0.010***	–
Lagged Variables				
Poverty St.	0.368***	0.001	0.021***	-0.006***
Employment St.	0.029***	0.491***	-0.000	0.004***
Children	0.011***	0.061***	0.855***	-0.001**
Partner in HH	0.037***	-0.000	0.071***	0.273***
Strictly Exogenous Variables				
Some School (D)	0.013***	0.027***	0.042***	0.003**
Higher Secondary (D)	0.033***	0.027***	0.051***	0.013***
Some Voc. Train. (D)	-0.008***	0.094***	0.031***	0.010***
University (D)	-0.006**	0.193***	0.041**	0.032***
Tech-Coll. (D)	0.030***	0.113***	0.032***	0.036***
Job Exp. (Full T.)	-0.001***	-0.013***	0.002***	0.001***
Age	0.003***	0.000***	-0.002***	-0.001***
East German (D)	-0.014***	0.022***	-0.012***	-0.052***
German (D)	0.007**	-0.005*	0.002	0.007***
Time Averages				
Some School (D)	-0.018***	0.023***	-0.013***	0.003**
Higher Secondary (D)	-0.046***	0.022***	-0.019***	-0.009***
Some Voc. Train. (D)	-0.019***	-0.040***	-0.003**	0.001
University (D)	-0.054***	-0.075***	-0.014***	-0.022***
Tech-Coll. (D)	-0.066***	-0.052***	-0.016***	-0.030***
Job Exp. (Full T.)	-0.000	0.018***	-0.002***	-0.002***
Age	-0.003***	-0.006***	-0.000***	0.001***
East German (D)	0.058***	-0.029***	0.001	0.049***
German (D)	-0.038***	-0.023***	-0.026***	-0.019***
N	60,544	61,140	106,544	61,011

Dummy variables are indicated by (D).
APEs for initial conditions from the Wooldridge-term are not reported. Standard errors underlying the reported significance are cluster robust on the individual level. The p-values for the APEs are approximated by the Delta-Method.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 9.16: Average Partial Effects for Headcount (Male) – Partial Likelihood Approach (1994 – 2010, Sampling Interval = 1 yr.)

Dependent Variables	Pov. St.	Emp. St.	Childn.	Partn.
Endogenous Variables				
Employment St.	-0.061***	–	–	–
Children	0.021***	0.025***	–	–
Partner in HH	-0.003**	0.022***	0.228***	–
Lagged Variables				
Poverty St.	0.339***	-0.003***	0.018***	-0.002***
Employment St.	0.035***	0.273***	0.027***	0.017***
Children	0.002**	-0.009***	0.801***	-0.003***
Partner in HH	-0.007***	0.025***	-0.038***	0.292***
Strictly Exogenous Variables				
Some School (D)	-0.009***	0.017***	0.006**	0.004***
Higher Secondary (D)	0.004***	0.037***	0.017***	0.012***
Some Voc. Train. (D)	0.006***	0.082***	0.015***	0.003***
University (D)	-0.030***	0.175***	-0.007**	0.017***
Tech-Coll. (D)	-0.026***	0.117***	-0.005*	0.017***
Job Exp. (Full T.)	-0.000	-0.007***	-0.003***	-0.002***
Age	0.003***	-0.000***	-0.001***	0.001***
East German (D)	0.003	-0.015***	0.011***	-0.023***
German (D)	0.002	-0.029***	0.035***	0.003*
Time Averages				
Some School (D)	0.014***	0.042***	0.002	0.001
Higher Secondary (D)	-0.001	0.018***	-0.009***	-0.002
Some Voc. Train. (D)	-0.016***	-0.027***	-0.005***	0.002**
University (D)	-0.028***	-0.073***	0.014***	-0.011***
Tech-Coll. (D)	-0.008***	-0.059***	0.010***	-0.005***
Job Exp. (Full T.)	-0.001***	0.011***	0.002***	0.003***
Age	-0.002***	-0.007***	-0.001***	-0.001***
East German (D)	0.023***	-0.017***	-0.026***	0.025***
German (D)	-0.029***	0.011***	-0.055***	-0.018***
N	58,089	58,617	100,681	58,469

Dummy variables are indicated by (D).
APEs for initial conditions from the Wooldridge-term are not reported. Standard errors underlying the reported significance are cluster robust on the individual level. The p-values for the APEs are approximated by the Delta-Method.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

9.10 Discussion

Summarizing the previous findings, some rather robust results that establish the existence of a lagged feedback from poverty experiences to perceived control are provided by the presented dynamic panel estimates. Compared to early interventional studies (see, e.g., Almlund et al., 2011), most of the retrieved impacts on perceived control are comparably low. They are, however, sufficiently substantial in order to claim that the assumption of complete invariance of personality traits in adulthood is inappropriate in some cases. Depending on the respective specification and gender, the negative effects of past poverty experiences range from 4 to 10 percent of a standard deviation of perceived control. Referring to the trait formation literature, the results thus far are largely in line with previous findings that advocate small impacts of trigger events in adulthood on the stability of personality traits like perceived control (see, e.g., Cobb-Clark and Schurer, 2013). As opposed to the setting at hand, these are one time occurrences, however. As such, this pattern seems to be slightly altered when the persistence of the event is considered. These findings support the hypothesis that there can be sizeable changes in attitudes when an individual experiences certain long lasting environmental changes. Whether this result also shows that other kinds of personality traits are susceptible to similar changes is not resolved by the empirical findings presented here. Considering the inverse causal association, it has been shown that control attitudes do not provide any substantial information on the probability of slipping into poverty in the first place. The poverty status seems to primarily result from large state dependencies in poverty and employment, as well as from the corresponding cross-effects. Taken together, the results suggest that, apart from the alleged channel via perceived control, poverty experiences are further associated with processes of depreciation of human capital, demoralization, and incentive reductions. These mediators seem to jointly increase the probability that individuals who become poor will remain so for extended periods.

With respect to the other entities that are incorporated into the model frameworks, some additional interesting insights can be obtained. From a methodological perspective, the existence of feedback effects across equations on the future values of the predetermined variables makes the use of traditional random or fixed effects models, which are based on the strict exogeneity assumption, questionable with respect to the current and related settings. Based on the framework of Wooldridge (2000), the empirical analysis at hand draws on dynamic models that explicitly allows for such feedbacks. Previous estimation results suggest that feedback effects indeed prevail in lots of panel data settings, for instance in case of low (wage-)incomes and employment (see Stewart, 2007). Given the estimates provided here for German data, however, low equivalent incomes in the previous period does not impinge on the employment probability of the next period. There are indications for feedbacks on other entities, though. For instance, there is evidence that poverty affects household constitution, though in different directions for females and males. However, as discussed in the previous sections, household and family formation are quite complex decision problems and the results presented here are indicative at best. The further reflected state dependence effects have been found in previous studies as well. For instance in case of (un)employment dynamics, Arulampalam et al. (2000) also show large degrees of state dependence. As for the present results, it has been additionally shown that the extent of state dependence is subject to the chosen observational interval in large parts.

10

Concluding Remarks*

This thesis has reviewed, and within the scope of two empirical applications, also added to a recent and influential strand of the economic literature that considers the role of personality traits as an aspect of human capital. A selection of empirical studies that highlight the determination of crucial achievements and outcomes as a result of these traits has been briefly sketched and discussed. Moreover, the notion of personality traits in light of the relevant psychological literature has been introduced in order to outline the most important conceptions in the fields. In terms of trait measurement, empirical research in economics strongly benefits from psychometric concepts. Nonetheless, economists should be aware of the underlying assumptions when applying these concepts. An additional caveat lies in the fact that the commonly used constructs to measure personality traits are not completely conclusive. On the one hand, overall measures tend to be too general in that they veil important variation, whereas on the other hand, measures of specific personality traits may put the researcher to a hard choice regarding their adequacy. As has been shown in Chapter 3, psychometric coefficients of validity and reliability, which are often used to assess the eligibility of constructs, have some limitations in their own right and should be interpreted with caution.

By the same token, it has been shown that personality measures applied within an econometric framework tend to suffer from measurement error, simultaneity bias, and spurious influences by other unobservables. Due to these issues, the relation between personality traits and economic preference parameters is still patchwork and leaves many unanswered questions. Drawing inference on correlations between traits and preferences is a necessary first step but provides only cursory results. A better understanding of the pathways between both concepts is inevitable in order to obtain more conclusive insights, and also in order to enable more comprehensive models of decision making. Most likely, this goal may be achieved by means of more interdisciplinary research on this topic.

* The discussions presented in this chapter are published in Thiel and Thomsen (2013).

As a consequence of the latter two points raised, Chapter 5 has highlighted that the empirical analysis of trait inventories also urges adequate methods. As a cursory introduction for some eligible methods, the corresponding part of the thesis has invoked two somewhat related approaches, namely factor models and item response theory. The literature reviewed provides some practical aspects and guidance, and, furthermore, makes clear what exactly the benefits of both approaches are. Whereas item response models seem to better fit the nature of trait-based items without being too computationally intense, factor structure models, especially Bayesian ones, can be simultaneously applied to a wide range of structural empirical problems.

As has been discussed furthermore, personality traits are important determinants of several outcomes, like educational achievements and labor market success. The revealed patterns for different personality traits are relatively unequivocal across studies. Educational achievements apparently are a major mediating pathway for later labor market merits. With regard to remuneration, however, it is yet unclear to a large extent in how far the compound of productivity enhancement, occupational sorting, wage premia due to social desirability, and self-selection interact in wage determination and differ between the various traits under study.

With regard to formation and stratification of traits and abilities, the role and the timing of educational and parental investments have been proven to be crucial in the empirical literature. Chapter 7 has reviewed the most important findings from the underlying research fields. Regardless of the particular effects, virtually all empirical studies suggest a joint conclusion: early investments are most crucial, but nonetheless, should be complemented later on. Early neglect, on the other hand, cannot be compensated in later stages of the life without prohibitively high costs. Hence, in terms of support for low skilled or disadvantaged individuals the focus should be on early preschool age. The Cunha-Heckman model, which has been briefly introduced in Chapter 7, formalizes this process by means of a dynamic production function and also provides parameter estimates, the main implications of which also have been summarized in course of the discussion. Though the estimation approach accounts for measurement error, the insights on pattern and transmission of parental investments are far from definite. Attributing parental traits to preferences like altruism, which would allow to model the investment behavior of parents with regard to their children's development process, is quite complex and probably infeasible yet. Nonetheless, retrieving such definite associations between both notions would be a highly desirable aim. Given the patterns known thus far and given the findings for the intertemporal allocation of resources, the role of schooling investments seems to be rather subordinate in comparison to home environments, most likely due to its absence in the very first years of life. Despite these preliminary findings, it may be the case that this apportionment is predominantly data driven, as the underlying data usually provide little information on schooling resources.

As a consequence, little is known about the impact of schooling characteristics on the expression of personality traits in late adolescence. Chapter 8 has analyzed the effects of an increase in scholastic intensity on seven dimensions of personality, namely the Big Five, Locus of Control, and Self-Control. The presented empirical analysis has explored a natural experiment induced by an education policy reform in the German federal state of Saxony-Anhalt, where the last year of higher secondary schooling was abolished for students in the ninth grade at the time of the implementation. Concurrently, students in the tenth grade were unaffected by the change. Based on data for the double graduation cohort in 2007, the differences in outcomes between the two groups have been construed to represent the causal effect of the reform. The empirical results suggest that there has not been a significant nor a sizeable impact on any of the personality traits involved. These findings add to the literature on the (non-)plasticity of personality traits with regard to schooling investments in late adolescence. Other than the sample that has been employed here, this period is not yet captured by other data sets used in the empirical literature on personality formation. The results are nonetheless in line with the previously discussed general findings in the economic and psychological field. They indicate that rather than very general traits, later secondary schooling promotes the acquisition of more specific competencies.

Even at later periods of life, this stability pattern seems to change to some extent when environmental changes are more severe, or at least, more long lasting. The analysis that has been presented in Chapter 9 seizes this suggestion. In the literature on poverty determinants, among others, there is one often alleged causal pathway with particular relevance for stability patterns in personality traits. Previous empirical investigations find a strong state dependence in individual poverty paths that is mostly assumed to be induced by stigma, disincentives, or demoralization and changes in attitudes. The empirical analysis in Chapter 9 has quantified the latter explanation by means of perceived control. This potential feedback effect from past poverty experiences to perceived control attitudes has been modeled within two dynamic panel data frameworks. The estimation results have shown that there is a sizeable effect with regard to this causal association, at least for perceived control. Whether these findings translate into other personality dimensions is not clear, on the contrary. Despite this limited external validity, the result suggest that even in adulthood, the assumption of completely stable personality traits can be too restrictive in particular settings.

In summary, the revealed empirical findings and those discussed in light of the relevant literature enrich the traditional view on human capital in economics by considering personality traits as an additional determinant of lifetime labor market and social outcomes. Moreover, the essential role of infancy and early childhood in producing these outcomes has been accentuated. This provides new policy implications. Good parenting is the major source of educational success. This is only indirectly driven by family income when other characteristics are accounted for. Therefore, intervention policies should be adopted already at preschool age and should primarily focus on home environment. The time interval for sufficient governmental influence is more limited in case of cognitive abilities than for personality traits. The malleability of personality traits throughout adolescence and beyond provides a powerful and instantaneous policy tool. Moreover, though later stages of life appear to be inefficient in terms of potential interventions, it should always be considered that very adverse environmental changes can still exert substantial influences on traits, and thus are likely to deteriorate human capital to some extent. The result that have been presented here seem to be quite consistent, but nonetheless are largely derived from a still evolving literature rather specific empirical settings. Hence, their generality remains to be determined.

Appendix A

Numerical Integration

The following section is self-contained, i.e., its notation is independent from that throughout the previous chapters, except where compliance is explicitly indicated. Throughout the discussion of the item response framework two occasions occur where integrals are part of the computational procedure, namely

$$p_i(r_{i1}, r_{i2}, \ldots, r_{im}) = \int p_{i1}(r_{i1}|\theta_i)p_{i2}(r_{i2}|\theta_i)\ldots p_{im}(r_{im}|\theta_i)f(\theta_i)d\theta_i, \tag{1}$$

with $\theta \sim \mathcal{N}(0,1)$ and

$$G(u) \propto \int_0^\theta e^{t_1\gamma_1(u)+t_2\gamma_2(u)+t_3\gamma_3(u)}du.^1 \tag{2}$$

For the first integral, the objective is to obtain the joint marginal $p_{ip}(\cdot)$ from the conditional response probabilities $p_{ip}(\cdot|\theta)$ by building the expectation over θ. In the second expression, which is used for the approximation of the response probabilities, the whole term is subject to the integral. This supplemental section derives the numerical approximation procedures used in both instances.[2]

Rationale of Gaussian Quadrature Methods

A numerical approach to integration is necessary if, as in the two above cases, a closed-form solution for the integral is not available. Numerical integration by quadrature rules, including the Gaussian case, is a special application of operator expansion for linear functionals (see Dahlquist and Björck, 2008).[3] The aim is to obtain an approximation of

[1] The denomminator is only a special case of this integral.

[2] It should be noted, that the integral contained in $G(u)$ can be computed by means of the error function $erf(z)$ as follows.

$$\frac{1}{2}\frac{\sqrt{\pi}e^{-\frac{1}{6}\frac{15t_2^2+t_1^2}{t_2}}\left[erf\left(\frac{t_1-3t_2}{\sqrt{-6t_2}}\right)-erf\left(\frac{6t_2\theta+t_1-3t_2}{\sqrt{-6t_2}}\right)\right]}{\sqrt{-6t_2}}.$$

Compared to quadrature interpolation, the error function is more efficiently implementable from the a computational point of view. As is obvious from the above result however, this alternative holds only for the two-parameter case. Generalizations to higher Legendre polynomials are intractable again.

[3] Operator expansions also comprise various other kinds of function approximations, like Newton interpolation.

the integral functional $I[g]$ by a weighted sum of function values $g(x_i)$ at various nodes $x_1 < x_2 < \ldots < x_n$ in order to obtain a form like

$$I[g] \approx \sum_{i=1}^{n} \omega_i g(x_i). \tag{3}$$

Problems may arise due to singularities along the domain of the function g to be approximated as well as for several forms of non-smoothness. Technically speaking, for any $g(x)$ to be approximated, one shall require the moments

$$m_k = \int_a^b x^k w(x)dx$$

to exist and be finite for all $k \geq 0$ and arbitrary positive and continuous weight functions $w(x)$. Weight functions are advantageous (and often necessary) for practical implementation of quadrature methods. In general, the integral $I[g] = \int_a^b w(x)g(x)dx$ is approximated by using nodes $x_1 < x_2 < \ldots < x_n \in [a, b]$ and a unique polynomial $p(x)_{n-1}$ of degree $n-1$, leading to the approximating integral

$$I_n[g] = \int_a^b w(x)p(x)_{n-1}dx.$$

For illustration, consider the case where the $p_{n-1}(x)$ that interpolates $g(x)$ in the integral expression is set up from Lagrange polynomials. Then

$$p_{n-1}(x) = \sum_{i=1}^{n} p_{n-1}(x_i)\mathcal{L}_{n-1,i}(x), \text{ with } \mathcal{L}_{n-1,i}(x) = \prod_{j=1,j\neq i}^{n} \frac{x - x_j}{x_i - x_j} \text{ for } j = 1, \ldots, n.$$[4]

It follows immediately that the approximating integral simplifies to

$$\begin{aligned}
I_n[g] &= \int_a^b w(x)p_{n-1}(x)dx \\
&= \int_a^b \sum_{i=1}^{n} w(x)p_{n-1}(x_i)\mathcal{L}_{n-1,i}(x)dx \\
&= \sum_{i=1}^{n} p_{n-1}(x_i) \int_a^b w(x)\mathcal{L}_{n-1,i}(x)dx \\
&= \sum_{i=1}^{n} p_{n-1}(x_i)\omega_i.
\end{aligned}$$

[4] Lagrange polynomials are of little practical relevance as they have to be recomputed every time the number of nodes is extended or reduced (see Harris and Stocker, 2006). They have some convenient properties in terms of theoretical derivations, however, including orthonormality, i.e. $\mathcal{L}_{n-1,i}(x_j) = \delta_{ij}$ with δ_{ij} being the Kronecker delta.

Apart from swapping $g(x_i)$ with $p_{n-1}(x_i)$, the last equation is the result claimed in equation (3). It also shows that when $g(x)$ is itself a polynomial of degree $n-1$ (or lower), then $I[g] = \sum_{i=1}^{n} \omega_i g(x_i) = I_n[g]$ and the approximation is exact. This leads to the following result.

Theorem 1. *Given* $\omega_i = \int_a^b w(x)\mathcal{L}_{n-1,i}(x)dx$ *for* $i = 1, \ldots, n$,

$$\int_a^b w(x)p_{n-1}(x)dx = \sum_{i=1}^{n} p_{n-1}(x_i)\omega_i.$$

The result for obtaining the weights ω_i derives from the particular choice of the Lagrange polynomials in this case. For practical application, more general and more efficient formulas for the computation of the weights are available (see Judd, 1998, for an alternative). Moreover, it turns out that the ratio of the sequences of the potential polynomial orders $\{g_n\}_{n=1}^{\infty}$ and $\{p_n\}_{n=1}^{\infty}$ is $\mathscr{O}(1)$, i.e., if the order of the $g(x)$ to be interpolated is increased by k, the order of $p_{n-1}(x_i)$ has to be increased by the same magnitude. A contrivance for improvement is in not assuming the interpolation nodes $x_1 < x_2 < \ldots < x_n \in [a, b]$ to be prescribed, but making their choice explicit to the exactness considerations. At best, each of the n derived nodes increases the exactness of the approximation result given in Theorem 1 by one. Such judicious choice of nodes can be obtained when considering the particular class of orthogonal polynomials. Let $\pi(x) = \prod_{i=1}^{n}(x - x_i)$ be the factorial of an orthogonal polynomial π made up from its n distinct and real zeros.[5] Let $q(x)$ be another polynomial of degree $n-1$.[6] Then, one has

Theorem 2. *An interpolatory quadrature rule for* $g(x)$ *based on* $p_{n-1}(x_i)$ *at the zeros of* $\pi(x)$ *has degree of exactness* $2n-1$ *if and only if the inner product*

$$\langle \pi, q \rangle = \int_a^b w(x)\pi(x)q(x)dx = 0.$$

Proof. To show the necessity of the result, it is easily derived that $\pi(x)q(x)$ is of degree $2n-1$, since the degrees for $\pi(x)$ and $q(x)$ add up to $n+n-1$. For any continuous interval $[a, b]$ and for real x, whatever result for $\langle \pi, q \rangle = \int_a^b w(x)\pi(x)q(x)dx$ is obtained, generally implies a discrete analogue $\langle \pi, q \rangle = \sum_{i=1}^{n} \omega_i \pi(x_i)q(x_i)$ (see Davis and Rabinowitz, 1984). At the zeros of the n-th order polynomial $\pi(x)$ the product $\pi(x)q(x)$ clearly vanishes.

[5] These requirements are known to be fulfilled for the class of orthogonal polynomials (see Dahlquist and Björck, 2008, for a general treatise).

[6] Note that it would be pointless to consider higher degrees, since for degree n, orthogonality of $q(x)$ and $\pi(x)$ and thus the following theorem do not apply, and for all higher degrees one would use the zeros of $q(x)$ instead of $\pi(x)$.

Thus
$$0=\sum_{i=1}^{n}\omega_i\pi(x_i)q(x_i)=\int_a^b w(x)\pi(x)q(x)dx=\langle\pi,q\rangle,$$
where the second equality follows from Theorem 1 and the fact that $x_1<x_2<\ldots<x_n\in[a,b]$ are zeros. The polynomials $\pi(x)$ and $q(x)$ are mutually orthogonal.

In order to show sufficiency, recall from above that $I_n[g]=\int_a^b w(x)p_{n-1}(x)dx$ is exact for $I[g]$ as long as $g(x)$ is also of degree $n-1$. For the proclaimed exactness of degree $2n-1$ the order exceeds $n-1$ by n. The resulting error amounts to
$$I[g]-I_n[g]=\int_a^b w(x)\left[g(x)-p_{n-1}(x)\right]dx>0.$$

As $g(x)$ is $2n-1$, the error function $e(x)=g(x)-p_{n-1}(x)$ is of the same degree, and thus can be expressed in terms of the above $\pi(x_i)q(x_i)$. It follows that $g(x)=e(x)+p_{n-1}(x)$ and correspondingly
$$\int_a^b w(x)g(x)dx=\int_a^b w(x)\pi(x)q(x)dx+\int_a^b w(x)p_{n-1}(x)dx,$$
where $\langle\pi,q\rangle=\int_a^b w(x)\pi(x)q(x)dx=0$ leads to a vanishing error $\langle\pi,q\rangle=\int_a^b w(x)e(x)dx=\int_a^b w(x)\pi(x)q(x)dx$, and thus to
$$\int_a^b w(x)g(x)dx=\int_a^b w(x)p_{n-1}(x)dx,$$
which establishes sufficiency. □

As a final and more practical relation one can reuse the result from Theorem 1 and approximate every $g(x)$ of degree $2n-1$ by
$$\int_a^b w(x)g(x)dx=\sum_{i=1}^{n}p_{n-1}(x_i)\omega_i,$$
whenever the nodes $x_1<x_2<\ldots<x_n\in[a,b]$ are the zeros of an orthogonal polynomial $\pi(x)$. The weights ω_i are computed at the respective x_i as before. As opposed to the general case covered by Theorem 1, however, the weights are always positive. This can be shown as follows.

Corollary 2.1. *The interpolation rule from Theorem 2 is exact for all polynomials of degree $2n-1$, so it is also exact for $\left(\mathcal{L}_{n-1,i}(x)\right)^2$, which has degree $2(n-1)<2n-1$.*

Moreover, $\mathcal{L}_{n-1,i}(x_j) = \delta_{ij}$, i.e. zero for all $i \neq j$, and therefore

$$\int_a^b w(x) \left(\mathcal{L}_{n-1,i}(x)\right)^2 dx = \sum_{i=1}^{n} \omega_i \left(\mathcal{L}_{n-1,i}(x)\right)^2 = \omega_i = \int_a^b w(x)\mathcal{L}_{n-1,i}(x)dx.$$

As $w(x) > 0$ by definition, ω_i is also positive.

Truncation Error for General Integrands

In case of a more general $g(x)$ which is not bound to be a polynomial of degree $2n-1$, a general truncation error of $\mid I\left[g\right] - I_n\left[g\right] \mid > 0$ arises. In order to derive a general result for it, first consider *Weierstrass' Approximation Theorem.* It states that for any $\epsilon > 0$ there exists an arbitrary polynomial p_N of degree N for which $\max_{x\in[a,b]} \mid g(x) - p_N(x) \mid \leq \epsilon$ (see Dahlquist and Björck, 2008, for one possible proof). By applying the results derived thus far to $p_N(x)$, one can rewrite the truncation error as

$$\begin{aligned} I\left[g\right] - I_n\left[g\right] =& I\left[g\right] - \int_a^b p_N(x)dx \\ &+ \int_a^b p_N(x)dx - I_n\left[p_N\right] \\ &+ I_n\left[p_N\right] - I_n\left[g\right]. \end{aligned} \tag{4}$$

Without loss of generality, a possible weighting function $w(x)$ is kept implicit for $\int_a^b p_N(x)dx$. For the first two terms in equation (4) it is obvious that

$$\mid I\left[g\right] - \int_a^b p_N(x)dx \mid = \mid \int_a^b \left[g(x) - p_N(x)\right] dx \mid \leq \int_a^b \mid g(x) - p_N(x) \mid dx \leq \epsilon(b-a),$$

where the moduli bars and the the upper bound ϵ are based on $\max_{x\in[a,b]} \mid g(x) - p_N(x) \mid \leq \epsilon$. The first inequality arises as the polynomial $p_N(x)$ may alternatingly lie above or below $g(x)$ for $x \in [a,b]$, whereas the second one derives from the result that $\max_{x\in[a,b]} \mid g(x) - p_N(x) \mid \leq \epsilon$. Multiplication with the range $[a,b]$ simply results from the fact that $\int_a^b dx = b - a$. For the two terms in the second row of equation (4), one has

$$\int_a^b p_N(x)dx - I_n\left[p_N\right] = 0$$

if there are at least n interpolation nodes in $I_n\left[p_N\right]$ such that $p_N(x)$ is exactly approximated for $N = 2n-1$ (or lower). For the last two terms in equation (4), a result very similar to the one above can be found, namely

$$| I_n [p_N] - I_n [g] |=| \sum_{i=1}^{n} [p_N(x_i) - g(x_i)] \omega_i |\leq \sum_{i=1}^{n} | [p_N(x_i) - g(x_i)] || \omega_i |\leq \epsilon \sum_{i=1}^{n} | \omega_i | .$$

As the weights ω_i are always positive by the above corollary, one can extend the modulus beyond the first equation to those that comprise sums over the weights. Otherwise, the inequalities follow the same logic as above. It is trivial to show that $\sum_{i=1}^{n} | \omega_i |$ is equal to the integral $\int_a^b f(x)dx$ with $f(x) \equiv 1$ and thus equal to $b-a$. Combining both inequalities for the overall difference in equation (4) leads to

$$I [g] - I_n [g] \leq 2\epsilon(b - a). \tag{5}$$

As discussed by Dahlquist and Björck (2008), the upper bound $2\epsilon(b - a)$ depends on the smoothness of g, on the narrowness of $b - a$, and on the number of interpolation nodes n. The latter is the relation with the most practical implication that may be written as

$$\lim_{n\to\infty} I_n [g] = I [g] .$$

As such, all well-behaved integrants $g(x)$ that are a polynomial of very high order or that do not belong to any polynomial family at all can be approximated arbitrarily close by choosing a sufficiently large number of interpolation nodes n. As n always remains finite, a so-called truncation error prevails in any case.

Bounds of Integration

For all Gauss quadratures, the choice of the family of node polynomials $\pi(x)$ and the corresponding weight function $w(x)$ predetermine the interval of integration, as specific polynomials are only defined for specific domains $[a, b]$ (see, e.g., Abromowitz and Stegun, 1972, for a general overview). More importantly, it often is the weight function that exhibits conjugate properties with $g(x)$ that constitute a particular convenient choice for the type of the node-polynomial. In order to adapt the bounds of integration accordingly, the following theorem taken from Judd (1998) is highly useful.

Theorem 3 (Change of Variables). *Let $\gamma : \mathbb{R} \to \mathbb{R}$ be monotonically increasing and C^1 on $[a, b]$, then for any integrable $g(x)$ on the same interval*

$$\int_a^b g(x)dx = \int_{\gamma^{-1}(a)}^{\gamma^{-1}(b)} g [\gamma(z)] \gamma'(z)dz.$$

Approximation of $G(u)$

First, recall the integral contained in the representation of the cumulated response probabilities

$$G(u) \propto \int_0^\theta e^{t_1\gamma_1(u)+t_2\gamma_2(u)+t_3\gamma_3(u)}du = \int_0^\theta e^{\Gamma(t,u)}du, \qquad \boxed{6}$$

where $\Gamma(t,u)$ is an implicit expression for the polynomial basis and θ is a placeholder for the corresponding $u = \Phi(\theta)$. For approximating the integral a Gauss-Legendre quadrature can be used, for which the weighting function is $w(x) = 1$ and the domain for the Legendre polynomials is $[-1, 1]$. As opposed to this, $g(x)$ has bounds $a = 0$ and $0 < b \leq 1$, i.e., with respect to Theorem 3 it is required that $\gamma^{-1}(b) \Leftrightarrow z = 1$ and $\gamma^{-1}(a) \Leftrightarrow z = -1$. Many different $\gamma(\cdot)$ can be found to come up with this result, that is why $x = a$ and $x = b$ only have to imply $\gamma^{-1}(a) = z = -1$ and $\gamma^{-1}(b) = z = 1$. A simple candidate would be $\gamma(z) = x = mz + n$. As $x = a$ and $x = b$ has to be fulfilled for $\int_a^b g(x)dx$, one has that $a = -m + n$ and $b = m + n$, which can be rearranged to $m = \frac{b-a}{2}$ and $n = \frac{b+a}{2}$. Thus for the Gauss-Legendre case the bounds of the integral can be changed arbitrarily by

$$\int_a^b g(x)dx = \int_{-1}^1 g\left(\frac{b-a}{2}z + \frac{b+a}{2}\right)\left(\frac{b-a}{2}\right)dz,$$

which changes the integral in the cumulated response probabilities to

$$\int_0^\theta e^{\Gamma(t,u)}du = \frac{\theta}{2}\int_{-1}^1 e^{\Gamma(t,\frac{\theta}{2}u+\frac{\theta}{2})}du \approx \frac{\theta}{2}\sum_{i=1}^n \omega_i e^{\Gamma(t,\frac{\theta}{2}u+\frac{\theta}{2})},$$

with Legendre weights ω_i. As the last relation suggests, the Gauss-Legendre quadrature always remains an approximation to the integral in $G(u)$ since the exponential term is not a polynomial and also cannot be expanded to a finite one. Thus, the optimal choice for the number of interpolation nodes cannot be analytically determined and some error estimates for different n have to be used instead. For this purpose, consider the following error estimates.

$$\text{Relative Error: } RE_n = \frac{\int_0^\theta e^{\Gamma(t,u)}du - \frac{\theta}{2}\sum_{i=1}^n \omega_i e^{\Gamma(t,\frac{\theta}{2}u+\frac{\theta}{2})}}{\int_0^\theta e^{\Gamma(t,u)}du}$$

$$\text{Significant Digits: } SD_n = -\log_{10} 2RE_n$$

As no analytical value for the term $\int_0^\theta e^{\Gamma(t,u)}du$ can be derived, a numerical solution for which the precision can be specified in advance is required. For this reason, the Romberg value (see, e.g., Harris and Stocker, 2006) with 19-digit precision is computed as a reference.

The accuracy results presented in the table below suggest that no less than $n = 21$ nodes should be used in order to yield sufficiently precise approximations for the cumulated response probabilities. However, even for 21 nodes, the accuracy slightly vanishes towards the right tail of the possible interval of integration $[0, 1]$ for some combinations in the parameter space. This also translates into the summary statistics μ and σ, where the average number of significant digits is only 12.7 when the full range of the integral is covered. Moreover, the dispersion of the inaccuracy over the tested parameter grid is quite high. The values for the full range of the integral normalize to that of the other columns just as $n = 31$ nodes are approached.

Accuracy of the Gauss-Legendre Quadrature

	$\mathbf{RE_n}$	$\mathbf{SD_n}$	$\mathbf{RE_n}$	$\mathbf{SD_n}$	$\mathbf{RE_n}$	$\mathbf{SD_n}$	$\mathbf{RE_n}$	$\mathbf{SD_n}$
	$[0, 0.25]$		$[0, 0.5]$		$[0, 0.75]$		$[0, 1.0]$	
	$t_1 = t_2 = t_3 = -3.0$							
$n = 5$	6.774724×10^{-5}	3	2.164433×10^{-3}	2	5.866239×10^{-3}	1	3.363760×10^{-2}	1
$n = 11$	4.316585×10^{-13}	12	5.483942×10^{-9}	7	3.217848×10^{-7}	6	2.295669×10^{-5}	4
$n = 21$	2.427146×10^{-15}	14	1.528022×10^{-15}	14	1.408391×10^{-15}	14	3.461939×10^{-12}	11
$n = 31$	1.493628×10^{-15}	14	8.227811×10^{-16}	14	1.564879×10^{-16}	15	2.457236×10^{-15}	14
	$t_1 = t_2 = t_3 = -1.0$							
$n = 5$	1.962013×10^{-7}	6	1.699031×10^{-5}	4	6.255079×10^{-5}	3	2.873593×10^{-3}	2
$n = 11$	1.554190×10^{-15}	14	2.393438×10^{-13}	12	3.914155×10^{-11}	10	3.993049×10^{-9}	8
$n = 21$	1.695480×10^{-15}	14	2.394935×10^{-15}	14	7.886910×10^{-16}	14	7.198701×10^{-16}	14
$n = 31$	1.271610×10^{-15}	14	2.394935×10^{-15}	14	1.380209×10^{-15}	14	3.599351×10^{-16}	15
	$t_1 = t_2 = t_3 = 1.0$							
$n = 5$	1.976597×10^{-8}	7	8.634096×10^{-6}	4	4.981245×10^{-5}	4	2.023932×10^{-2}	1
$n = 11$	9.892010×10^{-16}	14	1.048849×10^{-14}	13	2.828978×10^{-11}	10	9.439718×10^{-7}	5
$n = 21$	1.648668×10^{-16}	15	3.178330×10^{-16}	15	2.156611×10^{-16}	15	0.000000	–
$n = 31$	8.243342×10^{-16}	14	1.589165×10^{-16}	14	1.078306×10^{-15}	15	3.732239×10^{-15}	14
	$t_1 = t_2 = t_3 = 3.0$							
$n = 5$	9.017437×10^{-7}	5	2.729138×10^{-4}	3	3.065569×10^{-3}	2	4.968473×10^{-1}	0
$n = 11$	2.086081×10^{-15}	14	3.438677×10^{-10}	9	1.262497×10^{-7}	6	3.268156×10^{-3}	3
$n = 21$	1.604677×10^{-16}	15	3.015959×10^{-16}	15	8.496409×10^{-16}	14	4.498005×10^{-9}	8
$n = 31$	9.628064×10^{-16}	14	7.539897×10^{-16}	14	8.496409×10^{-16}	14	7.834999×10^{-16}	14
	$\mu([t_1 : -3.0, -2.5, \ldots, 3.0] \times [t_2 : -3.0, -2.5, \ldots, 3.0] \times [t_3 : -3.0, -2.5, \ldots, 3.0])$							
$n = 5$	2.491678×10^{-4}	5.6	5.803270×10^{-3}	3.4	1.932558×10^{-2}	2.5	6.793724×10^{-2}	1.5
$n = 11$	4.538452×10^{-11}	13.5	2.357598×10^{-7}	10.2	9.242347×10^{-6}	7.7	1.466063×10^{-4}	5.6
$n = 21$	8.724423×10^{-16}	14.9	1.185268×10^{-15}	14.8	2.995938×10^{-13}	14.4	9.629635×10^{11}	12.7
$n = 31$	1.095446×10^{-15}	14.8	1.228426×10^{-15}	14.7	1.383944×10^{-15}	14.7	2.189678×10^{15}	14.6
	$\sigma([t_1 : -3.0, -2.5, \ldots, 3.0] \times [t_2 : -3.0, -2.5, \ldots, 3.0] \times [t_3 : -3.0, -2.5, \ldots, 3.0])$							
$n = 5$	8.447181×10^{-4}	2.1	1.591008×10^{-2}	1.8	4.603207×10^{-2}	1.5	1.011226×10^{-1}	1.3
$n = 11$	4.395189×10^{-10}	1.8	1.179001×10^{-6}	2.7	4.293877×10^{-5}	2.5	4.395189×10^{-4}	2.4
$n = 21$	9.196976×10^{-16}	0.4	1.132619×10^{-15}	0.4	2.162188×10^{-12}	0.9	4.723080×10^{-10}	2.0
$n = 31$	7.535687×10^{-16}	0.3	1.040752×10^{-15}	0.4	1.091805×10^{-15}	0.4	2.346453×10^{-15}	0.5

The lower two panels give the mean and the standard deviation over a selection of possible parameter realizations for t_1, t_2, and t_3. $N = 343$ different parameter permutations over the grid $[t_1 : -3.0, -2.5, \ldots, 3.0] \times [t_2 : -3.0, -2.5, \ldots, 3.0] \times [t_3 : -3.0, -2.5, \ldots, 3.0]$ with 0.5-increments are used.

Approximation of $p(\cdot|\theta)$

For the integral that occurs in the likelihood contribution, steps quite similar to the above ones can be found in order to implement a quadrature approximation. Recall that the respective joint probability is given as

$$p_i(r_{i1}, r_{i2}, \ldots, r_{im}) = \int p_{i1}(r_{i1}|\theta_i)p_{i2}(r_{i2}|\theta_i)\ldots p_{im}(r_{im}|\theta_i)f(\theta_i)d\theta_i,$$

where $f(\theta_i)$ is normal distribution with $\mu = 0$ and $\sigma = 1$. As it turns out, a Gauss-Hermite quadrature with weighting function $w(x) = e^{-x^2}$ and domain $[-\infty, \infty]$ is a natural candidate for these kind of integrals. Since

$$f(\theta_i) = \frac{1}{\sqrt{2\pi}}e^{-\frac{\theta_i^2}{2}},$$

it evident that, apart from the constant $1/\sqrt{2\pi}$, the exponential term is quite similar to the claimed weighting function. Following the notion of Theorem 3, let $z = x/\sqrt{2} \Leftrightarrow x = \sqrt{2}z$. Then, it follows that for the above integral over θ_i

$$\begin{aligned}
p_{ip}(r_{i1}, r_{i2}, \ldots, r_{im}) &= \int_{-\infty}^{\infty} \prod_{j=1}^{m} p_{ij}(r_{ij}|\theta_i)f(\theta_i)d\theta_i \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \prod_{j=1}^{m} p_{ij}(r_{ij}|\theta_i)e^{-\frac{\theta_i^2}{2}}d\theta_i \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{\pi}} \prod_{j=1}^{m} p_{ij}(r_{ij}|\sqrt{2}\theta_i)e^{-\theta_i^2}d\theta_i \\
&\approx \sum_{i=1}^{n} \omega_i \frac{1}{\sqrt{\pi}} \prod_{j=1}^{m} p_{ij}(r_{ij}|\sqrt{2}\theta_i).
\end{aligned}$$

As indicated by the table below, the accuracy results are highly shape-dependent. In particular for those cumulated response curves that exhibit a sharp increase within their respective supports, as those for $t_1 = t_2 = t_3 = -3.0$ and $t_1 = t_2 = t_3 = -1.0$, the relative errors are still substantial under a comparably high number of Hermite nodes. As the corresponding summary statistics suggest, the dispersion of the errors and significant digits becomes acceptable for 21 and 31 nodes already. However, likelihood contributions containing cumulated response curves that obey very quickly changing slopes still remain a problem in those settings. This phenomenon just fades out for 41 Hermite nodes or more. As such, a number n=41 is should be chosen for approximating the outer integral.

Accuracy of the Gauss-Hermite Quadrature

	$n = 5$	
$t_1 = t_2 = t_3 = -3.0$	3.162821×10^{-4}	3
$t_1 = t_2 = t_3 = -1.0$	1.422656×10^{-6}	6
$t_1 = t_2 = t_3 = 1.0$	3.741423×10^{-7}	6
$t_1 = t_2 = t_3 = 3.0$	4.565988×10^{-9}	8
μ	5.553487×10^{-4}	6.5
σ	2.644992×10^{-2}	3.2
	$n = 11$	
$t_1 = t_2 = t_3 = -3.0$	1.962013×10^{-7}	6
$t_1 = t_2 = t_3 = -1.0$	1.554190×10^{-15}	14
$t_1 = t_2 = t_3 = 1.0$	1.695480×10^{-15}	14
$t_1 = t_2 = t_3 = 3.0$	1.271610×10^{-15}	14
μ	5.553487×10^{-5}	6.5
σ	2.644992×10^{-2}	3.9
	$n = 21$	
$t_1 = t_2 = t_3 = -3.0$	2.967753×10^{-7}	6
$t_1 = t_2 = t_3 = -1.0$	5.001314×10^{-13}	12
$t_1 = t_2 = t_3 = 1.0$	2.044795×10^{-14}	13
$t_1 = t_2 = t_3 = 3.0$	1.383461×10^{-15}	15
μ	2.590963×10^{-6}	11.2
σ	1.247417×10^{-3}	2.9
	$n = 31$	
$t_1 = t_2 = t_3 = -3.0$	2.967753×10^{-7}	6
$t_1 = t_2 = t_3 = -1.0$	5.001314×10^{-13}	12
$t_1 = t_2 = t_3 = 1.0$	2.044795×10^{-14}	13
$t_1 = t_2 = t_3 = 3.0$	1.383461×10^{-15}	15
μ	5.991310×10^{-8}	13.5
σ	3.123828×10^{-7}	2.3
	$n = 41$	
$t_1 = t_2 = t_3 = -3.0$	2.771490×10^{-15}	14
$t_1 = t_2 = t_3 = -1.0$	1.097849×10^{-15}	15
$t_1 = t_2 = t_3 = 1.0$	4.682736×10^{-16}	15
$t_1 = t_2 = t_3 = 3.0$	1.729326×10^{-16}	15
μ	6.846350×10^{-10}	13.8
σ	3.387870×10^{-9}	1.9
	$n = 51$	
$t_1 = t_2 = t_3 = -3.0$	3.132988×10^{-15}	14
$t_1 = t_2 = t_3 = -1.0$	1.463799×10^{-15}	15
$t_1 = t_2 = t_3 = 1.0$	7.804559×10^{-16}	15
$t_1 = t_2 = t_3 = 3.0$	1.729326×10^{-16}	16
μ	1.170095×10^{-11}	14.0
σ	5.477883×10^{-11}	1.3

The lower two entries of each panel give the mean and the standard deviation over a selection of possible parameter realizations for t_1, t_2, and t_3. $N = 343$ different parameter permutations over the grid $[t_1 : -3.0, -2.5, \ldots, 3.0] \times [t_2 : -3.0, -2.5, \ldots, 3.0] \times [t_3 : -3.0, -2.5, \ldots, 3.0]$ with 0.5-increments are used.

Appendix B

Identification of the Spady Model

A structure is said to be a predetermined set of hypotheses that are in line with observations. In the parametric case, these hypotheses concern parameters and at least a distributional family. For the item response model, I have stipulated an exponential parameter family that uses a linear series approximation for the actually unknown functional form of the response curves G. Such a linear *sieve* approximation (see Chen, 2007) makes the estimation problem one of the so-called regular parametric cases (see Cramér, 1962), and so is the corresponding model identification. Among other (subsidiary) assumptions, the "regular" case is characterized by.

Assumption 1. *The parameter space for all parameters should be open in* $\mathbb{R}^n$.[7]

Assumption 2. *The sample space of the response data, i.e. the responses, and of the transformed latent traits* u*, for which the* f *is strictly positive, is the same for all* $\mathbf{t}$.

With these presumptions, all a priori admissible structures constitute a model, each contained structure of which is uniquely associated with an observed distribution. Identification of the model concerns the question as to whether the inverse mapping of this association is also one-to-one. As one observes the distribution of responses $f(r, u, \mathbf{t})$ given a presumed distributional family $f(\cdot)$, the parameters (of interest) $\mathbf{t}$, and the nuisance parameter $u \in [0, 1]$, identification is at hand whenever

$$f(r, u, \mathbf{t}_1) = f(r, u, \mathbf{t}_2) \Leftrightarrow \mathbf{t}_1 = \mathbf{t}_2.$$[8]

Local Identification

For the given case, it is meaningful to consider criteria that do not impose to many analytic requirements, as assessing the recursively modeled threshold functions that constitute $f(r, u, \mathbf{t})$ can get quite tedious with increasingly large item scales. Consequently, I first examine local identification (a concept to be formalized below) of the response model as a necessary condition for global identification.

---

[7] This assumption is rather technical as it simply precludes cases along the closure of the parameter space for which the derivative of f is zero just due to its value being zero beyond these coordinates.

[8] Technically, the nuisance does not differ from actual data the distribution is conditional on.

Skrondal and Rabe-Hesketh (2004), in a slight overload of the terminology used in the identification literature, propose an approach that is based on *reduced form parameters* and their relation to the observed distribution. As opposed to the common notation (see, e.g., Intriligator, 1983), reduced form parameters are meant to be (also nonlinear) transformations of the structural parameters that completely characterize the distribution of endogenous model variables $f(r, u, \mathbf{t})$ due to moment conditions $\mathbf{m}(\cdot)$. For categorically distributed response variables, this characterization amounts to first moments only, as

$$f(r, u, \mathbf{t}) = \prod_{j=1}^{m} \prod_{k=1}^{K} p_{jk}(r = k)^{\mathbb{1}[r=k]}, \qquad \text{(7)}$$

such that $E_{jk}(r = k) = p_{jk}(r = k)$, where k is the response for item j. One therefore has to show that the observed expectation conditions can be unequivocally associated with a parameter vector $\mathbf{t}$ given the data r, that means, the moment expression is bijective with respect to its feasible parameter arguments. An intuitive approach is to apply the *implicit function theorem* to the observed moments conditional on the data and an admissible parameter space $\mathscr{A}$. For that purpose, some further definitions are in order.

Definition 1. *Let* $J[\mathbf{m}(\mathbf{t})]$ *the Jacobian matrix based on the moment equations* $\mathbf{m}(\mathbf{t})$ *of the reduced form distribution. Moreover, the matrix elements* $\mathbf{m}(\mathbf{t})$ *are continuous functions of* $\mathbf{t}$ *everywhere in the parameter space* $\mathscr{A}$.

Definition 2. *Suppose* $\mathbf{M}(\mathbf{t})$ *is an arbitrary matrix with elements that are continuous functions of* $\mathbf{t}$ *everywhere in* $\mathscr{A}$. *A point* $\mathbf{t}^0 \in \mathscr{A}$ *is denoted a regular point if there exists an open neighborhood around* $\mathbf{t}^0$ *where* $\mathbf{M}(\mathbf{t})$ *has constant rank.*[9]

Apparently, Wald (1950) first discussed this approach in context of parameter identification, though already under linear independence assumptions for the first moments and therefore with a focus on identifiability with respect to the second moments. However, with the above regularity definitions analogously established by Wald (1950), one can generalize the derivation of the following result to hold for moments of arbitrary order.

Theorem 4. *Given there are* k *different parameters in* $\mathbf{t}$ *(excluding nuisance parameters* u*), then* $\mathbf{t}^0$ *is locally identified iff the rank of* $J[\mathbf{m}(\mathbf{t})]$ *is at least* k *(or in other words,* $J[\mathbf{m}(\mathbf{t})]$ *has full column rank).*

The general proof of the *implicit function theorem* is quite involved. However, for a number of moment equations that generally exceeds the number parameter in $\mathbf{t}$, an alternative one can be provided.

9 The Jacobian matrix and the information matrix are instances of such a matrix M.

Proof. In general (see, e.g., Lang, 1973), the following result holds for continuous multivariate scalar functions $F : O \to \mathbb{R}$ defined on an open set $O \subset \mathbb{R}^n$.

$$F(\mathbf{y}) = F(\mathbf{x}) + \nabla F(\mathbf{x})(\mathbf{y} - \mathbf{x}) + o(\| \mathbf{y} - \mathbf{x} \|), \tag{8}$$

where $\nabla F(\mathbf{x})$ is the gradient of the function and last term is a local approximation error. One shall rewrite the error term as

$$o(\| \mathbf{y} - \mathbf{x} \|) = \psi(\mathbf{y} - \mathbf{x}) = \| \mathbf{y} - \mathbf{x} \| \, g(\mathbf{y} - \mathbf{x}),$$

where $g(\mathbf{y} - \mathbf{x})$ is a mapping that depends on the difference in the two points $\mathbf{y}$ and $\mathbf{x}$, but is not defined for $\mathbf{y} - \mathbf{x} = 0$. One shall state, however, that

$$\lim_{\|\mathbf{y}-\mathbf{x}\| \to 0} \frac{\psi(\mathbf{y} - \mathbf{x})}{\| \mathbf{y} - \mathbf{x} \|} = 0.$$ [10]

These results are preserved for arbitrary mappings $\mathbf{F} : O \to \mathbb{R}^m$ defined on an open set $O \subset \mathbb{R}^n$, as is the case for moment conditions in the above theorem. Equation (8) can be generalized accordingly.

$$\mathbf{m}(\mathbf{t}^1) - \mathbf{m}(\mathbf{t}^0) = J\left[\mathbf{m}(\mathbf{t}^0)\right](\mathbf{t}^1 - \mathbf{t}^0) + \| \mathbf{t}^1 - \mathbf{t}^0 \| \, G(\mathbf{t}^1 - \mathbf{t}^0) \tag{9}$$

Suppose for now, a candidate point $\mathbf{t}^0$ is locally not identifiable in an open neighborhood $O \subset \mathscr{A}$. Then $\mathbf{t}^1 \in O$ implies the same observed moments $\mathbf{m}[\cdot]$ and $\mathbf{m}(\mathbf{t}^1) - \mathbf{m}(\mathbf{t}^0) = 0$. One can thus derive from equation (9) the following quadratic form for the non-identified case.

$$G(\mathbf{t}^1 - \mathbf{t}^0)' G(\mathbf{t}^1 - \mathbf{t}^0) = \frac{(\mathbf{t}^1 - \mathbf{t}^0)'}{\| \mathbf{t}^1 - \mathbf{t}^0 \|} J\left[\mathbf{m}(\mathbf{t}^0)\right]' J\left[\mathbf{m}(\mathbf{t}^0)\right] \frac{\mathbf{t}^1 - \mathbf{t}^0}{\| \mathbf{t}^1 - \mathbf{t}^0 \|} \tag{10}$$

The quadratic form always exist as $J[\mathbf{m}(\mathbf{t}^0)]' J[\mathbf{m}(\mathbf{t}^0)]$ is symmetric. If the distance between the points $\mathbf{t}^0$ and $\mathbf{t}^1$ is arbitrarily decreased, the quadratic form in equation (10) converges to zero. The following lemma helps to relate this result to the Jacobian matrix $J[\mathbf{m}(\mathbf{t}^0)]$.

[10] For a more intuitve understanding of $g(\mathbf{y} - \mathbf{x})$, consider the case of F in one variable, where $F(y) - F(x) = F'(x)(y - x) + |y - x| g(y - x)$. In this instance, $g(y - x)$ is simply the difference between the Newton quotient and the derivative at x, i.e.

$$g(y - x) = \frac{F(y) - F(x)}{y - x} - F'(x).$$

One immediately obtains the result $\lim_{y-x \to 0} g(y - x) = 0$, that extents to all coordinates $\mathbf{y} - \mathbf{x}$ in the multivariate case.

Lemma (Lang (1987)). *Let A be a real symmetric matrix, and let $h(x) = x'Ax$ be the associated continuous quadratic form with $x \in \mathscr{R}^n$. Let P be a point on the unit sphere such that $h(P)$ is a maximum for h on that sphere. Then P is an eigenvector for A and the corresponding largest eigenvalue $\lambda_1 = h(P)$. Moreover, the smallest eigenvalue λ_ℓ corresponds to the minimum of $h(x)$.*[11]

For equation (10), if one defines $A = J\left[\mathbf{m}(\mathbf{t}^0)\right]' J\left[\mathbf{m}(\mathbf{t}^0)\right]$ and $x = (\mathbf{t}^1 - \mathbf{t}^0)/(\| \mathbf{t}^1 - \mathbf{t}^0 \|)$ as a set of points within the unit sphere, Lang's lemma is directly applicable. Since the minimum of $h(x) = x'Ax$ is always equal to the smallest eigenvalue λ_ℓ of A, λ_ℓ converges to zero as $\mathbf{t}^1 - \mathbf{t}^0$ converges to zero. Since the determinant $|A| = \prod_\ell \lambda_\ell$,

$$\lim_{\mathbf{t}^1 - \mathbf{t}^0 \to 0} |J\left[\mathbf{m}(\mathbf{t}^0)\right]' J\left[\mathbf{m}(\mathbf{t}^0)\right]| = 0,$$

which only holds if $J\left[\mathbf{m}(\mathbf{t}^0)\right]$ has deficient column rank. This establishes sufficiency.

By Definition 2 and the underlying assumption with respect to the parameter space, the converse is also true. To show this association, suppose $J\left[\mathbf{m}(\mathbf{t}^0)\right]$ does not have full column rank and therefore there exists an $\mathbf{n}(\mathbf{t})$ for which

$$J\left[\mathbf{m}(\mathbf{t}^0)\right]\mathbf{n}(\mathbf{t}) = \mathbf{0}.$$

Since $\mathbf{t}^0$ is a regular point (with constant rank), $J\left[\mathbf{m}(\mathbf{t}^0)\right]$ as well as $\mathbf{n}(\mathbf{t})$ is continuous in $O \subset \mathscr{A}$. For each coordinate in $\mathbf{t}$ one can define a differential equation with solution $t_i(s)$

$$\frac{\partial t_i(s)}{\partial s} = n_i(\mathbf{t}) \text{ with}$$
$$t_i(0) = t_i^0, \text{ and}$$
$$t_i(S) = t_i^1, \text{ as a terminal condition for } i = 1, \ldots, k.$$

Differentiating the moments $\mathbf{m}(\mathbf{t})$ w.r.t. s by the chain rule gives

$$\frac{\partial \mathbf{m}\left[\mathbf{t}(s)\right]}{\partial s} = J\left[\mathbf{m}(\mathbf{t}^0)\right]\frac{\partial \mathbf{t}(s)}{\partial s} = J\left[\mathbf{m}(\mathbf{t}^0)\right]\mathbf{n}(\mathbf{t}) = 0.$$

This implies that $\mathbf{t}$ is unidentified along all partial trajectories $\mathbf{t}(s)$ showing that deficient column rank in $J\left[\mathbf{m}(\mathbf{t}^0)\right]$ is sufficient for local non-identification. Thus, sufficiency for both associations has been shown. By contraposition, $J\left[\mathbf{m}(\mathbf{t}^0)\right]$ has full rank, iff $\mathbf{t}^0$ is locally identified. This completes the proof. □

[11] As both, x and A are real, and furthermore, A is symmetric, the eigenvalues of A are also real (see Theil, 1983, for a proof). Continuity of $h(x)$ follows from Definition 1.

Local identification everywhere is necessary but not sufficient for global identification (see Parthasarathy, 1983). The real virtue of Theorem 4 is, apart from the rather easily verifiable Jacobian criterion, that for not identified points $\mathbf{t}$ in the moment-systems, one can always find a corresponding basis of a null-space $\mathbf{n}(\mathbf{t})$ satisfying

$$J\left[\mathbf{m}(\mathbf{t})\right]\mathbf{n}(\mathbf{t}) = \mathbf{0},$$

that clearly shows for which parameters linear dependencies are evident in $J\left[\mathbf{m}(\mathbf{t})\right]$ (see Bekker, 1989, for a more specific application). All zero-valued coordinates in $\mathbf{n}(\mathbf{t})$ are partially identified.

For the exponentially tilted series approximation that is used to model the cumulated response curves G, for every item j, exactly one moment equation k depends on a linear mapping in the three Legendre coefficients t, whereas for the remaining cumulated response curves the moments are products in similar linear mappings (recall that this multiplicative recurrence is used for the sake of non-intersection). As to that, two adjacent moments differ in one "layer" of the factorials formed by the G-curves and therefore are linearly independent, except for the two moments that depend on the highest and second highest factorial. In what follows, I apply the above criteria to show conditions for just-identification. Consider the case of $K = 3$ and $J = 2$, i.e., two items in three response categories. Then one obtains a total of four moment conditions in 12 parameters, a measurement system that is clearly under-identified. In general, there are $J \times (k-1)$ (non-redundant) moments to identify $3 \times (K-1) \times J$ parameters. This indeterminacy can only be removed by introducing some variation on the right-hand side of the moment conditions. One may consider covariates that affect the locations of the latent traits that enter u for this purpose (as suggested in Spady, 2007). As an alternative approach, which is somewhat less ad hoc, one may slightly shift the means of the trait nuisances upward as the responses in the other trait specific items increase, and vice versa. This strategy does not violate the local independence assumption between the response probabilities and the latent traits. For the above example with $K = 3$ and $J = 2$ one would only require three out of $3^2 = 9$ possible item combinations in order to render the response model identified, as the Jacobian criterion of Theorem 4 is fulfilled for this number of item combinations and beyond. A minimum of three different item combination holds as a general rule of thumb and should be given for "well behaved" response data.

Global Identification

As already mentioned, global identification is not automatically established by showing local identifiability for every $\mathbf{t}^0 \in \mathscr{A}$. Local identification is nonetheless a necessary condition and furthermore, way more conveniently implementable. In order to establish a global result, first recall that the response probabilities are additive and multiplicative combinations of exponential families. As such, the global identification problem breaks down to the identification of the single contributions to the response probabilities. Denote these contributions to $f(r, u, \mathbf{t})$ as $\tilde{f}(u, \tilde{\mathbf{t}})$, where $\tilde{\mathbf{t}}$ is the parameter triplet that characterizes the respective $G(u)$. The exponential family representation becomes obvious by rearranging the definition of the cumulated response curves

$$G(u) = \frac{\int_0^u e^{t_1\gamma_1(u)+t_2\gamma_2(u)+t_3\gamma_3(u)}du}{\int_0^1 e^{t_1\gamma_1(u)+t_2\gamma_2(u)+t_3\gamma_3(u)}du}$$

as $\tilde{\mathbf{t}}'\boldsymbol{\gamma}(u) = \sum_{i=1}^3 \tilde{t}_i\gamma_i(u)$ and $B(\tilde{\mathbf{t}}) = \log \int_0^1 e^{\tilde{t}_1\gamma_1(u)+\tilde{t}_2\gamma_2(u)+\tilde{t}_3\gamma_3(u)}du$. Then one can rewrite $G(u)$ as a CDF that results from integration over the exponential family density

$$\tilde{f}(u, \tilde{\mathbf{t}}) = \int_0^u e^{\tilde{\mathbf{t}}'\boldsymbol{\gamma}(u)-B(\tilde{\mathbf{t}})}.$$

Thus, $B(\tilde{\mathbf{t}})$ is the so-called cumulant generating function and $\tilde{\mathbf{t}}'\boldsymbol{\gamma}(u)$ is the sufficient statistic.

Belonging to the exponential family greatly alleviates global identification. One shall use a result from Rothenberg (1971) to show this in what follows. First, let $g(u, \tilde{\mathbf{t}}) = \log \tilde{f}(u, \tilde{\mathbf{t}})$ and furthermore define the directional derivative

$$g_i(u, \tilde{\mathbf{t}}) = \frac{\partial g(u, \tilde{\mathbf{t}})}{\partial \tilde{t}_i}$$

and the gradient

$$\nabla g(u, \tilde{\mathbf{t}}) = \frac{\partial g(u, \tilde{\mathbf{t}})}{\partial \tilde{\mathbf{t}}}.$$

Furthermore, let

$$R(\tilde{\mathbf{t}}) = E\left[\frac{\partial g(u, \tilde{\mathbf{t}})}{\partial \tilde{\mathbf{t}}}\frac{\partial g(u, \tilde{\mathbf{t}})'}{\partial \tilde{\mathbf{t}}}\right]$$

be the information matrix. Using the information matrix as a criterion for identification is quite intuitive as it provides a sensitivity measure for changes in $\tilde{f}(u, \tilde{\mathbf{t}})$ given small changes in $\tilde{\mathbf{t}}$. If the structure is the same for some (if not all) parameters, this will result in non-univalence of the information matrix system with respect to parameters $\tilde{\mathbf{t}}$.

Theorem 5 (Rothenberg (1971))**.** *Let $f(u, \tilde{\mathbf{t}})$ be a member of the exponential family. If*

$R(\tilde{\mathbf{t}})$ is nonsingular in a convex set containing $\mathscr{A}$, then every $\tilde{\mathbf{t}}$ in $\mathscr{A}$ is globally identifiable.

Proof. Applying the mean value theorem to $g(u, \tilde{\mathbf{t}})$ for two arbitrary parameter vectors $\tilde{\mathbf{t}}^1$ and $\tilde{\mathbf{t}}^0$ leads to

$$g(u, \tilde{\mathbf{t}}^1) - g(u, \tilde{\mathbf{t}}^0) = \nabla g(u, \tilde{\mathbf{t}}^*)'(\tilde{\mathbf{t}}^1 - \tilde{\mathbf{t}}^0), \tag{11}$$

where $\tilde{\mathbf{t}}^*$ is any $\tilde{\mathbf{t}}$ between $\tilde{\mathbf{t}}^1$ and $\tilde{\mathbf{t}}^0$. If $\tilde{\mathbf{t}}^1$ and $\tilde{\mathbf{t}}^0$ were observationally equivalent, i.e. (globally) unidentified, the difference in equation (11) would be zero, implying that the variance at $\tilde{\mathbf{t}}^*$

$$E\left[\left(\nabla g(u, \tilde{\mathbf{t}}^*)'(\tilde{\mathbf{t}}^1 - \tilde{\mathbf{t}}^0)\right)^2\Big|_{\tilde{\mathbf{t}}^*}\right] = 0.$$

If the vector $\tilde{\mathbf{t}}^*$ implied by the mean value theorem is allowed to be anywhere in the interval $(\tilde{\mathbf{t}}^1, \tilde{\mathbf{t}}^0)$ irrespective of u, one could generally rewrite the above expectation as

$$(\tilde{\mathbf{t}}^1 - \tilde{\mathbf{t}}^0)'R(\tilde{\mathbf{t}}^*)(\tilde{\mathbf{t}}^1 - \tilde{\mathbf{t}}^0) = E\left[\left(\nabla g(u, \tilde{\mathbf{t}}^*)'(\tilde{\mathbf{t}}^1 - \tilde{\mathbf{t}}^0)\right)^2\Big|_{\tilde{\mathbf{t}}^*}\right] = 0. \tag{12}$$

For this case, it is obvious that $R(\tilde{\mathbf{t}}^*)$ has to be singular for Theorem 5 to hold.[12] In general, $\tilde{\mathbf{t}}^*$ will not vary independently of u, unless $\tilde{f}(u, \tilde{\mathbf{t}})$ belongs to the exponential family. One can easily derive this from equation (11), as

$$g(u, \tilde{\mathbf{t}}^1) - g(u, \tilde{\mathbf{t}}^0) = (\tilde{\mathbf{t}}^1 - \tilde{\mathbf{t}}^0)'\boldsymbol{\gamma}(u) - \left[B(\tilde{\mathbf{t}}^1) - B(\tilde{\mathbf{t}}^0)\right]$$

and application of the mean value theorem to the difference in B yields

$$\begin{aligned} 0 &= \left[\boldsymbol{\gamma}(u) - \nabla B(\tilde{\mathbf{t}}^*)\right]'(\tilde{\mathbf{t}}^1 - \tilde{\mathbf{t}}^0) \\ &= \nabla g(u, \tilde{\mathbf{t}}^*)'(\tilde{\mathbf{t}}^1 - \tilde{\mathbf{t}}^0), \end{aligned} \tag{13}$$

which is the separable quadratic form of equation (12). It has been shown that global non-identification implies singularity and, thus, that $R(\tilde{\mathbf{t}}^*)$ being nonsingular is sufficient for global identification. This is the above theorem.

□

[12] Except for the trivial case $\tilde{\mathbf{t}}^1 - \tilde{\mathbf{t}}^0 = 0$.

Resume with the above case of $K = 3$ and $J = 2$ for only three observed item combinations that has been shown to establish local just-identification, numerical evaluation of

$$R(\tilde{\mathbf{t}}) = E\left[\frac{\partial g(u, \tilde{\mathbf{t}})}{\partial \tilde{\mathbf{t}}} \frac{\partial g(u, \tilde{\mathbf{t}})'}{\partial \tilde{\mathbf{t}}}\right]$$

yields that it is nonsingular. Hence, the model is also globally identified.

Appendix C

This Appendix provides a possible implementation of a Fortran 90 `SUBROUTINE` that returns the response-specific probability contributions to the likelihood function

$$
\begin{aligned}
p_i(r_{i1}, r_{i2}, \ldots, r_{iJ}) &= \int p_{ip}(r_{i1}, r_{i2}, \ldots, r_{iJ}|\theta_i) f(\theta_i) d\theta_i \\
&= \int p_{i1}(r_{i1}|\theta_i) p_{i2}(r_{i2}|\theta_i) \ldots p_{iJ}(r_{iJ}|\theta_i) f(\theta_i) d\theta_i.
\end{aligned}
$$

discussed in Chapter 5, augmented by the corresponding Gauss-Hermite nodes. The code presented below is based on GNU Fortran 90, i.e., does not contain any inline directives as would be necessary for Intel or Salford Compilers.[13] It is intended to be compiled as a dynamically linked library (DLL or SO) that can subsequently be linked into R code for the likelihood function.

```
SUBROUTINE cprobs(Xj, m, p, shifts, nrow, Z1,k, Z2, GLw, l, pMat, prow, OutMat)
!****************************************************************************
! Matrix of response combinations, No. of items, mean shifts, rows of Xj, GH
! nodes, No of nodes, GL nodes, GL weights, No of nodes/weights, parameter
! matrix, row number, Return Matrix
!****************************************************************************
USE utils

IMPLICIT NONE
!****************************************************************************
! DIMENSION GLOBALS
!****************************************************************************
INTEGER :: p, k, l, nrow, prow, m
!****************************************************************************
! GLOBAL ARGUMENTS
!****************************************************************************
INTEGER       :: Xj (nrow, p)
REAL (KIND=8) :: shifts( nrow )
REAL (KIND=8) :: Z1 (k)
REAL (KIND=8) :: Z2 (l)
REAL (KIND=8) :: GLw (l)
REAL (KIND=8) :: pMat ( prow ,p)
REAL (KIND=8) :: OutMat (nrow, k)
!****************************************************************************
! AUXILIARY VARIABLES
!****************************************************************************
INTEGER :: nx

REAL (KIND=8) :: resMat (nrow, k)
```

[13] Apart from that, the routine can be compiled by any other compiler.

```
REAL (KIND=8) :: pVec(prow)
REAL (KIND=8) :: Z1Mat (m -1, k)
REAL (KIND=8) :: outV( k )
REAL (KIND=8) :: eps, u
REAL (KIND=8) :: lICC( k), uICC (k)
REAL (KIND=8) :: sZ1(k)

INTEGER :: mj, j, i, a, b

INTEGER, PARAMETER :: po = 3

!******************************************************************************

OutMat = 0.0D+00

eps = (EPSILON(Z1))**(1.0/3.0)


DO j=1, p
        ! find item specific highest response category
    mj = MAXVAL(Xj( :,j))
        ! Put parameter column into vector
    pVec = pMat(: , j)
        ! iterate over rows of X( ,j)
    DO i = 1, nrow
            ! Define number of ICCs necessary
        IF ( Xj(i,j) <= 2) THEN
            nx = mj - 1
        ELSE
            nx = mj - Xj(i,j) + 1
        END IF
            ! Add some mean shift to the GH approximation
        sZ1 = Z1 + shifts(i)
            ! transform to (0,1) by normal CDF; iterate over Z1()
        DO a=1, k
            CALL cumnor(sZ1(a), u)
            sZ1(a)=u
        END DO

            ! Compute the ICCs
        CALL transform( sZ1, Z2, glw, k, l, m, pVec, prow, po, Z1Mat)
            ! ensure openness in (0,1) for cummulated probabilities
        DO a = 1, m

            DO b= 1, k

                IF  (Z1Mat(a,b) - 1 >= 0) THEN

                    Z1Mat(a,b) = Z1Mat(a,b) - eps

                ELSE IF (Z1Mat(a,b) <= 0) THEN
```

```
                Z1Mat(a,b) = eps

            END IF


        END DO

    END DO
        ! Set up the recurrence relation for each ICC
        ! In log.out, lowest ICC P(r <= 1) corresponds
        ! to the first line, whereas P(r <= K) = 1 is the last

    IF (nx == 1) THEN

        Z1Mat = 1 - Z1Mat

        outV = LOG(Z1Mat(nx, :))

    ELSE IF (Xj(i, j) == 1) THEN

        Z1Mat = LOG(Z1Mat)

        outV = SUM( Z1Mat, 1)

    ELSE

            Z1Mat = LOG(Z1Mat)

            DO b= 1,k
                Z1Mat( :, b) = cumsum(Z1Mat( :, b), m)
            END DO
                ! lower ICC
            lICC = Z1Mat( nx, :)
                ! upper ICC
            uICC = Z1Mat(nx -1, : )

            lICC = EXP(lICC)
            uICC = EXP(uICC)

            outV = LOG(uICC - lICC)

    END IF

    resMat( i , : ) = outV

  END DO

  outMat = outMat + resMat

END DO
```

```
RETURN

END SUBROUTINE
```

This main routine, however, depends on a module containing several auxiliary routines that have to be considered in the built process. One of the subroutines is largely derived from William Cody's code from the CDFLIB90 library. Apart from these dependencies, it should be noted that on some systems it may be necessary to specify `-fpic` or `-fPIC` compiler flags for position independent code (at least for GNU compilers). Without this option, some systems may not be able to correctly allocate the memory required at runtime within the global offset table.

```
MODULE utils

!******************************************************************************
!
!! This module contains routines for computation of the Normal CDF

IMPLICIT NONE

CONTAINS

!******************************************************************************
SUBROUTINE swap ( x, y )

    IMPLICIT NONE

    REAL ( KIND = 8 ) :: x
    REAL ( KIND = 8 ) :: y
    REAL ( KIND = 8 ) :: z

    z = x
    x = y
    y = z

    RETURN

END SUBROUTINE

!******************************************************************************
SUBROUTINE cumnor ( arg, cum )

!******************************************************************************
!
!! This routine builds William Cody's implementation from the CDFLIB90 and
!  computes the cumulative normal distribution by rational
!  function approximations.
!
!
```

```
!  Parameters:
!
!    Input, real ( kind = 8 ) ARG, the upper limit of integration.
!
!    Output, real ( kind = 8 ) CUM, CCUM, the Normal density CDF and
!    complementary CDF.
!
!*****************************************************************************
  IMPLICIT NONE

!*****************************************************************************
! ARGUMENTS
!*****************************************************************************

  REAL ( KIND = 8 ) :: arg
  REAL ( KIND = 8 ) :: cum

!*****************************************************************************
! CONSTANTS
!*****************************************************************************

  REAL ( KIND = 8 ), PARAMETER, DIMENSION ( 5 ) :: a = (/ &
    2.2352520354606839287D+00, &
    1.6102823106855587881D+02, &
    1.0676894854603709582D+03, &
    1.8154981253343561249D+04, &
    6.5682337918207449113D-02 /)

  REAL ( KIND = 8 ), PARAMETER, DIMENSION ( 4 ) :: b = (/ &
    4.7202581904688241870D+01, &
    9.7609855173777669322D+02, &
    1.0260932208618978205D+04, &
    4.5507789335026729956D+04 /)

  REAL ( KIND = 8 ), PARAMETER, DIMENSION ( 9 ) :: c = (/ &
    3.9894151208813466764D-01, &
    8.8831497943883759412D+00, &
    9.3506656132177855979D+01, &
    5.9727027639480026226D+02, &
    2.4945375852903726711D+03, &
    6.8481904505362823326D+03, &
    1.1602651437647350124D+04, &
    9.8427148383839780218D+03, &
    1.0765576773720192317D-08 /)

  REAL ( KIND = 8 ), PARAMETER, DIMENSION ( 8 ) :: d = (/ &
    2.2266688044328115691D+01, &
    2.3538790178262499861D+02, &
    1.5193775994075548050D+03, &
    6.4855582982667607550D+03, &
    1.8615571640885098091D+04, &
```

```
    3.4900952721145977266D+04, &
    3.8912003286093271411D+04, &
    1.9685429676859990727D+04 /)

  REAL ( KIND = 8 ), PARAMETER, DIMENSION ( 6 ) :: p = (/ &
    2.1589853405795699D-01, &
    1.274011611602473639D-01, &
    2.2235277870649807D-02, &
    1.421619193227893466D-03, &
    2.9112874951168792D-05, &
    2.307344176494017303D-02 /)

  REAL ( KIND = 8 ), PARAMETER, DIMENSION ( 5 ) :: q = (/ &
    1.28426009614491121D+00, &
    4.68238212480865118D-01, &
    6.59881378689285515D-02, &
    3.78239633202758244D-03, &
    7.29751555083966205D-05 /)

  REAL ( KIND = 8 ), PARAMETER :: root32 = 5.656854248D+00
  REAL ( KIND = 8 ), PARAMETER :: sixten = 16.0D+00

  REAL ( KIND = 8 ), PARAMETER :: sqrpi = 3.9894228040143267794D-01
  REAL ( KIND = 8 ), PARAMETER :: thrsh = 0.66291D+00

!******************************************************************************
! AUXILIARY VARIABLES
!******************************************************************************

  REAL ( KIND = 8 ) :: ccum
  REAL ( KIND = 8 ) :: temp

  REAL ( KIND = 8 ) :: del
  REAL ( KIND = 8 ) :: eps
  INTEGER ( KIND = 4 ) :: i

  REAL ( KIND = 8 ) :: x
  REAL ( KIND = 8 ) :: xden
  REAL ( KIND = 8 ) :: xnum
  REAL ( KIND = 8 ) :: y
  REAL ( KIND = 8 ) :: xsq
!
!  Machine dependent constants
!
  eps = EPSILON ( 1.0D+00 ) * 0.5D+00

  x = arg
  y = ABS ( x )

  IF ( y <= thrsh ) THEN
!
```

```
!  Evaluate  anorm  for  |X| <= 0.66291
!
    IF ( eps < y ) THEN
      xsq = x * x
    ELSE
      xsq = 0.0D+00
    END IF

    xnum = a(5) * xsq
    xden = xsq

    DO i = 1, 3
      xnum = ( xnum + a(i) ) * xsq
      xden = ( xden + b(i) ) * xsq
    END DO

    cum = x * ( xnum + a(4) ) / ( xden + b(4) )
    temp = cum
    cum = 0.5D+00 + temp
    ccum = 0.5D+00 - temp
!
!  Evaluate ANORM for 0.66291 <= |X| <= sqrt(32)
!
  ELSE IF ( y <= root32 ) THEN

    xnum = c(9) * y
    xden = y

    DO i = 1, 7
      xnum = ( xnum + c(i) ) * y
      xden = ( xden + d(i) ) * y
    END DO

    cum = ( xnum + c(8) ) / ( xden + d(8) )
    xsq = AINT ( y * sixten ) / sixten
    del = ( y - xsq ) * ( y + xsq )
    cum = EXP ( - xsq * xsq * 0.5D+00 ) * EXP ( -del * 0.5D+00 ) * cum
    ccum = 1.0D+00 - cum

    IF ( 0.0D+00 < x ) THEN
      CALL swap ( cum, ccum )
    END IF
!
!  Evaluate ANORM for sqrt(32) < |X|.
!
  ELSE

    cum = 0.0D+00
    xsq = 1.0D+00 / ( x * x )
    xnum = p(6) * xsq
    xden = xsq
```

```
    DO i = 1, 4
      xnum = ( xnum + p(i) ) * xsq
      xden = ( xden + q(i) ) * xsq
    END DO

    cum = xsq * ( xnum + p(5) ) / ( xden + q(5) )
    cum = ( sqrpi - cum ) / y
    xsq = AINT ( x * sixten ) / sixten
    del = ( x - xsq ) * ( x + xsq )
    cum = EXP ( - xsq * xsq * 0.5D+00 ) &
      * EXP ( - del * 0.5D+00 ) * cum
    ccum = 1.0D+00 - cum

    IF ( 0.0D+00 < x ) THEN
      CALL swap ( cum, ccum )
    END IF

  END IF

  IF ( cum < TINY ( cum ) ) THEN
    cum = 0.0D+00
  END IF

  IF ( ccum < TINY ( ccum ) ) THEN
    ccum = 0.0D+00
  END IF

  RETURN

END SUBROUTINE

!*****************************************************************************
SUBROUTINE transform( uZ1, Z2, GLw, k, l, m, pVec, prow, po, Z1Mat)
!*****************************************************************************
!
!
!
!*****************************************************************************
    IMPLICIT NONE

!*****************************************************************************
! ARGUMENTS
!*****************************************************************************
    INTEGER :: k, l, prow, po
    INTEGER :: m

    REAL (KIND=8) :: uZ1 ( k )
    REAL (KIND=8) :: Z2 ( l )
    REAL (KIND=8) :: GLw ( l )
    REAL (KIND=8) :: pVec ( prow )
```

```
    REAL (KIND=8) :: Z1Mat ( m-1, k )

!******************************************************************************
! AUXILIARY VARIABLES
!******************************************************************************
    REAL (KIND=8) :: COV
    REAL (KIND=8) :: PolVec (3)
    REAL (KIND=8) :: jMat ( m -1, l )
    REAL (KIND=8) :: tempVec ( m -1 )
    REAL (KIND=8) :: tempMat ( m-1 , k )
    REAL (KIND=8) :: T (po, m-1)

    INTEGER :: i, j

    T = RESHAPE(pVec, (/po, m - 1 /))

        ! Build the integrals for the numerator
    DO i=1, k

        DO j=1, l

              !Change of Variables procedure
            COV = 0.5 * uZ1(i) * Z2(j) + 0.5 * uZ1(i)
              ! Set up the Legendre Polysnomials
            PolVec(1) = 2 * COV -1
            PolVec(2) = 6 * COV**2 - 6 * COV + 1
            PolVec(3) = 20* COV**3 - 30 * COV**2 + 12 *COV -1

            jMat( : , j) = 0.5 * uZ1(i) * EXP( MATMUL(PolVec, T) )

        END DO

        Z1Mat( : , i) = MATMUL( jMat, GLw)

    END DO

        ! Build the normalizing integrals of the denominator
    DO j=1, l

        COV =  0.5 * Z2(j) + 0.5

           ! Set up the Legendre Polysnomials
        PolVec(1) = 2*COV -1
        PolVec(2) = 6*COV**2 -6*COV + 1
        PolVec(3) = 20*COV**3 -30*COV**2 + 12*COV -1

        jMat( : ,j) = 0.5 * EXP( MATMUL( PolVec, T))

    END DO

        ! Complete the integral
```

```
    tempVec = MATMUL(jMat, GLw)

    tempMat = SPREAD(tempVec , 2, k)
        ! Normalize the numerators
    Z1Mat = 1- (Z1Mat / tempMat)

END SUBROUTINE
!******************************************************************************
FUNCTION cumsum(arr, i) RESULT(ans)
!******************************************************************************
    INTEGER :: i

    REAL(KIND=8), DIMENSION(i), INTENT(IN) :: arr
    REAL(KIND=8), DIMENSION(i) :: ans
    INTEGER :: j

    ans(1)=arr(1)

    DO j=2,i
        ans(j)=ans(j-1)+arr(j)
    END DO

END FUNCTION cumsum

END MODULE
```

Two different interfaces may be employed to dynamically link the above code into R. The first one, R's `.Fortran` interface, uses a pass-by-reference scheme that requires the complete list of arguments `Xj,..., OutMat` as indicated in the above header for `cprobs`. Alongside the actual arguments, this list also comprises scalar arguments passing the dimensions of the arrays involved, as other than native Fortran the interface does not support assumed-shape arguments. An alternative is the second possible implementation via R's `.Call` interface. It takes advantage of the fact that the R internals are written in C and provides full access to them on the C-side. Using the R-headers `R.h`, `Rinternals.h`, and `Rmath.h`, one may call a C wrapper function that does all the arrangements before passing everything to the Fortran subroutine `cprobs`. This could be done as follows.

```
#include<R.h>
#include<Rinternals.h>
#include<Rmath.h>

void cprobs_(double *, int *, int *, double *, int *, ..., double *);

#cprobs(Xj, m, p, shifts, nrow, Z1,k, Z2, GLw, l, pMat, prow, OutMat)
#*****************************************************************************
# Matrix of response combinations, No. of items, mean shifts, rows of Xj, GH
# nodes, No of nodes, GL nodes, GL weights, No of nodes/weights, parameter
# matrix, row number, Return Matrix
#*****************************************************************************

SEXP C_wrapper(SEXP Xj, SEXP shifts, SEXP Z1, SEXP Z2, SEXP GLw, SEXP pMat )
{
    int m, p, nrow, k, l, prow;
    int Rdim[2];
    SEXP OutMat;


    PROTECT(Xj = AS_NUMERIC(Xj));
    PROTECT(shifts = AS_NUMERIC(shifts));
    PROTECT(Z1 = AS_NUMERIC(Z1));
    PROTECT(Z2 = AS_NUMERIC(Z2));
    PROTECT(GLw = AS_NUMERIC(GLw));
    PROTECT(pMat = AS_NUMERIC(pMat));

    Rdim= getAttrib(Xj,R_DimSymbol);
    m= INTEGER(Rdim)[0];
    p= INTEGER(Rdim)[1];

    nrow = LENGTH(shifts);
    k = LENGTH(Z1);
    l = LENGTH(Z2);
    Rdim= getAttrib(pMat,R_DimSymbol);
    prow = INTEGER(pMat)[0];

    cprobs_(&Xj, &m, &p, &shifts, &nrow, &Z1,
            &k, &Z2, &GLw, &l, &pMat, &prow, &OutMat);
    UNPROTECT(6);
    return(OutMat);
}
```

Appendix D

Personality Tests included in the Questionnaire of Chapter 8

Big Five

I see myself as someone who...

1. is original, comes up with new ideas (Openness to Experience).
2. values artistic experiences (Openness to Experience).
3. has an active imagination (Openness to Experience).
4. does a thorough job (Conscientiousness).
5. does things effectively and efficiently (Conscientiousness).
6. tends to be lazy (Conscientiousness, reversed).
7. is communicative, talkative (Extraversion).
8. is outgoing, sociable (Extraversion).
9. is reserved (Extraversion, reversed).
10. is sometimes somewhat rude to others (Agreeableness, reversed).
11. has a forgiving nature (Agreeableness).
12. is considerate and kind to others (Agreeableness).
13. worries a lot (Neuroticism).
14. gets nervous easily (Neuroticism).
15. is relaxed, handles stress well (Neuroticism, reversed).

Locus of Control (LOC)

Using the scale provided, indicate what your attitudes towards life and towards your own future are.

1. How my life goes depends on me (Internal LOC, *discarded*).
2. If a person is socially or politically active, he/she can have an effect on social conditions (Internal LOC, *discarded*).
3. One has to work hard in order to succeed (Internal LOC, *discarded*).
4. If I run up against difficulties in life, I often doubt my own abilities (reversed, Internal LOC, *discarded*).
5. Compared to other people, I have not achieved what I deserve (External LOC).
6. What a person achieves in life is above all a question of fate or luck (External LOC).
7. I frequently have the experience that other people have a controlling influence over my life (External LOC).
8. The opportunities that I have in life are determined by the social conditions (External LOC).
9. Inborn abilities are more important than any efforts one can make (External LOC).
10. I have little control over the things that happen in my life (External LOC).

Brief Self-Control Scale

Using the scale provided, please indicate how much each of the following statements reflects how you typically are.

1. I am good at resisting temptation.
2. I have a hard time breaking bad habits (reversed).
3. I say inappropriate things (reversed).
4. I do certain things that are bad for me, if they are fun (reversed).
5. I refuse things that are bad for me.
6. I wish I had more self-discipline (reversed).
7. People would say that I have iron self-discipline.
8. Pleasure and fun sometimes keep me from getting work done (reversed).
9. I have trouble concentrating (reversed).
10. I am able to work effectively toward long-term goals.
11. Sometimes I can't stop myself from doing something, even if I know it is wrong (reversed).

Personality traits as surveyed in the GSOEP for Chapter 9

Perceived Control/ Locus of Control (LOC), as of 1999

Using the scale provided, indicate what your attitudes towards life and towards your own future are.

1. How my life goes depends on me (Internal LOC).
2. If a person is socially or politically active, he/she can have an effect on social conditions (Internal LOC).
3. One has to work hard in order to succeed (Internal LOC).
4. If I run up against difficulties in life, I often doubt my own abilities (reversed, Internal LOC).
5. Compared to other people, I have not achieved what I deserve (External LOC).
6. What a person achieves in life is above all a question of fate or luck (External LOC).
7. I frequently have the experience that other people have a controlling influence over my life (External LOC).
8. The opportunities that I have in life are determined by the social conditions (External LOC).
9. Inborn abilities are more important than any efforts one can make (External LOC).
10. I have little control over the things that happen in my life (External LOC).

Perceived Control, as of 1994

The following are various attitudes towards life and the future. Please indicate what most applies to you.

1. I determine what happens to me in life (Internal).
2. It is useless to make plans because they seldom work out (External).
3. My behavior determines my life (Internal).
4. No one can escape their fate, everything in life happens as it must happen (External).
5. If I get something I want then it's mostly due to luck (External).
6. Most plans I make are successful (Internal).
7. There is little sense in planing ahead because something unexpected always comes up (External).
8. Things always happen differently, one can't rely on anything (External).

Bibliography

Aassve, A., S. Burgess, A. Chesher, and C. Propper (2002): "Transitions From Home to Marriage of Young Americans," *Journal of Applied Econometrics*, 17(1), 1–23.

Aassve, A., S. Burgess, M. Dickson, and C. Propper (2006a): "Modelling Poverty by not Modelling Poverty: An Application of a Simultaneous Hazard Approach to the UK," CASEpaper 106, Centre for Analysis of Social Exclusion - London School of Economics, London.

Aassve, A., S. Burgess, C. Propper, and M. Dickson (2006b): "Employment, Family Union and Childbearing Decisions in Great Britain," *Journal of the Royal Statistical Society Series A*, 169(4), 781–804.

Abromowitz, M., and I. A. Stegun (1972): *Handbook of Mathematical Functions*. National Bureau of Standards - Applied Mathematics Series, Washington D.C.

Aigner, D. J., C. Hsiao, A. Kapteyn, and T. Wansbeek (1984): "Latent Variables Models in Econometrics," in *Handbook of Econometrics*, ed. by Z. Griliches, and M. D. Intriligator, vol. 2, chap. 23, pp. 1321–1393. Elsevier B.V., Amsterdam, North Holland.

Ainslie, G. (1975): "Specious Reward: A Behavioral Theory of Impulsiveness and Impulse Control," *Psychological Bulletin*, 82(4), 463–496.

——— (1991): "Derivation of "Rational" Economic Behavior from Hyperbolic Discount Curves," *American Economic Review*, 81(2), 334–340.

Allport, G. W. (1937): *Personality: A Psychogical Interpretation*. Holt, Rhinehart and Winston, New York.

——— (1961): *Pattern and Growth in Personality*. Holt, Rinehart and Winston, New York, NY.

Allport, G. W., and H. Odbert (1936): "Trait-Names, a Psycholexical Study," *Psychological Monographs*, 47(211).

Almlund, M., A. L. Duckworth, J. Heckman, and T. Kautz (2011): "Personality Psychology and Economics," in *Handbook of the Economics of Education*, ed. by E. A. Hanushek, S. J. Machin, and L. Woessmann, vol. 4, chap. 1, pp. 1–158. Elsevier B.V., Amsterdam: North-Holland.

Amemiya, T. (1985): *Advanced Econometrics*. Harvard University Press, Cambridge, MA.

Andersen, E. B. (1972): "The Numerical Solution of a Set of Conditional Estimation Equations," *Journal of the Royal Statistical Society, Series B*, 34(1), 42–54.

Anderson, T. W. (2003): *An Introduction to Multivariate Statistical Analysis*. John Wiley & Sons, Hoboken, NJ.

Anderson, T. W., and N. Kunitomo (1992): "Tests of Overidentification and Predeterminedness in Simultaneous Equation Models," *Journal of Econometrics*, 54, 49–78.

Anderson, T. W., and H. Rubin (1956): "Statistical Inference in Factor Analysis," in *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, ed. by J. Neyman, pp. 111–150. Berkeley - University of California Press, Berkeley, California.

Andrews, S. R., J. B. Blumenthal, D. L. Johnson, C. J. Ferguson, T. M. Lasater, P. E. Malone, and D. B. Wallace (1982): "The Skills of Mothering: A Study of Parent Child Development Centers," *Monographs of the Society for Research in Child Development*, 47(6), 1–83.

Antecol, H., and D. A. Cobb-Clark (2010): "Do Non-Cognitive Skills Help Explain the Occupational Segregation of Young People?," IZA Discussion Paper Series 5093, Institute for the Study of Labor (IZA), Bonn.

Appadurai, A. (2004): "The Capacity to Aspire," in *Culture and Public Action*, ed. by V. Rao, and M. Walton, pp. 59–84. The International Bank for Reconstruction and Development, The World Bank, Washington, DC.

Arellano, M., and B. Honoré (2001): "Panel Data Models: Some Recent Developments," in *Handbook of Econometrics*, ed. by J. J. Heckman, and E. E. Leamer, vol. 5, chap. 76, pp. 3229–3296. Elsevier B.V., Amsterdam, North Holland.

Arulampalam, W., A. L. Booth, and M. P. Taylor (2000): "Unemployment Persistence," *Oxford Economic Papers*, 52(1), 24–50.

Atkinson, A. B. (1970): "On the Measurement of Inequality," *Journal of Economic Theory*, 2(3), 244–263.

Atkinson, A. B., L. Rainwater, and T. M. Smeeding (1995): *Income Distribution in OECD Countries*. OECD, Paris.

Baker, F. B., and S.-H. Kim (2004): *Item Response Theory*. Dekker Media, New York.

Baltagi, B., and S. Song (2006): "Unbalanced Panel Data: A Survey," *Statistical Papers*, 47(4), 493–523.

Bandura, A. (1986): *Social Foundations of Thought and Action: A Social Cognitive Theory*. Prentice Hall, New Jersey.

Barron, A. R., and C.-H. Sheu (1991): "Approximation of Density Functions by Sequences of Exponential Families," *The Annals of Statistics*, 19(3), 1347–1369.

Becker, A., T. Deckers, T. Dohmen, A. Falk, and F. Kosse (2012): "The Relationship Between Economic Preferences and Psychological Personality Measures," *Annual Review of Economics*, 4(1), 453–478.

Becker, G. S. (1964): *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education*, The University of Chicago Press. National Bureau of Economic Research, New York, Third Edition from 1993.

——— (1993): *A Treatise on the Family*. Harvard University Press.

Bekker, P. A. (1989): "Identifiability in Restricted Factor Models and the Evaluation of Rank Conditions," *Journal of Econometrics*, 41, 5–16.

Ben-Porath, Y. (1967): "The Production of Human Capital and the Life Cycle of Earnings," *The Journal of Political Economy*, 75(4), 352–365.

Bernheim, B. D., D. Ray, and S. Yeltikin (2013): "Poverty and Self-Control," NBER Working Papers 18742, National Bureau of Economic Research (NBER), Cambridge, MA.

Biewen, M. (2004): "Measuring State Dependence in the Individual Poverty Status: Are There Feedback Effects to Employment Decisions and Household Composition?," IZA Discussion Paper Series 1138, Institute for the Study of Labor (IZA), Bonn.

Birnbaum, A. (1968): "Some Latent Trait Models and their Use in Infering an Exmaminee's Ability," in *Statistical Theories of Mental Test Scores*, ed. by F. M. Lord, and M. R. Novick, pp. 397–479. Addison-Wesley, Reading, MA.

Blackburn, M., and D. Neumark (1992): "Unobserved Ability, Efficiency Wages, and Interindustry Wage Differentials," *Quarterly Journal of Economics*, 107(4), 1421–1436.

Blomeyer, D., K. Coneus, M. Laucht, and F. Pfeiffer (2009): "Initial Risk Matrix, Home Resources, Ability Development and Children's Achievement," *Journal of the European Economic Association*, 7(2-3), 638–648.

Bock, R. R., and M. Aitken (1981): "Marginal Maximum Likelihood Estimation of Item Parameters: Application of an EM Algorithm," *Psychometrika*, 46(4), 443–459.

Bollen, K. A. (1989): *Structural Equations with Latent Variables*. John Wiley & Sons, New York, NY.

Borghans, L., A. L. Duckworth, J. Heckman, and B. ter Weel (2008a): "The Economics and Psychology of Personality Traits," *Journal of Human Resources*, 43(4), 972–1059.

Borghans, L., B. Golsteyn, J. J. Heckman, and J. E. Humphries (2011): "Identification Problems in Personality Psychology," *Personality and Individual Differences*, 51, 315–320.

Borghans, L., H. Meijers, and B. ter Weel (2008b): "The Role of Noncognitive Skills in Explaining Cognitive Test Scores," *Economic Inquiry*, 46(1), 2–12.

Borghans, L., B. ter Weel, and B. A. Weinberg (2008c): "Interpersonal Styles and Labor Market Outcomes," *Journal of Human Resources*, 43(4), 815–858.

Bouchard, T. J., and J. C. Loehlin (2001): "Genes, Evolution, and Personality," *Behavioral Genetics*, 31(3), 243–273.

Bound, J., Z. Griliches, and B. H. Hall (1986): "Wages, Schooling and IQ of Brothers and Sisters: Do the Family Factors Differ?," *International Economic Review*, 27(1), 77–105.

Bowles, S., H. Gintis, and M. Osborne (2001): "The Determinants of Earnings: A Behavioral Approach," *Journal of Economic Literature*, 38(4), 1137–1176.

Braakmann, N. (2009): "The Role of Psychological Traits for the Gender Gap in Full-Time Employment and Wages: Evidence from Germany," DIW Discussion Papers 162, German Institute for Economic Research (DIW), Berlin.

Browning, M., P.-A. Chiappori, and Y. Weiss (2011): *Economics of the Family*. Cambridge University Press, Cambridge.

Büttner, B., and S. L. Thomsen (forthcoming): "Are We Spending too Many Years in School? Causal Evidence of the Impact of Shortening Secondary School Duration," forthcoming in German Economic Review, `DOI:10.1111/geer.12038`.

Burdett, K., and M. G. Coles (1999): "Long-Term Partnership Formation: Marriage and Employment," *Economic Journal*, 109(456), F307–34.

Buss, A. H., and R. Plomin (1975): *A Temperament Theory of Personality Development*. John Wiley & Sons, New York.

Butler, J. S., and R. Moffitt (1982): "A Computationally Efficient Quadrature Procedure for the One-Factor Multinomial Probit Model," *Econometrica*, 50(3), 761–764.

Campbell, D. T., and D. W. Fiske (1959): "Convergent and Discriminant Validation by the Multitrait-Multimethod Matrix," *Psychological Bulletin*, 56(2), 81–105.

Campbell, F. A., and C. T. Ramey (1994): "Effects of Early Intervention on Intellectual and Academic Achievement: A Follow-Up Study of Children from Low-Income Families," *Child Development*, 65, 684–698.

Canli, T. (2006): *Biology of Personality and Individual Differences*. Guilford Press, New York.

CARD, D. (1999): "The Causel Effects of Education on Earnings," in *Handbook of Labor Economics*, ed. by O. Ashenfelter, R. Layard, and D. Card, vol. 3, chap. 30, pp. 1801–1863. Elsevier B.V., Amsterdam: North-Holland.

CARNEIRO, P., C. CRAWFORD, AND A. GOODMAN (2007): "The Impact of Early Cognitive and Non-Cognitive Skills on Later Outcomes," CEE Discussion Papers 0092, Centre for the Economics of Education at the London School of Economics (CEE), London.

CARNEIRO, P., K. HANSEN, AND J. HECKMAN (2003): "Estimating Distributions of the Treatment Effects with an Application to the Returns to Schooling and Measurement of the Effects of Uncertainty on College Choice," *International Economic Review*, 44(2), 361–421.

CASPI, A. (2000): "The Child Is Father of the Man: Personality Continuities From Childhood to Adulthood," *Journal of Child Psychology and Psychiatry*, 78(1), 158–172.

CASPI, A., J. MCCLAY, T. E. MOFFITT, J. MILL, J. MARTIN, I. W. CRAIG, A. TAYLOR, AND R. POULTON (2002): "Role of Genotype in Cycle of Violence Maltreated Children," *Science*, 297, 851–853.

CASPI, A., B. W. ROBERTS, AND R. L. SHINER (2005): "Personality Development - Stability and Change," *Annual Review of Psychology*, 56(1), 453–484.

CATTELL, R. (1971): *Abilities: Their Structure, Growth and Action.* Houghton Mifflin, Boston.

CEDAR, H., AND Y. BERGMAN (2012): "Programming of DNA Methylation Patterns," *Annual Review of Biochemistry*, 81, 97–117.

CHAMBERLAIN, G. (1982a): "The General Equvalence of Granger and Sims Causality," *Econometrica*, 50(3), 569–581.

——— (1982b): "Multivariate Regression Models for Panel Data," *Journal of Econometrics*, 18(1), 5–46.

CHEN, X. (2007): "Large Sample Sieve Estimation of Semi-Nonparametric Models," in *Handbook of Econometrics*, ed. by J. J. Heckman, and E. E. Leamer, vol. 6B, chap. 76, pp. 5549–5632. Elsevier B.V., Amsterdam, North Holland.

COASE, R. H. (1937): "The Nature of the Firm," *Economica*, 4(16), 386–405.

COBB-CLARK, D. A., AND S. SCHURER (2013): "Two Economists' Mursings on the Stability of Locus of Control," *Economic Journal*, 123(8), 358–400.

CONEUS, K., J. GERNANDT, AND M. SAAM (2011): "Noncognitive Skills, Internet Use and Educational Dropout," *Schmollers Jahrbuch - Journal of Applied Social Science Studies*, 131(4), 547–568.

Coneus, K., and M. Laucht (2011): "The Effect of Early Noncognitive Skills on Social Outcomes in Adolescence," *Education Economics*, iFirst, http://www.tandfonline.com/doi/pdf/10.1080/09645292.2010.547720.

Coneus, K., and F. Pfeiffer (2007): "Self-Productivity in Early Childhood," ZEW Discussion Paper 053, Centre for European Economic Research (ZEW), Mannheim.

Conti, G., S. Frühwirth-Schnatter, J. Heckman, and R. Piatek (2014): "Bayesian Exploratory Factor Analysis," *Journal of Econometrics*, 183(1), 31–57.

Costa, P., and R. McCrae (1992): *Revised Neo Personality Inventory (NEO PI-R) and Neo Five-Factor (NEO-FFI)*. Psychological Assessment Resources, Odessa, FL.

Costa, P. T., and R. R. McCrae (2008): "Revised NEO Personality Inventory (NEO PI-R) and NEO Five-Factor Inventory (NEO-FFI)," in *The SAGE Handbook of Personality Theory and Assessment: Personality Measurement and Testing*, ed. by G. J. Boyle, G. Matthews, and D. H. Saklofske, vol. 2, pp. 179–196. SAGE Publications Ltd., Thousand Oaks, California.

Costello, A. B., and J. W. Osborne (2005): "Best Practices in Exploratory Factor Analysis: Four Recommendations for Getting the Most From Your Analysis," *Practical Assessment, Research & Evaluation*, 10(7).

Cowell, F. A. (2011): *Measuring Inequality*. Oxford University Press, Oxford, NY.

Cramér, H. (1962): *Mathematical Methods of Statistics*. GLS Press, Bombay.

Cronbach, L. J. (1951): "Coefficient Alpha and the Internal Structure of Tests," *Psychometrika*, 16(3), 297–334.

Cunha, F., and J. Heckman (2006): "Investing in our Young People," unpublished, available at http://www-news.uchicago.edu/releases/06/061115.education.pdf.

——— (2007): "The Technology of Skill Formation," *American Economic Review*, 97(2), 31–47.

——— (2008): "Formulating, Identifying and Estimating the Technology of Cognitive and Noncognitive Skill Formation," *Journal of Human Resources*, 43(4), 738–782.

——— (2009): "The Economics and Psychology of Inequality and Human Development," *Journal of the European Economic Association*, 7(2/3), 320–364.

Cunha, F., J. Heckman, L. Lochner, and D. Masterov (2006): "Interpreting the Evidence on Life Cycle Skill Formation," in *Handbook of the Economics of Education*, ed. by E. A. Hanushek, and F. Welch, vol. 1, chap. 12, pp. 697–812. Elsevier B.V., Amsterdam: North-Holland.

Cunha, F., J. Heckman, and S. Schennach (2010): "Estimating the Technology of Cognitive and Noncognitive Skill Formation," *Econometrica*, 78(3), 883–991.

Currie, J. (2009): "Healthy, Wealthy, and Wise: Socioeconomic Status, Poor Health in Childhood, and Human Capital Development," *Journal of Economic Literature*, 47(1), 87–122.

Currie, J., and D. Thomas (1995): "Does Head Start Make a Difference?," *American Economic Review*, 85(3), 341–364.

Dagsvik, J. K., S. Strom, and Z. Jia (2006): "Utility of income as a random function: Behavioral characterization and empirical evidence," *Mathematical Social Sciences*, 51(1), 23–57.

Dahl, R. E. (2004): "Adolescent Brain Development: A Period of Vulnerabilities and Opportunities," in *Annals of the New York Academy of Sciences 1021*, pp. 1–22. New York Academy of Sciences.

Dahlquist, G., and A. Björck (2008): *Numerical Methods in Scientific Computing - Volume 1*. Society for Industrial and Applied Mathematics, Philadelphia.

Dalton, P. S., S. Ghosal, and A. Mani (2013): "Poverty and Aspirations Failure," Discussion paper, Tilburg University and University of Warwick, mimeo.

Davidson, J. (1994): *Stochastic Limit Theory: An Introduction for Econometricians*. Oxford University Press.

Davis, P. J., and P. Rabinowitz (1984): *Methods of Numerical Intergration*. Academic Press, London.

Dehne, M., and J. Schupp (2007): "Persönlichkeitsmerkmale im Sozio-oekonomischen Panel (SOEP) - Konzept, Umsetzung und empirische Eigenschaften," Research Note 26, German Institute for Economic Research (DIW), Berlin.

Dempster, A. P., N. M. Laird, and D. B. Rubin (1977): "Maximum Likelihood from Incomplete Data via the EM Algorithm," *Journal of the Royal Statistical Society, Series B*, 39(1), 1–38.

Diebolt, J., and C. P. Robert (1984): "Estimation of Finite Mixture Distributions through Bayesian Sampling," *Journal of the Royal Statistical Society, Serie B*, 56(2), 363–375.

Digman, J. M. (1997): "Higher-Order Factors of the Big Five," *Journal of Personality and Social Psychology*, 73(6), 1246–1256.

Dohmen, T., A. Falk, D. Huffman, and U. Sunde (2008): "Representative Trust and Reciprocity: Prevalence and Determinants," *Economic Inquiry*, 46(1), 84–90.

——— (2010): “Are Risk Aversion and Impatience Related to Cognitive Ability?,” *American Economic Review*, 100(3), 1238–60.

Dohmen, T., A. Falk, D. Huffman, U. Sunde, J. Schupp, and G. G. Wagner (2011): “Individual Risk Attitudes: Measurement, Determinants, And Behavioral Consequences,” *Journal of the European Economic Association*, 9(3), 522–550.

Doksum, K. (1974): “Empirical Probability Plots and Statistical Inference for Nonlinear Models in the Two-Sample Case,” *The Annals of Statistics*, 2(2), 267–277.

Duckworth, A. L., and M. Seligman (2005): “Self-Discipline Outdoes IQ in Predicting Academic Performance of Adoloscents,” *American Psychological Society*, 16(12), 939–944.

Duttweiler, P. (1984): “The Internal Control Index: A Newly Developed Measure of Locus of Control,” *Educational and Psychological Measurement*, 44(2), 209–221.

Engle, R. F., D. F. Hendry, and J.-F. Richard (1983): “Exogeneity,” *Econometrica*, 51(2), 277–304.

Eysenck, H. J. (1991): “Dimensions of Personality: Sixteen, Five or Three? Criteria for a Taxonomic Paradigm,” *Personality and Individual Differences*, 8(12), 773–790.

Fehr, E., and S. Gächter (2000): “Fairness and Retaliation: The Economics of Reciprocity,” *Journal of Economic Perspectives*, 14(3), 159–181.

Fehr, E., and K. M. Schmidt (2006): “The Economics of Fairness, Reciprocity and Altruism - Experimental Evidence and New Theories,” in *Handbook of the Economics of Giving, Altruism and Reciprocity*, ed. by S.-C. Kolm, and J. M. Ythier, vol. 1, pp. 615–699. Wiley-Interscience, Amsterdam: North-Holland.

Firpo, S. (2007): “Efficient Semiparametric Estimation of Quantile Treatment Effects,” *Econometrica*, 75(1), 259–276.

Firpo, S., N. Fortin, and T. Lemieux (2009): “Unconditional Quantile Regression,” *Econometrica*, 77(3), 953–973.

Flossmann, A., R. Piatek, and L. Wichert (2007): “Going Beyond Returns to Education: The Role of Noncognitive Skills on Wages in Germany,” unpublished, available at `http://www.afse.fr/docs/congres2007/docs/619716flopiwi.pdf`.

Forero, C. G., and A. Maydeu-Olivares (2009): “Estimation of IRT Graded Response Models: Limited Versus Full Information Methods,” *Psychological Methods*, 14(3), 275–299.

Fortin, N. M. (2008): “The Gender Wage Gap among Young Adults in the United States - The Importance of Money versus People,” *Journal of Human Resources*, 43(4), 884–918.

Foster, J., and A. Sen (1997): *On Economic Inequality*. Oxford University Press, New York.

Foster, J., S. Seth, M. Lokshin, and Z. Sajaia (2013): *A Unified Approach to Measuring Poverty and Inequality - Theory and Practice*. The World Bank, Washington, DC.

Fraga, M., E. Ballestar, M. Paz, S. Ropero, F. Setien, M. Ballestar, D. Heine-Suner, J. Cigudosa, M. Urioste, J. Benitez, M. Boix-Chornet, A. Sanchez-Aguilera, C. Ling, E. Carlsson, P. Poulsen, A. Vaag, Z. Stephan, T. Spector, Y.-Z. Wu, C. Plass, and M. Esteller (2005): "Epigenetic Differences Arise during the Lifetime of Monozygotic Twins," *Proceedings of the National Academy of Sciences of the United States of America*, 102(30), 10604–10609.

Frölich, M. (2006): "Non-Parametric Regression for Binary Dependent Variables," *Econometrics Journal*, 9, 511–540.

Fuerst, J. S., and D. Fuerst (1993): "Chicago Experience with an Early Childhood Programme: The Special Case of the Child Parent Center Program," *Educational Research*, 35(3), 237–253.

Fuleky, P. (2012): "On the Choice of the Unit Period in Time Series Models," *Applied Economics Letters*, 19(12), 1179–1182.

Gale, D., and H. Nikaido (1965): "The Jacobian Matrix and Global Univalence of Mappings," *Math. Annalen*, 159, 81–93.

Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin (1995): *Bayesian Data Analysis*. Chapman and Hall, London.

Geman, S., and D. Geman (1984): "Stochastic Relaxation, Gibbs Distributions and the Bayesian Restoration of Images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(5), 609–628.

Geman, S., and C.-R. Hwang (1982): "Nonparametric Maximum Likelihood Estimation by the Method of Sieves," *The Annals of Statistics*, 10(2), 401–414.

Genicot, G., and D. Ray (2013): "Aspirations and Inequality," Discussion paper, New York University, mimeo.

Geweke, J. (2005): *Contemporary Bayesian Econometrics and Statistics*. John Wiley & Sons, New York, NY.

Geweke, J., and G. Zhou (1996): "Measuring the Pricing Error of the Arbitrage Pricing Theory," *Review of Financial Studies*, 9(2), 557–587.

Gilks, W. R., S. Richardson, and D. Spiegelhalter (1996): *Markov Chain Monte Carlo in Practice*. Chapman & Hall, London.

Goldberg, L. (1971): "A Historical Survey of Personality Scales and Inventories," in *Advances in Psychological Assessment*, ed. by P. McReynolds, pp. 293–336. Science and Behavior Books, Palo Alto, California.

Goldsmith, H. H., A. H. Buss, R. Plomin, M. K. Rothbart, A. Thomas, S. Chess, R. A. Hinde, and R. B. McCall (1987): "Rounddtable: What is temperament? Four approaches," *Child Development*, 58, 505–529.

Gorman, W. M. (1980): "A Possible Procedure for Analysing Quality Differentials in the Egg Market," *The Review of Economic Studies*, 47(5), 843–856.

Gouriéroux, C., and A. Monfort (1991): "Simulation Based Inference in Models with Heterogeneity," *Annales d'Economie et de Statistique*, 20/21, 69–107.

Grenander, U. (1981): *Abstract Inference*. John Wiley & Sons, New York, NY.

Griliches, Z. (1977): "Estimating the Returns to Schooling: Some Econometric Problems," *Econometrica*, 45(1), 1–22.

Grossman, J., and J. Tierney (1998): "Does Mentoring Work? An Impact Study of the Big Brothers Big Sisters Program," *Evaluation Review*, 22(3), 403–426.

Hamilton, J. D. (1994): *Time series analysis*. Princeton Univ. Press, Princeton, NJ.

Hampel, F. R. (1974): "The Influence Curve and its Role in Robust Estimation," *Journal of the American Statistical Association*, 69(346), 383–393.

Hansen, K., J. Heckman, and K. Mullen (2004): "The Effect of Schooling and Ability on Achievement Test Scores," *Journal of Econometrics*, 121(1/2), 39–98.

Hanushek, E. A., and L. Woessmann (2011): "The Economics of International Differences in Educational Achievements," in *Handbook of the Economics of Education*, ed. by E. A. Hanushek, S. J. Machin, and L. Woessmann, vol. 3, chap. 2, pp. 89–200. Elsevier B.V., Amsterdam: North-Holland.

Harris, J. W., and H. Stocker (2006): *Handbook of Mathematics and Computational Science*. Springer, New York.

Hause, J. C. (1972): "Earnings Profile: Ability and Schooling," *Journal of Political Economy*, 80(3, Part 2), 108–138.

Haushofer, J., and E. Fehr (2014): "On the Psychology of Poverty," *Science*, 344(6186), 862–867.

Heckman, J. (2008): "Schools, Skills, and Synapses," *Economic Inquiry*, 46(3), 289–324.

Heckman, J., and T. Kautz (2012): “Hard Evidence on Soft Skills,” IZA Discussion Paper 6580, Institute for the Study of Labor (IZA), Bonn.

Heckman, J., L. Lochner, and P. Todd (2006): “Fifty Years of Mincer Earinings Regression,” in *Handbook of the Economics of Education*, ed. by E. A. Hanushek, and F. Welch, vol. 1, chap. 7, pp. 307–451. Elsevier B.V., Amsterdam: North-Holland.

Heckman, J., S. H. Moon, R. Pinto, P. A. Savelyev, and A. Yavitz (2010): “The Rate of Return to the High Scope Perry Program,” *Journal of Public Economics*, 94(1/2), 114–128.

Heckman, J., and R. Pinto (2013a): “Econometric Mediation Analyses: Identifying the Sources of Treatment Effects from Experimentally Estimated Production Technologies with Unmeasured and Mismeasured Inputs,” NBER Working Papers 19314, National Bureau of Economic Research (NBER), Cambridge, MA.

——— (2013b): “Causel Analysis after Haavelmo,” NBER Working Papers 19453, National Bureau of Economic Research (NBER), Cambridge, MA.

Heckman, J., R. Pinto, and P. Savelyev (2013): “Understanding the Mechanisms Through Which an Influential Early childhood Programm Boosted Adult Outcomes,” *American Economic Review*, 104(6), 2053–2086.

Heckman, J., and Y. Rubinstein (2001): “The Importance of Noncognitive Skills: Lessons from the GED Testing Program,” *American Economic Review*, 91(2), 145–149.

Heckman, 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(3), 411–482.

Heckman, J. J., and E. J. Vytlacil (2006): “Econometric Evaluation of Social Programs, Part I: Causal Models, Structural Models and Econometric Policy Evaluation,” in *Handbook of Econometrics*, ed. by J. J. Heckman, and E. E. Leamer, vol. 6B, chap. 70, pp. 4780–4877. Elsevier B.V., Amsterdam: North-Holland.

Heineck, G., and S. Anger (2010): “The Returns to Cognitive Abilities and Personality Traits in Germany,” *Labour Economics*, 17(3), 535–546.

Hill, J. L., J. Brooks-Gunn, and J. Waldfogel (2002): “Sustained Effects of High Participation in Early Intervention for Low-Birth-Weight Premature Infants,” *Developmental Psychology*, 39(4), 730–744.

Hojman, D., and F. Kast (2009): “On the Measurement of Poverty Dynamics,” Working paper series, Harvard University, John F. Kennedy School of Government.

Holtz-Eakin, D., W. Newey, and H. S. Rosen (1988): "Estimating Vector Autoregressions with Panel Data," *Econometrica*, 56(6), 1371–95.

Hopkins, K., and G. Bracht (1975): "Ten-Year Stability of Verbal and Nonverbal IQ Scores," *American Educational Research Journal*, 12(4), 469–477.

Horn, J. L. (1965): "A Rationale and Test for the Number of Factors in Factor Analysis," *Psychometrika*, 30(2), 179–185.

Horowitz, J. L. (2009): *Semiparametric and Nonparametric Methods in Econometrics*. Springer, Heidelberg.

Hough, L. M. (1992): "The "Big Five" Personality Variables - Construct Confusion: Description versus Prediction," *Human Performance*, 5(1), 139–155.

Härdle, W. (1995): *Applied Nonparametric Regression*. Cambridge University Press, New York, NY.

Härdle, W., M. Müller, A. Werwatz, and S. Sperlich (2004): *Nonparametric and Semiparametric Models*. Springer, Heidelberg.

Hsiao, C. (2003): *Analysis of Panel Data*. Cambridge University Press., New York, NY.

Imbens, G. W., and J. M. Wooldridge (2009): "Recent Developments in the Econometrics of Program Evaluation," *Journal of Economic Literature*, 47(1), 5–86.

Intriligator, M. D. (1983): "Economic and Econometric Models," in *Handbook of Econometrics*, ed. by Z. Griliches, and M. D. Intriligator, vol. 1, chap. 3, pp. 181–221. Elsevier B.V., Amsterdam: North-Holland.

Jackson, D. (1976): *Jackson Personality Inventory Test Manual*. Research Psychologists Press, Goshen, New York.

John, K., and S. L. Thomsen (2014): "Heterogeneous Returns to Personality - The Role of Occupational Choice," *Empirical Economics*, 47(2), 553–592.

Johnson, A. (1996): "An Evaluation of Long-Term Impacts of the Sponsor-a-Scholar Program on Student Performance," *Mathematica Policy Research Inc.*

Johnson, D. L., and T. Walker (1991): "A Follow-up Evaluation of Houston Parent-Child Development Center: School Performance," *Journal of Early Intervention*, 15(3), 226–236.

Johnson, J. S., and E. L. Newport (1989): "Criticle Period Effects in Second Language Learning: The Influence of Maturational State on the Acquisition of English as a Second Language," *Cognitive Psychology*, 21, 60–99.

Jöreskog, K. (1977): "Structural Equation Models in the Social Sciences: Specification, Estimation and Testing," in *Applications of Statistics*, ed. by P. R. Krishnaih, pp. 265–287. Elsevier B.V., Amsterdam, North Holland.

Jöreskog, K. G. (1967): "Some Contributions to Maximum Likelihood Factor Analysis," *Psychometrika*, 32(4), 443–482.

Jöreskog, K. G., and I. Moustaki (2001): "Factor Analysis of Ordinal Variables: A Comparison of Three Approaches," *Multivariate Behavioral Research*, 36(3), 347–387.

Judd, K. (1998): *Numerical Methods in Economics*. MIT Press, Cambridge, Massachusetts.

Kendall, P., and L. Wilcox (1979): "Self-Control in Children: Development of a Rating Scale," *Journal of Consulting and Clinical Psychology*, 47, 1020–1029.

Kilian, L., and C. Vega (2011): "Do Energy Prices Respond to U.S. Macroeconomic News? A Test of the Hypothesis of Predetermined Energy Prices," *The Review of Economics and Statistics*, 93(2), 660–671.

Kirby, K. N. (1997): "Bidding on the Future: Evidence Against Normative Discounting of Delayed Rewards," *Journal of Experimental Psychology. General*, 126(1), 54–70.

Knudsen, E. (2004): "Sensitive Periods in the Development of the Brain and Behavior," *Journal of Cognitive Neuroscience*, 16(4), 1412–1425.

Knudsen, E., J. Heckman, J. Cameron, and J. Shonkoff (2006): "Economic, Neurobiological and Behavioral Perspectives on Building America's Future Workforce," *World Economics*, 7(3), 14–41.

Koenker, R., and G. Basset (1978): "Regression Quantiles," *Econometrica*, 46(1), 33–50.

Kosse, F., and F. Pfeiffer (2012): "Impatience Among Preschool Children and Their Mothers," *Economics Letters*, 115(3), 493–495.

——— (2013): "Quasi-Hyperbolic Time Preferences and their Intergenerational Transmission," *Applied Economic Letters*, 20(10), 983–986.

Laibson, D. I. (1997): "Golden Eggs and Hyperbolic Discounting," *Quaterly Journal of Economics*, 112, 443–477.

Lambert, P. J. (2001): *The Distribution and Redistribution of Income*. Manchester University Press., New York, NY.

Lancaster, K. J. (1966): "A New Approach to Consumer Theory," *Journal of Political Economy*, 74(2), 132–157.

LANG, S. (1973): *Calculus of Several Variables.* Addison-Wesley, Reading, MA.

——— (1987): *Linear Algebra.* Springer, New York, NY.

LAYARD, R., AND G. M. ANDN S. NICKELL (2008): "The Marginal Utility of Income," *Journal of Public Economics*, 92(8-9), 1846–1857.

LAZARSFELD, P. F. (1950): "The Logical and Mathematical Foundation of Latent Structure Analysis," in *Measurement and Prediction*, ed. by S. Stouffer, L. Guttman, E. A. Suchman, P. F. Lazarsfeld, S. A. Stas, and J. A. Clausen, pp. 362–412. Princeton University Press, Princeton, NJ.

LEIBOWITZ, A. (1974): "Home Investments in Children," *Journal of Political Economy*, 82(2, Part 2), 111–131.

LORD, F. M. (1952): *A Theory of Test Scores.* Psychometric Monograph Nr. 7, Psychometric Society.

LORD, F. M., AND M. R. NOVICK (1968): *Statistical Theories of Mental Test Scores.* Addison-Wesley, Reading, MA.

MADDALLA, G. (1983): *Limited-Dependent and Qualitative Variables in Econometrics.* Cambridge University Press., New York, NY.

MADDEN, G. J., A. M. BEGOTKA, B. R. RAIFF, AND L. L. KASTERN (2003): "Delay Discounting of Real and Hypothetical Rewards," *Experimental and Clinical Psychopharmacology*, 11(2), 139–145.

MCCLURE, S. M., D. I. LAIBSON, G. LOEWENSTEIN, AND J. D. COHEN (2004): "Separate Neural Systems Value Immediate and Delayed Monetary Rewards," *Economics Letters*, 306(5695), 503–507.

MCCULLAGH, P., AND J. A. NELDER (1989): *Generalized Linear Models.* Chapman and Hall, London, United Kingdom.

MEREDITH, W. (1993): "Measurement Invariance, Factor Analysis and Factorial Invariance," *Psychometrika*, 58(4), 525–543.

MEYER, T., AND S. THOMSEN (2013): "How Important is Secondary School Duration for Post-School Education Decision? – Evidence from a Natural Experiment," NIW Discussion Paper Series 6, NIW, Hannover.

MILLIMET, D. L., AND I. K. MCDONOUGH (2013): "Dynamic Panel Data Models with Irregular Spacing: With Applications to Early Childhood Development," Iza discussion papers, Institute for the Study of Labor (IZA).

Mischel, W. (1968): *Personality and Assessment*. Wiley, New York.

——— (1973): "Toward a Cognitive Social Learning Reconceptualization of Personality," *Psychological Review*, 80(4), 246–268.

Mischel, W., and Y. Shoda (1995): "A Cognitive-Affective System Theory of Personality: Reconceptualizing Situations, Dispositions, Dynamics, and Invariance in Personality Structure," *Psychological Review*, 102(2), 246–268.

Modigliani, F., and R. Brumberg (1954): "Utility Analysis and the Consumption Function: An Interpretation of Cross-Section Data," in *Post-Keynesian Economics*, ed. by K. K. Kurihara, pp. 388–436. Rutgers University Press, New Brunswick, New Jersey.

Morgeson, F., M. Campion, R. Dipboye, J. Hollenbeck, K. Murphy, and N. Schmitt (2007): "Reconsidering the use of Personality Tests in Personnel Selection Contexts," *Personal Psychology*, 60, 683–729.

Mueller, G., and E. Plug (2006): "Estimating the Effect of Personality on Male and Female Earnings," *Industrial and Labor Relations Review*, 60(1), 3–22.

Mulaik, S. A. (2009): *Linear Causal Modelling with Structural Equations*. Chapman & Hall/CRC Press, Boca Raton, FL.

——— (2010): *Foundations of Factor Analysis*. CRC Press, Boca Raton, FL.

Mundlak, Y. (1978): "On the Pooling of Time Series and Cross Section Data," *Econometrica*, 46(1), 69–85.

Muraven, M., and R. F. Baumeister (2000): "Self-Regulation and Depletion of Limited Ressources: Does Self-Control resemble a muscle," *Psychological Bulletin*, 126(2), 247–257.

Murnane, R. J., J. B. Willett, M. J. Braatz, and Y. Duhaldeborde (2000): "Do Different Dimensions of Male High School Students' Skills Predict Labor Market Success a Decade Later? Evidence from the NLSY," *Economics of Education Review*, 20(4), 311–320.

Newport, E. (1990): "Maturational Constraints on Language Learning," *Cognitive Science*, 14(1), 11–28.

Neyman, J., and E. L. Scott (1948): "Consistent Estimates Based on Partially Consistent Observations," *Econometrica*, 16(1), 1–32.

O'Connor, T. G., M. Rutter, C. Beckett, L. Keaveney, J. M. Kreppner, and the English and Romanian Adoptees Study Team (2000): "The Effects of Global Severe Privation on Cognitive Competence: Extension and Longitudinal Follow-Up," *Child Development*, 71(2), 376–390.

Orom, H., and D. Cervone (2009): "Personality Dynamics, Meaning, and Idiosyncrasy: Assessing Personality Architecture and Coherence Idiographically," *Journal of Research in Personality*, 43(2), 228–240.

Osborne-Groves, M. (2005): "How Important is your Personality? Labor Market Returns to Personality for Women in the US and UK," *Journal of Economic Psychology*, 26, 827–841.

Parthasarathy, T. (1983): *On Global Univalence Theorems*. Springer, New York, NY.

Pearl, J. (2009): *Causality: Models, Reasoning, and Inference*. Cambridge University Press, New York.

Pervin, L., D. Cervone, and O. John (2005): *Theories of Personality*. John Wiley & Sons, New York, NY.

Piatek, R. (2010): "Bayesian Inference for Factor Structure Models via Gibbs Sampling," CMS Discussion Paper 2010-03, Center for Quantitative Methods and Survey Research (CMS), Konstanz.

Piatek, R., and P. Pinger (2010): "Maintaining (Locus of) Control? Assessing the Impact of Locus of Control on Education Decisions and Wages," IZA Discussion Paper Series 5289, Institute for the Study of Labor (IZA), Bonn.

Rabe-Hesketh, S., A. Skrondal, and X. Zheng (2007): "Multilevel Structural Equation Modeling," in *Handbook of Computing and Statistics with Applications*, ed. by E. Kontoghiorghes, vol. 1, pp. 209–227. Elsevier B.V., Amsterdam, North Holland.

Rencher, A. (2004): *Methods of Multivariate Analysis*. Wiley Interscience.

Rijmen, F., F. Tuerlinckx, P. D. Boeck, and P. Kuppens (2003): "A Nonlinear Mixed Model Framework for Item Response Theory," *Psychological Methods*, 8(2), 185–205.

Roberts, B., P. Harms, J. 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, chap. 22, pp. 321–335. American Psychological Association, Washington, DC.

Roberts, B., N. Kuncel, R. Shiner, A. Caspi, and L. Goldberg (2007): "The Power of Personality," *Perspectives on Psychological Science*, 2(4), 313–345.

Roberts, B. W. (2009): "Back to the Future: Personality and Assessment and Personality Development," *Journal of Research in Psychology*, 43(2), 137–145.

Roberts, B. W., and W. F. Delvecchio (2000): "The Rank-Order Consistency of Personality from Childhood to Old Age: A Quantitative Review of Longitudinal Studies," *Psychological Bulletin*, 126, 3–25.

ROBERTS, B. W., W. VIECHTBAUER, AND K. E. WALTON (2006): "Patterns of Mean-Level Change in Personality Traits Across the Life Course: A Meta-Analysis of Longitudial Studies," *Psychological Bulletin*, 132(1), 1–25.

ROSENBAUM, P. R., AND D. B. RUBIN (1983): "The Central Role of the Propensity Score in Observational Studies for Causal Effect," *Biometrika*, 70(1), 41–55.

ROSENBERG, M. (1965): *Society and the Adolescent Self-Image*. Princeton University Press, Princeton, New Jersey.

ROTHBART, M. K. (1981): "Measurement of Temperament in Infancy," *Child Development*, 52, 569–578.

ROTHBART, M. K., S. A. AHADI, AND D. E. ADAMS (2000): "Temperament and Personality: Origins and Outcomes," *Journal of Personality and Social Psychology*, 78(1), 122–135.

ROTHBART, M. K., AND M. BATES (2006): "Temperament," in *Handbook of Child Psychology Vol. 3 - Social, Emotional and Personality Development*, ed. by W. Damon, N. Eisenberg, and R. M. Lerner, pp. 99–166. John Wiley & Sons, New York.

ROTHENBERG, T. J. (1971): "Identification in Parametric Models," *Econometrica*, 39(3), 577–591.

ROTTER, J. (1966): *Generalized Expectancies for Internal versus External Control of Reinforcement*. American Psychological Association, Washington, DC.

SALOVEY, P., M. BRACKETT, AND J. D. MAYER (2004): *Emotional Intelligence - Key Readings on the Mayer and Salovey Model*. National Professional Ressources Inc., Port Chester, NY.

SAMEJIMA, F. (1952): *"Estimation of Latent Ability Using a Response Pattern of Graded Scores"*. Psychometric Monograph Nr. 17, Psychometric Society.

SCHMIDT, P. (1981): "Constraints on the Parameters in Simultaneous Tobit and Probit Models," in *Structural Analysis of Discrete Data with Econometric Applications*, ed. by C. Manski, and D. McFadden, pp. 422–434. The M.I.T. Press.

SCHNEEWEISS, H. (1976): "Consistent Estimation of a Regression with Errors in Variables," *Metrika*, 23, 101–115.

SCHOKKAERT, E. (2007): "The capabilities approach," Center for Economic Studies - Discussion papers ces0734, Katholieke Universiteit Leuven, Centrum voor Economische Studiën.

SCHUMPETER, J. A. (1934): *The Theory of Economics Development*. Oxford University Press, London.

Schweinhart, L. J., H. V. Barnes, and D. P. Weikart (1993): *Significant Benefits - The High/Scope Perry Preschool Study Through Age 27*. High/Scope Educational Research Foundation, Ypsilanti, Michigan.

Sen, A. (1976): "Poverty: An Ordinal Approach to Measurement," *Econometrica*, 44(2), 219–231.

——— (1982): "Equality of What?," in *Choice, Welfare and Measurement*, ed. by A. Sen, chap. 16, pp. 353–369. Blackwell, Oxford.

——— (1999): "The Possibility of Social Choice," *American Economic Review*, 89(3), 349–378.

Sher, J. (1977): *Education in Rural America*. Westview: Boulder.

Shiner, R., and A. Caspi (2003): "Personality Differences in Childhood and Adolescence: Measurement, Development, and Consequences," *Journal of Child Psychology and Psychiatry*, 44(1), 2–32.

Sijtsma, K., and B. W. Junker (2006): "Item Response Theory: Past Performance, Present Developments, and Future Expectations," *Behaviormetrika*, 33(1), 75–102.

Skrondal, A., and S. Rabe-Hesketh (2004): *Generalized Latend Variable Modelling: Multilevel, Longitudinal, and Structural Equation Models*. Chapman & Hall/CRC, Boca Raton, FA.

Spady, R. H. (2006): "Identification and Estimation of Latent Attitudes and their Behavioral Implications," CEMMAP Working Paper CWP12/06, Centre for Micro Data Methods and Practice (CEMMAP), London.

——— (2007): "Semiparametric Methods for the Measurement of Latent Attitudes and the Estimation of their Behavioral Consequences," CEMMAP Working Paper CWP26/07, Centre for Micro Data Methods and Practice (CEMMAP), London.

Stevens, A. H. (1999): "Climbing out of Poverty, Falling Back in: Measuring the Persistence of Poverty Over Multiple Spells," *Journal of Human Resources*, 34(3), 557–588.

Stewart, M. B. (2007): "The Interrelated Dynamics of Unemployment and Low-Wage Employment," *Journal of Applied Econometrics*, 22(3), 511–531.

Tangney, J., R. Baumeister, and M. Boone (2004): "High Self-Control Predicts Good Adjustment, Less Pathology, Better Grades, and Interpersonal Success," *Journal of Personality*, 72(2), 271–322.

Tanner, M. A., and W. H. Wong (1987): "The Calculation of Posterior Distributions by Data Augmentation," *Journal of the American Statistical Association*, 82(398), 528–540.

Tellegen, A. (1985): "Structures of Mood and Personality and their Relevance to Assessing Anxiety, with an Emphasis on Self-Report," in *Anxiety and the Anxiety Disorders*, ed. by H. Tuma, and J. Maser, pp. 681–706. Lawrence Erlbaum Associates, Hillsdale, New Jersey.

Thaler, R. H. (1994): "Psychology and Savings Policies," *American Economic Review*, 84(2), 186–92.

Thaler, R. H., and S. Benartzi (2004): "Save More Tomorrow: Using Behavioral Economics to Increase Employee Saving," *Journal of Political Economy*, 112(S1), 164–S187.

Theil, H. (1967): *Economics and Information Theory*. Elsevier, Amsterdam, North Holland.

Theil, H. (1983): "Economic and Econometric Models," in *Handbook of Econometrics*, ed. by Z. Griliches, and M. D. Intriligator, vol. 1, chap. 1, pp. 5–65. Elsevier B.V., Amsterdam: North-Holland.

Thiel, H., and S. Thomsen (2013): "Noncognitive Skills in Economics: Models, Measurement, and Empirical Evidence," *Research in Economics*, 67(2), 189–214.

——— (2015): "Individual Poverty Paths and the Stability of Control-Perception," Discussion paper, Leibniz Universität Hannover, mimeo.

Thiel, H., S. L. Thomsen, and B. Büttner (2014): "Variation of learning intensity in late adolescence and the effect on personality traits," *Journal of the Royal Statistical Society Series A*, 177(4).

Thomas, A., and S. Chess (1977): *Temperament and Development*. Brunner/Mazel, New York.

Thomas, A., S. Chess, and H. G. Birch (1968): *Temperament and Behavior Disorders in Children*. New York University Press, New York.

Thompson, B. (2004): *Exploratory and Confirmatory Factor Analysis - Understanding Concepts and Applications*. American Psychological Association, Washington D.C.

Todd, P., and K. Wolpin (2007): "The Production of Cognitive Achievement in Children: Home, School and Racial Test Score Gaps," *Journal of Human Capital*, 1(1), 91–136.

Todd, P. E., and K. I. Wolpin (2003): "On The Specification and Estimation of The Production Function for Cognitive Achievement," *Economic Journal*, 113(485), F3–F33.

Turkheimer, E., A. Haley, M. Waldron, B. D'Onofrio, and I. Gottesman (2003): "Socioeconomic Status Modifies Heritability of IQ in Young Children," *Psychological Science*, 14(6), 623–628.

Urzua, S. (2008): “Racial Labor Market Gaps: The Role of Abilities and Schooling Choices,” *Journal of Human Resources*, 43(4), 919–971.

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

Wagner, G. G., J. R. Frick, and J. Schupp (2007): “The German Socio-Economic Panel Study (SOEP) - Scope, Evolution and Enhancements,” *Schmollers Jahrbuch*, 127, 139–169.

Wald, A. (1950): “Note on the Identification of Economic Relations,” in *Statistical Inference in Dynamic Economic Models*, ed. by T. C. Koopmans, pp. 238–244. Cowles Commission Monograph 10, Wiley, New York.

Wansbeek, T., and E. Meijer (2000): “Measurement Error and Latent Variables in Econometrics,” in *Advanced Textbooks in Economics*, ed. by C. J. Bliss, and M. D. Intriligator, vol. 37. Elsevier B.V., Amsterdam, North Holland.

Watts, H. (1968): “An Economic Definition of Poverty,” in *On Understanding Poverty: Perspectives from the Social Sciences*, ed. by D. P. Moynihan. Basic Books, New York, NY.

Webb, E. (1915): “Character and Intelligence,” *British Journal of Psychology*, 1(3), 1–99.

White, H. (1982): “Maximum Likelihood Estimation of Misspecified Models,” *Econometrica*, 50(1), 1–25.

Wolfe, R., and S. Johnson (1995): “Personality as a Predictor of College Performance,” *Educational and Psychological Measurement*, 55(2), 177–185.

Wood, D. (2007): “Using PRISM to Compare the Explanatory Value of General and Role-Contextualized Trait Ratings,” *Journal of Personality*, 76(6), 1103–1126.

Wooldridge, J. M. (1997): “Multiplicative Panel Models without the Strict Exogeneity Assumption,” *Econometric Theory*, 13(5), 667–678.

——— (2000): “A Framework for Estimating Dynamic, Unobeserved Effects Panel Data Models with Possible Feedback to Future Explanatory Variables,” *Economic Letters*, 68, 245–250.

——— (2002): *Econometric Analysis of Cross Section and Panel Data*. MIT Press, Hong Kong.

——— (2005): “Simple Solutions to the Initial Conditions Problem in Dynamic, Nonlinear Panel Data Models with Unobserved Heterogeneity,” *Journal of Applied Econometrics*, 20(1), 39–54.

Zheng, B. (1992): “An Axiomatic Characterization of the Watts Poverty Index,” *Economic Letters*, 42, 81–86.

——— (1994): "Can a Poverty Index be Both Relative and Absolute," *Econometrica*, 62(6), 1453–1458.

——— (1997): "Aggregate Poverty Measures," *Journal of Economic Surveys*, 11(2), 123–162.

Zhu, C., R. H. Byrd, and J. Nocedal (1997): "L-BFGS-B: Algorithm 778: L-BFGS-B, FORTRAN Routines for Large Scale Bound Constrained Optimization," *ACM Transactions on Mathematical Software*, 23(4), 550–560.