Web Appendix

Maintaining (Locus of) Control?

Data Combination for the Identification and Inference of Factor Structure Models

Rémi Piatek[1] and Pia Pinger[2]

[1]Department of Economics, University of Copenhagen, Denmark.
[2]Department of Economics, University of Bonn, Germany.

January 29, 2015

Contents

A A Theoretical Framework for Locus of Control 1
A.1 The concept of locus of control 1
A.2 Stability of locus of control in the literature 2
A.3 A theoretical framework 2

B General Identification of the Factor Model 6

C Bayesian Inference 10
C.1 Prior specification 10
C.2 MCMC sampling: special features of our algorithm 11
C.3 MCMC sampler 13

D Monte Carlo Study: Experimental Setup 20

E Data Addendum 23
E.1 Combining samples 23
E.1.1 Youth sample 23
E.1.2 Adult sample 24
E.2 'Pre-market' locus of control 24
E.3 School choice 25
E.4 Wage construction and labor market participation 25
E.5 Covariates 26

F Exploratory Factor Analysis 30

G Empirical Application: Additional Posterior Results 32

H Simulation of the Model and Goodness of Fit 36
H.1 Model simulation 36
H.2 Assessing goodness of fit 36

I Robustness Checks 40
I.1 Model specification. 40
I.2 Distributional assumptions. 41

J The Schooling Equation among Youths and Adults 45

K Comparison to more Traditional Methods of Estimation 50

L The Stability of Locus of Control 53

References 56

List of Tables

C.1 Prior Parameter Specification 11
E.2 Samples and included covariates 27
E.3 Descriptive statistics: covariates in the measurement system 29
E.4 Descriptive statistics: covariates in the outcome equations 29
F.5 Locus of control, youth sample 30
G.6 Regression coefficients in measurement system, males 32
G.7 Regression coefficients in measurement system, females 33
G.8 Regression coefficients in outcome system, males 34
G.9 Regression coefficients in outcome system, females 35
I.10 Robustness to model specification 42
I.11 Robustness to distributional assumptions 43
J.12 p-values of a two-sided test in means 49
K.13 Wage regressions with locus of control factor (*contemporaneous* measures) 51
K.14 Comparisons of wage model estimates with predicted factor scores and raw scores (*contemporaneous* measures) 52

List of Figures

F.1 Scree plot: all measurements versus 5 ‘external’ items only 31
F.2 Scatterplot of loadings: all measurements versus 6 ‘external’ items only 31
H.3 Goodness-of-fit check for wages, 2-mixture-component case 37
H.4 Goodness-of-fit check for wages, 3-mixture-component case 38
H.5 Goodness-of-fit check for wages, normal case 39
J.6 Schooling equation, testing for differences youth vs. adults, males . . . 47
J.7 Schooling equation, testing for differences youth vs. adults, females . . 48
L.8 Kernel density distributions of latent youth and adult locus of control factors 54
L.9 Histogram of percentile differences in latent youth and adult locus of control factors 55

A A Theoretical Framework for Locus of Control

A.1 The concept of locus of control

Since the seminal works of Mincer (1958) and Becker (1964), human capital is defined as the stock of knowledge and personal abilities an individual possesses, and is perceived as a factor of production that can be improved through education, training and experience. The focus usually lies on estimating returns to education, training, experience or cognitive skills (Psacharopoulos, 1981; Card, 1999; Heckman *et al.*, 2006a).[1] However, this concept mainly refers to the cognitive abilities, while more recently other facets of human capital have come to the forefront. Bowles and Gintis (1976) were among the first to point out what seems intuitively obvious: economic success is only partly determined by cognitive abilities and knowledge acquired in schools. Personality, incentive-enhancing preferences and socialization are other important components of human capital (Heckman *et al.*, 2006b; Heineck and Anger, 2010).[2] Furthermore, a vast literature in behavioral economics is currently emerging, which analyzes the economic impact of risk aversion, reciprocity, self-confidence and time preference (Dohmen *et al.*, 2010; Falk *et al.*, 2006; Frey and Meier, 2004).

Originally, locus of control is a psychological concept, generally attributed to Rotter (1966), that measures the attitude regarding the nature of the causal relationship between one's own behavior and its consequences. In this concept, which is related to self-efficacy, people who believe that they have control over their lives are called *internalizers*. People who believe that fate, luck, or other people determine their lives, are termed *externalizers*. Generally, externalizers (in this taxonomy, the low return personality types) do not have much confidence in their ability to influence their environment, and do not see themselves as responsible for their lives. Therefore, these individuals are generally less likely to trust their own abilities or to push themselves through difficult situations. Conversely, internalizers (the high return personality types) perceive themselves as more capable of altering their economic situation.

[1]See Gebel and Pfeiffer (2010), Pischke and Von Wachter (2008), Lauer and Steiner (2000), Flossmann and Pohlmeier (2006) for estimates of returns to education or skills in the German context.

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

A.2 Stability of locus of control in the literature

Endogeneity bias only arises if an individual's personality is not fixed at birth, but responsive to positive or negative life experiences.[3] Behavioral geneticists argue that about 50% of the phenotypic variation in personality can be ascribed to genes, while the other half is shaped by environmental factors (Krueger and Johnson, 2008). The fact that personality traits can be altered has also been shown in a recent body of literature, which argues that personality is much more malleable than for example cognition (for a summary of the literature see Almlund *et al.*, 2011; Heckman and Kautz, 2013). Concerning the evolution of locus of control, Furnham and Steele (1993) explain that it is "partially but not wholly the product of causal attributional beliefs about past events" and Furnham *et al.* (1992) point out that positive (negative) experiences reinforce tendencies toward internal control, which in turn increase (decrease) initiative and motivation for success. Cobb-Clark and Schurer (2013) show that overall stability of locus of control is high at adult ages, but lower for young and very old individuals. They also find that both men and women who improve their finances become more internal, and that females become more internal in response to a job promotion. However, they also report that even large shocks to locus of control only change wage returns by about 5% for men and by 3% for women. Furthermore, Boyce *et al.* (2013) and Roberts *et al.* (2003) have confirmed that personality changes in general and that it does so especially for young individuals and in response to early labor market experiences. Similarly, Gottschalk (2005) provides evidence from a randomized control trial which shows that working at a job has positive effects on locus of control.

A.3 A theoretical framework

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

In our study, we focus on the external dimension of locus of control (see Section 6.1.1 in paper). We assume that external locus of control is represented by a latent variable θ that is continuously distributed in the range $(-\infty, +\infty)$, where smaller values represent

[3] With "changes" in personality, we mean intra-individual changes as opposed to mean-level consistency or rank-order consistency.

a more external locus and larger values a less external locus of control. We assume that an individual's psychic costs of education and wage are both functions of θ. Hence, individuals with low levels of θ are likely to have higher psychic costs of education and earn lower wages, while individuals with high levels of θ incur lower costs of obtaining a degree and earn more.

In a typical model of human capital investment, individuals decide on the level of education based on the expected returns to the respective choice, net of the costs associated with this choice. In this framework, external locus of control likely increases the perceived psychic costs of education independent of the actual work effort. One reason for that may be that individuals with a more external locus of control believe ex ante that they would need to work harder than internalizers to feel well-prepared for the exams. Another reason may be that externalizers believe that no matter how hard they work, their education outcomes will depend on fate or luck. As a consequence this may induce them to work less than internalizers, but will increase their fear of failure.[4] We call differences in education investments due to the perceived psychic costs of education the *behavioral impact* of locus of control. Furthermore, external locus of control may be viewed as a skill with a direct impact on wages, for example because employers value having employees who exhibit a higher locus of control. This is what we term the *productive impact* of locus of control.

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

$$U^* \geq 0,$$

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

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

[4] Individuals who score higher on neuroticism, a concept related to locus of control, are also more likely to exhibit test anxiety (Moutafi *et al.*, 2006).

$$C = X_C\beta_C + \theta\alpha_C + \varepsilon_C,$$

with $\mathrm{E}(\varepsilon_1 \mid X_{wt}, \theta) = \mathrm{E}(\varepsilon_0 \mid X_{wt}, \theta) = \mathrm{E}(\varepsilon_C \mid X_C, \theta) = 0$. Here α_s, β_s (with $s \in \{0, 1\}$) and α_C, β_C measure the impact of pre-market locus of control θ and observable characteristics (X_{wt}, X_C) on wages and education costs, respectively. Since locus of control is determined before the individual enters the labor market, it does not depend on time t in our model. Moreover, ε_{st} and ε_C are random and independent idiosyncratic shocks. The total net present value from education, accounting for the discounted flow of ex post earnings, is then:

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

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

By differentiating Eq. (1) with respect to θ, it appears that a ceteris paribus change in locus of control affects education decisions as follows:

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

Given that α_1 and α_0 are independent of t, and making use of revealed education choices, our goal is to identify α_1, α_0 and α_C. More precisely, we are investigating whether locus of control enters the education decision and outcomes both directly as a skill, in which case we would have $\alpha_1 > 0$ and $\alpha_0 > 0$, or only indirectly via the costs of education, in which case $\alpha_C < 0$. We cannot identify α_C directly, because we do not observe education costs. However, we can make inference on the overall impact of locus of control on education choices, and given the identification of α_1 and α_0, we can retrieve α_C. More specifically, if we find for example that $\alpha_1 = \alpha_0 = 0$ or if we find that $\alpha_1 = 0$ and $\alpha_0 > 0$, we know that any impact of locus of control on education choices must work through α_C.

The empirical model we estimate is an approximation to this very simple theoretical framework. By combining different subsamples and using revealed schooling decisions, we are able to identify the impact of pre-market locus of control on wages, and thus to make inferences about its productive or behavioral impact, respectively.

B General Identification of the Factor Model

The combination of continuous and discrete variables in the framework of a factor structure model raises nontrivial problems. This section outlines the main steps of the identification strategy, and discusses the type of information that can be extracted from the observed variables to properly identify the model. To facilitate the exposition, we assume that all variables are observed for all the individuals of the sample. The missing data problem and its implications for identification are then discussed in Section 3.2 of the paper.

Although techniques to deal with ordinal variables in latent variable models have a long history in statistics (see Jöreskog and Moustaki, 2001, for a survey of different approaches), a widespread approach in empirical research consists of ignoring ordinality and treating the manifest items as continuous. This simplification, however, may distort the results, especially when the number of categories is limited, and/or the distribution of the answers is skewed or shows high kurtosis (Muthén and Kaplan, 1985; Rhemtulla *et al.*, 2012). In this paper, on the contrary, we assume that the discrete variables are generated by an underlying latent process. Additional identifying restrictions are therefore required to fully identify the model.

The identification of factor models is well documented in the literature (Anderson and Rubin, 1956) and many different strategies have been applied in practice to achieve it (see, for instance, Carneiro *et al.*, 2003). First, some assumptions are required on the latent factor:

$$\mathrm{E}(\theta) = 0, \qquad\qquad \mathrm{V}(\theta) = \sigma_\theta^2 \ll \infty, \qquad\qquad \theta \perp\!\!\!\perp X,$$

where the first two ones ensure that the factor is centered with finite variance, while the last one completes the independence assumptions (see Eq. (5) in paper) to guarantee that θ represents the only source of unobserved correlation between the observed variables. Using the independence assumptions, the identification of the factor loadings is straightforward to achieve if the variables $(S^\star, \{D_s^\star, Y_s\}_{s=0,1}, \{M_k^\star\}_{k=1,\dots,K})$ are observed for all individuals. In this case, following Carneiro *et al.* (2003), identification comes from the covariance matrix:

$$\Omega^\star \equiv \mathrm{V}(S^\star, D^\star, Y, M^\star \mid X)\,, \tag{2}$$

where each single covariance depends on the corresponding factor loadings and on the variance of the factor, e.g., $\mathrm{Cov}(S^\star; M_1^\star \mid X) = \alpha_S \alpha_{M_1} \sigma_\theta^2$. Assuming nonzero covariances, the ratios of the observed covariances therefore identify the ratios of the corresponding factor loadings, as for example:[5]

$$\frac{\mathrm{Cov}(M_1^\star; M_2^\star \mid X)}{\mathrm{Cov}(M_2^\star; M_3^\star \mid X)} = \frac{\alpha_{M_1}}{\alpha_{M_3}}. \tag{3}$$

As a consequence, the factor loadings are identified up to a common proportionality constant. This scaling issue arises because it is always possible to transform the model by multiplying all factor loadings by a constant $\delta \in \mathbb{R}$ and simultaneously dividing the latent factor by the same constant, without changing the structure of the covariance matrix. For instance, $\mathrm{Cov}(M_1^\star; M_2^\star) = (\delta\alpha_{M_1})(\delta\alpha_{M_2})(\sigma_\theta^2/\delta^2) = \alpha_{M_1}\alpha_{M_2}\sigma_\theta^2$ is unchanged, and the unrestricted variance of the factor is scaled by the inverse of δ^2, i.e., $\mathrm{V}(\theta/\delta) = \sigma_\theta^2/\delta^2$. This is the well-known *rotation problem*. To set the scale of the latent factor and thereby solve this issue, either a factor loading or the variance of the factor has to be set to a fixed value. The latter approach (setting $\sigma_\theta^2 = 1$) does not completely solve the rotation problem, as it is still possible to switch the signs of the factor loadings and of the latent factors simultaneously (using $\delta = -1$) without affecting the likelihood function. Using the covariance matrix $\Omega^\star$, it is straightforward to see that only the squares of the factor loadings can be identified in that case, as for example:

$$\frac{\mathrm{Cov}(M_1^\star; M_2^\star \mid X)\,\mathrm{Cov}(M_1^\star; M_3^\star \mid X)}{\mathrm{Cov}(M_2^\star; M_3^\star \mid X)} = (\alpha_{M_1})^2,$$

which identifies the factor loading α_{M_1} up to its sign. The former identification approach (setting $\alpha_{M_1} = 1$) addresses this *sign-switching problem*. In Eq. (3), fixing α_{M_1} to 1 automatically identifies the inverse of α_{M_3}. It also has the benefit of giving a metric to the latent factor, by anchoring it to a given variable of interest (Cunha *et al.*, 2010).

It is important to emphasize that both normalizations are innocuous for the results. In our application, we fix the variance of the factor to 1 and address the sign-switching issue a posteriori in the framework of our Bayesian analysis.[6] Hence, we implicitly assume that the scale of the factor is the same across different groups, but that the

[5] Note that in the following, the identification strategy is described for the general case. Identification would be easier in our framework, since the variances of the error terms are set to 1 in the measurement system, implying that $\mathrm{V}(M_k^\star \mid X_M) = \alpha_{M_k}^2 + 1$ (assuming $\sigma_\theta^2 = 1$). This would only be possible, however, if the latent variables $M_k^\star$ were observed.

[6] See Section C.2 for details.

impact of the factor on the variables might differ. For instance, females and males might react differently to a given psychometric question measuring locus of control. This restriction will thus facilitate comparisons across groups.

Since we are working with a combination of continuous and discrete variables, the covariance matrix $\Omega^\star$ defined in Eq. (2) cannot be directly observed. It can, however, be estimated if additional distributional assumptions are made on the unobservables of the model. When multivariate normality is assumed, it can be shown that the correlations between all underlying latent variables and manifest continuous variables are identified. More specifically, the *polychoric* correlations are identified in the measurement system (Olsson, 1979a;b; Jöreskog, 1994) and can thus be used to identify the factor loadings based on the strategy outlined above. How well this identification strategy works in practice depends on the amount of information that can be extracted from the observed discrete variables. Intuitively, the larger the number of categories of the Likert scale used for the measurements, and the more scattered the answers on this scale, the more informative the data. This information allows a better empirical identification of the factor model. Section 5 of the paper (see also Section D for the experimental setup) investigates how much information can be extracted from the observed variables to identify the model.

The presence of discrete variables does not usually hinder the nonparametric identification of the factor model, as long as some minimal conditions are fulfilled. Cunha *et al.* (2010) show how the distribution of the latent factor and of the error terms can be nonparametrically identified in a general framework. Unfortunately, at least two continuous variables are required to achieve nonparametric identification, whereas most of the variables are discrete in our model. This is why we have to rely on a parametric approach, and assume normality of the unobservables for this purpose:

$$\theta \sim \mathcal{N}\big(0;\, \sigma_\theta^2\big)\,, \qquad (\varepsilon_S, \varepsilon_D^0, \varepsilon_D^1, \varepsilon_{M_1}, \ldots, \varepsilon_{M_K})' \sim \mathcal{N}(0;\, \Sigma)\,, \tag{4}$$

where $\sigma_\theta^2 = 1$ and $\Sigma = I_{K+3}$. For the error terms of the wage equation, we relax normality by specifying a mixture of H_s normal distributions with zero mean:

$$\varepsilon_Y^s \sim \sum_{h=1}^{H_s} \pi_h^s \mathcal{N}\big(\mu_h^s;\, (\omega_h^s)^2\big)\,, \qquad \mathrm{E}(\varepsilon_Y^s) = \sum_{h=1}^{H_s} \pi_h^s \mu_h^s = 0, \tag{5}$$

for $s = 0, 1$, where $\vartheta_h^s = (\pi_h^s, \mu_h^s, \omega_h^s)$ denotes, respectively, the weight, mean and standard deviation of mixture component h for outcome Y_s, and $\vartheta = (\vartheta^0, \vartheta^1)$ with $\vartheta^s = (\vartheta_1^s, \ldots, \vartheta_{H_s}^s)$. Mixtures of normals are widely used as a flexible semiparametric approach to density estimation (Ferguson, 1983; Escobar and West, 1995). These mixtures allow us to capture unobserved heterogeneity that arises because individuals work in different sectors of the labor market.[7]

[7]In a frequentist approach, Dagsvik *et al.* (2011) find that Gaussian mixtures improve the fit of heavy-tailed log earnings distributions compared to normal distributions.

C Bayesian Inference

This section presents the prior specification of our model (Section C.1), discusses some challenging aspects of the sampling scheme in our particular framework (Section C.2), and provides a complete description of the algorithm (Section C.3).

C.1 Prior specification

The posterior distributions of the parameters are obtained by combining the likelihood function to their prior distributions through the application of Bayes' theorem. Therefore, the first step of the analysis consists of specifying prior distributions for all parameters. All the distributions we rely on are conjugate priors and are commonly used in the literature.

Conjugate normal prior distributions are used for the factor loadings and the regression coefficients:

$$\alpha \sim \mathcal{N}(a_0;\, A_0)\,, \qquad \text{for } \alpha = \alpha_S, \alpha_D^0, \alpha_D^1, \alpha_Y^0, \alpha_Y^1, \alpha_{M_1}, \ldots, \alpha_{M_K},$$
$$\beta \sim \mathcal{N}(b_0\, \iota;\, B_0\, I)\,, \qquad \text{for } \beta = \beta_S, \beta_D^0, \beta_D^1, \beta_Y^0, \beta_Y^1, \beta_{M_1}, \ldots, \beta_{M_K},$$

with $A_0 > 0$ and $B_0 > 0$, and where I and ι denote, respectively, the identity matrix and the vector of ones of dimensions corresponding to the length of the vector β.

The latent factor θ has been specified to be normally distributed in Eq. (4) for identification purposes, which, in our Bayesian framework, represents the prior distribution of this latent variable.

For the mixture of normals in the wage equations, we use a Dirichlet distribution for the weights, a normal for the means and an inverse-Gamma distribution for the variances of the mixture components:

$$\pi^s = (\pi_1^s, \ldots, \pi_{H_s}^s) \sim \mathcal{D}ir(p, \ldots, p)\,, \quad \mu_h^s \sim \mathcal{N}(m_0;\, M_0)\,, \quad (\omega_h^s)^2 \sim \mathcal{G}^{-1}(g_0;\, G_0)\,,$$

with $p > 0$, $M_0 > 0$, $g_0 > 0$ and $G_0 > 0$, for $h = 1, \ldots, H_s$, except for the mean of the last component that is assumed to be $\mu_{H_s}^s = -\sum_{h=1}^{H_s-1} \pi_h^s \mu_h^s / \pi_{H_s}^s$ to guarantee that the mean of the mixture is equal to zero (see Eq. (5)).

The last ingredient is the prior on the thresholds of the ordinal variables for the measurement equations. For each cutoff point $\gamma_{k,c}$, we assume a uniform prior on the interval delimited by the previous and the next thresholds, to guarantee the ordering:

$$\gamma_{k,c} \sim \mathcal{U}(\gamma_{k,c-1};\, \gamma_{k,c+1})\,, \qquad \text{for } k = 1, \ldots, K \text{ and } c = 2, \ldots, C_k - 1,$$

where $\gamma_{k,1}$ is set to zero for identification purpose, while $\gamma_{k,0} = -\infty$ and $\gamma_{k,C_k} = +\infty$.

Table C.1 shows the values of the prior parameters specified in our Monte Carlo study presented in Section 5 of the paper. None of these distributions impose a lot of prior knowledge on the model, so as to remain as general as possible.

Table C.1: Prior Parameter Specification

Model parameters	Prior parameters	
Factor loadings	$a_0 = 0.0$	$A_0 = 3.0$
Regression coefficients	$b_0 = 0.0$	$B_0 = 10.0$
Mixture components for wage equations:		
Weights	$p = 1.0$	
Means	$m_0 = 0.0$	$M_0 = 1.0$
Variances	$g_0 = 2.0$	$G_0 = 1.0$

C.2 MCMC sampling: special features of our algorithm

We implement a Gibbs sampler that draws the parameters and the latent variables iteratively from their posterior distributions (Casella and George, 1992; Smith and Roberts, 1993). Starting with random values, each parameter is sampled conditional on the current values of the other parameters, and the procedure is repeated until convergence to the stationary distribution is achieved. The values obtained from the first iterations are discarded ("burn-in period"), and only the last ones corresponding to the target distribution are saved for posterior inference. The detailed steps of the algorithm with the corresponding conditional distributions are presented in Section C.3.

Bayesian inference for ordinal variable models. The standard approach proposed by Albert and Chib (1993), which consists of drawing the thresholds sequentially from their uniform conditional distributions, generates Markov chains that are highly correlated, which can prevent the algorithm to explore the whole posterior distribution of the parameters. As noted by Cowles (1996), the high correlation between the cutpoints γ and the latent response variable $M^\star$ results in a slow convergence and in a poor mixing of the Markov chain for the parameters of the ordinal equations. In the end, this can bias the inference if the algorithm fails to generate a sample that is representative of the posterior distribution of interest.

To remedy this problem, several technical improvements have been proposed. Cowles (1996) introduced a Hastings-within-Gibbs step in the algorithm to draw the cut-points and the latent response variable simultaneously, while Nandram and Chen (1996) proposed a simple reparameterization that proves to be particularly effective, especially in the three-category case. More recent approaches based on parameter expansion (Liu and Wu, 1999) or marginal data augmentation (van Dyk and Meng, 2001; van Dyk, 2010) can be applied as well for this purpose. Alternatively, we opt for the approach introduced by Liu and Sabatti (2000a), which consists of applying group transformations in the sample space of the parameters to improve the efficiency of the sampler, without affecting the target distribution of the parameters.[8] This procedure speeds up convergence and enhances the mixing of the chain, while being less computationally burdensome than other methods.

Missing data problem: To sample or not to sample? The missing data problem results in an unbalanced sample, where for some individuals only the measurements and the schooling decision are observed, while for others only schooling and outcomes are available. Section 3.2 of the paper explains how to integrate out the unobserved variables to deal with this issue. From a Bayesian perspective, two different approaches can be adopted to achieve this goal: The missing variables can either be integrated out analytically or numerically. In the former case, any left-hand side variable that is not observed disappears from the likelihood function, and as a consequence the sampler only uses the available information to draw the corresponding parameters of each equation. For example, to sample the factor loadings in the outcome equations, only the individuals with non-missing outcomes – but possibly missing measurements – are used. In the latter case, the missing variables are simulated during sampling to do the integration numerically. This is done by adding one step to the algorithm, where each of the missing variables is sampled from its conditional distribution for the corresponding individuals. This procedure represents another application of data augmentation methods. It restores the balance of the sample, as it makes the number of available values the same for all variables (a combination of observed and simulated values).

[8]In our case, one separate group transformation is applied for each ordinal equation M_k, for $k = 1, \dots, K$.

Which of the two approaches should be favored in practice? The answer to this question depends on the objectives of the analyst. Both methods are equivalent and provide the same numerical results, but the resulting Markov chains have different statistical properties. The analytical integration is more efficient, as it gets rid of any missing information that would otherwise have to be simulated. The simulation of the missing values can indeed be unattractive from a statistical point of view, as it increases the level of autocorrelation between the parameters. Intuitively, during the iterative process the missing variables are simulated conditional on the parameters, and the parameters are subsequently sampled conditional on the simulated variables. This can hinder convergence, and compromise the mixing of the Markov chain. For these reasons, some authors advocate to "collapse" the Gibbs sampler, i.e., to integrate out analytically any parameters or variables that are not relevant for the analysis, whenever possible (Liu, 1994; Liu *et al.*, 1994; van Dyk and Park, 2008).

In some cases, however, the analyst might be interested in recovering the missing values, for instance to study the distribution of counterfactual outcomes.[9] In this scenario, the numerical integration based on data augmentation should be used, as the benefits of simulating the missing values clearly outweigh the loss of efficiency introduced by the procedure. Longer Markov chains (i.e., more MCMC iterations) are one solution to address this issue.

C.3 MCMC sampler

The Gibbs sampler draws the parameters and the augmented data of the model sequentially from their conditional distributions (see Casella and George, 1992; Smith and Roberts, 1993). At each step, each parameter is updated conditionally on the data and on the current values of the other parameters. We provide below the corresponding posterior distributions, omitting to specify the conditioning sets to simplify the notation.

As discussed in Section C.2, the unobserved measurements and outcomes can either be integrated out analytically or numerically. In the latter case, additional steps are added to the algorithm to sample the missing values. In both cases, we denote $\mathcal{I}_Y$ the set of indices corresponding to the individuals included in the analysis to update the

[9] Counterfactuals are the potential outcomes the individuals would have received had they been in the other schooling group, which can never be observed in practice, but simulated in the framework of our analysis.

corresponding parameters of equation Y (i.e., $\mathcal{I}_Y = \{1, \ldots, N\}$ when data augmentation is implemented for the numerical integration of the missing values, and $\mathcal{I}_Y$ is the subset of individuals for whom Y is observed in the case of the analytic integration).

Starting with some initial values (either fixed to some pre-specified values or drawn randomly), the Gibbs sampler proceeds by drawing the parameters and the augmented variables sequentially from the following conditional distributions, until practical convergence is achieved:

1. **Update the mixtures of normals for the error terms of the wage equations.** Data augmentation is applied through the introduction of binary variables z_{ih}^s to indicate group membership $h = 1, \ldots, H_s$ of each individual $i = 1, \ldots, N$, for each potential wage $s = 0, 1$ (see Diebolt and Robert, 1994). Let $z_{ih}^s = 1$ if individual i belongs to mixture group h, and to zero otherwise.

   (a) **Update the group indicators.** Each individual $i \in \mathcal{I}_{Y_s}$ is allocated to mixture component $h = 1, \ldots, H_s$ with probability:

   $$\Pr(z_{ih}^s = 1) = \frac{\pi_h^s \, \phi\left(\widetilde{Y}_{si}; \mu_h^s, (\omega_h^s)^2\right)}{\sum_{l=1}^{H_s} \pi_l^s \, \phi\left(\widetilde{Y}_{si}; \mu_l^s, (\omega_l^s)^2\right)},$$

   where $\phi\left(\widetilde{Y}_{si}; \mu_h^s, (\omega_h^s)^2\right)$ is the probability density function of the normal distribution with mean μ_h^s and variance $(\omega_h^s)^2$ evaluated at $\widetilde{Y}_{si}$, with $\widetilde{Y}_{si} = Y_{si} - X_{Yi}\beta_Y^s - \theta_i \alpha_Y^s$, denoting X_{Yi} the ith row of the matrix of covariates X_Y.

   (b) **Mixture weights:**

   $$\pi_s \sim \mathcal{D}ir\left(p + n_1^s, \ldots, p + n_{H_s}^s\right),$$

   where $n_h^s = \sum_{i \in \mathcal{I}_{Y_s}} z_{ih}^s$ is the number of individuals in mixture group h.

   (c) **Update the mixture variances:**

   $$(\omega_h^s)^2 \sim \mathcal{G}^{-1}\left(g_0 + \frac{n_h^s}{2}; \, G_0 + \frac{1}{2} \sum_{\substack{i \in \mathcal{I}_{Y_s} \\ z_{ih}^s = 1}} (\widetilde{Y}_{si})^2\right).$$

(d) **Update the mixture means.** To satisfy the zero mean restriction of the mixture, sample the first $H_s - 1$ mixture means as:

$$(\mu_1^s, \ldots, \mu_{H_s-1}^s) \sim \mathcal{N}\left(\widetilde{\mu}_s; \widetilde{\Omega}_s\right),$$

with:

$$\widetilde{\Omega}_s^{-1} = \operatorname{diag}_{h=1}^{H_s-1} \left\{ n_h^s (\omega_h^s)^{-2} + M_0^{-1} \right\} + \frac{\pi_{-H_s} \pi'_{-H_s}}{\pi_{H_s}} \left\{ n_{H_s}^s (\omega_{H_s}^s)^{-2} + M_0^{-1} \right\},$$

$$\widetilde{\mu}_s = \widetilde{\Omega}_s \operatorname{vec}_{h=1}^{H_s-1} \left\{ (\omega_h^s)^{-2} \sum_{\substack{i \in \mathcal{I}_{Y_s} \\ z_{ih}^s = 1}} \widetilde{Y}_{si} + \frac{m_0}{M_0} - \frac{\pi_h^s}{\pi_{H_s}^s} \left[(\omega_{H_s}^s)^{-2} \sum_{\substack{i \in \mathcal{I}_{Y_s} \\ z_{iH_s}^s = 1}} \widetilde{Y}_{si} + \frac{m_0}{M_0} \right] \right\},$$

where $\pi_{-H_s} = (\pi_1^s, \ldots, \pi_{H_s-1}^s)'$, and diag $\{\cdot\}$ (resp., vec $\{\cdot\}$) is the matrix operator that creates a diagonal matrix (resp., a column vector) with the corresponding elements specified.

Compute the last mixture mean as $\mu_{H_s}^s = -\sum_{h=1}^{H_s-1} \pi_h^s \mu_h^s / \pi_{H_s}^s$.

(e) **Transform the variables** of the wage equations as follows, for all $i \in \mathcal{I}_{Y_s}$:

$$Y_{si}^{\text{mix}} \leftarrow (Y_{si} - \mu_{g^{s_i}}^s)/\omega_{g^{s_i}}^s, \qquad X_{Yi}^{\text{mix}} \leftarrow X_{Yi}/\omega_{g^{s_i}}^s, \qquad \theta_i^{\text{mix}} \leftarrow \theta_i/\omega_{g^{s_i}}^s, \qquad (6)$$

where $g_i^s = \sum_{h=1}^{H_s} h\, z_{ih}^s$ denotes the group membership of individual i. With these transformed variables, the conditional distribution of wages (Y_s^{mix}) becomes a normal distribution conditional on group membership of the individuals, which simplifies the sampling of the other parameters.

2. **Update the factor loadings.** For each equation $W = S, D_0, D_0, Y_0, Y_1, M_1, \ldots, M_K$:

$$\alpha_W \sim \mathcal{N}(a_{\alpha_W}; A_{\alpha_W}), \qquad A_{\alpha_W}^{-1} = \sum_{i \in \mathcal{I}_W} \theta_i^2 + A_0^{-1},$$

$$a_{\alpha_W} = A_{\alpha_W} \left(\sum_{i \in \mathcal{I}_W} \theta_i \widetilde{W}_i + \frac{a_0}{A_0} \right),$$

for each loading $\alpha_W = \alpha_S, \alpha_D^0, \alpha_D^1, \alpha_Y^0, \alpha_Y^1, \alpha_{M_1}, \ldots, \alpha_{M_K}$, where $\widetilde{W}_i = W_i^\star - X_{Wi}\beta_W$ for the corresponding covariates $X_W = X_S, X_D, X_Y, X_M$. Note that in the wage equations, the transformed variables Y_s^{mix}, X_Y^{mix} and θ^{mix} obtained from Eq. (6) are used to deal with the mixture of normals.

3. **Update the regression coefficients.** For each equation $W = S, D_0, D_0, Y_0, Y_1, M_1, \ldots, M_K$:

$$\beta_W \sim \mathcal{N}(b_{\beta_W};\, B_{\beta_W}), \qquad B_{\beta_W}^{-1} = \sum_{i \in \mathcal{I}_W} X'_{Wi} X_{Wi} + B_0^{-1} I,$$

$$b_{\beta_W} = B_{\beta_W} \left(\sum_{i \in \mathcal{I}_W} X'_{Wi}(W_i^\star - \theta_i \alpha_W) + \frac{a_0}{A_0} \iota \right),$$

where $\beta_W = \beta_S, \beta_D^0, \beta_D^1, \beta_Y^0, \beta_Y^1, \beta_{M_1}, \ldots, \beta_{M_K}$ is the vector of regression coefficients corresponding to the covariates $X_W = X_S, X_D, X_Y, X_M$, and in the wage equations the transformed variables Y_s^{mix}, X_Y^{mix} and θ^{mix} are used to deal with the mixture of normals (see Eq. (6)).

4. **Update the latent variables underlying the variables:**

$$W_i^\star \sim \mathcal{TN}_{[\mathcal{W}_i]}\left(X_{Wi}\beta_W + \theta_i \alpha_W;\, 1\right), \qquad \text{for } i \in \mathcal{I}_W,$$

for each equation $W = S, D_0, D_0, M_1, \ldots, M_K$, using the corresponding factor loadings $\alpha_W = \alpha_S, \alpha_D^0, \alpha_D^1, \alpha_{M_1}, \ldots, \alpha_{M_K}$, the covariates $X_W = X_S, X_D, X_M$, and the regression coefficients $\beta_W = \beta_S, \beta_D^0, \beta_D^1, \beta_{M_1}, \ldots, \beta_{M_K}$. The normal distribution is truncated to the interval $\mathcal{W}_i$ that depends on the type and on the observed value of the corresponding variable W_i, for each individual $i = 1, \ldots, N$: In the binary case (schooling and labor market participation), $\mathcal{W}_i = (-\infty; 0]$ if $W_i = 0$ and $\mathcal{W}_i = (0; +\infty)$ if $W_i = 1$. In the ordinal case (measurements), $\mathcal{W}_i = (\gamma_{c-1}, \gamma_c)$ if $W_i = c$.

Note that if the missing values are integrated out numerically, they are automatically sampled from their conditional distributions (since $\mathcal{I}_W$ is the set of indices of all included individuals, including those with missing values in the case of numerical integration). In this case, there is no truncation, i.e., $\mathcal{W}_i = (-\infty, +\infty)$.

If necessary (i.e., if the missing values of the wages are integrated out numerically), impute the missing wages from the following mixture of normals, for $s = 0, 1$:

$$Y_{si} \sim \sum_{h=1}^{H_s} \pi_h^s \mathcal{N}\big(\mu_h^s + X_{Yi}\beta_Y^s + \theta_i \alpha_Y^s;\, (\omega_h^s)^2\big).$$

5. **Update the thresholds of the ordinal equations in the measurement system.** For each measurement M_k, $k = 1, \ldots, K$, update each cutoff points as follows, conditional on all the other cutoff points (see Albert and Chib, 1993):

$$\gamma_{k,c} \sim \mathcal{U}\left(\max \left\{ \max_{\substack{i \in \mathcal{I}_{M_k} \\ M_{ki}=c}} (M^\star_{ki}); \gamma_{k,c-1} \right\}; \min \left\{ \min_{\substack{i \in \mathcal{I}_{M_k} \\ M_{ki}=c+1}} (M^\star_{ki}); \gamma_{k,c+1} \right\} \right),$$

for each $c = 2, \ldots, C_k - 1$ (remember that $\gamma_{k,1} = 0$ for identification, and $\gamma_{k,0} = -\infty$ and $\gamma_{k,C_k} = +\infty$).

6. **Apply the Liu and Sabatti (2000b) transformation to the ordinal equations** to boost sampling. For each equation M_k of the measurement system, $k = 1, \ldots, K$, consider the following group transformation that rescales the underlying latent part:

$$\Gamma_{\nu_k} = \left\{ \left([\delta_k M^\star_{ki}]_{i \in \mathcal{I}_{M_k}}, \delta_k \beta_{M_k}, \delta_k \alpha_{M_k}, \delta_k \gamma_k \right), \delta_k > 0 \right\},$$

where ν_k is the number of parameters to transform, i.e., $\nu_k = N_k + p_k + C_k - 1$, where $N_k = \text{card}(\mathcal{I}_{M_k})$ is the number of included individuals for whom the latent variable $M^\star_{ki}$ needs to be rescaled, p_k is the number of regression coefficients in β_{M_k}, $C_k - 2$ is the number of unrestricted cutoff points, and there is a single factor loading.

The (squared) transformation parameter δ_k^2 is sampled from the following Gamma distribution:

$$\delta_k^2 \sim \mathcal{G}\left(\frac{\nu_k + 1}{2}; \frac{1}{2} \left[\sum_{i \in \mathcal{I}_{M_k}} (M^\star_{ki} - X_{Mi}\beta_{M_k} - \theta_i \alpha_{M_k})^2 + \frac{\alpha^2_{M_k}}{A_0} + \frac{\beta'_{M_k}\beta_{M_k}}{B_0} \right] \right),$$

and the parameters and latent variables are then rescaled as follows:

$$M^\star_{ki} \leftarrow \delta_k M^\star_{ki}, \qquad \beta_{M_k} \leftarrow \delta_k \beta_{M_k}, \qquad \alpha_{M_k} \leftarrow \delta_k \alpha_{M_k}, \qquad \gamma_k \leftarrow \delta_k \gamma_k.$$

7. **Update the latent factor.** Denote $\widetilde{Y}_i^\star$ the vector of demeaned variables and α_i the vector of factor loadings for individual i:

$$\widetilde{Y}_i^\star = \begin{pmatrix} S_i^\star - X_{Si}\beta_S \\ D_{0i}^\star - X_{Di}\beta_{D_0} \\ D_{1i}^\star - X_{Di}\beta_{D_1} \\ (Y_{0i} - X_{Yi}\beta_{Y_0} - \mu^0_{g_i^0})/\omega^0_{g_i^0} \\ (Y_{1i} - X_{Yi}\beta_{Y_1} - \mu^0_{g_i^1})/\omega^1_{g_i^1} \\ M_{1i}^\star - X_{Mi}\beta_{M_1} \\ \vdots \\ M_{Ki}^\star - X_{Mi}\beta_{M_K} \end{pmatrix}, \qquad \widetilde{\alpha}_i = \begin{pmatrix} \alpha_S \\ \alpha_{D_0} \\ \alpha_{D_1} \\ \alpha_{Y_0}/\omega^0_{g_i^0} \\ \alpha_{Y_1}/\omega^1_{g_i^1} \\ \alpha_{M_1} \\ \vdots \\ \alpha_{M_K} \end{pmatrix},$$

where for the wage equations the variables and factor loadings are transformed appropriately to take into account the mixtures of normals, considering that individual i belongs to mixture group g_i^s, for $s = 0, 1$. Using this notation, the latent factor θ_i is sampled from:

$$\theta_i \sim \mathcal{N}\left(\frac{\widetilde{\alpha}_i' \widetilde{Y}_i^\star}{\widetilde{\alpha}_i'\widetilde{\alpha}_i + 1/\sigma_\theta^2}; \frac{1}{\widetilde{\alpha}_i'\widetilde{\alpha}_i + 1/\sigma_\theta^2} \right),$$

where in our case $\sigma_\theta^2 = 1$ for identification purpose (see Section B of the paper).

When the missing values are integrated out numerically (e.g., missing measurements for the adults), only the subvectors of $\widetilde{Y}_i^\star$ and $\widetilde{\alpha}_i$ corresponding to the observed variables for individual i are considered to sample the latent factor. For example, the latent factor of an individual from the adult sample who achieved higher education would be updated using the three corresponding equations:

$$\widetilde{Y}_i^\star = \begin{pmatrix} S_i^\star - X_{Si}\beta_S \\ D_{1i}^\star - X_{Di}\beta_{D_1} \\ (Y_{1i} - X_{Yi}\beta_{Y_1} - \mu^0_{g_i^1})/\omega^1_{g_i^1} \end{pmatrix}, \qquad \widetilde{\alpha}_i = \begin{pmatrix} \alpha_S \\ \alpha_{D_1} \\ \alpha_{Y_1}/\omega^1_{g_i^1} \end{pmatrix}.$$

8. **Update the variance of the latent factor.** *Note that this step is not required in our analysis, since we are assuming $\sigma_\theta^2 = 1$ to set the scale of the factor (see Section B of the paper). We provide it here for the sake of completeness.*

Assuming that the variance σ_θ^2 is a priori following an inverse-Gamma distribution with shape parameter g_0^θ and scale parameter G_0^θ, the posterior distribution is:

$$\sigma_\theta^2 \sim \mathcal{G}^{-1}\left(g_0^\theta + \frac{N}{2};\, G_0^\theta + \frac{1}{2}\sum_{i=1}^{N}\theta_i^2\right).$$

Note on the potential label-switching problem affecting the mixture of normals in the wages equations. The prior specification on the mixture of normals used in the paper does not secure the identification of the individual mixture parameters, as the mixture components might be switched without affecting the likelihood. This label-switching problem has drawn a lot of interest in the literature on mixture modeling (Celeux, 1998; Stephens, 2000a;b), and solutions have been proposed to solve it (for a review, see Frühwirth-Schnatter, 2006, Section 3.5.5). However, it is not a concern in our case, as we only use the mixture to estimate the distribution of the error terms in a flexible ways. This distribution is identified, even if the single mixture components are not. For this reason, we do not add any additional restrictions.

D Monte Carlo Study: Experimental Setup

Data generation. We run several experiments using synthetic data generated from the model described in Section 3.1 of the paper. Each of the data sets contains five ordinal measurements (M), one schooling equation (S) and two potential outcome equations (Y). The variables are censored to create two distinct subsamples: a *youth sample* where only M and S are observed, and an *adult sample* where only S and Y are observed. Therefore, the schooling equation links the two subsamples. We also allow the subsamples to overlap, and look at four scenarios: full overlap (i.e., no missing data problem), no overlap (no individuals with M, S and Y observed simultaneously), and partial overlap (all variables available for 40% or 13% of the individuals, where the latter case corresponds to our empirical application).[10] The schooling equation is specified either as binary or as continuous,[11] so as to better apprehend the role played by this equation in linking the two data sets.

In this simplified model, the *youth sample* allows to fully identify the measurement system M and the schooling equation S, but not the outcome system Y. The *adult sample*, on the contrary, does not allow to identify the measurement system. Remarkably, this second sample does not identify the outcome system either, as two covariances are not enough to identify three factor loadings: only $\mathrm{Cov}(S^\star; Y_0 \mid X)$ and $\mathrm{Cov}(S^\star; Y_1 \mid X)$ are available, since the covariance between the two potential outcomes $\mathrm{Cov}(Y_0; Y_1 \mid X)$ can neither be observed nor estimated, but this is not enough to identify α_S, α_Y^0, α_Y^1. However, the combination of both samples, as well the overlap sample, provide enough information to solve the missing data problem.

We vary in different ways the amount of information that is available from the data to identify the distribution of the latent factor. The more overlap between the two samples, or the larger the number of observations, the easier it should be to extract information to proxy the factor, thus the more precisely we should be able to measure its impact on the outcomes. Similarly, the inference should be facilitated when the schooling equation is continuous rather than binary, as continuous variables bring more information to the table.

[10]The subsamples are balanced (same number of observations in both subsamples) in all cases except the 13% overlap case that is simulated to mimic the data of our empirical application: 39% in the adult sample only, 48% in the youth sample only, and the remaining 13% in the overlap sample (see Table 2 in paper).

[11]In the continuous schooling case, the potential outcomes are observed based on the sign of the schooling variable.

For the latent part of the model, we specify the factor loadings as follows:

$$\alpha = \begin{pmatrix} \alpha_S & \alpha_Y^0 & \alpha_Y^1 & \alpha_{M_1} & \dots & \alpha_{M_5} \end{pmatrix} = \begin{pmatrix} 0.4 & -0.3 & 0.3 & 1.0 & 0.9 & 0.8 & 0.7 & 0.6 \end{pmatrix},$$

and simulate the unobservables from normal distributions, for $s = 0, 1$ and $k = 1, \dots, 5$:

$$\theta \sim \mathcal{N}(0;\, 1)\,, \qquad \varepsilon_S \sim \mathcal{N}(0;\, 1)\,, \qquad \varepsilon_Y^s \sim \mathcal{N}(0;\, 0.2)\,, \qquad \varepsilon_{M_k} \sim \mathcal{N}(0;\, 1)\,.$$

The same variance is used for the error term of the schooling equation in the binary case and in the continuous case, even if in most applications the idiosyncratic variance would typically be much lower in the continuous case. Keeping the same signal-to-noise ratio allows us to better compare the results across the two specifications.

In each equation, we include an intercept term and four covariates simulated independently from standard normal distributions, i.e., $X_S = X_Y = X_M \equiv X = \begin{pmatrix} \iota & \widetilde{X} \end{pmatrix}$ where $\widetilde{X}_i \sim \mathcal{N}(0;\, I_4)$ for $i = 1, \dots, N$. The corresponding regression coefficients are simulated as:

$$\begin{aligned} \beta_S &= \begin{pmatrix} 0.5 & \widetilde{\beta}_S \end{pmatrix}, & \beta_Y^s &= \begin{pmatrix} 0.5 & \widetilde{\beta}_Y^s \end{pmatrix}, & \beta_{M_k} &= \begin{pmatrix} 2.0 & \widetilde{\beta}_{M_k} \end{pmatrix}, \\ \widetilde{\beta}_S &\sim \mathcal{N}(0;\, I_4)\,, & \widetilde{\beta}_Y^s &\sim \mathcal{N}(0;\, I_4)\,, & \widetilde{\beta}_{M_k} &\sim \mathcal{N}(0;\, I_4)\,, \end{aligned}$$

for $s = 0, 1$ and $k = 1, \dots, 5$. The latent utilities of the measurements are discretized using the thresholds $\gamma_k = (-\infty, 0, 2, 4, +\infty)$, for $k = 1, \dots, 5$, creating ordinal variables on a 4-category Likert scale ($C_k = 4$).

We generate data sets of dimensions $N = 1,500$ (similar to our empirical application) and $N = 10,000$, and replicate the simulations 100 times for each Monte Carlo experiment. The model parameters specified above (factor loadings, regression coefficients, idiosyncratic variances and thresholds) are fixed across Monte Carlo replications, only the latent factor, the covariates and the error terms are resampled, thus generating new measurements, schooling and potential outcomes. With this parameter specification, the proportion of treated individual is equal to 0.58 in the binary schooling case, and to 0.50 in the continuous case.

MCMC Sampling. The formal identification of the factor model is achieved by fixing the variance of the latent factor to 1. The MCMC sampler is run without any additional restrictions, which implies that a sign-switching problem of the factor loadings can arise.

We solve this problem *a posteriori* by processing the draws to restore the correct signs. This post-processing is based on the factor loading of the first measurement equation, α_{M_1}, which is assumed to be always positive.[12] The prior parameters are specified in Table C.1. Markov chains with $30,000$ iterations are simulated for each experiment, where the first $10,000$ iterations are discarded as burn-in period.

In some cases, the factor loading of the schooling equation converges to zero, and the loadings in the outcome system are nonsensical. The same phenomenon can be observed in our empirical application. This appears to be a symptom that the data combination strategy fails to empirically identify the full model. In these particular cases, the algorithm gets trapped in an area of the parameter space where the latent factor captures two different traits in the two subsamples, and these two interpretations cannot not be reconciled in the common schooling equation – thus resulting in a zero factor loading α_S. Clearly, the bridge between the two subsamples is broken and the data combination fails. Fortunately, these ill-behaved cases can easily be detected and discarded by re-running the algorithm with different starting values, and checking that α_S converges to a nonzero value.

[12]A posteriori, for each MCMC iteration $r = 1, \ldots, R$, set $\alpha^{(r)} \leftarrow -\alpha^{(r)}$ if $\alpha_{M_1}^{(r)} < 0$. Frühwirth-Schnatter and Lopes (2010), Conti *et al.* (2014) use a similar posterior approach to solve the sign-switching problem.

E Data Addendum

Our data come from the German Socioeconomic Panel (SOEP), a representative longitudinal micro-data set that contains a wide range of socio-economic information on individuals in Germany, comprising follow-ups for the years 1984-2011. Information was first collected from about 12,200 randomly selected adult respondents in West Germany in 1984. After German reunification in 1990, the SOEP was extended to around 4,500 persons from East Germany, and subsequently supplemented and expanded by additional samples. The data are well suited for our analysis in that they allow us to exploit information on a wide range of background variables, locus of control and wages, for a representative panel of individuals. Furthermore, the inclusion of a special youth survey, comprising information on 17-year-old individuals, allows us to obtain background variables and locus of control measures for individuals who have not yet entered the labor market.

E.1 Combining samples

Our focus is to analyze the impact of locus of control and to purge our estimates of measurement error and endogeneity problems. Hence, to investigate how locus of control affects schooling decisions and wages, respectively, we would ideally need a sample of individuals for whom locus of control measures are collected at several points in time: first, prior to schooling, then at the time when individuals make education decisions, and third, at a time just before they start working on the labor market. However, we only have access to one measure of what we term 'pre-market' locus of control. This measure is taken when individuals are 17 years of age, just after compulsory schooling, but before they enter the labor market.[13] We then combine the sample of youth for which we have 'pre-market' locus of control measures with a sample of young adults for whom we observe labor market outcomes. We draw our samples on the basis of selection criteria that are explained in the following.

E.1.1 Youth sample

Our youth sample is composed of 1,901 individuals born between 1984 and 1994, all of which are children of SOEP panel members. A comprehensive set of background

[13]Locus of control measures have also been collected for a cross section of young adults in 2005 and 2010. We use this information to compare distributions of locus of control and to investigate how much these locus of control measures are flawed by previous labor market experience.

variables, schooling choices, as well as locus of control measures of these individuals, have been collected in the years 2001-2011, when the subjects were 17 years of age. After the first interview at age 17, all subjects are subsequently interviewed on a yearly basis until early adulthood. For example, in 2011, the oldest youth are 27 years of age. An exception to the age rule was made for the 2001 wave, such that some subjects were already 18 or 19 years of age when first completing the questionnaire. We exclude these individuals from our sample. To ensure that our results are not flawed by post 1991 schooling and labor market adjustments, all individuals who went to school in East Germany (the former German Democratic Republic) have been excluded. Last, we exclude all individuals with missing locus of control measures, missing schooling information, or missing information among the covariates.

E.1.2 Adult sample

The adult sample used for our analysis comprises information on 1,606 individuals, aged 26-35, who are drawn from all West German representative subsamples We construct a cross-section of individuals based on the most recent information available from the waves 2004-2011. Hence, most of our information on the adult sample stems from the 2011 wave. However, if some important pieces of information on certain individuals in that wave are missing, they are filled up with information from 2010. If the information in the 2009 wave is also missing, information from 2006 is used, and so on.

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

E.2 'Pre-market' locus of control

In the SOEP, locus of control is measured by a 10-item questionnaire. However, the number of possible answers differs between the years 2001-2005, where a 7-point scale was used, and the years 2006-2011, where a 4-point item scale was used. To make the questionnaire comparable across samples, we transform the 7-point scale into a 4-point scale by assigning the middle category (4) either to category 2 or 3 of the 4-item scale, depending on the most probable answer. For example, if in the 2005 sample most youth answered "completely agree," people who answered "indifferent" in the 2006 sample are assumed to tend toward the "slightly agree" answer. After transforming answers to have the same scale, each question is answered on a Likert scale ranging from 1 ("completely disagree ") to 4 ("completely agree").

E.3 School choice

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

E.4 Wage construction and labor market participation

Wages are constructed by using most recent wage information available from the SOEP. Whenever occurring, missing wage information was substituted by wage information obtained in one of the earlier years. Wages have been inflation adjusted to match 2011 wage levels (inflation rates obtained from Eurostat). Wages are assigned a missing whenever the respective individual is indicating not to have a regular (full time or part time) job. We exclude other types of employment such as marginal employment, to ensure that we are not including typical student jobs.

Hourly wages have been constructed by dividing gross monthly wages by the actual number of hours worked in the last month before the interview. Log hourly wages are then obtained by taking the natural logarithm of the hourly wage variable. To account for outliers, we trim hourly wages below the first and above the ninety ninth percentiles. All individuals who indicate a positive wage are classified as labor market participants. The low levels of labor market participation arise because many individuals still participate in education or training. In fact, education measures are part of active labor market policies in Germany, such that almost all young individuals not currently employed are enrolled in education or training programs.

E.5 Covariates

In our measurements system, schooling equation and outcome equations, we control for a large set of background variables (see Table E.2). The locus of control factor distribution is identified from the covariance structure of the unobservables of the model. Hence, any controls in the measurement system purge our measures of locus of control of any effects which are captured by the covariates. Thus, the covariates in place should be uncorrelated with the latent trait we want to capture, since in our model the latent factor has to be uncorrelated with these covariates by construction. In the following, a brief description of the different categories of covariates is provided. Descriptive statistics are shown in Tables E.3 and E.4.

Parental education and investment

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

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

Table E.2: Samples and included covariates

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

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

Region dummies and city size

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

- North: Berlin, Bremen, Hamburg, Lower Saxony, Schleswig-Holstein,

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

Unemployment rates

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

Marital status and number of children

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

Track recommendation after elementary school

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

Table E.3: Descriptive statistics: covariates in the measurement system

Variables	Mean and (sd)			
	Males		**Females**	
Broken home	0.22	(0.41)	0.22	(0.41)
Number of siblings	0.98	(1.22)	1.01	(1.22)
Father grammar school	0.30	(0.46)	0.33	(0.47)
Father dropout	0.03	(0.16)	0.03	(0.18)
Mother grammar school	0.24	(0.43)	0.26	(0.44)
Mother dropout	0.01	(0.10)	0.03	(0.16)
Recommendation: grammar school	0.39	(0.49)	0.45	(0.50)
Recommendation: general secondary school	0.16	(0.37)	0.13	(0.34)
Childhood in large city	0.21	(0.41)	0.22	(0.42)
Childhood in medium city	0.18	(0.39)	0.20	(0.40)
Childhood in small city	0.28	(0.45)	0.25	(0.43)
North	0.23	(0.42)	0.24	(0.43)
South	0.33	(0.47)	0.35	(0.48)
N	962		949	

Source: SOEP, cross section using most recent information from
the waves 2004-2011. Own calculations.

Notes: p-values of a two-sided t-test for differences in means are reported.

Table E.4: Descriptive statistics: covariates in the outcome equations (by schooling)

Variables	**Males**			**Females**		
	Low	**High**	**Pval**	**Low**	**High**	**Pval**
Age	26.63	25.85	0.00	26.52	25.88	0.01
Broken home	0.19	0.16	0.12	0.22	0.16	0.00
Father grammar school	0.10	0.47	0.00	0.10	0.46	0.00
Father dropout	0.03	0.01	0.03	0.04	0.01	0.00
Mother grammar school	0.10	0.35	0.00	0.09	0.35	0.00
Mother dropout	0.01	0.00	0.04	0.03	0.01	0.00
Married	0.14	0.11	0.03	0.19	0.11	0.00
Number of Children	0.24	0.13	0.00	0.32	0.14	0.00
Childhood in large city	0.16	0.26	0.00	0.18	0.25	0.00
Childhood in medium city	0.18	0.19	0.58	0.19	0.20	0.43
Childhood in small city	0.25	0.28	0.18	0.24	0.25	0.74
North	0.20	0.23	0.10	0.20	0.22	0.28
South	0.32	0.28	0.04	0.36	0.27	0.00
Local unemployment rate	7.62	7.50	0.46	7.58	7.63	0.77
N		1584			1532	

Source: SOEP, cross section using most recent information.
from the waves 2004-2011. Own calculations.

Notes: p-values of a two-sided t-test for differences in means are reported.

F Exploratory Factor Analysis

To determine the underlying dimensionality of locus of control, as measured by the items displayed in Table F.5, we conduct a scree plot analysis displayed in Figure F.1. With two eigenvalues larger than one this analysis suggests two underlying factors. A scatter plot of the respective factor loadings (Fig. F.2), with the first two principal factors on the axis, shows that the external locus of control measures (Q3, Q4, Q6, Q7 and Q9) load very highly on the external locus of control factor (factor 1), while the other items have a loading close to zero (mainly Q1, Q5, Q8 and Q10). Furthermore, the items with a close to zero loading are items that were designed to capture an internal attitude, while the other items mostly capture the external dimension of locus of control.

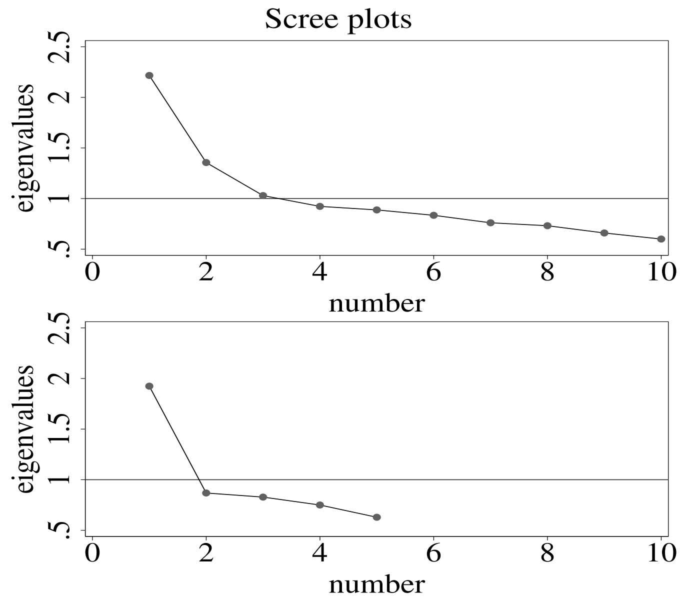
We can draw two conclusions from the item design and from our exploratory factor analysis. First, researchers who use an index, constructed for example as the standardized mean of the items, instead of a latent factor, force each of the measurement items to enter the index with an equal weight. Doing this yields a locus of control measure that is potentially flawed by measurement error, and to coefficients that are biased downward due to attenuation bias. Second, a unidimensional factor of locus of control captures mostly the external attitude dimension of locus of control. This motivates our decision to focus on the external dimension of locus of control in our paper.

Table F.5: Locus of control, youth sample

	Variables	Mean and (sd)			
		Males		**Females**	
Q1	My life's course depends on me	3.57	(0.61)	3.51	(0.60)
Q2	I have not achieved what I deserve	2.06	(0.85)	1.92	(0.80)
Q3	**Success is a matter of fate or luck**	2.22	(0.82)	2.33	(0.78)
Q4	**Others decide about my life**	2.16	(0.83)	2.12	(0.83)
Q5	Success is a matter of hard work	3.50	(0.60)	3.52	(0.57)
Q6	**In case of difficulties, doubts about own abilities**	2.03	(0.81)	2.32	(0.86)
Q7	**Possibilities in life depend on social conditions**	2.70	(0.78)	2.71	(0.75)
Q8	Abilities are more important than effort	3.05	(0.70)	3.06	(0.68)
Q9	**Little control over what happens to me**	1.91	(0.76)	1.95	(0.76)
Q10	Social involvement can help influence social cond	2.49	(0.86)	2.51	(0.78)
N		962		949	

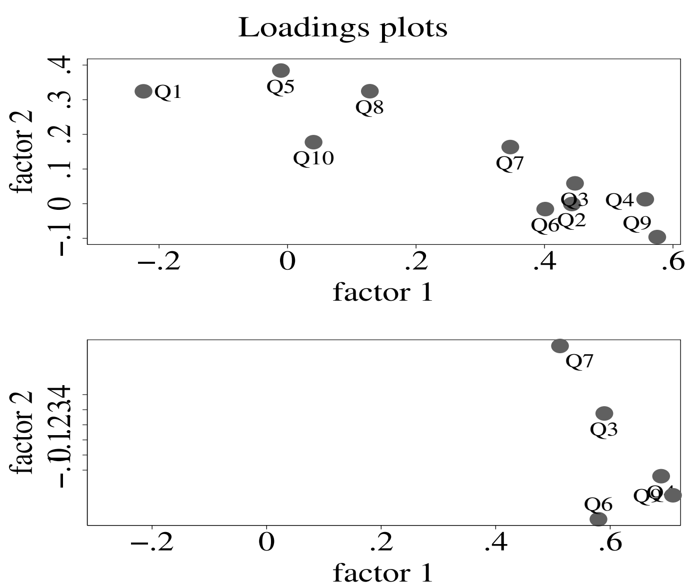
Means and (sd) displayed. Locus of control Answers: 1 (disagree completely) to 4 (agree completely).
External Locus of Control items in bold.
Source: SOEP youth sample 2000-2011.

Figure F.1: Scree plot: all measurements versus 5 'external' items only



SOEP, own calculations.

Figure F.2: Scatterplot of loadings: all measurements versus 6 'external' items only



GSOEP, own calculations.

G Empirical Application: Additional Posterior Results

This section provides posterior results for the regression coefficients of the measurement and outcome systems.

Table G.6: Regression coefficients in measurement system, males (locus of control reverse coded)

Variables	Loc 3		Loc 4		Loc 6		Loc 7		Loc 9	
Constant	0.99	(0.12)	0.97	(0.14)	0.65	(0.11)	1.65	(0.12)	0.63	(0.12)
Childhood in large city	0.05	(0.12)	-0.07	(0.14)	-0.04	(0.12)	-0.15	(0.11)	-0.15	(0.13)
Childhood in medium city	0.00	(0.12)	0.10	(0.14)	-0.02	(0.12)	0.03	(0.12)	0.11	(0.13)
Childhood in small city	0.02	(0.11)	0.02	(0.12)	-0.06	(0.10)	-0.05	(0.10)	0.04	(0.12)
North	0.04	(0.11)	0.01	(0.12)	0.06	(0.10)	0.19	(0.10)	-0.07	(0.12)
South	-0.02	(0.10)	0.14	(0.11)	0.19	(0.09)	-0.02	(0.09)	0.22	(0.11)
Rec: grammar school	0.00	(0.10)	-0.01	(0.12)	-0.19	(0.10)	0.26	(0.09)	0.09	(0.11)
Rec: general secondary school	0.29	(0.12)	0.12	(0.14)	0.09	(0.12)	0.02	(0.11)	0.11	(0.13)
Number of siblings	0.00	(0.03)	-0.01	(0.04)	0.01	(0.03)	-0.07	(0.03)	-0.00	(0.04)
Broken home	0.29	(0.10)	0.07	(0.11)	0.12	(0.10)	0.12	(0.09)	0.08	(0.11)
Father grammar school	-0.17	(0.10)	0.02	(0.12)	0.05	(0.10)	-0.10	(0.10)	-0.30	(0.12)
Father dropout	0.08	(0.26)	0.11	(0.31)	0.50	(0.26)	-0.07	(0.25)	-0.11	(0.29)
Mother grammar school	-0.15	(0.11)	-0.01	(0.12)	-0.01	(0.11)	0.01	(0.10)	-0.05	(0.12)
Mother dropout	0.72	(0.41)	-1.40	(0.52)	-0.19	(0.40)	-0.21	(0.38)	-0.40	(0.46)
Locus of control	-0.58	(0.06)	-0.88	(0.10)	-0.51	(0.06)	-0.43	(0.05)	-0.77	(0.08)

Means and standard deviations (in brackets) of the posterior coefficient distributions displayed.

Source: SOEP youth sample 2000-2011.

Table G.7: Regression coefficients in measurement system, females (locus of control reverse coded).

Variables	Loc 3		Loc 4		Loc 6		Loc 7		Loc 9	
Constant	1.29	(0.12)	0.91	(0.13)	1.16	(0.12)	1.58	(0.12)	0.61	(0.17)
Childhood in large city	0.12	(0.11)	-0.09	(0.12)	0.04	(0.11)	-0.10	(0.11)	0.11	(0.17)
Childhood in medium city	0.22	(0.11)	0.26	(0.13)	0.15	(0.12)	0.24	(0.11)	0.27	(0.17)
Childhood in small city	0.03	(0.10)	0.23	(0.12)	0.04	(0.11)	0.14	(0.10)	0.19	(0.15)
North	-0.05	(0.10)	0.02	(0.12)	0.05	(0.10)	0.01	(0.10)	0.06	(0.15)
South	0.04	(0.09)	0.00	(0.10)	0.22	(0.09)	0.03	(0.09)	0.25	(0.13)
Rec: grammar school	-0.07	(0.09)	-0.15	(0.11)	-0.11	(0.09)	0.25	(0.09)	-0.00	(0.14)
Rec: general secondary school	0.20	(0.12)	-0.21	(0.14)	-0.07	(0.13)	0.03	(0.12)	0.06	(0.18)
Number of siblings	-0.02	(0.03)	0.03	(0.04)	-0.08	(0.03)	0.00	(0.03)	0.08	(0.05)
Broken home	0.07	(0.10)	0.11	(0.11)	-0.09	(0.10)	-0.02	(0.09)	0.07	(0.14)
Father grammar school	-0.20	(0.10)	-0.09	(0.11)	-0.13	(0.10)	-0.15	(0.10)	-0.11	(0.15)
Father dropout	-0.05	(0.23)	0.57	(0.26)	0.06	(0.24)	-0.13	(0.23)	0.15	(0.34)
Mother grammar school	-0.17	(0.10)	-0.03	(0.11)	-0.08	(0.10)	0.06	(0.10)	-0.10	(0.15)
Mother dropout	0.52	(0.26)	-0.28	(0.29)	0.14	(0.27)	0.34	(0.26)	0.66	(0.37)
Locus of control	-0.45	(0.05)	-0.74	(0.07)	-0.56	(0.06)	-0.40	(0.05)	-1.23	(0.18)

Means and standard deviations (in brackets) of the posterior coefficient distributions displayed.

Source: SOEP youth sample 2000-2011.

Table G.8: Regression coefficients in outcome system, males (locus of control reverse coded)

Variables	**S**		**E (low edu)**		**E (high edu)**		**Y (low edu)**		**Y (high edu)**	
Constant	-0.37	(0.32)	-3.59	(0.43)	-8.07	(0.67)	1.49	(0.27)	0.98	(0.50)
Age	0.25	(0.10)	0.17	(0.02)	0.34	(0.03)	0.04	(0.01)	0.06	(0.02)
Local unemployment rate	0.16	(0.10)	-0.09	(0.02)	-0.08	(0.04)	-0.02	(0.01)	-0.01	(0.01)
Childhood in large city	0.17	(0.09)	-0.06	(0.17)	0.18	(0.25)	0.03	(0.05)	-0.08	(0.08)
Childhood in medium city	0.08	(0.09)	0.23	(0.17)	-0.06	(0.26)	0.01	(0.04)	0.04	(0.08)
Childhood in small city	-0.09	(0.08)	0.16	(0.15)	0.04	(0.22)	-0.06	(0.04)	-0.01	(0.07)
North	-0.23	(0.11)	-0.07	(0.15)	-0.01	(0.20)	0.04	(0.05)	-0.05	(0.07)
South	-0.04	(0.08)	0.13	(0.17)	-0.56	(0.25)	0.08	(0.05)	0.15	(0.09)
Married	-0.03	(0.03)	0.53	(0.25)	0.54	(0.40)	0.05	(0.05)	0.18	(0.08)
Number of children	-0.15	(0.09)	0.03	(0.14)	0.01	(0.25)	-0.04	(0.03)	-0.09	(0.05)
Father grammar school	1.03	(0.09)	-0.43	(0.18)	-0.25	(0.17)	0.08	(0.06)	0.00	(0.06)
Father dropout	-0.15	(0.28)	-0.34	(0.36)	1.14	(0.88)	-0.05	(0.14)	-0.09	(0.32)
Mother grammar school	0.57	(0.10)	-	-	-	-	-0.03	(0.07)	0.05	(0.12)
Mother dropout	-0.54	(0.46)	-	-	-	-	-0.08	(0.10)	-0.04	(0.17)
Locus of control	0.19	(0.07)	0.14	(0.10)	-0.07	(0.13)	0.06	(0.04)	-0.07	(0.06)

Means and standard deviations (in brackets) of the posterior coefficient distributions displayed. Maternal education dummies were excluded from the employment equation because of too little variation.

Source: SOEP youth sample 2000-2011.

Table G.9: Regression coefficients in outcome system, females (locus of control reverse coded)

Variables	**S**		**E (low edu)**		**E (high edu)**		**Y (low edu)**		**Y (high edu)**	
Constant	0.08	(0.32)	-5.12	(0.52)	-7.60	(0.63)	2.07	(0.31)	1.36	(0.38)
Age	0.22	(0.10)	0.21	(0.02)	0.33	(0.02)	0.01	(0.01)	0.04	(0.01)
Local unemployment rate	0.05	(0.10)	-0.00	(0.02)	-0.09	(0.04)	-0.03	(0.01)	-0.02	(0.01)
Childhood in large city	0.13	(0.09)	-0.17	(0.17)	0.05	(0.20)	0.04	(0.06)	-0.07	(0.06)
Childhood in medium city	-0.10	(0.09)	-0.22	(0.18)	-0.36	(0.21)	-0.02	(0.05)	0.01	(0.06)
Childhood in small city	-0.35	(0.08)	-0.04	(0.16)	-0.24	(0.21)	-0.00	(0.05)	-0.04	(0.06)
North	-0.20	(0.11)	-0.30	(0.16)	0.14	(0.19)	-0.02	(0.06)	0.03	(0.06)
South	0.00	(0.09)	0.36	(0.18)	-0.24	(0.22)	-0.01	(0.05)	-0.00	(0.07)
Married	-0.04	(0.03)	0.30	(0.18)	1.37	(0.54)	0.07	(0.05)	0.03	(0.06)
Number of children	-0.25	(0.09)	-0.51	(0.10)	0.68	(0.42)	-0.07	(0.03)	-0.02	(0.04)
Father grammar school	1.07	(0.09)	-0.05	(0.21)	-0.31	(0.15)	0.09	(0.07)	0.06	(0.04)
Father dropout	-0.54	(0.26)	-0.17	(0.31)	0.09	(0.95)	-0.02	(0.11)	0.37	(0.35)
Mother grammar school	0.61	(0.10)	-	-	-	-	0.03	(0.07)	0.09	(0.09)
Mother dropout	-0.32	(0.28)	-	-	-	-	0.18	(0.11)	0.02	(0.13)
Locus of control	0.22	(0.07)	0.19	(0.11)	-0.00	(0.11)	0.08	(0.04)	0.01	(0.05)

Means and standard deviations (in brackets) of the posterior coefficient distributions displayed. Maternal education dummies were excluded from the employment equation because of too little variation.

Source: SOEP youth sample 2000-2011.

H Simulation of the Model and Goodness of Fit

H.1 Model simulation

We simulate the model by computing the expected wage for different quantiles of the distribution of the factor, conditional on a given set of covariates X_Y and conditional on schooling S. The Gibbs algorithm we implement to estimate our model generates a sample of the model parameters from their conditional distribution that can be used as follows to approximate the expected wage for each quantile q_θ of the factor distribution:

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

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

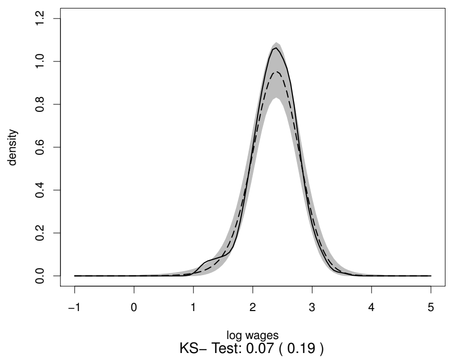
$$\frac{1}{M}\sum_{m=1}^{M}\Phi\left(X_S\beta_S^{(m)}+q_\theta^{(m)}\alpha_S^{(m)}\right),\qquad \frac{1}{M}\sum_{m=1}^{M}\Phi\left(X_E\beta_D^{(m)}+q_\theta^{(m)}\alpha_D^{(m)}\right),$$

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

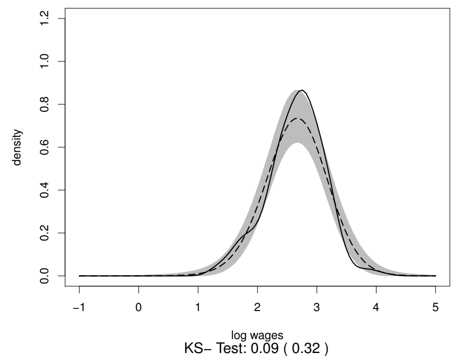
H.2 Assessing goodness of fit

Our model provides a good fit to the data, and especially to the distribution of wages. Figures H.3 to H.5 display the observed distribution of wages, along with their posterior predictive distribution for the different specifications. The actual distribution is quite well approximated by the posterior predictive distribution. The Kolmogorov-Smirnov tests we conduct to compare the actual distribution and the posterior predictive distribution never reject the null hypothesis of equal distribution. This result is in great part due to the use of normal mixtures for the error term, allowing for a flexible approximation of the true distribution.

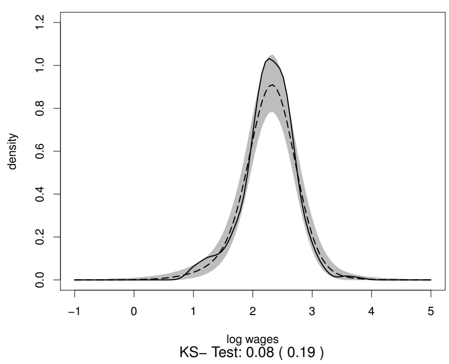
Figure H.3: Goodness-of-fit check for wages: posterior predictive (dashed) vs. actual distribution (solid) and Kolmogorov-Smirnov test for equal distributions (**2 mixture components**).



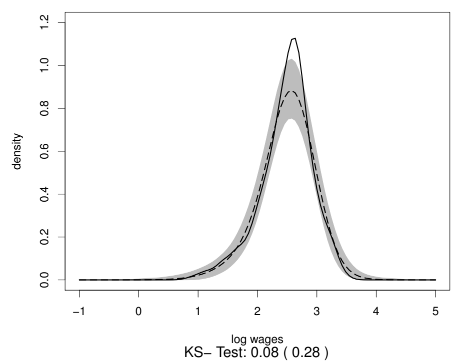
(a) Males, no higher education



(b) Males, with higher education



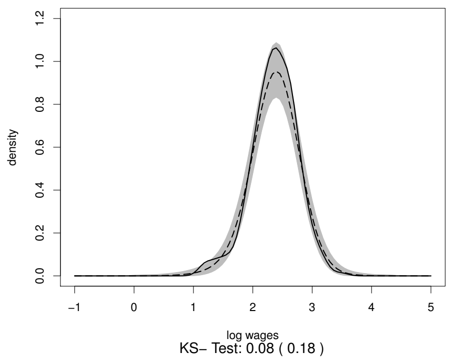
(c) Females, no higher education



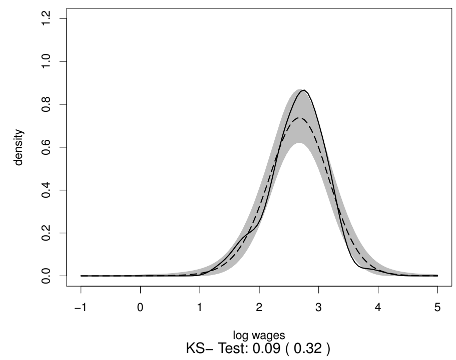
(d) Females, with higher education

Notes: Model estimated using external locus of control measures and a 2-component mixture for the error term. Figures display posterior predictive distributions and actual wages. Kernel density estimation implemented using a Gaussian kernel with bandwidth selected using Silverman's rule of thumb (Silverman, 1986) with the variation proposed by Scott (1992). Wages predicted from their posterior distribution using 20,000 replications of the sample. Shaded area represents 95% confidence interval of posterior predictive distribution. Kolmogorov-Smirnov test: Two-sample KS-test with null hypothesis that the actual sample and the posterior predictive sample have the same distribution. p-values in brackets, computed using Monte Carlo simulations to determine the proper p-value in the presence of ties in the actual distribution of wages (see Sekhon, 2011).

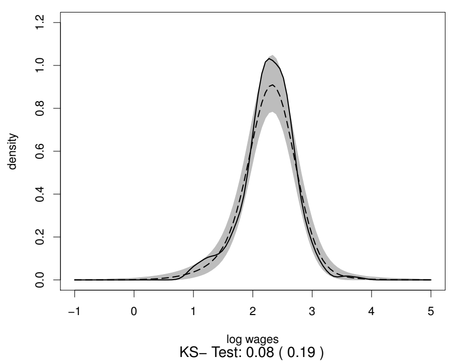
Figure H.4: Goodness-of-fit check for wages: posterior predictive (dashed) vs. actual distribution (solid) and Kolmogorov-Smirnov test for equal distributions (**3 mixture components**).



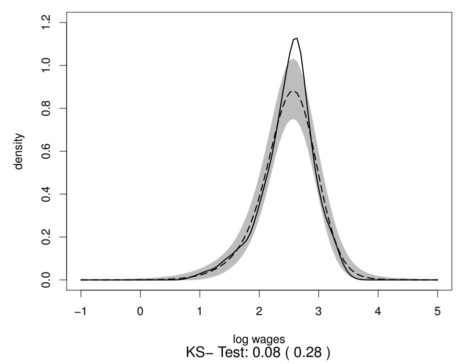
(a) Males, no higher education



(b) Males, with higher education



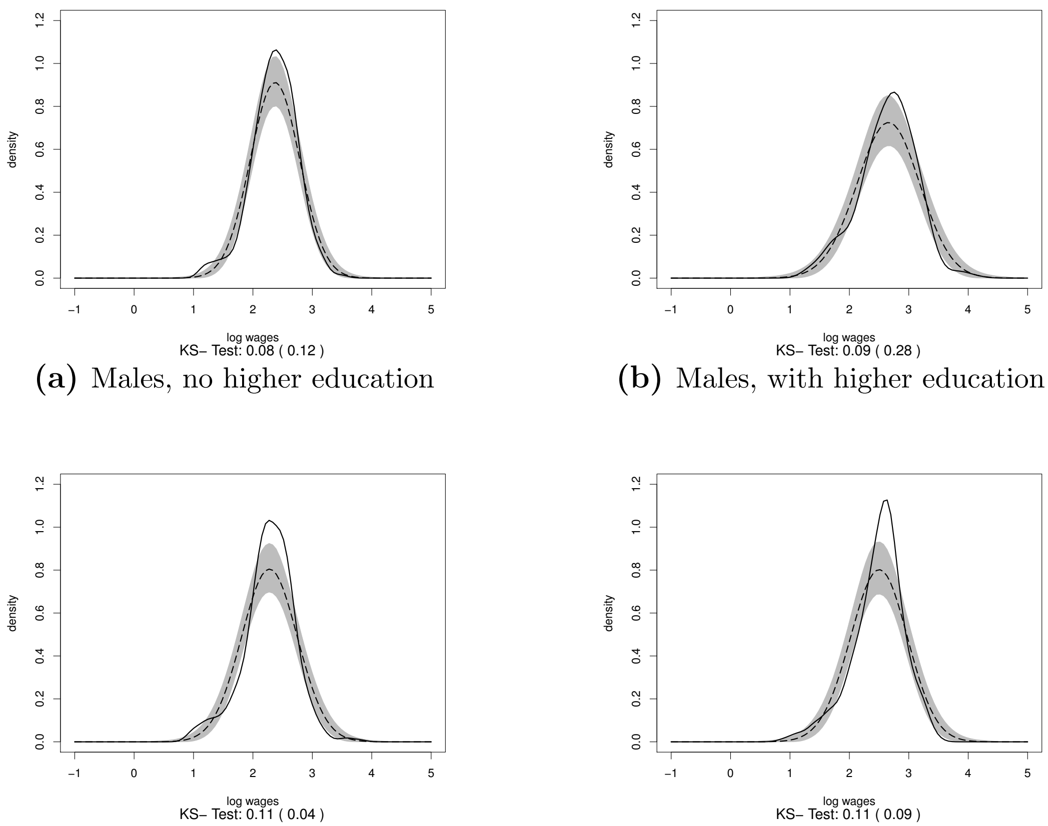
(c) Females, no higher education



(d) Females, with higher education

Notes: Model estimated using external locus of control measures and a 3-component mixture for the error term. Figures display posterior predictive distributions and actual wages. Kernel density estimation implemented using a Gaussian kernel with bandwidth selected using Silverman's rule of thumb (Silverman, 1986) with the variation proposed by Scott (1992). Wages predicted from their posterior distribution using 20,000 replications of the sample. Shaded area represents 95% confidence interval of posterior predictive distribution. Kolmogorov-Smirnov test: Two-sample KS-test with null hypothesis that the actual sample and the posterior predictive sample have the same distribution. p-values in brackets, computed using Monte Carlo simulations to determine the proper p-value in the presence of ties in the actual distribution of wages (see Sekhon, 2011).

Figure H.5: Goodness-of-fit check for wages: posterior predictive (dashed) vs. actual distribution (solid) and Kolmogorov-Smirnov test for equal distributions (**normal error term**).



(a) Males, no higher education

(b) Males, with higher education

(c) Females, no higher education

(d) Females, with higher education

Notes: Model estimated using external locus of control measures and a gaussian normal for the error term. Figures display posterior predictive distributions and actual wages. Kernel density estimation implemented using a Gaussian kernel with bandwidth selected using Silverman's rule of thumb (Silverman, 1986) with the variation proposed by Scott (1992). Wages predicted from their posterior distribution using 20,000 replications of the sample. Shaded area represents 95% confidence interval of posterior predictive distribution. Kolmogorov-Smirnov test: Two-sample KS-test with null hypothesis that the actual sample and the posterior predictive sample have the same distribution. p-values in brackets, computed using Monte Carlo simulations to determine the proper p-value in the presence of ties in the actual distribution of wages (see Sekhon, 2011).

I Robustness Checks

In this section, we check the robustness of our model with respect to several aspects of the specification: the number and type of locus of control measures, the covariates of the model, the distributional assumptions on the latent locus of control factor and on the error terms in the wage equations. Second, our identification hinges on the assumption of common parameters in the schooling equation among the youth and adult samples. To assess whether this assumption is indeed reasonable, we perform several checks concerning differences in the distribution of parameters and the covariates across the two samples.

I.1 Model specification.

First, our results may depend on the model specification and distributional assumptions. Most importantly, cognitive abilities may play a role in the school choice model. As a robustness check, we therefore use four alternative ways of accounting for cognitive abilities: (i) omitting track recommendation from the measurement system, (ii) adding track recommendation to the schooling equation, (iii) adding school grades to the measurement system (iv) introducing a second cognitive factor with a separate measurement system which enters the schooling equation. Table I.10 compares the results of five alternative models. Most importantly, cognitive abilities are likely to be part of the structural school choice model. We therefore purge the locus of control estimates by using track recommendation in the measurement system of locus of control. The top four panels of Table I.10 present four alternative ways of accounting for cognitive abilities. Panel 1 omits the track recommendation dummies from the measurement system and panel 2 adds them to the schooling equation. Our results are invariant to these changes and the posterior mean of the estimated coefficients remains almost identical, as does the estimated variance of the latent factor. Panel 3 displays the results of a model where the most recent school grades obtained in math, German and the first foreign language were included with track recommendation in the measurement system. Again, this does not have a large impact on the results except that for females the factor loadings in the schooling and wage equations decrease slightly. We proceed by testing the robustness of our results with respect to the direct inclusion of cognitive abilities in the model. Hence, we add a second measurement system with cognitive measures and a second latent *cognitive factor* to the schooling equation. To do this, we use cognitive test measures that were administered to a subsample of our youth sample, namely to all those who turned 17 in or after the year 2006. These cognitive measures are part of the I-S-T 2000 R intelligence test, which measures the cognitive potential of an individual on hands of three subscales: verbal numeric and figural cognitive potential (von Rosenbladt and Stimmel, 2005). The results of this exercise are displayed in panel 4 of Table I.10 and again do not change our main results and qualitative implications. If at all the factor loadings slightly increase in size and significance. The analysis with two factors, however, faces two potential limitations that are related to the small sample for which we observe cognitive abilities. Specifically, all individuals for which we observe cognitive ability measures are still relatively young and not yet part of the labor force. Hence, there is no overlap sample for this group and including the cognitive ability factor in the wage equations renders the model more unstable. Because of the limited sample size, we can also not allow the cognitive and locus of control factors to be correlated. Empirically, this is however not problematic because the correlation of cognitive and locus of control measures is always less than 0.1 and very close to zero for most item pairs. This result, together with our results from the previous robustness checks, show that cognitive abilities and locus of control are two very distinct orthogonal concepts and that the inclusion of cognitive abilities is not crucial for the identification of our parameter estimates. Panel 5 of Table I.10 presents a last robustness check of our model specification where we change the specification of the wage equation to include not only the age of an individual but also a quadratic term in age as a proxy for the squared experience term of a Mincer wage equation.[14] Again, we find that this change in model specification does not change our results.

I.2 Distributional assumptions.

In addition to the sensitivity analyses discussed so far, we perform a range of estimations to assess the robustness of the results with respect to the assumptions we make regarding the dimensionality and shape of the locus of control factor and the error term in the wage equations. First, the results of a factor model tend to be sensitive to the types, validity and dimensionality of the underlying measurements used. To see how the inclusion of additional items affects our results we re-estimated our model with the entire set of ten locus of control measurements provided by the SOEP youth questionnaire. The

[14]We also have information about total labor force experience which however tends to be highly correlated ($\sim$0.8) with age.

Table I.10: Robustness to model specification, factor loadings of outcome system for different specifications

	Males		**Females**	
No track recommendation in measurement system				
S	0.204***	(0.059)	−0.225***	(0.060)
E_0	0.140	(0.101)	−0.189*	(0.109)
E_1	−0.062	(0.138)	0.012	(0.110)
Y_0	0.063*	(0.036)	−0.082**	(0.038)
Y_1	−0.068	(0.064)	−0.013	(0.050)
Track recommendation in education choice				
S	0.182***	(0.070)	0.217***	(0.067)
E_0	0.151	(0.100)	0.185*	(0.110)
E_1	−0.060	(0.133)	0.021	(0.115)
Y_0	0.060	(0.036)	0.080**	(0.038)
Y_1	−0.067	(0.062)	0.003	(0.050)
With grades in measurement system				
S	0.193***	(0.069)	0.215***	(0.065)
E_0	0.143	(0.098)	0.197*	(0.113)
E_1	−0.072	(0.135)	−0.009	(0.112)
Y_0	0.063*	(0.036)	0.082**	(0.037)
Y_1	−0.068	(0.063)	0.012	(0.051)
With cognitive factor in schooling equation				
S	0.254**	(0.115)	0.355***	(0.123)
E_0	0.193	(0.132)	0.277*	(0.162)
E_1	−0.107	(0.179)	−0.018	(0.167)
Y_0	0.077	(0.047)	0.118**	(0.056)
Y_1	−0.080	(0.082)	0.016	(0.074)
Mincer equation for wages				
S	0.191***	(0.070)	0.215***	(0.068)
E_0	0.141	(0.099)	0.198*	(0.108)
E_1	−0.069	(0.137)	−0.003	(0.114)
Y_0	0.056	(0.036)	0.081**	(0.038)
Y_1	−0.066	(0.061)	−0.010	(0.048)

Notes: Standard errors in brackets. Significance check: */**/*** if zero lies outside the 90%/95%/99% confidence interval of the posterior distribution of the corresponding parameter. Factor variances are set to one except for panel 4 where for computational reasons the normalization is on factor loading 2.

Table I.11: Robustness to distributional assumptions, factor loadings of outcome system for different assumptions

	Males		**Females**	
10-item locus of control scale				
S	0.184***	(0.066)	0.235***	(0.066)
E_0	0.148	(0.096)	0.242**	(0.112)
E_1	−0.087	(0.124)	−0.016	(0.105)
Y_0	0.046	(0.036)	0.078**	(0.039)
Y_1	−0.063	(0.060)	0.019	(0.050)
Normal distribution for ε in the wage equation				
S	0.193***	(0.070)	0.217***	(0.067)
E_0	0.140	(0.096)	0.192*	(0.110)
E_1	−0.063	(0.131)	−0.001	(0.111)
Y_0	0.069*	(0.035)	0.043	(0.044)
Y_1	−0.073	(0.066)	0.015	(0.051)
3-component mixture for ε in the wage equation				
S	0.194***	(0.068)	0.215***	(0.068)
E_0	0.137	(0.099)	0.194*	(0.110)
E_1	−0.061	(0.135)	−0.002	(0.110)
Y_0	0.058	(0.035)	0.080**	(0.037)
Y_1	−0.063	(0.061)	0.009	(0.048)
2-component mixture for locus of control				
S	0.193***	(0.069)	0.212**	(0.100)
E_0	0.140	(0.103)	0.199	(0.153)
E_1	−0.071	(0.134)	−0.058	(0.159)
Y_0	0.060	(0.036)	0.113**	(0.054)
Y_1	−0.063	(0.064)	0.016	(0.075)

Notes: Standard errors in brackets. Significance check: */**/*** if zero lies outside the 90%/95%/99% confidence interval of the posterior distribution of the corresponding parameter. Factor variances are set to one.

top panel of Table I.11 displays the means and standard deviations of the posterior distributions of the estimated parameters in the outcome system. Comparing these results to the main results displayed in Table 3 of the paper, we can infer that all coefficient estimates and results stay almost exactly the same.[15] Another concern with our results could be that a 2-component mixture is too restrictive or that precision of the estimates could be improved by using a normal distribution for wages in the error term. Figure H.4, Figure H.5, and Figure H.3 described in the previous section, assess which of these distributional assumptions generates the best fit to to the data by comparing the posterior predictive wage distribution generated by the model to the actual distribution of wages. Both eyeballing and a formal Kolmogorov-Smirnov test tell us that a 2-component mixture fits the data better than a normal distribution, while a 3-component mixture does not further improve model fit.[16] Moreover, Table I.11 shows that neither a normal distribution nor a 3-component mixture for the error term in the wage equation has a large effect on our results. Last, the use of a normal distribution for the latent factor might be too restrictive. The bottom panel of Table I.11 displays the results of a model where we use a 2-component mixture for the latent factor instead and we find that our results are invariant to this change in model specification.

[15]In an earlier working paper version of this article, we obtain very similar results using the full battery of locus of control measures.

[16]In earlier version of this paper we used a smaller sample size with slightly more heterogeneity among individuals. For that sample a 3-component mixture provided the best fit to the data.

J The Schooling Equation among Youths and Adults

As explained in Section 3.2 of the paper, the schooling equation represents the link between the two subsamples. Therefore, it plays a central role for model identification. Our main assumption postulates that the parameters of the schooling equation are the same across the youth and the adult samples, after controlling for observables and cohort fixed effects. In other words, we assume that the schooling decision is generated by the same structural model for both the youths and the adults. This section takes a closer look at the credibility of this assumption.

First, we test if the parameters β_S of the schooling equation can be assumed to be the same across the two subsamples. To this end, we estimate the schooling equation separately for the youths and the adults. We exclude the individuals from the overlap sample from this analysis, to make the separation between the two subsamples more clear-cut. We use a probit model similar to the schooling equation used in the paper, with the only difference that the latent factor is relegated to the error term (i.e., $\nu_S = \alpha_S\theta + \varepsilon_S$), as we can only control for the observables in X_S in this single-equation model.[17] Each of these two models is estimated using a Gibbs sampler with non-informative priors ($\beta_{Sk} \sim \mathcal{N}(0;\, 10)$ for each covariates k), where $40,000$ MCMC iterations are saved for posterior inference after a burn-in of $2,000$ iterations.

To compare the posterior distributions of the parameters across the two samples, we compute the statistic $\xi_k = \beta_{Sk}^{\text{adult}} - \beta_{Sk}^{\text{youth}}$ for each covariate k and look at its distribution. Figures J.6 and J.7 show the results for females and males. For a given covariates k, there is evidence that $\beta_{Sk}^{\text{adult}} = \beta_{Sk}^{\text{youth}}$ if 0 is contained in the 95% highest posterior density interval (i.e., credible interval for ξ_k corresponding to covariate k).

Overall, the graphs show that we cannot discard the assumption that the parameters of the schooling equation are the same across the two subsamples for most of the covariates. The only exceptions appear to be for 'local unemployment rate' for females and 'mother grammar school' for males, which are both borderline. Note that 'Mother dropout' is outside the 95% credible interval for both genders. However, the sample mean of this dummy variable is so small in the adult subsample (0.0048 for males and 0.0266 for females) that it makes it very difficult to get sensible results for the corresponding coefficient, thus deteriorating the test for this variable.

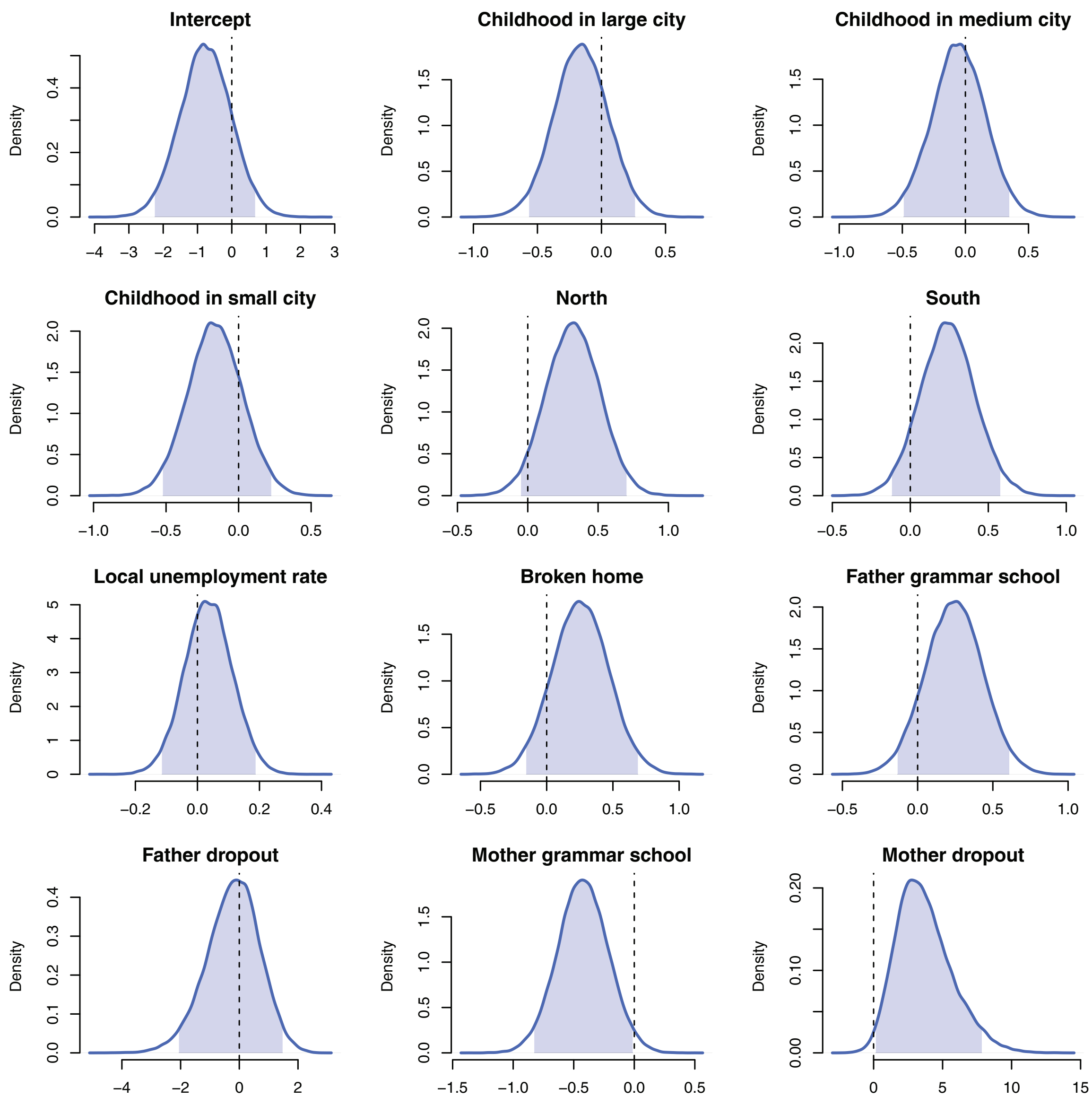
However, these encouraging results should not hide the fact that testing the assumption of equal parameters across the two subsamples is challenging in this data set,

[17]This has no impact on β_S, however, since $X_S \perp\!\!\!\perp \theta$ by assumption.

because of the relatively small number of observations in each of the subsamples. This is why we perform additional tests, in order to further investigate the similarity of the two subsamples.

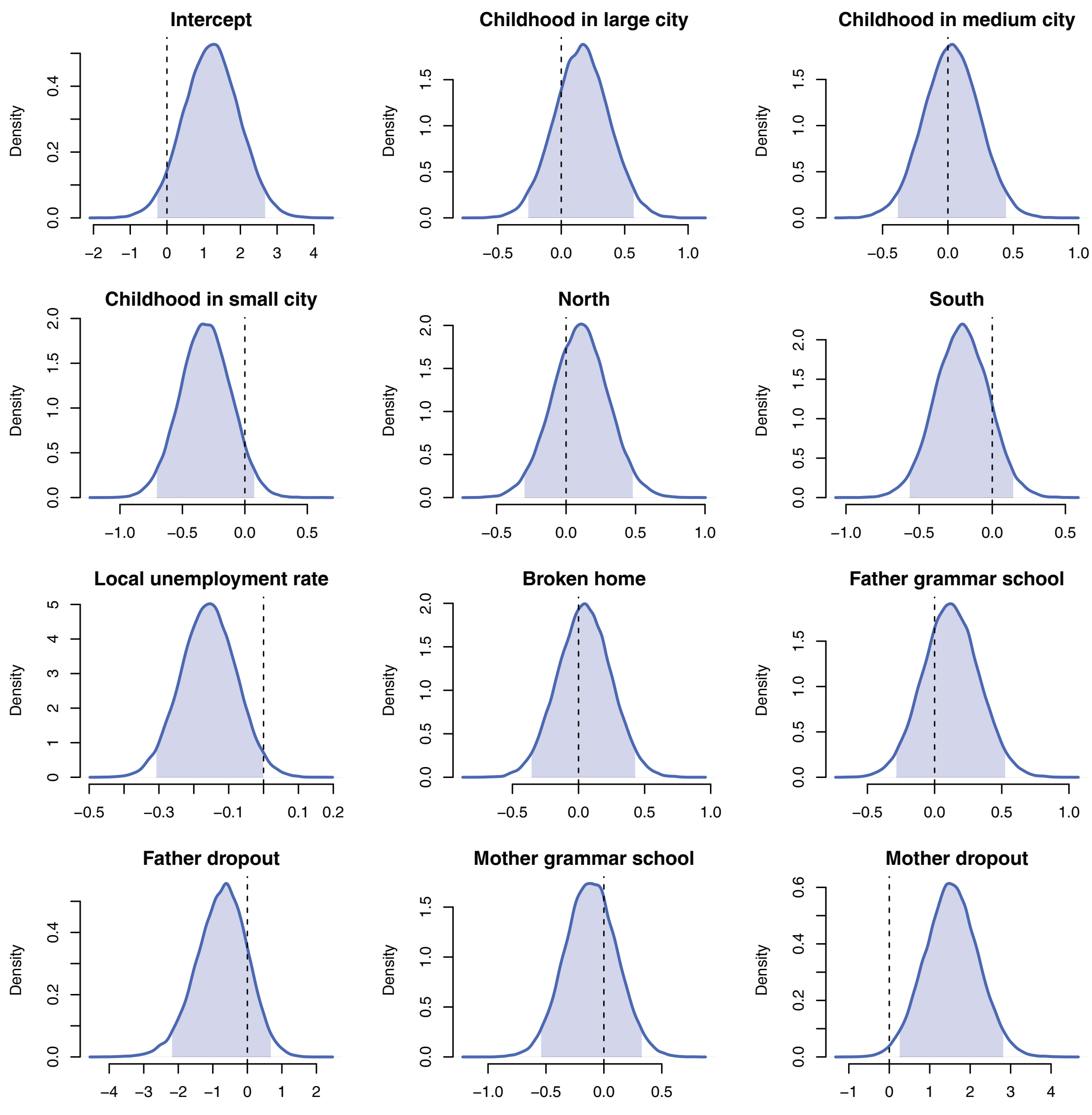
It is common practice in the literature on data combination to provide evidence that the variables have the same distribution across the different samples that are combined. Although this assumption is not strictly required for our analysis, it is important to check that the differences in levels that we observe for some variables can be explained by age and cohort effects. One concern, in our case, may be that these age and cohort effects are not sufficient to account for the rising share of individuals with higher education displayed in Table 2 of the paper. We test for this by checking whether any of the variables that are relevant to the schooling equation still differ significantly in their means after age and cohort effects have been accounted for. To this end, we first 'residualize' the higher education variable and all covariates that are part of the schooling equation by estimating the residuals of a linear regression of these variables on a linear age trend and the cohort dummies used in our main specification. We then test for significant differences in these residuals. The results, displayed in Table J.12, show that the differences between these two samples seem to be generated by level-effects that can be accounted for by a linear age trend and cohort fixed effects and that the p-values of the test on the residualized higher education variable are particularly high.

Figure J.6: Testing for differences between the youth and the adult samples in the schooling equation for males. Each graph shows the posterior distribution of the statistic $\xi_k = \beta_{Sk}^{\text{adult}} - \beta_{Sk}^{\text{youth}}$, for each of the corresponding coefficients β_{Sk}. Shaded areas show the 95% highest posterior density intervals. Zero inside shaded area means $\beta_{Sk}^{\text{adult}} = \beta_{Sk}^{\text{youth}}$ is credible.



Notes: Distribution of ξ_k computed from the posterior distributions of the corresponding parameters $\beta_{Sk}^{\text{adult}}$ and $\beta_{Sk}^{\text{youth}}$. These posteriors are obtained from probit models estimated separately for the youth and the adult samples, using a Gibbs sampler with prior $\mathcal{N}(0; 10)$ on the coefficients. For the adult sample, a dummy variable is included for the ages 31 to 35 to control for potential cohort effects.

Figure J.7: Testing for differences between the youth and the adult samples in the schooling equation for females. Each graph shows the posterior distribution of the statistic $\xi_k = \beta_{Sk}^{\text{adult}} - \beta_{Sk}^{\text{youth}}$, for each of the corresponding coefficients β_{Sk}. Shaded areas show the 95% highest posterior density intervals. Zero inside shaded area means $\beta_{Sk}^{\text{adult}} = \beta_{Sk}^{\text{youth}}$ is credible.



Notes: Distribution of ξ_k computed from the posterior distributions of the corresponding parameters $\beta_{Sk}^{\text{adult}}$ and $\beta_{Sk}^{\text{youth}}$. These posteriors are obtained from probit models estimated separately for the youth and the adult samples, using a Gibbs sampler with prior $\mathcal{N}(0; 10)$ on the coefficients. For the adult sample, a dummy variable is included for the ages 31 to 35 to control for potential cohort effects.

Table J.12: p-values of a two-sided test in means of detrended covariates among youth and adult samples (by gender).

Variables	**Males**	**Females**
	p-value	**p-value**
Higher education	0.99	0.75
Father grammar school	0.61	0.40
Father dropout	0.97	0.94
Mother grammar school	0.79	0.68
Mother dropout	0.72	0.77
Childhood in large city	0.94	1.00
Childhood in medium city	0.64	0.75
Childhood in small city	0.58	0.82
North	0.80	0.87
South	0.78	0.67
Broken home	0.71	0.28
Unemployment at schooling decision	0.07	0.12
N	1584	1532

Source: SOEP waves 2001-2011. Own calculations.
Notes: p-values of a two-sided t-test for differences in means are reported.
Variables are detrended using a linear age trend
and a cohort dummy for the age groups 26-30 and 31-35, respectively.

K Comparison to more Traditional Methods of Estimation

We conduct several wage regressions to obtain an idea of the estimates we would have obtained without the methodological contributions of the paper.[18] First, to assess endogeneity bias, we conduct a standard wage regression of the log hourly wage on contemporaneous locus of control factor values as predicted from a standard confirmatory factor model. The results are displayed in Table K.13 and show that, with the exception of highly educated men, the use of contemporaneous measurements yields significant coefficients both in a pooled analysis that conditions on education outcomes and in separate analyses by education level.[19] Moreover, the coefficients obtained in this regression are very similar in magnitude to the ones reported in Heineck and Anger (2010, Table 1). Last, comparing the magnitude of the effect to the effects of pre-market locus of control on wages, we find that most of the differences occur for highly educated individuals. One interpretation of this result is that working in a high status environment increases locus of control, and the effect may be stronger for women (this interpretation is also supported by the results reported in Trzcinski and Holst, 2010). Second, to see how much standard results are affected by measurement error, we replace the estimated factor with a simple standardized score, generated by summing up the external locus of control measures. Columns 1 and 2 (5 and 6 for females) of Table K.14 show the difference in estimated coefficients when only external locus of control items are used and columns 3 and 4 (7 and 8 for females) when the full battery of locus of control items is used. If only the external items are used, we find the coefficients are not attenuated in the estimations for males and by only 5% in the estimations for females. This difference in coefficients is small because a factor analysis on the external locus of control items returns similar loadings (weights) for each of the items. To see what happens if this is not the case, we report estimated coefficients for the case where the full battery of locus of control items is used for the analysis. In this case the estimated coefficients from using raw scores are reduced by around 25% for both males and females.

[18] For results on the stability of locus of control over time see Section L.

[19] These results where obtained using simple ordinary least squares to make the analysis comparable to what is usually done in the literature. Bayesian inference results (not displayed) are very similar.

Table K.13: Wage regressions with locus of control factor (*contemporaneous* measures)

Log hourly wages	Males			Females		
	All	$S = 1$	$S = 0$	All	$S = 1$	$S = 0$
External loc factor (std)	0.046***	0.013	0.061***	0.063***	0.080**	0.049**
	(0.016)	(0.032)	(0.017)	(0.017)	(0.031)	(0.022)
Education	0.234***			0.117***		
	(0.036)			(0.035)		
Observations	692	252	440	660	308	352
R-squared	0.246	0.242	0.163	0.194	0.208	0.174

Robust standard errors in parentheses. The covariates are the same as the main specification (See Table E.2). Factor scores were predicted from a standard confirmatory factor model.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table K.14: Comparisons of wage model estimates with predicted factor scores and raw scores (*contemporaneous* measures)

Log hourly wages	Males				Females			
External locus of control *factor*	0.046***				0.063***			
	(0.016)				(0.017)			
External locus of control *score*		0.046***				0.060***		
		(0.016)				(0.017)		
Locus of control *factor*			0.052***				0.071***	
			(0.016)				(0.018)	
Locus of control *score*				0.040**				0.054***
				(0.018)				(0.017)
Education	0.234***	0.233***	0.229***	0.238***	0.117***	0.119***	0.115***	0.127***
	(0.036)	(0.036)	(0.036)	(0.036)	(0.035)	(0.035)	(0.035)	(0.035)
Observations	692	692	692	692	660	660	660	660
R-squared	0.246	0.246	0.249	0.243	0.194	0.193	0.199	0.189

Robust standard errors in parentheses. The covariates are the same as the main specification (See Table E.2).
Factor scores were predicted from a standard confirmatory factor model.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

L The Stability of Locus of Control

In this section, we use contemporaneous measures of locus of control for the adult sample to compare locus of control distributions for individuals over time. To be able to do so, we have to overcome several shortcomings of the data. Most importantly, the battery of contemporaneous locus of control measures uses the same questions but a Likert scale with seven instead of four answers as in the youth questionnaire. We therefore have to make sure that the underlying latent construct measured by both test batteries is indeed comparable over time. To address this concern and to be able to use both *pre-market* and *contemporaneous* locus of control constructs for later analysis we follow a simple three step approach. In a first instance, we estimate polychoric correlation matrices among (i) pre-market locus of control measures for the youth sample and (ii) contemporaneous locus of control measures administered in 2010 for the adult sample. Here, the polychoric correlation among two measurements is given by the correlation among two latent measurements M_1^* and M_2^* which follow a bivariate normal distribution and who can be categorized to obtain the observe measures M_1 and M_2. In a second instance, we use the estimated polychoric correlation matrix to perform a simple factor analysis and to predict factor values for each individual. Last, we use these factor values to compare distributions of locus of control and individual locus of control measures over time. A second limitation of the data arises at this point because we only obtain factor values at two different points in time for a small overlap sample. Note that this overlap sample is distinct and much smaller from the one we discuss in the paper, which comprises those individuals that are part of the youth sample, but for which we also have measures on adult wages. In what follows we will therefore refer to the *youth sample* the *adult sample* and the *loc-overlap sample*.

Figure L.8 compares latent external locus of control distributions for the youth and adult sample. The left panel shows that there are no significant differences in distributions when all observations from the youth and adult samples are used. The p-value of a Kolmogorov-Smirnov test for significant differences in distributions is 0.15, thus not rejecting the null assumption of equal distributions at a 5% level. The right panel compares *pre-market* youth locus of control to *market* adult locus of control for individuals from the overlap sample. It shows that the locus of control distribution of these same individuals differs significantly over time (the p-value of the Kolmogorov-Smirnov test is 0.002) and that the distribution widens as individuals age. Hence, the graph shows that some individuals become more internal (higher locus of control) and some individuals become more external (lower locus of control) as they enter the labor market. This result is further confirmed by Figure L.9 which displays the difference in percentile ranks from estimated external locus of control for the overlap sample. Again, the graph shows that some individuals become much more external while some others become more internal over time.

Figure L.8: Kernel density distributions of latent youth and adult locus of control factors

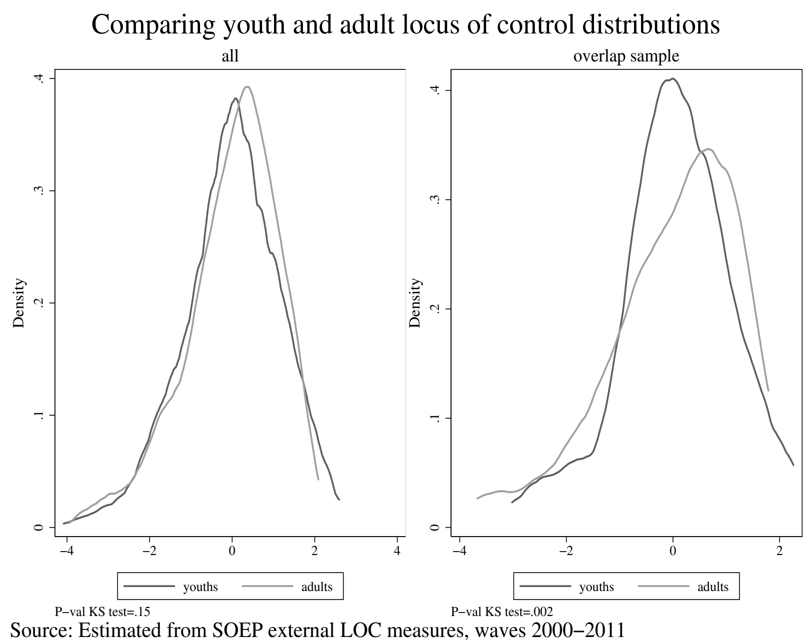
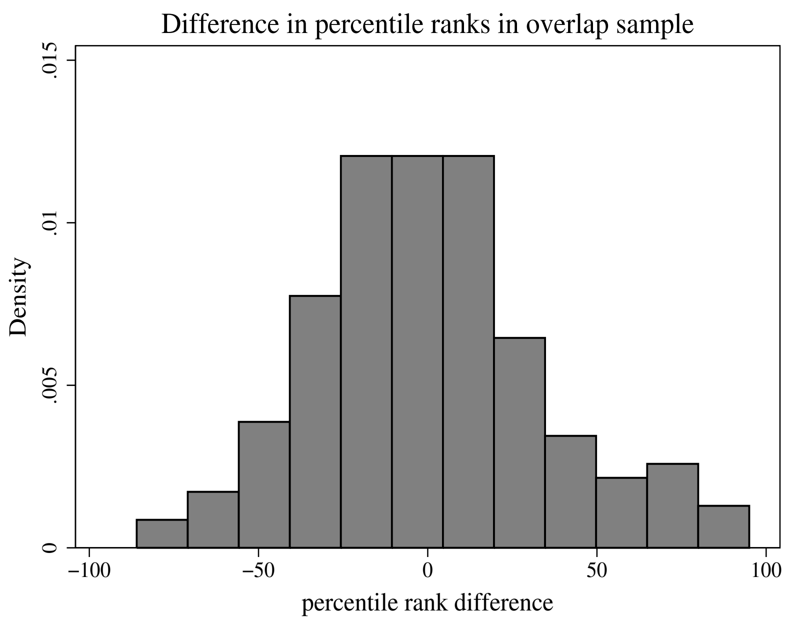


Figure L.9: Histogram of percentile differences in latent youth and adult locus of control factors



References

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

Almlund M, Duckworth AL, Heckman J, Kautz T. 2011. Personality psychology and economics. *Handbook of the Economics of Education* **4**. DOI: 10.1016/B978-0-444-53444-6.00001-8

Anderson T, Rubin H. 1956. Statistical inference in factor analysis. In Neyman J (ed.) *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, 5*. Berkeley: University of California Press, 111–150.

Becker GS. 1964. *Human Capital: A Theoretical and Empirical Analysis, With Special Reference to Education*. New York: National Bureau of Economic Research.

Borghans L, Duckworth AL, Heckman JJ, ter Weel B. 2008. The economics and psychology of personality traits. *Journal of Human Resources* **43**: 972–1059. DOI: 10.1353/jhr.2008.0017

Bowles S, Gintis H. 1976. *Schooling in Capitalist America: Educational Reform and the Contradictions of Economic Life*. Routledge.

Boyce CJ, Wood AM, Powdthavee N. 2013. Is personality fixed? personality changes as much as 'variable' economic factors and more strongly predicts changes to life satisfaction. *Social Indicators Research* **111**: 287–305. DOI: 10.1007/s11205-012-0006-z

Card D. 1999. The causal effect of education on earnings. *Handbook of Labor Economics* **3**: 1801–1863. DOI: 10.1016/S1573-4463(99)03011-4

Carneiro P, Hansen K, Heckman JJ. 2003. Estimating distributions of treatment effects with an application to the returns to schooling and measurement of the effects of uncertainty on college choice. *International Economic Review* **44**: 361–422. DOI: 10.1111/1468-2354.t01-1-00074

Casella G, George EI. 1992. Explaining the Gibbs Sampler. *The American Statistician* **46**: 167–174.

Celeux G. 1998. Bayesian inference for mixture: The label switching problem. In Green PJ, Rayne R (eds.) *COMPSTAT*. Heidelberg: Physica-Verlag, 227–232. DOI: 10.1007/978-3-662-01131-7_26

Cobb-Clark DA, Schurer S. 2013. Two economists' musings on the stability of locus of control. *The Economic Journal* **123**: 358–400. DOI: 10.1111/ecoj.12069

Conti G, Frühwirth-Schnatter S, Heckman JJ, Piatek R. 2014. Bayesian Exploratory Factor Analysis. *Journal of Econometrics* **183**: 31–57. DOI: 10.1016/j.jeconom.2014.06.008

Cowles MK. 1996. Accelerating Monte Carlo Markov chain convergence for cumulative-link generalized linear models. *Statistics and Computing* **6**: 101–111. DOI: 10.1007/BF00162520

Cunha F, Heckman JJ, Schennach SM. 2010. Estimating the Technology of Cognitive and Noncognitive Skill Formation. *Econometrica* **78**: 883–931. DOI: 10.3982/ECTA6551

Dagsvik JK, Hægeland T, Raknerud A. 2011. Estimating the returns to schooling: a likelihood approach based on normal mixtures. *Journal of Applied Econometrics* **26**: 613–640. DOI: 10.1002/jae.1172

Diebolt J, Robert CP. 1994. Estimation of finite mixture distributions through Bayesian sampling. *Journal of the Royal Statistical Society Series B: Methodological* **56**: 363–375.

Dohmen T, Sunde U, Falk A, Huffman D. 2010. Are risk aversion and impatience related to cognitive ability? *American Economic Review* **100**: 1238–1260. DOI: 10.1257/aer.100.3.1238

Escobar MD, West M. 1995. Bayesian density estimation and inference using mixtures. *Journal of the American Statistical Association* **90**: 577–588. DOI: 10.1080/01621459.1995.10476550

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

Ferguson TS. 1983. Bayesian density estimation by mixtures of normal distributions. In Chernoff H, Rizvi M, Rustagi J, Siegmund D (eds.) *Recent Advances in Statistics: Papers in Honor of Herman Chernoff on his Sixtieth Birthday*. New York: Academic Press, 287–302.

Flossmann A, Pohlmeier W. 2006. Causal returns to education: A survey on empirical evidence for Germany. *Journal of Economics and Statistics* **226**: 6–23. DOI: 10.2139/ssrn.917325

Frey BS, Meier S. 2004. Pro-social behavior, reciprocity or both? *Journal of Economic Behavior and Organization* **54**: 65–68. DOI: 10.1016/j.jebo.2003.10.001

Frühwirth-Schnatter S. 2006. *Finite Mixture and Markov Switching Models*. New York: Springer. DOI: 10.1007/978-0-387-35768-3

Frühwirth-Schnatter S, Lopes HF. 2010. Parsimonious Bayesian Factor Analysis when the Number of Factors is Unknown. *Working Paper* : The University of Chicago

Booth School of Business.

Furnham A, Sadka V, Brewin C. 1992. The development of an occupational attributional style questionnaire. *Journal of Organizational Behavior* **13**: 27–39. DOI: 10.1002/job.4030130104

Furnham A, Steele H. 1993. Measuring locus of control: A critique of general, children's, health-and work-related locus of control questionnaires. *British Journal of Psychology* **84**: 443–479. DOI: 10.1111/j.2044-8295.1993.tb02495.x

Gebel M, Pfeiffer F. 2010. Educational expansion and its heterogeneous returns for wage workers. *Schmollers Jahrbuch* **130**: 19–42. DOI: 10.3790/schm.130.1.19

Gottschalk P. 2005. Can work alter welfare recipients' beliefs? *Journal of Policy Analysis and Management* **24**: 485–498. DOI: 10.1002/pam.20111

Heckman JJ, Kautz T. 2013. Fostering and measuring skills: Interventions that improve character and cognition. Working paper 19656, NBER.

Heckman JJ, Lochner L, Todd P. 2006a. Earnings functions, rates of return and treatment effects: The Mincer equation and beyond. *Handbook of the Economics of Education* **1**: 307–458. DOI: 10.1016/S1574-0692(06)01007-5

Heckman JJ, Stixrud J, Urzua S. 2006b. The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. *Journal of Labor Economics* **24**: 411–482. DOI: 10.1086/504455

Heineck G, Anger S. 2010. The returns to cognitive abilities and personality traits in Germany. *Labour Economics* **17**: 535–546. DOI: 10.1016/j.labeco.2009.06.001

Hill MS. 1979. The wage effects of marital status and children. *Journal of Human Resources* **14**: 579–594. DOI: 10.2307/145325

Jöreskog KG. 1994. On the Estimation of Polychoric Correlations and their Asymptotic Covariance Matrix. *Psychometrika* **59**: 381–389. DOI: 10.1007/BF02296131

Jöreskog KG, Moustaki I. 2001. Factor analysis of ordinal variables: A comparison of three approaches. *Multivariate Behavioral Research* **36**: 347–387. DOI: 10.1207/S15327906347-387

Krueger RF, Johnson W. 2008. Behavioral genetics and personality. In *Handbook of personality: Theory and research*. New York: The Guilford Press, 287–310.

Lauer C, Steiner V. 2000. Private Erträge von Bildungsinvestitionen in Deutschland. *Applied Economics Quarterly* **51**: 71–101.

Liu JS. 1994. The Collapsed Gibbs Sampler in Bayesian Computations with Applications to a Gene Regulation Problem. *Journal of the American Statistical Association* **89**: 958–966. DOI: 10.2307/2290921

Liu JS, Sabatti C. 2000a. Generalised Gibbs sampler and multigrid Monte Carlo for Bayesian computation. *Biometrika* **87**: 353–369. DOI: 10.1093/biomet/87.2.353

Liu JS, Sabatti C. 2000b. Generalised Gibbs sampler and multigrid Monte Carlo for Bayesian computation. *Biometrika* **87**: 353–369.

Liu JS, Wong WH, Kong A. 1994. Covariance Structure of the Gibbs Covariance Structure of the Gibbs Sampler with Applications to the Comparisons of Estimators and Augmentation Schemes. *Biometrika* **81**: 27–40.

Liu JS, Wu YN. 1999. Parameter expansion for data augmentation. *Journal of the American Statistical Association* **94**: 1264–1274.

Mincer J. 1958. Investment in human capital and personal income distribution. *Journal of Political Economy* **66**: 281–302. DOI: 10.1086/258055

Moutafi J, Furnham A, Tsaousis I. 2006. Is the relationship between intelligence and trait neuroticism mediated by test anxiety? *Personality and Individual Differences* **40**: 587–597. DOI: 10.1016/j.paid.2005.08.004

Muthén BO, Kaplan D. 1985. A comparison of some methodologies for the factor analysis of non- normal Likert variables. *British Journal of Mathematical and Statistical Psychology* **38**: 171–180.

Nandram B, Chen MH. 1996. Reparameterizing the generalized linear model to accelerate Gibbs sampler convergence. *Journal of Statistical Computation and Simulation* **54**: 129–144. DOI: 10.1080/00949659608811724

Olsson U. 1979a. Maximum Likelihood Estimation of the Polychoric Correlation Coefficient. *Psychometrika* **44**: 443–460.

Olsson U. 1979b. On The Robustness Of Factor Analysis Against Crude Classification Of The Observations. *Multivariate Behavioral Research* **14**: 485–500.

Pischke J, Von Wachter T. 2008. Zero returns to compulsory schooling in Germany: Evidence and interpretation. *The Review of Economics and Statistics* **90**: 592–598. DOI: 10.1162/rest.90.3.592

Psacharopoulos G. 1981. Returns to education: an updated international comparison. *Comparative Education* **17**: 321–341.

Rhemtulla M, Brosseau-Liard PE, Savalei V. 2012. When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. *Psychological Methods* **17**: 354–373. DOI: 10.1037/a0029315

Roberts BW, Caspi A, Moffitt TE. 2003. Work experiences and personality development in young adulthood. *Journal of personality and social psychology* **84**: 582. DOI:

10.1037/0022-3514.84.3.582

Rotter JB. 1966. Generalized expectancies for internal versus external control of reinforcement. *Psychological Monographs: General & Applied* **80**: 1–28.

Scott DW. 1992. *Multivariate Density Estimation: Theory, Practice, and Visualization*. Wiley Interscience. DOI: 10.1002/9780470316849

Sekhon JS. 2011. Multivariate and propensity score matching software with automated balance optimization. *Journal of Statistical Software* **42**: 1–52.

Silverman BW. 1986. *Density Estimation for Statistics and Data Analysis*. CRC Press. DOI: 10.1007/978-1-4899-3324-9

Smith AFM, Roberts GO. 1993. Bayesian Computation Via the Gibbs Sampler and Related Markov Chain Monte Carlo Methods. *Journal of the Royal Statistical Society: Series B* **55**: 3–23.

Stephens M. 2000a. Bayesian Analysis of Mixture Models with an Unknown Number of Components - An Alternative to Reversible Jump Methods. *The Annals of Statistics* **28**: 40–74.

Stephens M. 2000b. Dealing with label switching in mixture models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **62**: 795–809. DOI: 10.1111/1467-9868.00265

Trzcinski E, Holst E. 2010. Interrelationship among locus of control and years in management and unemployment: Differences by gender. Discussion Papers of DIW Berlin 974, DIW Berlin, German Institute for Economic Research.

van Dyk DA. 2010. Marginal markov chain monte carlo methods. *Statistica Sinica* **20**: 1423–1454.

van Dyk DA, Meng XL. 2001. The art of data augmentation. *Journal of Computational & Graphical Statistics* **10**: 1–50. DOI: 10.1198/10618600152418584

van Dyk DA, Park T. 2008. Partially Collapsed Gibbs Samplers: Theory and Methods. *Journal of the American Statistical Association* **103**: 790–796. DOI: 10.1198/016214508000000409

von Rosenbladt B, Stimmel S. 2005. Pretest Jugend zum SOEP 2006, Fragebogen und kognitiver Test. Working paper PB20-44899, DIW.