EARLY ENDOWMENTS, EDUCATION AND HEALTH [*]

Gabriella Conti
University of Chicago

James J. Heckman
University of Chicago,
University College Dublin
Cowles Foundation, Yale University
and the American Bar Foundation

Sergio Urzúa
Northwestern University

First version: October 10, 2008
This version: May 6, 2009

Abstract

In this paper we revisit the relationship between education and health. We disentangle the causal effect of education from the role played by cognitive and socio-emotional abilities, and early-life health in determining adult outcomes. We show that early endowments are important determinants of adult health and success. Using models with unobservable components generated by factor structures, we compute distributions of treatment effects and we allow responses to education to vary among observationally equivalent agents. We show that heterogeneity matters and that the individuals with the poorest endowment of capabilities are the ones who benefit the most from education. Our results also indicate that, in general, the average marginal effects of education on health ($AMTE$) are larger than the estimated average effects (ATE).

Keywords: health, education, early endowments, factor models, treatment effects

JEL codes:

---

1 Introduction

The positive correlation between health and schooling is one of the most well-established findings in the social sciences (Kolata, 2007). However, whether and to what extent this correlation reflects causality is still subject of debate.[1] Three explanations are offered in the literature: that causality runs from schooling to health,[2] that it runs from health to schooling, and that both are instead determined by a third factor, such as time or risk preferences (Fuchs, 1982). Understanding the importance of these mechanisms in generating observed differences in health by education is of a clear policy relevance.

Health gaps among education groups are rising (Meara, Richards, and Cutler, 2008). While many authors (Grossman, 1975; Perri, 1984; Wolfe, 1985), have noted that better health early in life is associated with higher educational attainment, and that more educated individuals, in turn, have better health later in life and better labor market prospects (Grossman and Kaestner, 1997; Cutler and Lleras-Muney, 2007), the exact mechanisms that produce this relationship remain to be identified. Education may be a proxy, in part, of capabilities developed in the early years. In addition to this, most of the literature in epidemiology and public health decomposes health disparities by education without taking into account the fact that people make different educational choices on the basis of capabilities which are also determinants of health behaviors. The literature in economics is aware of the problem, but it usually estimates mean effects.

In this paper we develop an empirical model of schooling choice and post-schooling outcomes, where both may be influenced by latent capabilities (early cognitive, socio-emotional and health endowments). We follow (and extend) the empirical approach developed by Heckman, Stixrud, and Urzua (2006) and Urzua (2008). In this framework, and following Heckman and Vytlacil (2007a,b), we analyze the impact of education on health and adult outcomes when responses to treatment vary among observationally identical persons and agents select into the treatment on the basis of their idiosyncratic responses. Using factor models to identify and estimate the distributions of counterfactuals, we identify various mean treatment effects (average treatment effect, treatment effect on the treated and average marginal treatment effect) as well as distributions of treatment effects. We are able to assess the gains (and the losses) from participating in post-compulsory

---

[1]See Grossman (2000) and Grossman (2006) for comprehensive reviews of the literature.

[2]Either because of productive efficiency, or allocative efficiency, see Grossman (1972) and Grossman (2008).

education for people at different deciles of the initial endowments' distributions to identify the groups of individuals who are more likely to benefit from policy.

We find that early endowments explain a significant part of the observed correlation between health and schooling, and we are able to exactly quantify the proportion of the health gap by education which can be attributed to these "third factors". We also uncover substantial heterogeneity in the causal effect of education on health: the effectiveness of education policy as health policy critically depends on which segment of the population is going to be affected by it.

2 Brief Literature Review

This paper joins together different strands of the literature in economics, epidemiology and psychology.

The first strand of research refers to the relationship between health and cognitive ability. While the importance of the 'ability bias' has long been recognized in labor economics (see Griliches, 1977), the effect of cognitive ability on health has received relatively less attention (Grossman, 1975; Shakotko, Edwards, and Grossman, 1982; Hartog and Oosterbeek, 1998; Elias, 2004; Auld and Sidhu, 2005; Kenkel, Lillard, and Mathios, 2006; Cutler and Lleras-Muney, 2007; Kaestner, 2009, are the only exceptions). Instead, it has been flourishing recently the field of cognitive epidemiology: large epidemiological studies have found that intelligence in childhood predicts substantial differences in adult morbidity and mortality (Whalley and Deary, 2001; Gottfredson and Deary, 2004; Batty, Deary, Schoon, and Gale, 2007).

The second strand of research refers to the relationship between socio-emotional skills and health. While there is already an established tradition in psychology on the importance of personality traits on health and mortality (see, for example, Hampson and Friedman (2008)), economists have just started to explore the effect of socio-emotional skills on health (Kaestner, 2009) and health-related behaviors (Heckman, Stixrud, and Urzua, 2006; Cutler and Lleras-Muney, 2007).

Our framework also relates to the literature on biological programming (Barker, 1997) and on the role of early-life conditions on subsequent cognitive (Edwards and Grossman, 1979) and socio-emotional development, and on health outcomes (Case, Fertig, and Paxson, 2005), and to life-course epidemiology (Kuh and Ben-Shlomo, 1997). We go beyond the current literature which

3

usually looks at the effect of birth weight on later outcomes, and model health as a latent factor to fully capture its multi-dimensionality and the possibility that is measured with error.[3]

The final strand of literature that we refer to is that on the non-market returns to education. The positive correlation between education and health has long been recognized in the economic, epidemiologic and medical literature, and several attempts at disentangling correlation from causality have been made.[4] Most of the studies looking at the non-pecuniary effects of education have used instrumental variables, with instruments ranging from changes in compulsory schooling laws (Adams, 2002; Arendt, 2005; Lleras-Muney, 2005; Silles, 2009; Spasojevic, 2003), to local unemployment rate (Arkes, 2003), or parental level of education (Auld and Sidhu, 2005). As noted in Grossman (2008), the IV estimates sometimes exceed the OLS estimates, and the effect of the instrument in the first stage in many studies is typically modest: "it is possible that small exogenous changes in schooling can have larger changes on health in certain settings while having small or no effects on health in other settings". Understanding the role of heterogeneity in the effect of education policy is one of the main aims of this paper.

3 Empirical Model of Endogenous Schooling Decisions and Post-Schooling Outcomes

We first introduce the schooling choice model. In this paper, we study the schooling decision of whether or not to continue education beyond compulsory education. We model this binary decision using a latent index structure. Let S_i^* denote the net utility of the individual from choosing post-compulsory education, and D_i a binary variable indicating individual's decision ($D_i = 1$ if individual continues beyond compulsory education, and $D_i = 0$ otherwise). Thus, we assume:

$$D_i = 1 \text{ if } S_i^* \geq 0, \ D_i = 0 \text{ otherwise,} \tag{1}$$

---

[3]As Currie (2006) put it, "Birthweight is only an imperfect indicator of child health".
[4]In an extensive review of the literature, Grossman (2005) concludes that there seems to be evidence of a causal effect of education on health.

We assume that net utility S_i^* is determined by observed and unobserved individual's characteristics. Specifically, we assume

$$S_i^* = \mu_S(Z_i) + V_i$$

where Z_i is a vector of observed characteristics determining individual's net utility level, and V_i is an unobserved random variable also affecting utility. (Z_i, V_i) are assumed to be independent. In our empirical implementation of the model, we assume a linear structure for $\mu_S(Z_i)$, i.e., $\mu_S(Z_i) = \gamma Z_i$.

Once the individual has decided his schooling level, all future outcomes (labor market outcomes and health behaviors) are observed conditional on the decision. In the context of our model, unobserved characteristics (at least) partially drive the schooling decision process. Additionally, to the extent that these unobserved components correlate with unobservables determining future outcomes, we need to control for the potential consequences of selection when comparing outcomes across schooling levels. We deal with this issue by modeling post-schooling variables using potential outcome models in which we allow the unobserved components to be correlated across schooling levels and outcomes. We distinguish between continuous and discrete outcomes.

3.1 Continuous Outcomes

Let (Y_{0i}, Y_{1i}) denote the potential outcomes for individual i corresponding, respectively, to the event of dropping out once reached the compulsory schooling level and continuing education beyond it. The model assumes that each of the potential outcomes is determined by individual's observable and unobservable characteristics. Specifically, we write the potential outcome associated with post-compulsory education as:

$$Y_{i1} = \mu_1\left(X_i, U_{i1}\right) \tag{2}$$

and the potential outcome under compulsory education as:

$$Y_{i0} = \mu_0\left(X_i, U_{i0}\right) \tag{3}$$

where X_i is a vector of observed characteristics and (U_{i1}, U_{i0}) denote the unobserved components. On theoretical grounds, an additive separable structure for $\mu_0(X_i, U_{i0})$ and $\mu_1(X_i, U_{i1})$ is not required. However, in our empirical implementation of the model we assume additive separability, i.e., $\mu_0(X_i, U_{i0}) = \beta_0 X_i + U_{i0}$ and $\mu_1(X_i, U_{i1}) = \beta_1 X_i + U_{i1}$. Notice that we do not impose any assumptions on the correlations between U_{i1}, U_{i0}, and V_i. We allow the unobserved components from outcomes and schooling choices to be correlated, and as a consequence of this, any comparison of outcomes across schooling groups should take into account the potential selection problem.

Expressions (4), (2), and (3) can be used to define observed outcome Y_i. Observed outcome Y_i can be written as :[5]

$$Y_i = D_i Y_{i1} + (1 - D_i) Y_{i0}.$$

3.2 Discrete Outcomes

As discussed below, many of the outcomes in our data set are dichotomous. In these cases, we use a model of potential outcomes with an underlying latent index structure. Let B_{i0}^* and B_{i1}^* denote the net latent utilities with an outcome in each of the two regimes: compulsory and post-compulsory education, respectively. These latent utilities are assumed to be a function of observed (Q_i) and unobserved $(\epsilon_{i1}, \epsilon_{i0})$ characteristics. Specifically, we assume:

$$B_{i1}^* = \kappa_1(Q_i, \epsilon_{i1})$$
$$B_{i0}^* = \kappa_0(Q_i, \epsilon_{i0})$$

where we assume $Q_i \perp\!\!\!\perp (\epsilon_{i0}, \epsilon_{i1})$. Associated with each B_{is}^* $(s = \{0, 1\})$, we define the binary variable B_{is} such that:

$$B_{is} = 1 \text{ if } B_{is}^* \geq 0, \; B_{is} = 0 \text{ otherwise.}$$

As in the case of continuous outcomes, in our empirical implementation of the model we assume linear specifications for the functions $\kappa_0(Q_i, \epsilon_{i0})$ and $\kappa_1(Q_i, \epsilon_{i1})$, i.e., $\kappa_0(Q_i, \epsilon_{i0}) = \lambda_0 Q_i + u_{i0}$ and

---

[5]This is the Neyman (1923) - Fisher (1935) - Cox (1958) - Rubin (1974) model of potential outcomes. It is also the switching regression model of Quandt (1972) or the Roy model of income distribution (Roy, 1951; Heckman and Honoré, 1990).

$\kappa_1 (Q_i, \epsilon_{i1}) = \lambda_1 Q_i + u_{i1}$. We also note that we allow for correlations between $\epsilon_{i1}, \epsilon_{i0}, U_{i1}, U_{i0}$, and V_i.

In this context, the observed outcome B_i can be written as:

$$B_i = B_{i1} D_i + B_{i0}(1 - D_i).$$

We also study the effect of schooling on multinomial outcomes. Suppose the individual is deciding among J different alternatives. We follow the same strategy as in the binary case, and denote by H^*_{ij1} and H^*_{ij0} the latent utilities associated with alternative j under post-compulsory and compulsory education, respectively. We assume:

$$
\begin{aligned}
H^*_{ij1} &= \gamma_1 (P_{ij}, \xi_{ij1}) \\
H^*_{ij0} &= \gamma_0 (P_{ij}, \xi_{ij0}) \text{ with } j = 1, \ldots, J,
\end{aligned}
$$

where P_{ij} represents the observed characteristics of individual i determining the utility associated with the j-th alternative, and (ξ_{ij0}, ξ_{ij1}) are the respective unobserved components. Given his schooling decision s, the agent selects the optimal alternative by solving $\max\{H^*_{i1s} \ldots H^*_{iJs}\}$, and we can define H_{is} – the decision conditional on schooling level s (with $s = \{0, 1\}$) – as:

$$H_{is} = j \quad \text{if} \quad H^*_{ijs} = \max\{H^*_{i1s}, \ldots, H^*_{iJs}\}.$$

As in the previous cases, during estimation we assume $\gamma_1 (P_{ij}, \xi_{ij1}) = \gamma_1 P_{ij} + \xi_{ij1}$ and $\gamma_0 (P_{ij}, \xi_{ij0}) = \gamma_0 P_{ij} + \xi_{ij0}$ for $j = 1, \ldots, J$, and we allow the unobserved components to be correlated across schooling levels and across outcomes.

Finally, the observed outcome H_i can be written as:

$$H_i = H_{i1} D_i + H_{i0}(1 - D_i).$$

3.3 Unobserved Endowments

An important feature of our model is that it allows for general correlations among the unobserved components, namely V_i, U_{i1}, U_{i0}, ϵ_{i0}, ϵ_{i1}, ξ_{ij0}, and ξ_{ij1} (for $j = 1, \ldots, J$). Formally, we assume:

$$V_i \not\perp\!\!\!\perp U_{i1} \not\perp\!\!\!\perp U_{i0} \not\perp\!\!\!\perp \epsilon_{i0} \not\perp\!\!\!\perp \epsilon_{i1} \not\perp\!\!\!\perp \xi_{ij0} \not\perp\!\!\!\perp \xi_{ij1} \not\perp\!\!\!\perp \ldots \not\perp\!\!\!\perp \xi_{iJ0} \not\perp\!\!\!\perp \xi_{iJ1} | \left(X_i, Z_i, \{P_{ij}\}_{j=1}^J, Q_i \right)$$

where $A \not\perp\!\!\!\perp B | C$ denotes "A and B are not independent conditional on C". Below we analyze the underlying structure causing the error terms to be correlated.

We model these general correlations by assuming that the error terms are governed by a factor structure which we interpret as unobserved endowments. That is, we posit the existence of a vector of latent endowments, which, as we show below, include cognitive, socio-emotional and health endowments. Specifically, and suppressing the sub-index i for a better exposition, if we let θ denote a vector of unobserved endowments, i.e., $\theta = (\theta_C, \theta_S, \theta_H)$, we assume

$$
\begin{aligned}
V &= \alpha_V \theta + \upsilon_V \\
U_1 &= \alpha_{U_1} \theta + \upsilon_{U_1} \\
U_0 &= \alpha_{U_0} \theta + \upsilon_{U_0} \\
\epsilon_0 &= \alpha_{\epsilon_0} \theta + \upsilon_{\epsilon_0} \\
\epsilon_1 &= \alpha_{\epsilon_1} \theta + \upsilon_{\epsilon_1} \\
\xi_{j0} &= \alpha_{\xi_{j0}} \theta + \upsilon_{\xi_{j0}} \\
\xi_{j1} &= \alpha_{\xi_{j1}} \theta + \upsilon_{\xi_{j1}} \\
&\vdots \\
\xi_{J0} &= \alpha_{\xi_{J0}} \theta + \upsilon_{\xi_{J0}} \\
\xi_{J1} &= \alpha_{\xi_{J1}} \theta + \upsilon_{\xi_{J1}}
\end{aligned}
$$

where $\upsilon_V \perp\!\!\!\perp \upsilon_{U_1} \perp\!\!\!\perp \upsilon_{U_0} \perp\!\!\!\perp \upsilon_{\epsilon_0} \perp\!\!\!\perp \upsilon_{\epsilon_1} \perp\!\!\!\perp \upsilon_{\xi_{j0}} \perp\!\!\!\perp \upsilon_{\xi_{j1}} \perp\!\!\!\perp \ldots \perp\!\!\!\perp \upsilon_{\xi_{J0}} \perp\!\!\!\perp \upsilon_{\xi_{J1}}$. Using this structure we can analyze what is the effect of each of the components of θ on each of the outcomes (including schooling decisions). However, without further structure the model is not identified. We must supplement our schooling choice model and outcome models with additional information. Importantly, the new source of information cannot be affected by the schooling decisions, otherwise it would also be contaminated by selection.

8

3.4 Measurement System as Identification Device

Following Carneiro, Hansen, and Heckman (2003), we posit a linear measurement system to identify the joint distribution of the unobserved endowments $\theta = (\theta_C, \theta_S, \theta_H)$. Appendix B presents the identification argument in detail. Here we introduce the main ingredients.

We supplement the model introduced in section 3 with a set of equations linking early cognitive, socio-emotional and health measures with our unobserved endowments. These equations allow the interpretation of unobserved θ as early cognitive, socio-emotional and health endowments. Specifically, if denote by M_C, M_S, M_H the set of early cognitive, socio-emotional and health variables, and assuming they are "dedicated" measures, we have:

$$
\begin{aligned}
M_{C1} &= \delta_{C1} X + \alpha_{C1}\theta_C + v_{C1} \\
&\vdots \\
M_{CN_C} &= \delta_{CN_C} X + \alpha_{CN_C}\theta_C + v_{CN_C} \\
M_{S1} &= \delta_{S1} X + \alpha_{S1}\theta_S + v_{S1} \\
&\vdots \\
M_{SN_C} &= \delta_{SN_S} X + \alpha_{SN_S}\theta_S + v_{SN_S} \\
M_{H1} &= \delta_{H1} X + \alpha_{H1}\theta_H + v_{H1} \\
&\vdots \\
M_{HN_H} &= \delta_{HN_H} X + \alpha_{HN_H}\theta_H + v_{HN_H}
\end{aligned}
$$

where X denotes the set of observed characteristics determining the measures, we assume $v_{C1} \perp\!\!\!\perp \ldots \perp\!\!\!\perp v_{CN_C} \perp\!\!\!\perp v_{S1} \perp\!\!\!\perp \ldots \perp\!\!\!\perp v_{SN_C} \perp\!\!\!\perp v_{H1} \perp\!\!\!\perp \ldots \perp\!\!\!\perp v_{HN_C}$, and N_C, N_S, and N_H denote the number of cognitive, socio-emotional, and health measures available, respectively. We require $N_C > 2$, $N_S > 2$, and $N_H > 2$ to secure the identification of the model (see appendix B for details). Finally, since there are no natural units for latent endowments, for some M_{Ck}, M_{Sk}, and M_{Hk} we set $\alpha_{Ck} = \alpha_{Sk} = \alpha_{Hk} = 1$.

4 Data and Empirical Implementation

We use data from the British Cohort Study (BCS70), a survey of all babies born (alive or dead) after the 24th week of gestation from 00.01 hours on Sunday, 5th April to 24.00 hours on Saturday,

11 April, 1970 in England, Scotland, Wales and Northern Ireland.[6] There have been six follow-ups so far to trace all members of the birth cohort: in 1975, 1980, 1986, 1996, 2000, 2004. We draw information from the birth survey, the second sweep (age 10) and the fifth sweep (age 30).

After removing children born with congenital abnormalities (1,035 observations), attrition (by wave 3) and non-whites (2,739 observations), and deleting responses with missing information on relevant variables (2,441 men and 2,205 women), we are left with a sample of 4,383 men (and 4,159 women).

4.1   Schooling and Post-Schooling Outcomes

The outcomes considered in our model are:

Schooling Variable. Our schooling variable represents whether or not the individual stayed on in school after reaching the minimum school-leaving age. For the individuals in our data, the minimum school-leaving age was 16 years.

Labor Market Outcomes. We analyze two labor market outcomes: (log) hourly wages by age 30, and whether or not the individual is a manager. All of them are measured by age 30.[7]

Healthy Behaviors. We consider five healthy behaviors: eating food fried in fat, eating fish, cannabis use, smoking, and exercise. All measured at age 30.[8]

Health and Marriage. We include three variables characterizing individual's health status by age 30. These are: Poor health, high blood pressure, and BMI.[9] Our empirical model also includes a binary outcome for whether or not the individual was married at age 30.

Summary statistics for our sample are displayed in Table 1.

---

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

[7]We generate hourly wages combining information from net income and usual hours worked per week (excluding meal breaks and including paid overtime). We deflate hourly wages using the cpi for january 2000. The variable "manager" refers to the main job in which the individual was employed at age 30.

[8]The variable "eating food fried in fat" is equal to 1 if the individual eats food fried in hard fat "more than once a day", "once a day", "3-6 days a week", "1-2 days a week"; it is equal to 0 if the frequency is "less than one day a week", "occasionally" or "never". The variable "eating fish" is an index for the frequency of eating fish of any kind, which takes the following values: 1=never; 2=occasionally; 3=less than one day a week; 4=1-2 days a week; 5=3-6 days a week or more. The variable "cannabis use" takes value 1 if the individual has ever tried cannabis. The variable "smoking" takes value 1 if the individual smokes cigarettes every day. The variable "exercise" takes value 1 if the individual does any regular exercise.

[9]The variable "poor health" takes the value 1 if the individual reports his/her health to be generally "fair" or "poor". The variable "high blood pressure" takes the value of 1 if the individual reports to have ever had high blood pressure. The variable "BMI" is constructed in the standard way as weight in kilograms divided by height in meters squared.

10

4.2 Measurement System

As measures of health endowment at birth we use birth weight, gestational age, an indicator of whether the baby was jaundiced, and an indicator of whether the baby had any other illness at birth (such as being cyanotic, having breathing difficulties or displaying cerebral signs).[10]

As indicators of cognitive skills we use seven test scores administered to the children at age ten (the Picture Language Comprehension Test, the Friendly Math Test, the Shortened Edinburgh Reading Test, and the four British Ability Scales). See appendix A.1 for further details.

Finally, as measurements of socio-emotional endowments we use six scales, two administered to the child (locus of control scale and Lawrence self-esteem scale), and four to the teacher (perseverance, popularity, boldness and cooperativeness). Appendix A.2 presents a detailed description of these scales.

Table 2 presents summary statistics for the variables included in the measurement system.

4.3 Exogenous Observed Characteristics

The set of exogenous characteristics used as covariates in the outcome equations includes: mother's age at birth, mother's education at birth (whether or not mother continued education beyond the minimum school-leaving age), parity (number of children born by one woman), father's high social class at birth, and total gross family income at age 10.

We use the additional following exogenous characteristics in the measurement system: mother lived in Scotland at birth, mother lived in Wales at birth, contraception (ever used contraception after January 1968), father's middle social class at birth.

Finally, the schooling choice model also includes as covariate the claimant counts (seasonally-adjusted rate of unemployment-related benefit claims) as observed in January 1986.

Table 3 presents summary statistics for the covariates included in our model.

---

[10]The measurements for health at birth are derived from Parts II and III of the Birth Questionnaire, which were completed, respectively, soon after birth and within the first week of life, using all the available records, discussing with the doctor or anyone else concerned with the care of the mother or the baby (or from personal communication where necessary). Babies weighing more than 5,000 grams (excessively overgrown fetuses) and less than 1,500 (very low birth weight) at birth have been excluded. The variable "gestational age" was constructed by the study team, on the basis of the question "What was the first day of the last menstrual period?"; unbelievable dates were also identified. The variable "jaundiced" takes the value 1 if the baby was jaundiced either less than 24 hours of life, or 24 hours or later, and 0 if the baby was not jaundiced. The variable "other" takes the value 1 if the baby experienced at least one of the following conditions: i) developed any breathing difficulties, including respiratory distress syndrome; ii) had cyanotic attacks; iii) had any "cerebral" signs; iv) suffered other conditions, such as feeding deficiencies, vomiting, fail to thrive, other superficial inflammation, hemorrhages.

4.4 Distributional Assumption and Estimation Strategy

We use mixture of normal approximations to the underlying endowments' distributions. Specifically:

$$\begin{bmatrix} \theta_C \\ \theta_S \\ \theta_H \end{bmatrix} \sim p_1 \Phi\left(\mu_1, \boldsymbol{\Sigma}_1\right) + (1 - p_1)\Phi\left(\mu_2, \boldsymbol{\Sigma}_2\right)$$

where μ_1 and μ_2 are vectors of dimension 3×1, and Σ_1 and Σ_2 are matrices of dimension 3×3. We do not restrict the variance-covariance matrices to be diagonal matrices, so we allow the underlying factors to be correlated.

For the idiosyncratic components associated with the binary and multinomial choice models $(\upsilon_V, \upsilon_{\epsilon_0}, \upsilon_{\epsilon_1}, \upsilon_{\xi_{j0}}, \upsilon_{\xi_{j1}}, \ldots, \upsilon_{\xi_{J0}}, \upsilon_{\xi_{J1}})$ we assume independent normal distribution with mean 0 and variance 1. For the idiosyncratic components associated with the continuous outcomes $(\upsilon_{U_0}, \upsilon_{U_1})$ we assume independent normal distributions with means equal to zero and unknown variances.

Given the structure of the model, the sample likelihood can be presented as:

$$\prod_{i=1}^{N} f(Y_i, D_i, B_i, H_i, M_{Ci}, M_{Si}, M_{Hi}|X_i, Z_i, P_i, Q_i)$$

where $f(\cdot)$ is the joint density of continuous and discrete outcomes, schooling choices, cognitive measures, socio-emotional scales, and early health variables. Importantly, since we can identify the joint distribution of the unobserved endowments driving the correlation across all the ingredients of the model, the sample likelihood can be written as:

$$\prod_{i=1}^{N} \int \int \int_{(\theta_C, \theta_S, \theta_h) \in \Theta} f(Y_i, D_i, B_i, H_i, M_{Ci}, M_{Si}, M_{Hi}|X_i, Z_i, P_i, Q_i, t_C, t_S, t_H) dF_\theta(t_C, t_S, t_H)$$

where $F_\theta(\cdot)$ denotes the joint cumulative density associated with unobserved cognitive, socio-emotional and health endowments. Notice that conditional on unobserved endowments (and observed characteristics) $Y_i, D_i, B_i, H_i, M_{Ci}, M_{Si}$, and M_{Hi} are independent, and the sample likelihood simplifies accordingly. This fact clearly illustrates the empirical convenience of using latent endowments to account for the correlation across outcomes, schooling decisions, and measurement systems. We use Bayesian MCMC methods to compute the sample likelihood. Details on the

12

implementation of Gibbs Sampling for a Roy model with correlated factors are presented in our web appendix.

5   Estimating the Causal Effects of Education

We use this model to generate mean treatment parameters and distributions of treatment parameters from a common set of structural parameters. This allows us to identify where in the overall distribution of endowments an increase in the educational level of the population is the most effective mean to improve adult outcomes.

Let $\Delta_i = Y_{i1} - Y_{i0}$ denote the person-specific treatment effect for a given individual i and outcome Y. As before, we denote by Y_{i1} and Y_{i0} the outcomes associated with post-compulsory education $D_i = 1$ and compulsory education $D_i = 0$, respectively.

Clearly, Δ_i involves factual and counterfactual regimes: for a given individual, what would be his or her outcome if he or she continued after compulsory education compared to the case where the person had not received it? Since our model deals with the estimation of counterfactual outcomes, we can use it to estimate the distribution of person-specific treatment effects. With this distribution in hand, we can compute different treatment parameters. We omit the subindex i for simplicity. Furthermore, without loss of generality, throughout this section we denote by Y and X any outcome variable and its associated covariates.

The first parameter that we consider is the average effect of the treatment on a person drawn randomly from the population of individuals. The average treatment effect is:

$$\Delta^{ATE} \equiv \int \int E(Y_1 - Y_0 | X = x, \theta = t) dF_{X,\theta}(x, t),$$

where we integrate $E(Y_1 - Y_0 | X = x, \theta = t)$ (the average treatment effect given $X = x$ and $\theta = t$) with respect to the distributions of X and θ.

The second parameter that we consider is the average effect of the treatment on the treated, i.e. on a person drawn randomly from the population of individuals who entered the treatment:

$$\Delta^{TT} \equiv \int \int E(Y_1 - Y_0 | X = x, \theta = t, D = 1) dF_{X,\theta|D=1}(x, t),$$

13

The third parameter that we consider is the average effect of the treatment on the untreated, i.e. on a person drawn randomly from the population of individuals who did not enter the treatment:

$$\Delta^{TUT} \equiv \int \int E(Y_1 - Y_0 | X = x, \theta = t, D = 0) dF_{X,\theta|D=0}(x,t),$$

Finally, we consider the average marginal treatment effect, defined in Heckman (1997) and Carneiro, Heckman, and Vytlacil (2008) as the average effect of participating in the treatment for the individuals who are at the margin of indifference between participating or not. Specifically,

$$\Delta^{AMTE} \equiv \int \int E(Y_1 - Y_0 | X = x, m(Z,V) < \varepsilon) dF_{X,\theta|m(Z,V)<\varepsilon}(x,t) \tag{4}$$

where $m(Z,V) = |\gamma Z + V|$ and ε is close to zero.

For the question addressed in this paper, knowledge of the distributional parameters is fundamental. Does anybody benefit from post-compulsory education? Among those who stay on after 16, what fraction benefits? The factor structure setup allows us to estimate these distributional parameters, following Aakvik, Heckman, and Vytlacil (2005) and Carneiro, Hansen, and Heckman (2003).

6   Results

Full estimates of the model parameters are available in the Web Appendix. As evident from Table 4, the model fits the data very well.

6.1   Early Endowments

We first start looking at the distributions of endowments. Figure 1 plots the distributions of the endowments for males. We notice that, while the correlation between the cognitive and the socio-emotional endowments is of a relevant magnitude (0.44), the correlation between both the cognitive and the socio-emotional endowments and the health endowment is very small (0.09 and 0.08). Figure 2 plots the densities of the estimated factors by schooling level for men. We find significant evidence of sorting under cognitive and socio-emotional traits, but not on the basis of early health.

6.2 The Non-market Benefits of Early Endowments

We now turn to the effect of the early endowments on later outcomes, conditional on schooling. Notice that both background characteristics and early endowments enter both the demand for schooling and the outcomes.

Tables 5 reports the average of the marginal effects of the early endowments on the outcomes we consider, for males. Higher cognitive ability leads to better performance in the labor market (but not in the marriage market) and to healthier nutritional choices, but not to better health *per se*. Somewhat surprisingly, we find no effect of cognitive ability on the probability of being a daily smoker at age 30. On the other hand, the positive effect of cognitive ability on experimentation with cannabis has been documented elsewhere (Conti, 2009). We now turn to the effect of socio-emotional endowment. We start noticing that it displays a strong and significant effect in the expected direction for the same outcomes for which cognitive ability plays no role: it increases the probability of being married by age 30 and reduces that of being a daily smoker, and has also a negative effect on the probability of reporting a series of bad health indicators.

Figures 3 to 12 graphically summarize our main results for the effects of early endowments on later outcomes. While the main results reproduce those already discussed with reference to Table 5 above, the figures allow us to see how the effects vary across educational levels along the endowments distribution.

6.3 The Non-Market Benefits of Education

6.3.1 Mean Treatment Effects

Table 6 presents the treatment effects obtained from our model. Table 7 presents the treatment effects obtained for individuals with endowments below the median. Table 8 presents the treatment effects obtained for individuals with endowments above the median. Table 9 presents the results of test for the equality of parameters across schooling levels. The results show an in-ambiguous rejection of the null hypothesis that the parameters are statistically identical across schooling levels.

15

6.3.2 Treatment Effects Heterogeneity

Figure 13 presents the results for labor market outcomes. Figure 14 presents the results for healthy behaviors. Figure 15 presents the results for health status and marriage.

6.3.3 Distribution of Treatment Effects

Figure 16 presents the results for labor market outcomes. Figures 17 presents the results for healthy behaviors. Figure 18 presents the results for health status and marriage.

6.4 Understanding Non-Market Gradients by Education

6.4.1 Decomposing Educational Disparities

We decompose the observed mean differences in outcomes across schooling levels, i.e., $E(Y_1|D = 1) - E(Y_0|D = 0)$ into $ATE = E(Y_1 - Y_0)$, sorting on gains $E(U_1 - U_0|D = 1)$, and sorting on levels (or selection bias) $E(Y_0|D = 1) - E(Y_0|D = 0)$.

Figure 19 presents the observed disparities in outcomes associated with schooling.

6.4.2 Understanding the Selection Mechanism

Let Y denote the outcome of interest. We use the sub-index i to denote the schooling level. Thus, Y_1 (Y_0) denotes the outcomes in the schooling level 1 (0). Our model assumes that the outcome of interest is determined by observable characteristics X and unobserved characteristics θ. Finally, the schooling choice model D also depends on observed and unobserved variables. In this context, and given the assumptions in our model, we can write

$$\Pr(Y_1 = 1|D = 1) = \int\int_{(X,\theta)\in\Omega_1} \Pr(Y_1 = 1|D = 1, x, t)\, f_{X,\theta|D=1}(x,t)\, dxdt$$

where $f_{X,\theta|D=1}(x,t)$ denotes the distribution of observed and unobserved characteristics in the population of individuals selecting schooling level $D = 1$. Likewise, we can define

$$\Pr(Y_0 = 1|D = 0) = \int\int_{(X,\theta)\in\Omega_0} \Pr(Y_0 = 1|D = 0, x, t)\, f_{X,\theta|D=0}(x,t)\, dxdt$$

16

where $f_{X,\theta|D=0}(x,t)$ denotes the distribution of observed and unobserved characteristics in the population of individuals selecting schooling level $D = 0$.

We then write

$$OLS = \Pr(Y_1 = 1|D = 1) - \Pr(Y_0 = 1|D = 0)$$

and we decompose OLS as:

$$OLS = TT + \text{Selection bias}$$

where

$$
\begin{aligned}
TT &= \Pr(Y_1 = 1|D = 1) - \Pr(Y_0 = 1|D = 1) \\
\text{Selection bias} &= \Pr(Y_0 = 1|D = 1) - \Pr(Y_0 = 1|D = 0).
\end{aligned}
$$

Finally, in order to investigate the role of θ and X in sorting, we use Bayes' Theorem and write:

$$f_{X,\theta|D=1}(x,t) = \frac{\Pr(D = 1|X = x, \theta = t)\, f_X(x)\, f_\theta(t)}{\Pr(D = 1)}$$

and since $f_{X,\theta|D=1}(x,t) = f_{X|D=1,\theta=t}(x)\, f_{\theta|D=1}(t)$ we form

$$f_{X|D=1,\theta=t}(x) = \frac{\Pr(D = 1|X = x, \theta = t)\, f_X(x)\, f_\theta(t)}{\Pr(D = 1)\, f_{\theta|D=1}(t)}$$

so finally,

$$\Pr(Y_0 = 1|D = 1) = \int\int_{(X,\theta)\in\Omega_1} \Pr(Y_0 = 1|D = 1, x, t) \frac{\Pr(D = 1|X = x, \theta = t)\, f_X(x)\, f_\theta(t)}{\Pr(D = 1)\, f_{\theta|D=1}(t)} f_{\theta|D=1}(t)\, dx\, dt$$

In this context, we evaluate the effect of observable characteristics by computing:

$$\text{Selection}_X = \widetilde{\Pr}(Y_0 = 1|D = 1) - \Pr(Y_0 = 1|D = 0)$$

where

$$\widetilde{\Pr}(Y_0 = 1|D = 1) = \int\int_{(X,\theta)\in\Omega_1} \Pr(Y_0 = 1|D = 1, x, t) \frac{\Pr(D = 1|X = x, \theta = t)\, f_X(x)\, f_\theta(t)}{\Pr(D = 1)\, f_{\theta|D=1}(t)} f_{\theta|D=0}(t)\, dx\, dt.$$

so that we use the conditional distribution of unobserved factors in schooling level 0 when integrating out the unobserved components.

The formula analyzing the effect of the unobserved characteristics is analogous to this last expression.

It is worth mentioning that our Bayesian approach requires also to integrate out with respect to the parameters in the model. For sake of simplicity we omit this integral.

Table 10 presents the results from this decomposition.

7 Conclusions

In this paper we develop an empirical model of schooling choice and post-schooling outcomes, where both dimensions are influenced by latent capabilities (early cognitive, socio-emotional and health endowments). We allow these latent capabilities to be correlated, extending in this way the previous literature (see Heckman, Stixrud, and Urzua, 2006; Urzua, 2008). Our results suggest that while the correlation between cognitive and the socio-emotional endowments is of a relevant magnitude (0.44), the correlations between cognitive and health endowments, and socio-emotional and health endowments are small (0.09 and 0.08).

We find evidence of positive impacts of early endowments on schooling decisions, and on post-schooling labor market outcomes and healthy behaviors.

Using the structure of the estimated model, we then analyze the treatment effect of education on health and adult outcomes when responses to treatment vary among observationally identical persons. We identify various mean treatment effects (average treatment effect, treatment effect on the treated and average marginal treatment effect) as well as distributions of treatment effects. We assess the gains (and the losses) from participating in post-compulsory education for people at different deciles of the initial endowments' distributions. By doing this, we can identify the groups of individuals who are more likely to benefit from the treatment (education).

We find that early endowments explain a significant part of the observed correlation between health and schooling, and we are able to quantify the proportion of the health gap by education which can be attributed to these "third factors". We also uncover substantial heterogeneity in the causal effect of education on health: the effectiveness of education policy as health policy critically

18

depends on which segment of the population is going to be affected by it.

Specifically, our results suggest that the larger gains associated with schooling are observed among individuals with low levels of endowments. Therefore, education seems to play a compensatory role. When decomposing the observed educational disparities in health into sorting and causal effects, we estimate that selection bias explains 28% of the observed educational disparities in "high-blood pressure", 42% of the disparities in "poor-health", and 76% of the disparities in BMI. Importantly, for each of these cases, we estimate that a significant fraction of the selection bias can be attributed to the presence of latent endowments.

Our results also indicate that, in the overall population, the average marginal effects of education on health status and healthy behaviors ($AMTE$) are larger than the estimated average effects (or ATE). This suggests significant non-market (average) gains for individuals indifferent between continuing or not their education.

A Measurement System

A.1 Cognitive Ability Tests - BCS70 age 10

Test	Content	# items	Items coverage
PLCT: Pictorial Language Comprehension Test	New test specifically developed for the BCS70 on the basis of the American Peabody Picture Vocabulary Test and the English Picture Vocabulary Test.	71	vocabulary, sequence and sentence comprehension
SERT: Shortened Edinburgh Reading Test	Shortened version of the Edinburgh Reading Test, a test of word recognition particularly designed to capture poor readers. Reliability=0.87.	75	vocabulary, syntax, sequencing, comprehension, and retention
FMT: Friendly Math Test	New test specifically designed for the BCS70. Reliability=0.92.	72	arithmetic, fractions, algebra, geometry and statistics
BAS-WD: British Ability Scale - Word Definition	One of the two verbal scales of the British Ability Scale, which measures a mental construct similar to IQ.	37	
BAS-WS: British Ability Scale - Word Similarities	One of the two verbal scales of the British Ability Scale.	21	
BAS-RD: British Ability Scale - Recall Digits	One of the two non-verbal scales of the British Ability Scale.	34	
BAS-M: British Ability Scale - Matrices	One of the two verbal scales of the British Ability Scale.	28	

A.2 Non-Cognitive Ability Tests - BCS70 age 10

LOCUS OF CONTROL (CARALOC) SCALE

1. Do you feel that most of the time it's not worth trying hard because things never turn out right anyway?

2. Do you feel that wishing can make good things happen?

3. Are people good to you no matter how you act towards them?

4. Do you like taking part in plays or concerts? *(distractor item)*

5. Do you usually feel that it's almost useless to try in school because most children are cleverer than you?

6. Is a high mark just a matter of "luck" for you?

7. Are you good at spelling? *(distractor item)*

8. Are tests just a lot of guesswork for you?

9. Are you often blamed for things which just aren't your fault?

10. Are you the kind of person who believes that planning ahead makes things turn out better?

11. Do you find it easy to get up in the morning? *(distractor item)*

12. When bad things happen to you, is it usually someone else's fault?

13. When someone is very angry with you, is it impossible to make him your friend again?

14. When nice things happen to you is it only good luck?

15. Do you feel sad when it is time to leave school each day?

16. When you get into an argument is it usually the other person's fault?

17. Are you surprised when your teacher says you've done well?

18. Do you usually get low marks, even when you study hard?

19. Do you like to read books? *(distractor item)*

20. Do you think studying for tests is a waste of time?

Possible answers: yes, no, don't know.

LAWRENCE SELF-ESTEEM (LAWSEQ) SCALE

1. Do you think that your parents usually like to hear about your ideas?

2. Do you often feel lonely at school?

3. Do other children often break friends or fall out with you?

4. Do you like team games? *(distractor item)*

5. Do you think that other children often say nasty things about you?

6. When you have to say things in front of teachers, do you usually feel shy?

7. Do you like writing stories or doing other creative writing? *(distractor item)*

8. Do you often feel sad because you have nobody to play with at school?

9. Are you good at mathematics? *(distractor item)*

10. Are there lots of things about yourself you would like to change?

11. When you have to say things in front of other children, do you usually feel foolish?

12. Do you find it difficult to do things like woodwork or knitting? *(distractor item)*

13. When you want to tell a teacher something, do you usually feel foolish?

14. Do you often have to find new friends because your old friends are playing with somebody else?

15. Do you usually feel foolish when you talk to your parents?

16. Do other people often think that you tell lies?

Possible answers: yes, no, don't know.

PERSEVERANCE SCALE

"How much perseverance does the child show in face of difficult tasks?" [assessed by the teacher]
The scale ranges from 1=none to 47=unlimited.

POPULARITY SCALE

"Please make an estimate of how you see the child in regard to the following issues [...] Is highly popular with his peers [...] Is not at all popular with peers." [assessed by the teacher]
The scale ranges from 1=not popular to 47=highly popular.

BOLDNESS SCALE

"Please make an estimate of how you see the child in regard to the following issues [...] Shows extreme boldness of behaviour towards peers [...] Shows extreme shyness in the company of peers." [assessed by the teacher]
The scale ranges from 1=extreme shyness to 47=extreme boldness.

COOPERATIVENESS SCALE

"Please make an estimate of how you see the child in regard to the following issues [...] Very cooperative with peers [...] Unwilling to cooperate with peers." [assessed by the teacher]
The scale ranges from 1=uncooperative to 47=very cooperative.

B Identification of the three-factor model

This section provides a brief discussion of the strategy used to identify our model. For notational simplicity, we keep the conditioning on X implicit. Consider a set of K variables such that:

$$Y = A\theta + \epsilon \tag{5}$$

where θ are factors, ϵ uniquenesses, Y is $k \times 1$, A is $k \times 3$, θ is 3×1 and ϵ is $k \times 1$. First, assume that:

$$E(\epsilon) = 0$$

$$Var(\epsilon\epsilon') = \Omega = \begin{pmatrix} \sigma_{\epsilon 1}^2 & 0 & \ldots & 0 \\ 0 & \sigma_{\epsilon 2}^2 & 0 & \vdots \\ \vdots & 0 & \ldots & \vdots \\ 0 & \ldots & 0 & \sigma_{\epsilon K}^2 \end{pmatrix}$$

$$E(\theta) = 0$$

$$Var(Y) = A\Sigma_\theta A' + \Omega$$

$$\Sigma_\theta = \begin{pmatrix} \sigma_{\theta_1}^2 & \sigma_{\theta_1\theta_2} & \sigma_{\theta_1\theta_3} \\ \sigma_{\theta_1\theta_2} & \sigma_{\theta_2}^2 & \sigma_{\theta_2\theta_3} \\ \sigma_{\theta_1\theta_3} & \sigma_{\theta_2\theta_3} & \sigma_{\theta_3}^2 \end{pmatrix}$$

The only source of information on A and Σ_θ that we use is from covariances. We have $\frac{K(K-1)}{2}$ covariance terms from the data. With these we want to identify:

- $\sigma_{\epsilon_k}^2$ for $k = 1, \ldots, K$ (K unknowns)

- $3K$ factor loadings contained in the matrix A

- Nine elements of Σ_θ

It is a well-known result from factor analysis that this model is not identified against orthogonal transformations. In order to identify the model, we start imposing a normalization assumption.
Assumption 1: Since the scale of each factor is arbitrary, one loading devoted to each factor is normalized to unity to set the scale:

$$A = \begin{pmatrix} 1 & \alpha_{12} & \alpha_{13} \\ \alpha_{21} & 1 & \alpha_{23} \\ \alpha_{31} & \alpha_{32} & 1 \\ \vdots & \vdots & \vdots \\ \alpha_{K1} & \alpha_{K2} & \alpha_{K3} \end{pmatrix}$$

Notice that, differently from Heckman, Stixrud, and Urzua (2006), we do not assume that the factors are independent, so:

$$\theta_1 \not\perp \theta_2 \not\perp \theta_3$$

With these assumptions, working only with covariance information, we require that:

$$\frac{K(K-1)}{2} \geq 3K - 3 + 6$$

where $\frac{K(K-1)}{2}$ is the number of covariances computed from the data, $3K - 3$ is the number of unrestricted parameters in A and 6 is the number of elements in Σ_θ. Hence $K \geq 8$ is a necessary condition for identification. Our empirical model satisfies it. To give greater interpretability to the three factors, consider the following structure for the system (5):

$$Y = \begin{bmatrix} C \\ S \\ H \\ R \end{bmatrix} = A\Theta + \epsilon$$

where C is a vector of dimension $n_C(\geq 3)$, N is a vector of dimension $n_S(\geq 3)$, H is a vector of dimension $n_H(\geq 3)$, and R is a vector of dimension $n_K = K - n_C - n_S - n_H(> 0)$. The vectors C, S and H represent, respectively, the sets of cognitive, socio-emotional and health measurements, while R contains our outcomes of interest. We now make a further assumption.

Assumption 2:

$$A = \begin{pmatrix}
1 & 0 & 0 \\
\alpha_2^C & 0 & 0 \\
\vdots & \vdots & \vdots \\
\alpha_{nC}^C & 0 & 0 \\
0 & 1 & 0 \\
0 & \alpha_2^S & 0 \\
\vdots & \vdots & \vdots \\
0 & \alpha_{nS}^S & 0 \\
0 & 0 & 1 \\
0 & 0 & \alpha_2^H \\
\vdots & \vdots & \vdots \\
0 & 0 & \alpha_{nH}^H \\
\alpha_1^{C,R} & \alpha_1^{S,R} & \alpha_1^{H,R} \\
\vdots & \vdots & \vdots \\
\alpha_{nR}^{C,R} & \alpha_{nR}^{S,R} & \alpha_{nR}^{H,R}
\end{pmatrix}$$

We now prove how identification is achieved in our estimated model.

Remark: A sufficient condition for identification is the existence of at least three measurements for each factor. Note this is a necessary condition to identify the parameters of each of the factors out of its measurement system, as it is clear from the following. The measurement systems for, respectively, cognitive, socio-emotional and health capability is:

$$
\begin{aligned}
C_1 &= & \theta^C &+ & \epsilon_1^C \\
C_2 &= & \alpha_2^C \theta^C &+ & \epsilon_2^C \\
C_3 &= & \alpha_3^C \theta^C &+ & \epsilon_3^C \\
S_1 &= & \theta^S &+ & \epsilon_1^S \\
S_2 &= & \alpha_2^S \theta^S &+ & \epsilon_2^S \\
S_3 &= & \alpha_3^S \theta^S &+ & \epsilon_3^S \\
H_1 &= & \theta^H &+ & \epsilon_1^H \\
H_2 &= & \alpha_2^H \theta^H &+ & \epsilon_2^H \\
H_3 &= & \alpha_3^H \theta^H &+ & \epsilon_3^H
\end{aligned}
$$

25

By taking ratios of covariances, we can identify the elements of A and Σ_θ.

$$
\begin{aligned}
Cov(C_1, C_2) &= \alpha_2^C \sigma_{\theta_C}^2 \\
Cov(C_1, C_3) &= \alpha_3^C \sigma_{\theta_C}^2 \\
Cov(C_2, C_3) &= \alpha_2^C \alpha_3^C \sigma_{\theta_C}^2
\end{aligned}
$$

$$
\begin{aligned}
\frac{Cov(C_2, C_3)}{Cov(C_1, C_2)} &= \alpha_3^C \\
\frac{Cov(C_2, C_3)}{Cov(C_1, C_3)} &= \alpha_2^C \\
\frac{Cov(C_1, C_2)}{\alpha_2^C} &= \sigma_{\theta_C}^2
\end{aligned}
$$

Repeating the same reasoning for the measurement system for socio-emotional ability and health, we identify α_2^S, α_3^S, α_2^H, α_3^H, $\sigma_{\theta_S}^2$, $\sigma_{\theta_H}^2$. Then, by taking covariances between the measurements on which the factors are normalized, we identify the factor covariances:

$$
\begin{aligned}
Cov(C_1, S_1) &= \sigma_{\theta_C \theta_S} \\
Cov(C_1, H_1) &= \sigma_{\theta_C \theta_H} \\
Cov(S_1, H_1) &= \sigma_{\theta_S \theta_H}
\end{aligned}
$$

Then, by using the variances of Y_k for $k = 1, ..., K$, we can identify the elements of Ω. Finally, by taking covariances between outcomes and measurements, we can identify the parameters of the state-contingent outcomes, such as, for example:

$$
W_1 = \alpha_{W1}^C \theta^C + \alpha_{W1}^S \theta^S + \alpha_{W1}^H \theta^H + \epsilon_{W1}
$$

Acknowledgments

References

Aakvik, A., J. J. Heckman, and E. J. Vytlacil (2005): "Estimating Treatment Effects for Discrete Outcomes When Responses to Treatment Vary: An Application to Norwegian Vocational Rehabilitation Programs," *Journal of Econometrics*, 125(1-2), 15–51.

Adams, S. J. (2002): "Educational Attainment and Health: Evidence from a Sample of Older Adults," *Education Economics*, 10(1), 97–109.

Arendt, J. N. (2005): "Does education cause better health? A panel data analysis using school reforms for identification," *Economics of Education Review*, 24(2), 149–160.

Arkes, J. (2003): "Does Schooling Improve Adult Health?," *Working Paper*.

Auld, M. C., and N. Sidhu (2005): "Schooling, Cognitive Ability and Health," *Health Economics*, 14(10), 1019–1034.

Barker, D. J. (1997): "Intrauterine programming of coronary heart disease and stroke.," *Acta Paediatr Suppl*, 423, 178–82.

Batty, G. D., I. J. Deary, I. Schoon, and C. R. Gale (2007): "Mental Ability across Childhood in Relation to Risk Factors for Premature Mortality in Adult Life: The 1970 British Cohort Study," *Journal of Epidemiology and Community Health*, 61(11), 997–1003.

Carneiro, P., K. Hansen, and J. J. Heckman (2003): "Estimating Distributions of Treatment Effects with an Application to the Returns to Schooling and Measurement of the Effects of Uncertainty on College Choice," *International Economic Review*, 44(2), 361–422.

Carneiro, P., J. J. Heckman, and E. J. Vytlacil (2008): "Evaluating Marginal Policy Changes and the Average Effect of Treatment for Individuals at the Margin," Unpublished manuscript, University of Chicago, Department of Economics.

Case, A., A. Fertig, and C. Paxson (2005): "The Lasting Impact of Childhood Health and Circumstance," *Journal of Health Economics*, 24(2), 365–389.

Conti, G. (2009): "Cannabis, Cognition and Wages," Unpublished manuscript, University of Chicago, Department of Economics.

Cox, D. R. (1958): *Planning of Experiments*. Wiley, New York.

Currie, J. (2006): "Healthy, Wealthy, and Wise? The Link Between SES, Children's Health, and Human Capital Development," Presented at the IZA Seminar, April 7, 2006. Bonn, Germany. Forthcoming *Journal of Economic Literature*.

Cutler, D. M., and A. Lleras-Muney (2007): "Understanding Differences in Health Behaviors by Education," Unpublished working paper.

Edwards, L. N., and M. Grossman (1979): "The Relationship between Children's Health and Intellectual Development," in *Health: What is it Worth?*, ed. by S. Mushkin, and D. Dunlop. Pergamon Press, Elmsford, NY.

Elias, J. (2004): "The effects of ability and family background on non-monetary returns to education," *University of Chicago mimeograph, November*.

Fisher, R. A. (1935): *The Design of Experiments*. Oliver and Boyd, London.

Fuchs, V. R. (1982): *Time Preference and Health: An Exploratory Study*. National Bureau of Economic Research Cambridge, Mass., USA.

Gottfredson, L. S., and I. J. Deary (2004): "Intelligence Predicts Health and Longevity, but Why?," *Current Directions in Psychological Science*, 13(1), 1–4.

Griliches, Z. (1977): "Estimating the Returns to Schooling: Some Econometric Problems," *Econometrica*, 45(1), 1–22.

Grossman, M. (1972): "On the Concept of Health Capital and the Demand for Health," *Journal of Political Economy*, 80(2), 223–255.

———— (1975): "The Correlation Between Health and Schooling," in *Household Production and Consumption*, ed. by N. E. Terleckyj, pp. 147–211. Columbia University Press, New York.

———— (2000): "The Human Capital Model," in *Handbook of Health Economics*, ed. by A. J. Culyer, and J. P. Newhouse, vol. 1, pp. 347–408. Elsevier, Amsterdam.

———— (2005): "Education and Nonmarket Outcomes," *NBER Working Paper*.

——— (2006): "Education and Nonmarket Outcomes," *Handbook of the Economics of Education*, p. 577.

GROSSMAN, M. (2008): "The Relationship Between Health and Schooling," *Eastern Economic Journal*, 34, 281–292.

GROSSMAN, M., AND R. KAESTNER (1997): "Effects of Education on Health," in *The Social Benefits of Education*, ed. by J. R. Behrman, and N. Stacey, pp. 69–124. University of Michigan Press, Ann Arbor, MI.

HAMPSON, S. E., AND H. S. FRIEDMAN (2008): "Personality and health: A lifespan perspective," in *The handbook of personality: Theory and research*, ed. by O. P. John, R. Robins, and L. Pervin, pp. 770–794. Guilford, New York, third edn.

HARTOG, J., AND H. OOSTERBEEK (1998): "Health, wealth and happiness: why pursue a higher education?," *Economics of Education Review*, 17(3), 245–256.

HECKMAN, J. J. (1997): "Instrumental Variables: A Study of Implicit Behavioral Assumptions Used in Making Program Evaluations," *Journal of Human Resources*, 32(3), 441–462, Addendum published vol. 33 no. 1 (Winter 1998).

HECKMAN, J. J., AND B. E. HONORÉ (1990): "The Empirical Content of the Roy Model," *Econometrica*, 58(5), 1121–1149.

HECKMAN, J. J., J. STIXRUD, AND S. URZUA (2006): "The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior," *Journal of Labor Economics*, 24(3), 411–482.

HECKMAN, J. J., AND E. J. VYTLACIL (2007a): "Econometric Evaluation of Social Programs, Part I: Causal Models, Structural Models and Econometric Policy Evaluation," in *Handbook of Econometrics*, ed. by J. Heckman, and E. Leamer, vol. 6B, pp. 4779–4874. Elsevier, Amsterdam.

——— (2007b): "Econometric Evaluation of Social Programs, Part II: Using the Marginal Treatment Effect to Organize Alternative Economic Estimators to Evaluate Social Programs and to Forecast Their Effects in New Environments," in *Handbook of Econometrics*, ed. by J. Heckman, and E. Leamer, vol. 6B, pp. 4875–5144. Elsevier, Amsterdam.

KAESTNER, R. (2009): "Adolescent Cognitive and Non-cognitive Correlates of Adult Health," NBER Working Papers 14924, National Bureau of Economic Research, Inc.

KENKEL, D., D. LILLARD, AND A. MATHIOS (2006): "The Roles of High School Completion and GED Receipt in Smoking and Obesity," *Journal of Labor Economics*, 24(3), 635–660.

KOLATA, G. (2007): "A surprising secret to a long life: Stay in school," .

KUH, D., AND Y. BEN-SHLOMO (1997): *A Lifecourse Approach to Adult Disease*. Oxford University Press, Projected Date: 9710, New York.

LLERAS-MUNEY, A. (2005): "The Relationship Between Education and Adult Mortality in the United States," *Review of Economic Studies*, 72(1), 189–221.

MEARA, E. R., S. RICHARDS, AND D. M. CUTLER (2008): "The Gap Gets Bigger: Changes in Mortality and Life Expectancy, by Education, 1981-2000," *Health Affairs*, 27(2), 350–360.

NEYMAN, J. (1923): "Statistical Problems in Agricultural Experiments," *Journal of the Royal Statistical Society*, II (Supplement)(2), 107–180.

PERRI, T. J. (1984): "Health Status and Schooling Decisions of Young Men," *Economics of Education Review*, 3(3), 207–213.

QUANDT, R. E. (1972): "A New Approach to Estimating Switching Regressions," *Journal of the American Statistical Association*, 67(338), 306–310.

ROY, A. (1951): "Some Thoughts on the Distribution of Earnings," *Oxford Economic Papers*, 3(2), 135–146.

RUBIN, D. B. (1974): "Estimating causal effects of treatments in randomized and nonrandomized studies," *Journal of Educational Psychology*, 66(5), 688–701.

SHAKOTKO, R. A., L. N. EDWARDS, AND M. GROSSMAN (1982): "An Exploration of the Dynamic Relationship Between Health and Cognitive Development in Adolescence," .

SILLES, M. A. (2009): "The causal effect of education on health: Evidence from the United Kingdom," *Economics of Education Review*, 28(1), 122–128.

31

SPASOJEVIC, J. (2003): "Effects of Education on Adult Health in Sweden: Results from a Natural Experiment," *New York: City University of New York Graduate Center.*

URZUA, S. (2008): "Racial Labor Market Gaps: The Role of Abilities and Schooling Choices," Forthcoming, *Journal of Human Resources.*

WHALLEY, L., AND I. DEARY (2001): "Longitudinal cohort study of childhood IQ and survival up to age 76," *British Medical Journal*, 322(7290), 819.

WOLFE, B. (1985): "The influence of health on school outcomes. A multivariate approach," *Medical Care*, 23(10), 1127–1138.

Figure 1: Joint Distributions of Unobserved Endowments

Males

i. Cognitive and Socio-Emotional Endowments
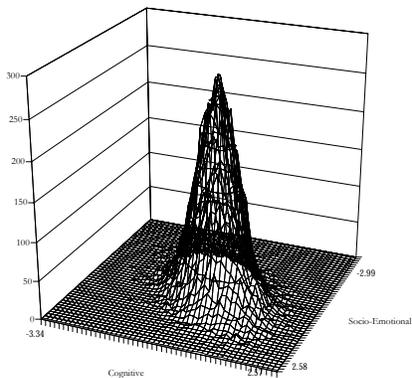
ii. Socio-Emotional and Health Endowments
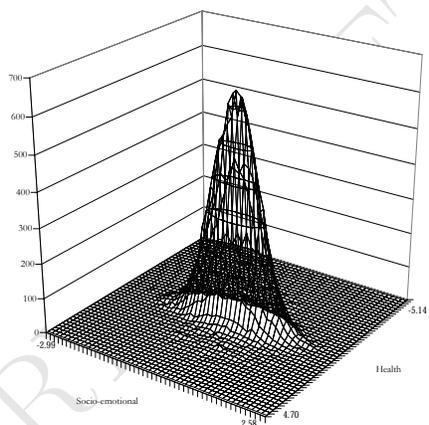
iii. Health and Cognitive Endowments

33
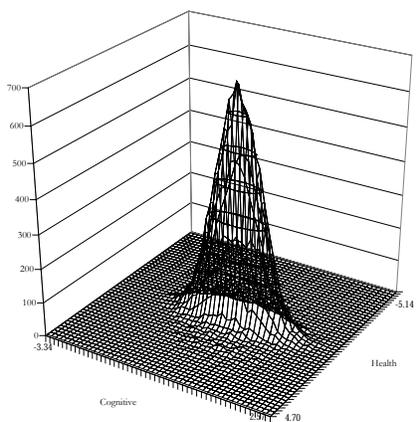
Figure 2: Marginal Distributions of Endowments by Schooling Level

Males
i. Cognitive Ability

ii. Socio-emotional Endowment

iii. Health Endowment
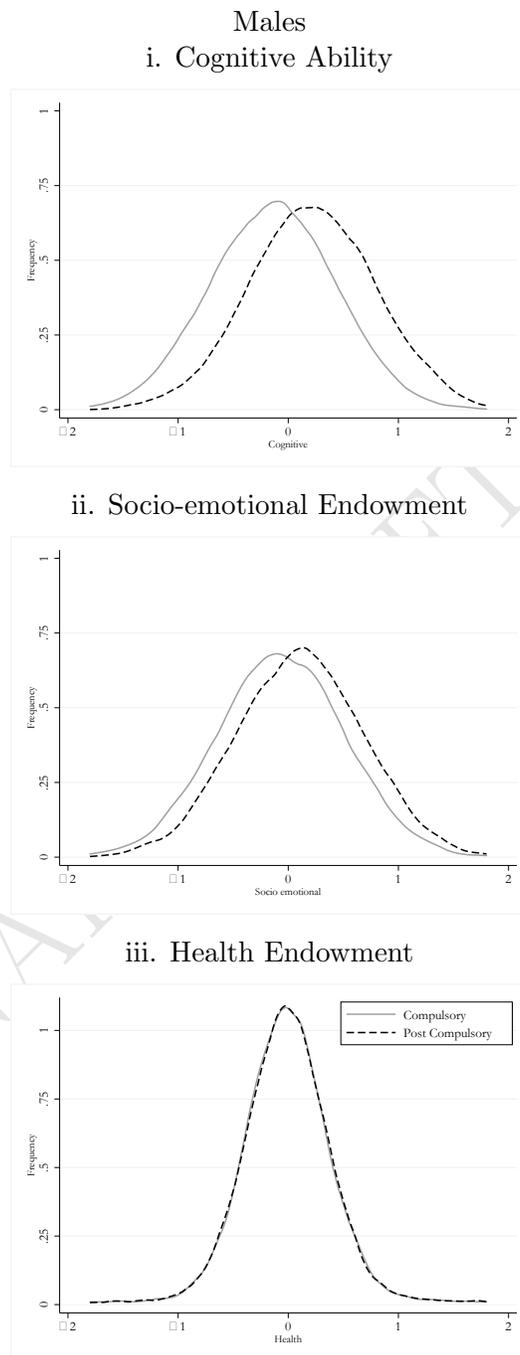
Figure 3: Effects of Endowments: Log of hourly wage

(A) Compulsory Education

(B) Post-Compulsory Education

i. Cognitive Ability
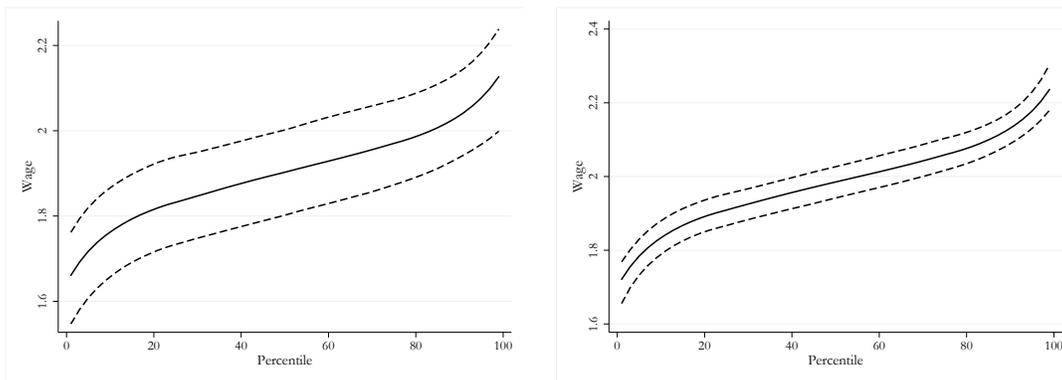
ii. Socio-emotional Endowment
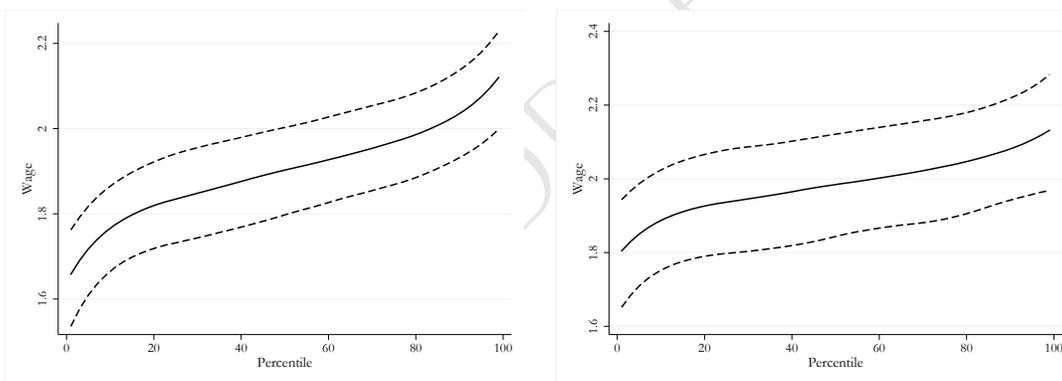
iii. Health Endowment
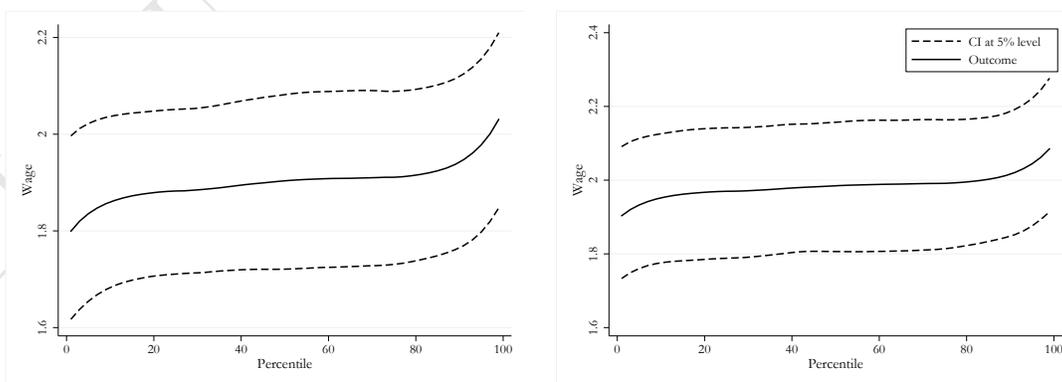
Figure 4: Effects of Endowments: Occupational choice: manager by age 30

(A) Compulsory Education   (B) Post-Compulsory Education

i. Cognitive Ability
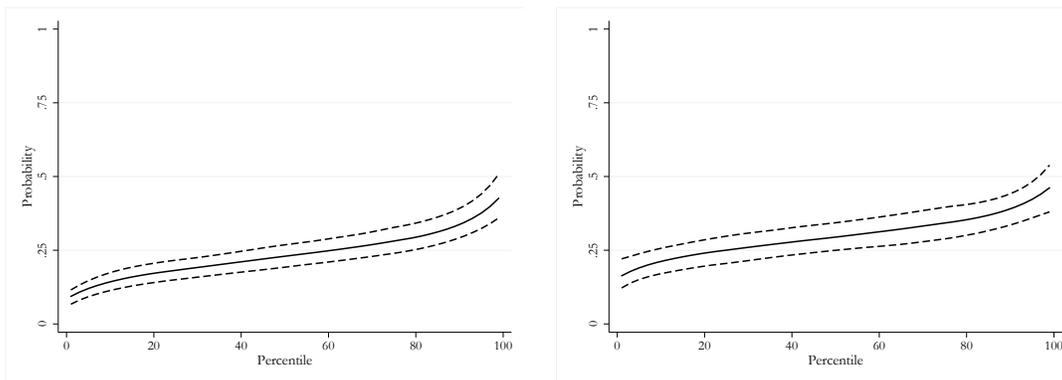
ii. Socio-emotional Endowment
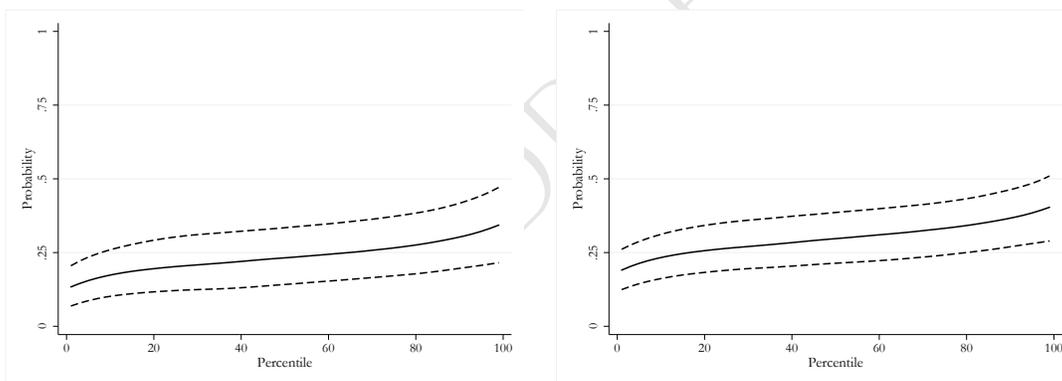
iii. Health Endowment
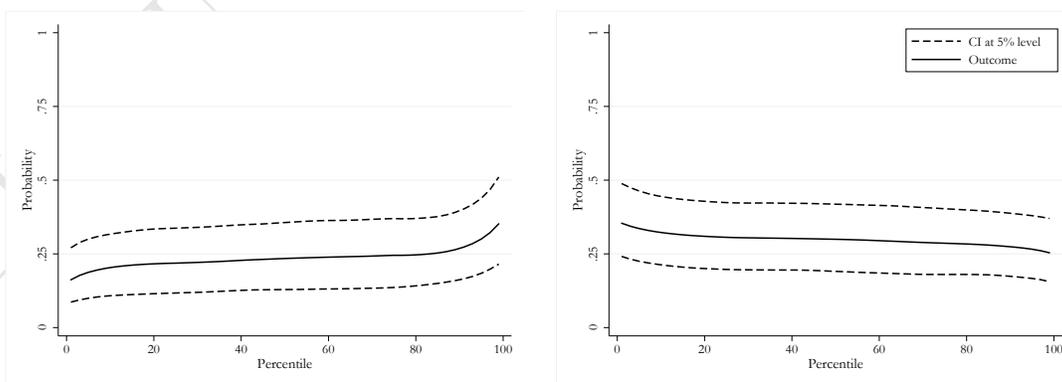
Figure 5: Effects of Endowments: Marriage by age 30

(A) Compulsory Education   (B) Post-Compulsory Education

i. Cognitive Ability
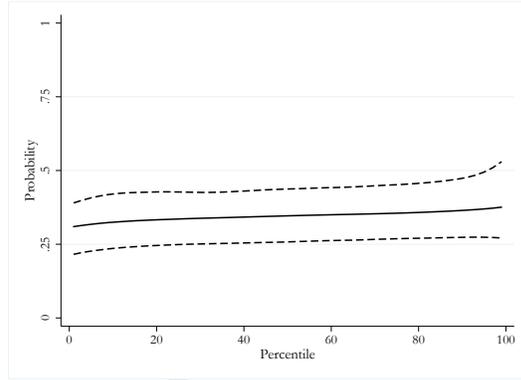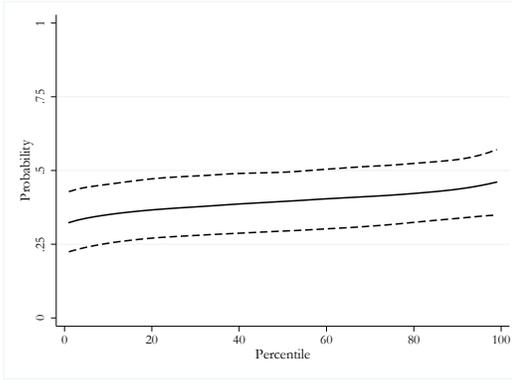
ii. Socio-emotional Endowment
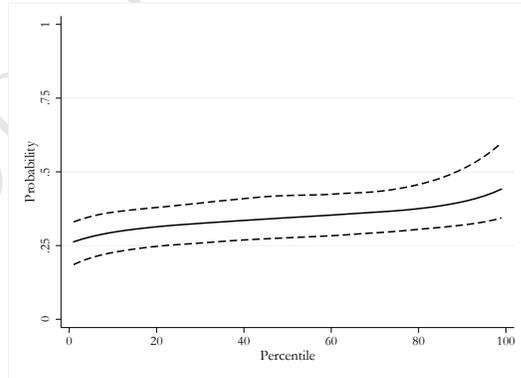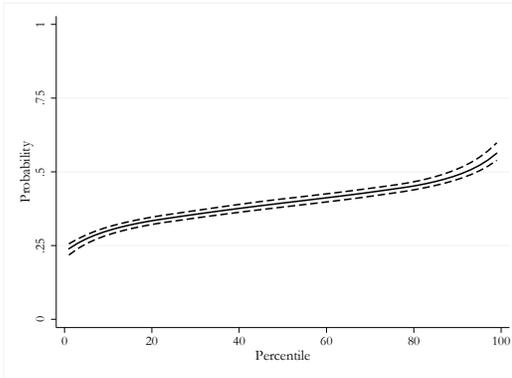
iii. Health Endowment
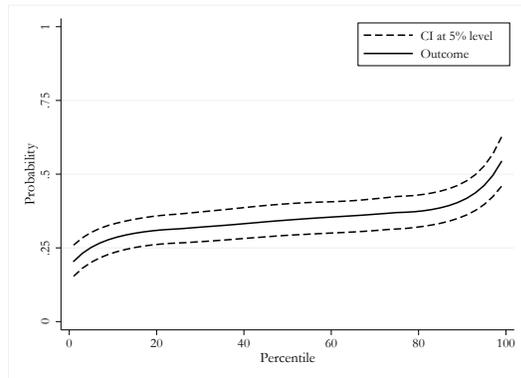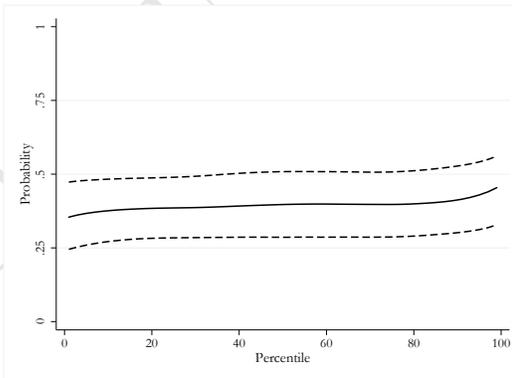
Figure 6: Effects of Endowments: Consumption of Fat and Fried Food

(A) Compulsory Education　　　　(B) Post-Compulsory Education

i. Cognitive Ability
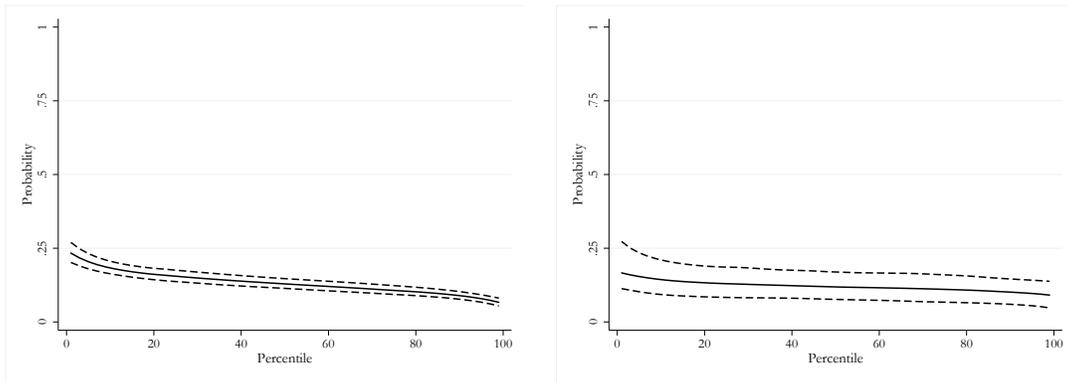
ii. Socio-emotional Endowment
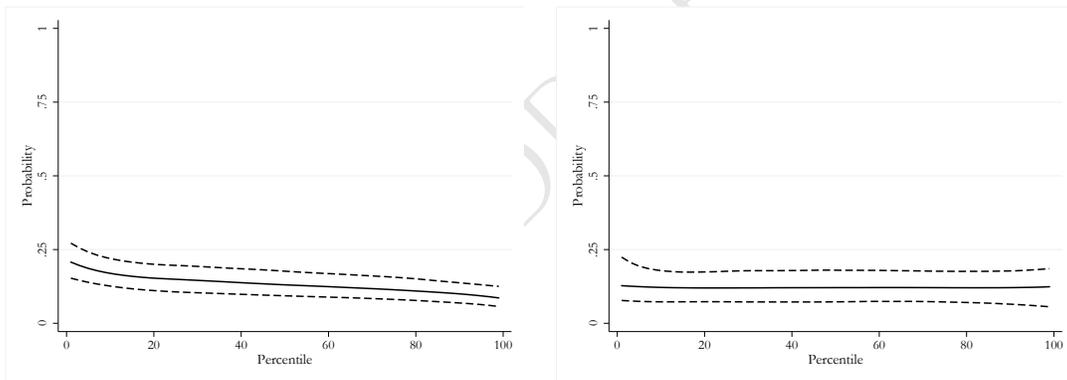
iii. Health Endowment
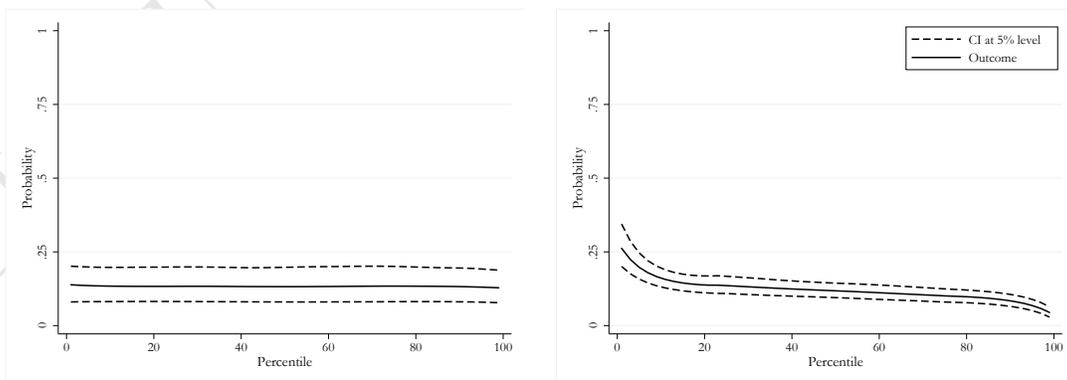
Figure 7: Effects of Endowments: Consumption of Fish

(A) Compulsory Education

(B) Post-Compulsory Education

i. Cognitive Ability
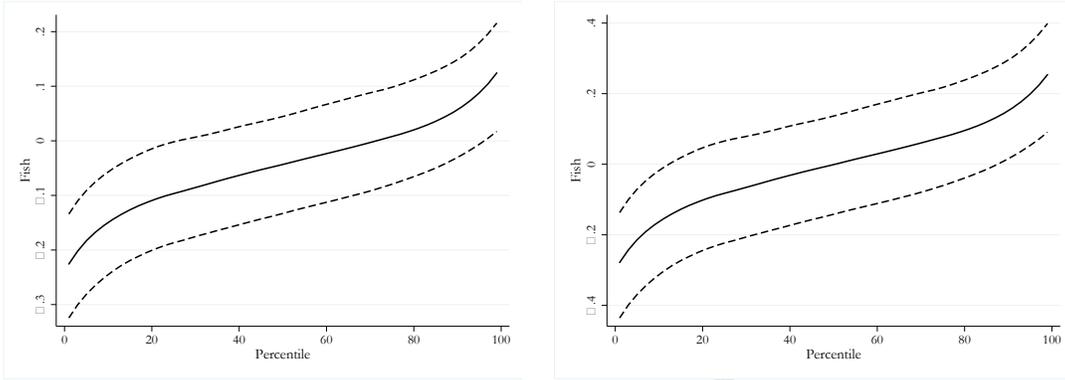
ii. Socio-emotional Endowment
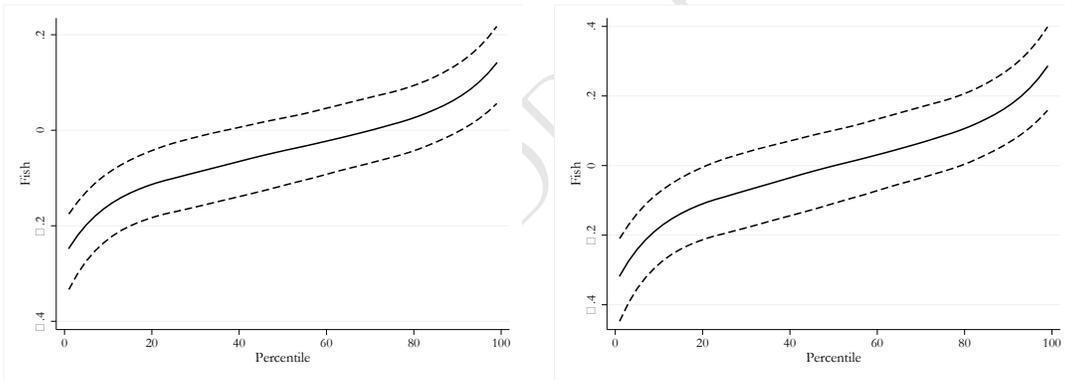
iii. Health Endowment
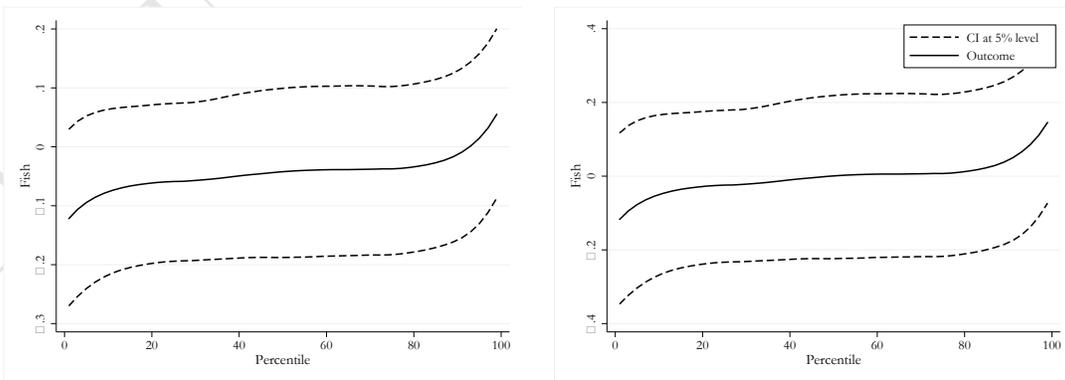
Figure 8: Effects of Endowments: Cannabis Use

(A) Compulsory Education (B) Post-Compulsory Education

i. Cognitive Ability
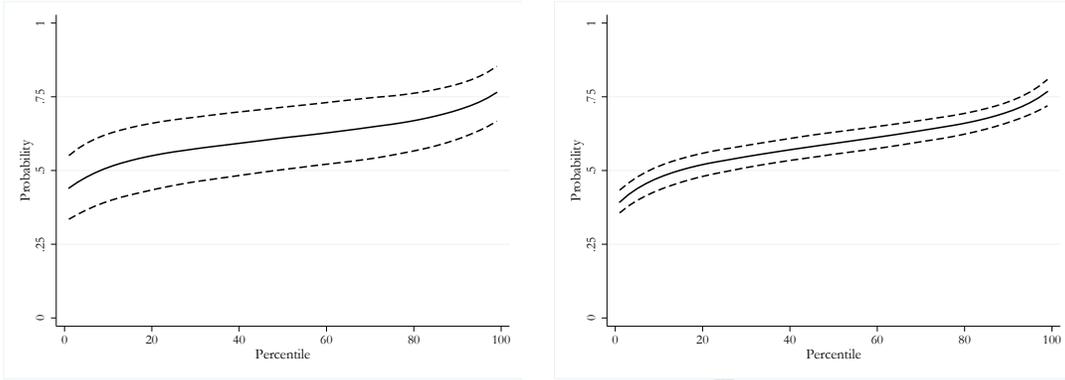
ii. Socio-emotional Endowment
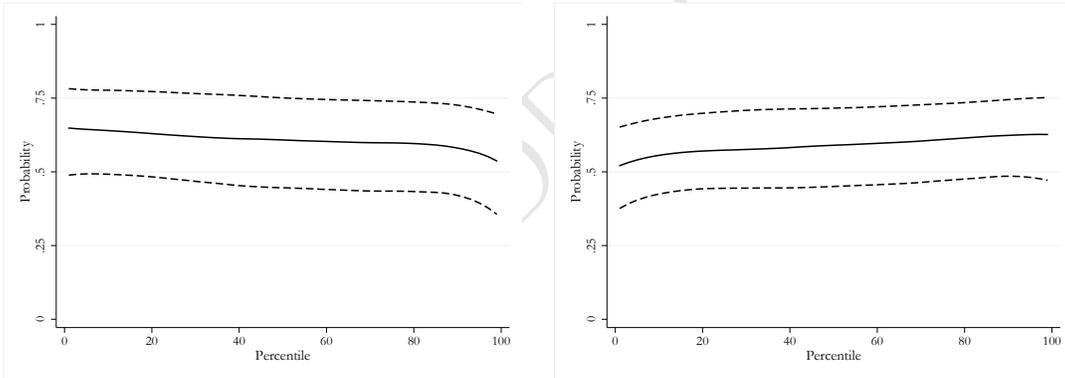
iii. Health Endowment
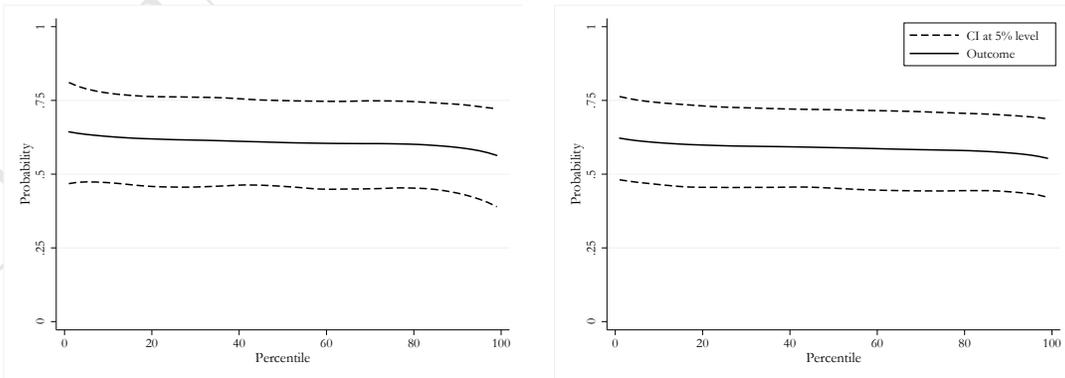
Figure 9: Effects of Endowments: Smoking

(A) Compulsory Education   (B) Post-Compulsory Education

i. Cognitive Ability

ii. Socio-emotional Endowment

iii. Health Endowment

41
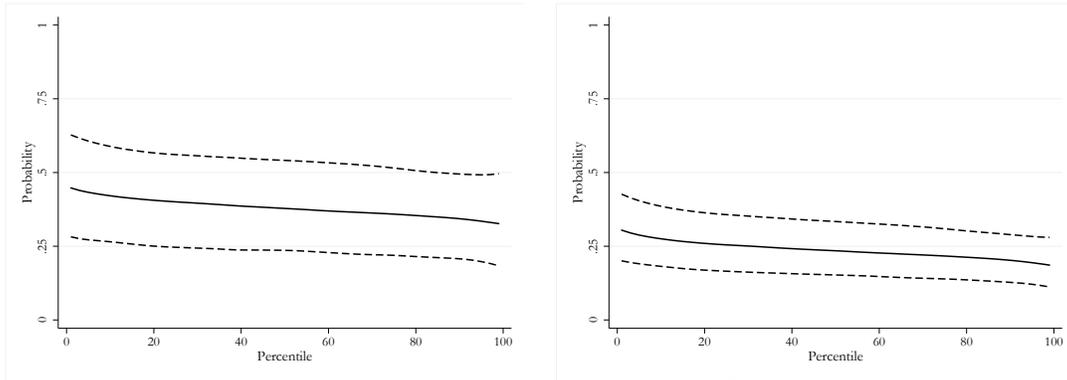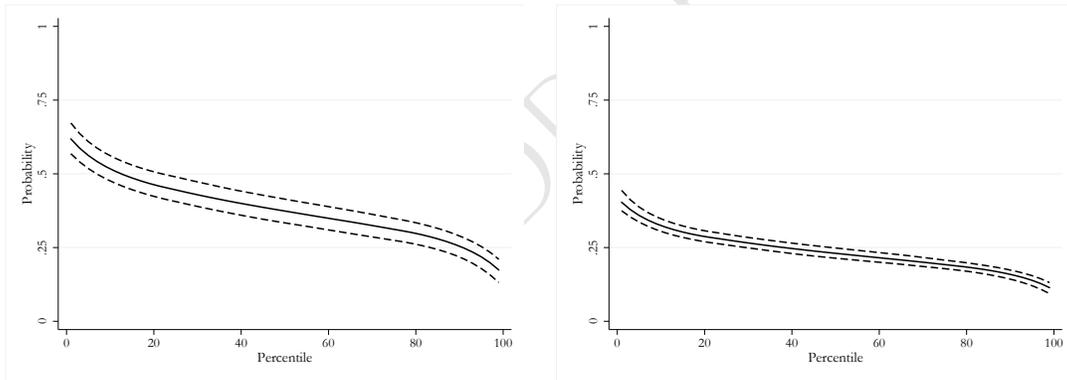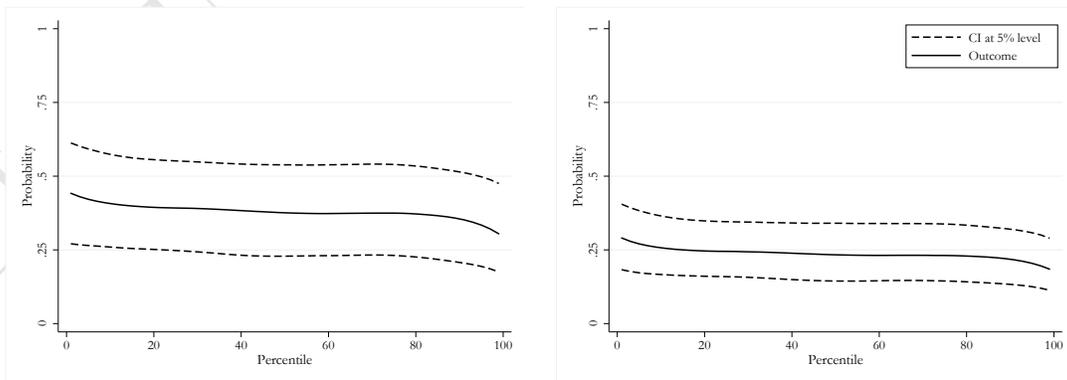
Figure 10: Effects of Endowments: Exercise

(A) Compulsory Education (B) Post-Compulsory Education

i. Cognitive Ability
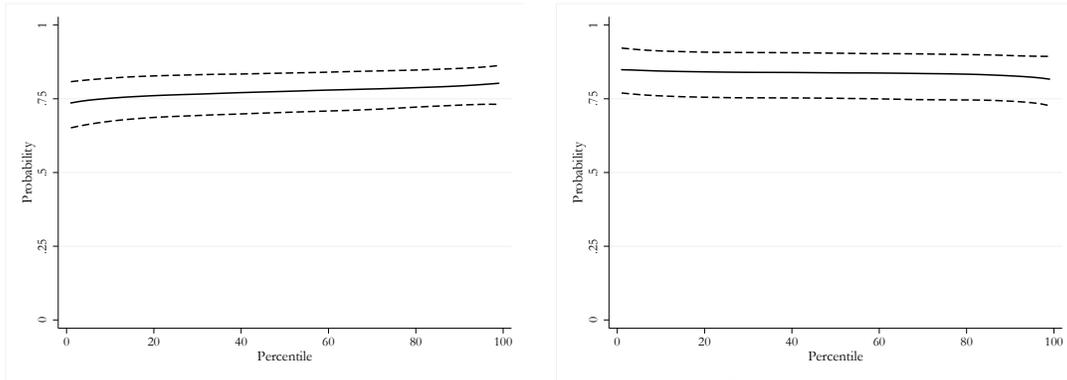
ii. Socio-emotional Endowment
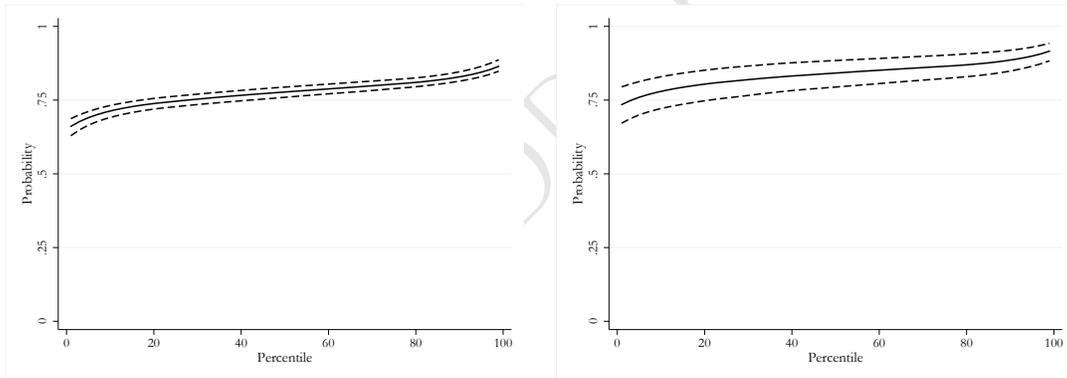
iii. Health Endowment
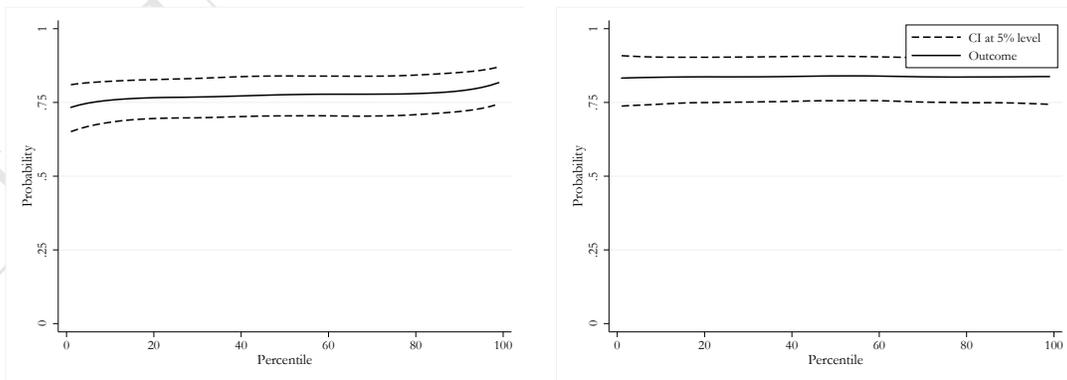
Figure 11: Effects of Endowments: Poor Health

(A) Compulsory Education   (B) Post-Compulsory Education

i. Cognitive Ability
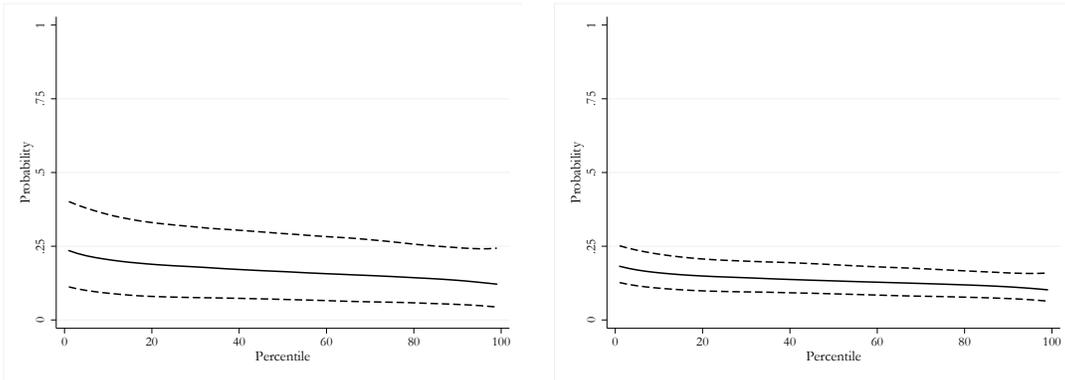
ii. Socio-emotional Endowment
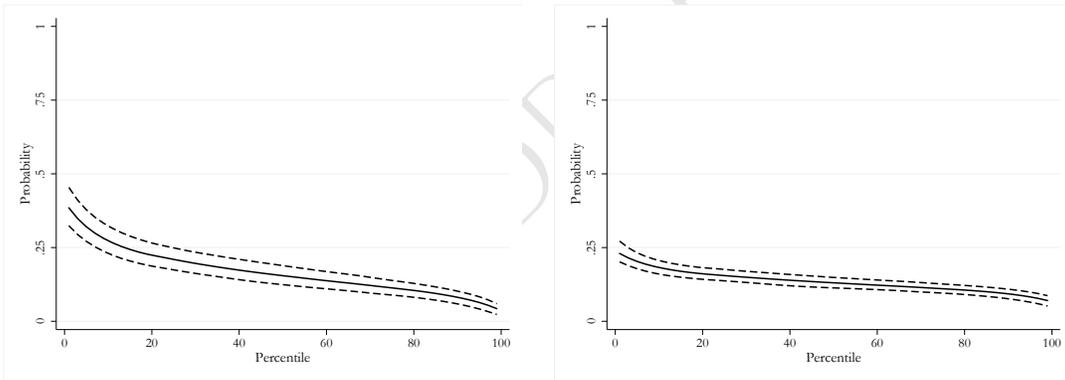
iii. Health Endowment
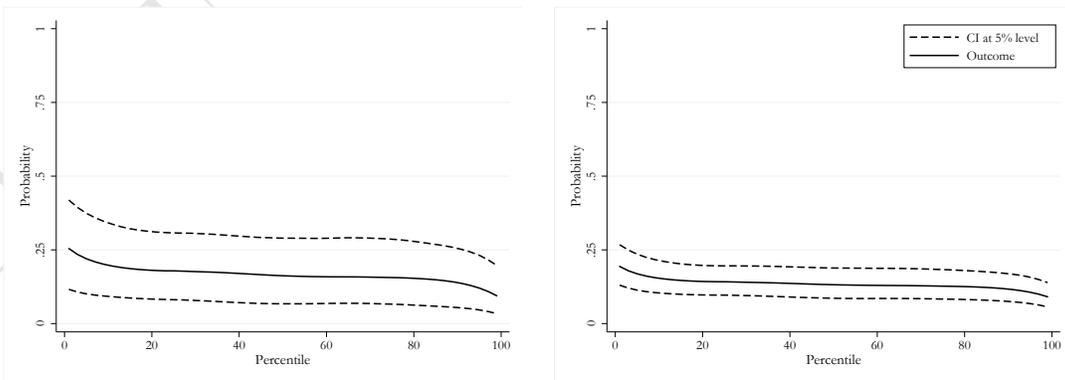
Figure 12: Effects of Endowments: High Blood Pressure

(A) Compulsory Education　　　　(B) Post-Compulsory Education

i. Cognitive Ability
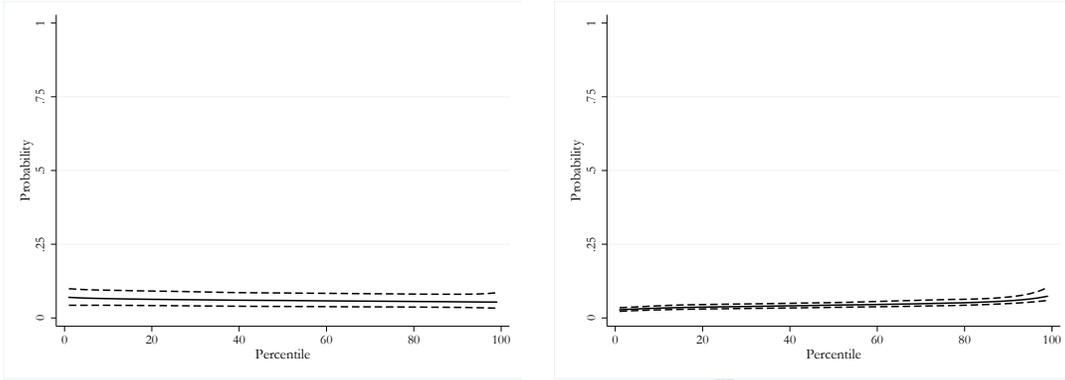
ii. Socio-emotional Endowment
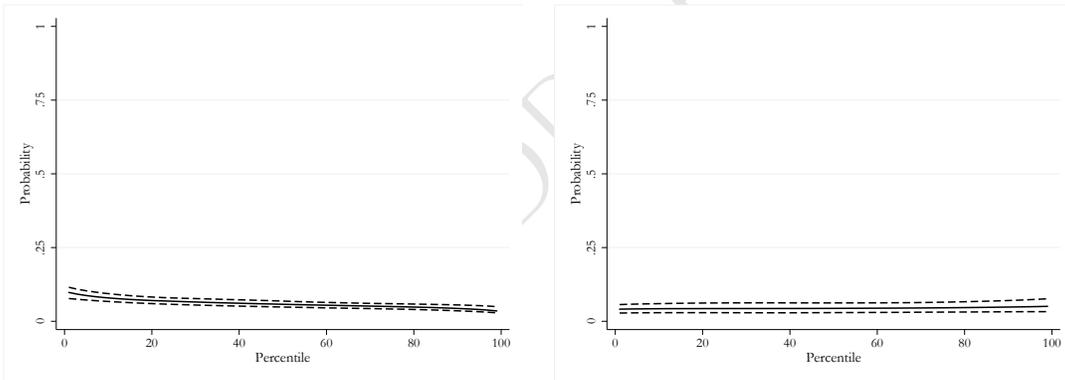
iii. Health Endowment
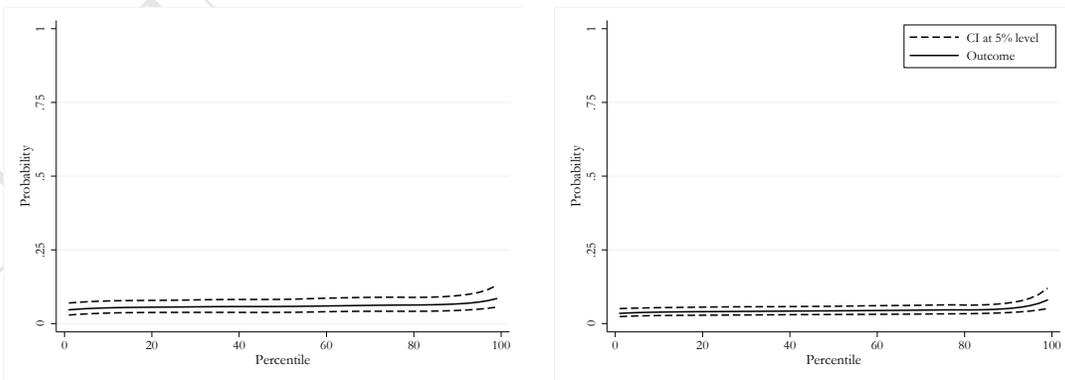
Figure 13: Treatment Effect Heterogeneity - Labor Market Outcomes

(A) Wage
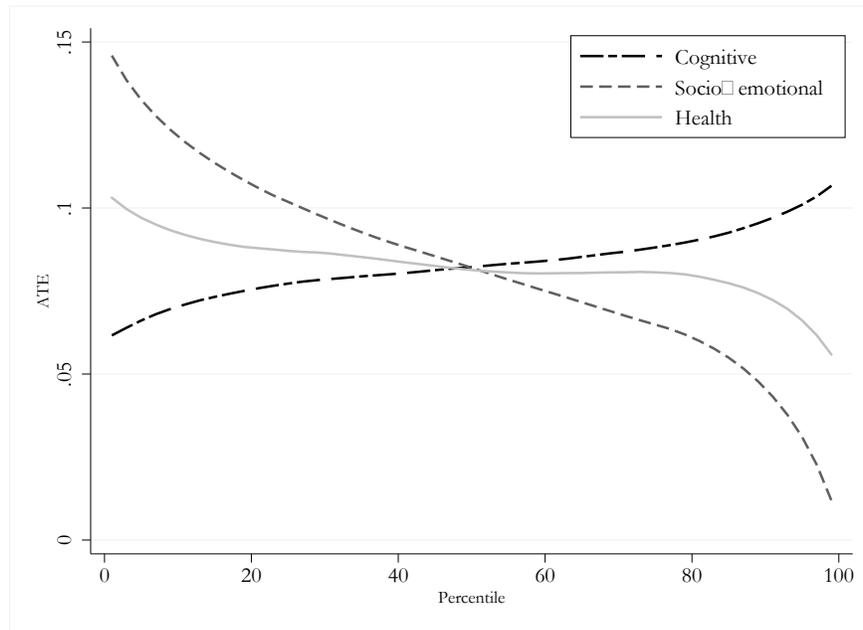
(B) Occupational choice: Manager
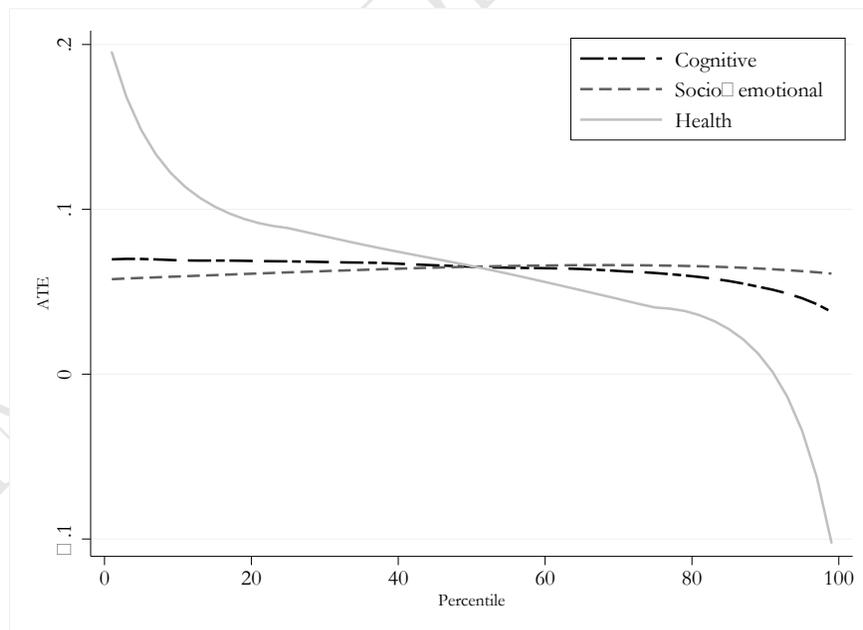
Figure 14: Treatment Effect Heterogeneity - Healthy Behaviors

(A) Fat and Fried Food
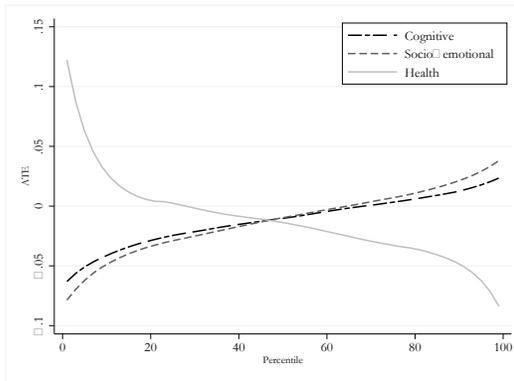
(B) Eating Fish
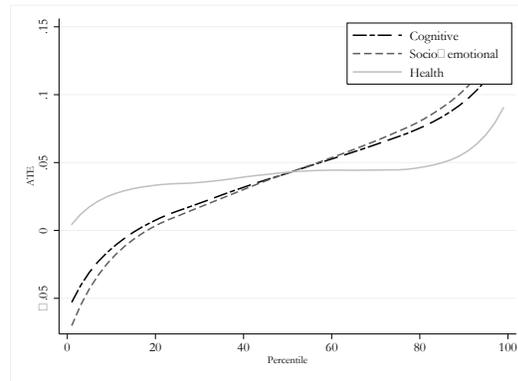
(C) Cannabis Use
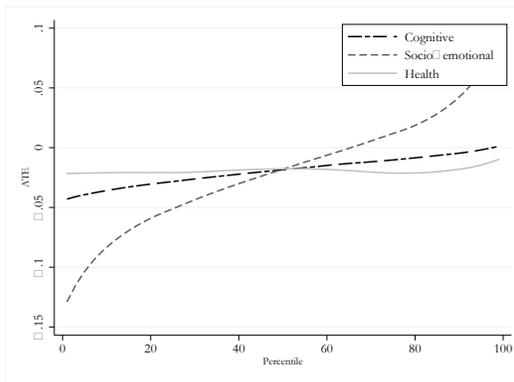
(D) Exercise
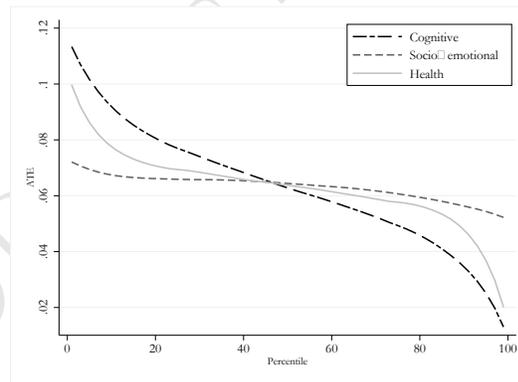
(E) Smoking
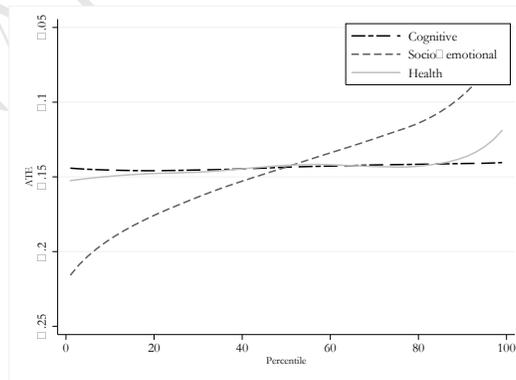
Figure 15: Treatment Effect Heterogeneity - Health and Marriage

(A) BMI

(B) High Blood Pressure

(C) Poor Health

(D) Married
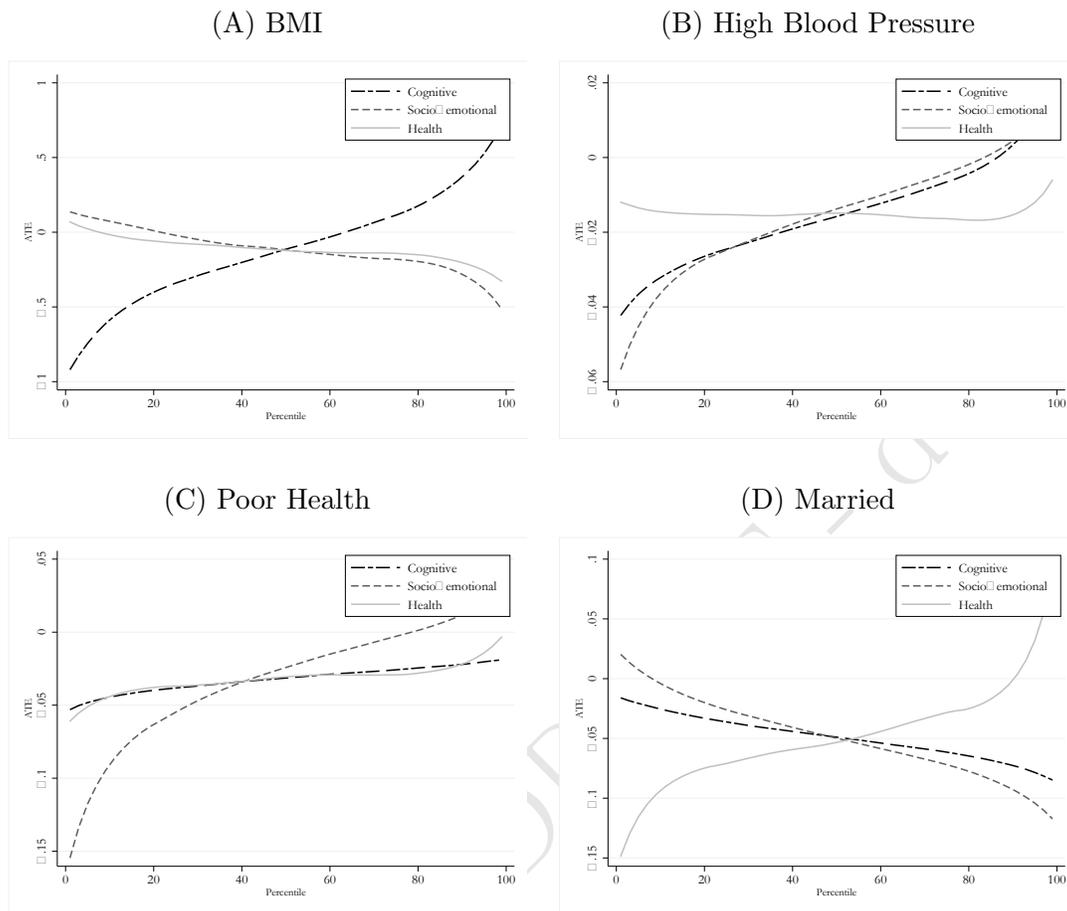
Figure 16: Distribution of Treatment Effects - Labor Market Outcomes

(A) Wage

(B) Occupational choice: Manager
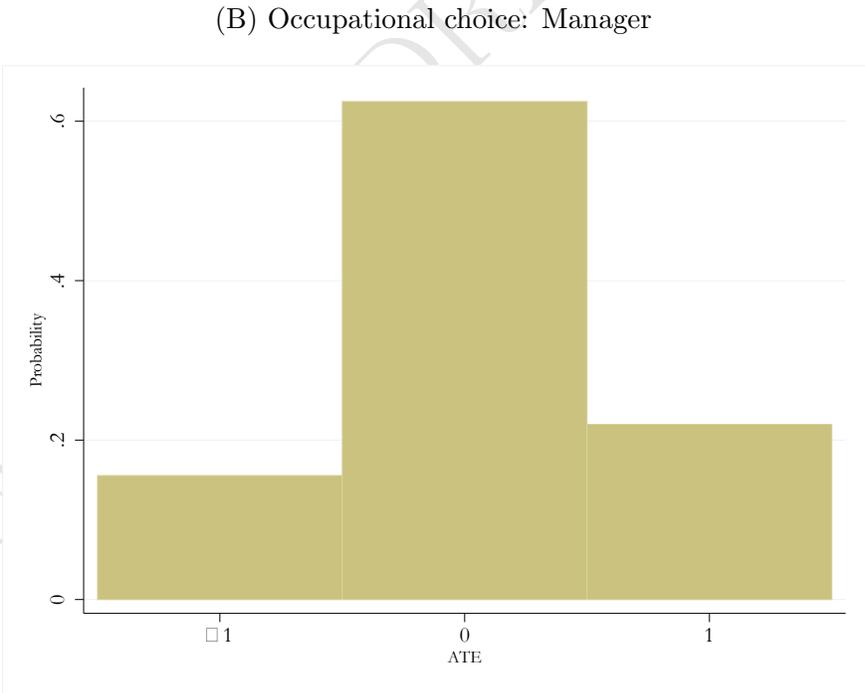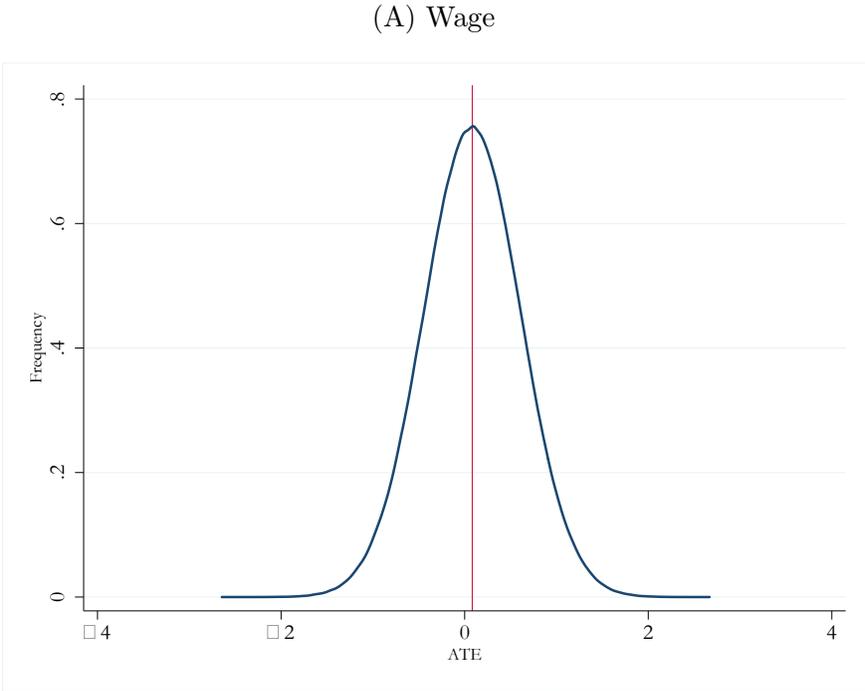
Figure 17: Distribution of Treatment Effects - Healthy Behaviors

(A) Fat and Fried Food
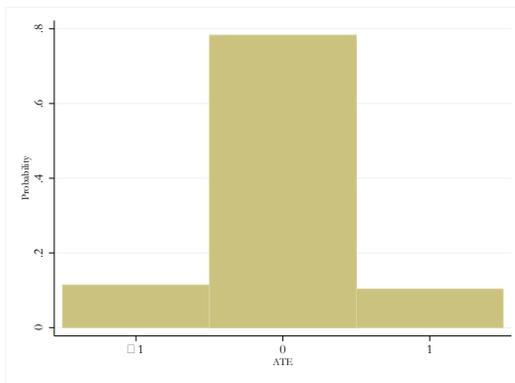
(B) Eating Fish
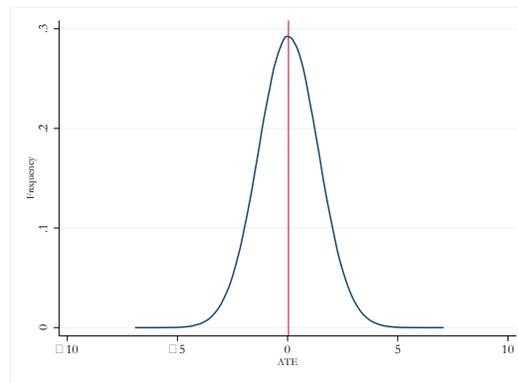
(C) Cannabis Use
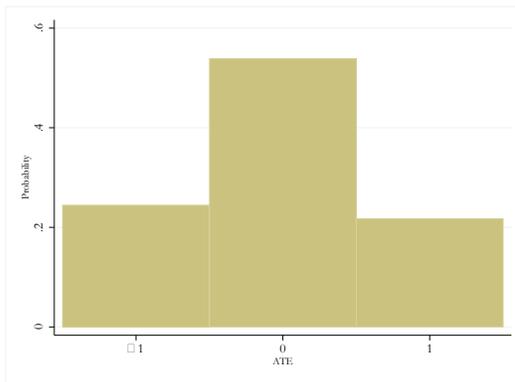
(D) Exercise
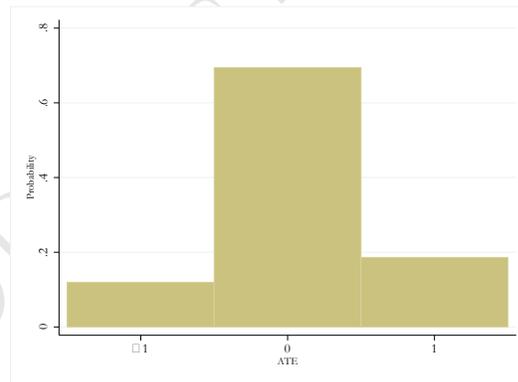
(E) Smoking
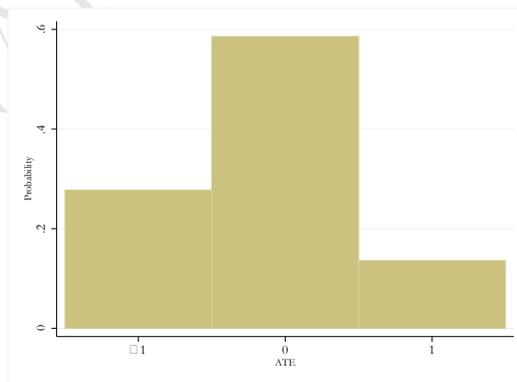
Figure 18: Distribution of Treatment Effects - Health and Marriage

(A) BMI

(B) High Blood Pressure

(D) Poor Health

(E) Married
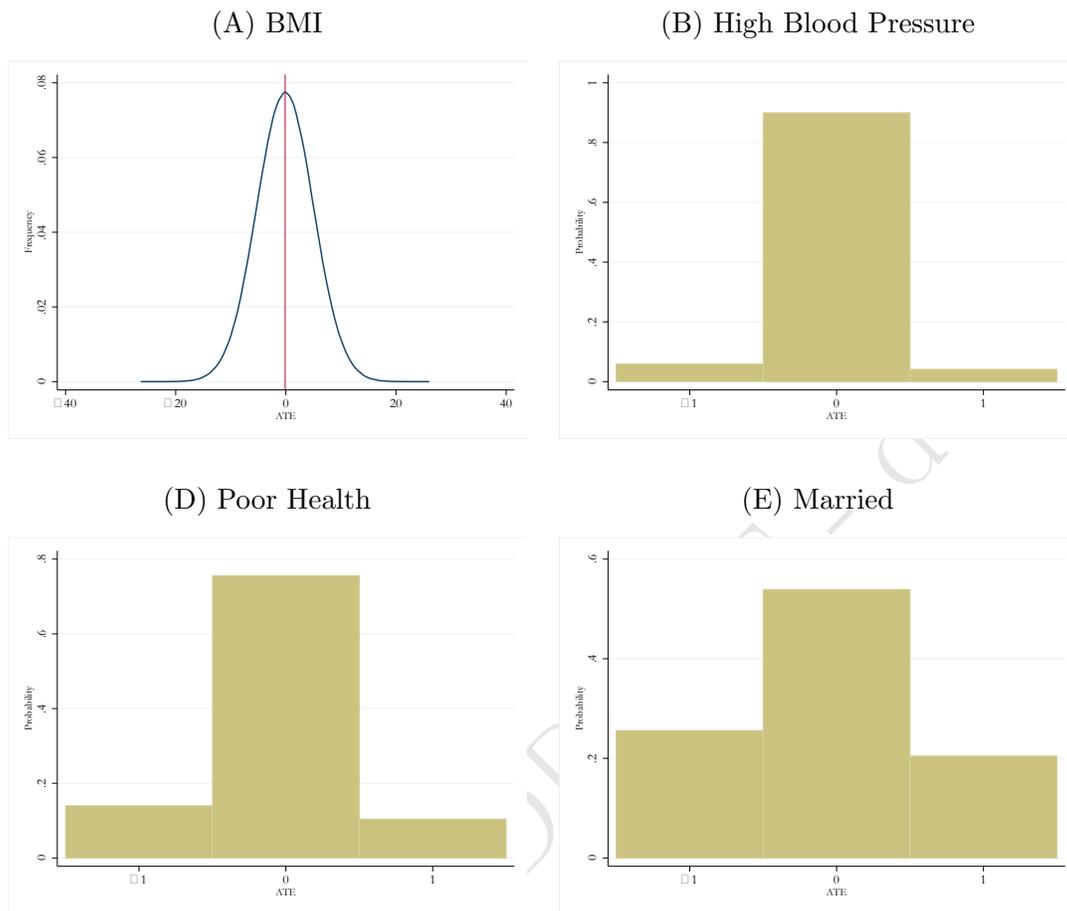
Figure 19: Decomposing the Observed Differences in Outcomes as a Function of ATE, Sorting on gains, and Selection Bias
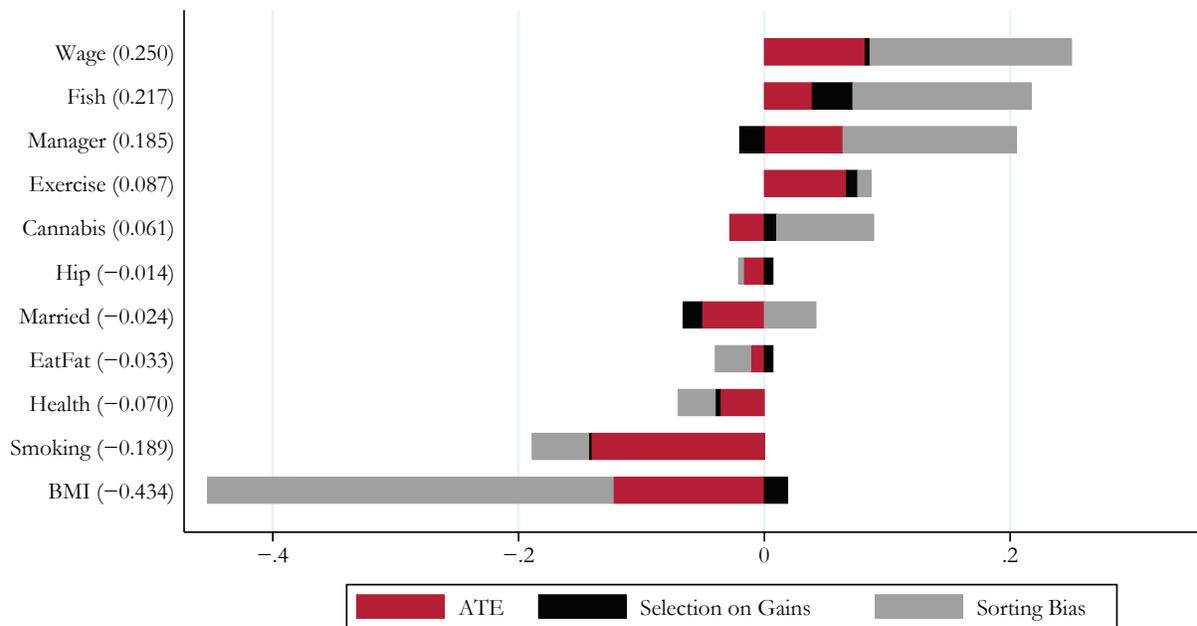
Note: We decompose the observed mean differences in outcomes across schooling levels, i.e., $E(Y_1|D = 1) - E(Y_0|D = 0)$ into $ATE = E(Y_1 - Y_0)$, sorting on gains $TT - ATE$ (where $TT = E(Y_1 - Y_0|D = 1)$), and sorting on levels (or selection bias) $E(Y_0|D = 1) - E(Y_0|D = 0)$.

Table 1: Summary Statistics for the Outcomes

| | Compulsory Schooling | | | | | Post-Compulsory Schooling | | | | |
| | Males | | | | | | | | | |
	Mean	S.D.	Min.	Max.	n	Mean	S.D.	Min.	Max.	n
	Labor Market Outcomes									
Log of hourly wage	1.86	0.38	0.83	3.28	1,357	2.10	0.40	0.84	3.22	988
Manager	0.19	0.39	0	1	1,418	0.37	0.48	0	1	1,048
	Healthy Behaviors									
Eating food fried in fat	0.14	0.34	0	1	1,821	0.11	0.31	0	1	1,213
Eating fish	-0.09	0.99	-1.76	1.66	1,823	0.12	0.94	-1.76	1.66	1,213
Cannabis use	0.58	0.49	0	1	1,803	0.64	0.48	0	1	1,207
Smoking	0.39	0.49	0	1	1,823	0.20	0.40	0	1	1,212
Exercise	0.77	0.42	0	1	1,823	0.86	0.35	0	1	1,213
	Health and Marriage									
Poor Health	0.17	0.38	0	1	1,825	0.10	0.31	0	1	1,214
High Blood Pressure	0.06	0.24	0	1	1,824	0.05	0.21	0	1	1,214
BMI	25.65	3.91	16.14	49.70	1,788	25.24	3.32	16.13	41.00	1,183
Married	0.38	0.49	0	1	1,825	0.35	0.48	0	1	1,214

Table 2: Summary Statistics for the Measurements

| | **Males** | | | | |
	Mean	S.D.	Min.	Max.	n
Health (Birth)					
Birth weight	0.11	1.01	-3.39	3.18	4,383
Gestational age	-0.02	0.97	-5.17	3.48	3,582
Jaundiced	0.23	0.42	0	1	4,328
Other	0.08	0.27	0	1	4,104
Cognitive Scores (Age 10)					
Picture Language Comprehension Test	0.10	0.99	-5.52	4.05	4,375
Friendly Math Test	0.07	1.03	-3.55	2.30	4,375
Shortened Edinburgh Reading Test	-0.05	1.02	-2.91	1.84	4,377
BAS Matrices	-0.03	1.00	-2.87	2.34	4,383
BAS Recall Digits	-0.03	1.02	-5.02	2.71	4,368
BAS Similarities	0.08	1.02	-4.76	3.08	4,351
BAS Word Definition	0.12	1.03	-2.05	3.96	4,365
Noncognitive Scales (Age 10)					
Lawrence self-esteem	0.12	0.95	-3.43	1.75	4,084
Locus of control	0.04	1.01	-3.35	2.57	4,056
Perseverance	-0.12	1.02	-2.40	1.64	4,383
Popularity	-0.01	1.01	-2.93	1.73	4,359
Boldness	0.10	1.00	-2.68	2.29	4,347
Cooperativeness	-0.04	1.01	-3.39	1.70	4,352

BAS = British Ability Scales.

Other: cyanotic, breathing difficulties, cerebral signs, other illness

Table 3: Summary Statistics for the Covariates

	Males				
	Mean	S.D.	Min.	Max.	n
Covariates in the Outcome Equations					
Mother's age at birth	25.83	5.35	14	52	4,383
Mother's education at birth > the msla	0.32	0.47	0	1	4,383
Parity	1.13	1.27	0	13	4,383
High SC at birth	0.28	0.45	0	1	4,383
Total gross family income at age 10[1]	4	1.24	1	7	4,383
Additional Covariates in the Measurement System					
Mother resident in Scotland at birth	0.11	0.32	0	1	4,383
Mother resident in Wales at birth	0.06	0.24	0	1	4,383
Contraception used since Jan. 1968	0.48	0.50	0	1	4,383
Middle SC at birth	0.46	0.50	0	1	4,383
Additional Covariate in the Schooling Choice Equation					
Claimant count[2]	12.89	3.41	7.8	19.8	4,383

Note: SC = Social Class. High Social Class comprises SCI, SCII and SCIIINM (Non-Manual).

Middle SC comprises SCIIIM (Manual).

[1] 1=under £35 pw; 2=£35-49 pw; 3=£50-99 pw; 4=£100-149 pw; 5=£150-199 pw; 6=£200-249 pw;

7=£250 or more pw.

[2] Seasonally-adjusted rate of claimant counts in January 1986, by region. Source: ONS.

Table 4: Goodness of Fit

Outcome	Simulated		Actual	
	Males			
	$E(Y_0\|D=0)$	$E(Y_1\|D=1)$	$E(Y_0\|D=0)$	$E(Y_1\|D=1)$
	Labor Market Outcomes			
Log of hourly wage ft	1.84	2.09	1.82	2.08
Manager	0.18	0.37	0.18	0.36
	Healthy Behaviors			
Food fried in fat	0.14	0.11	0.14	0.11
Eating fish	-0.09	0.13	-0.10	0.17
Cannabis use	0.58	0.64	0.59	0.67
Smoking	0.39	0.20	0.41	0.21
Exercise	0.77	0.86	0.77	0.86
	Health and Marriage			
Poor Health	0.17	0.10	0.19	0.09
High Blood Pressure	0.06	0.05	0.06	0.04
BMI	25.61	25.28	25.79	25.38
Married	0.38	0.37	0.38	0.35
Education	0.39		0.40	

Table 5: Marginal Effects of Endowments on Outcomes, by Educational Level

Outcome	Compulsory Education			Post-Compulsory Education		
	Cognitive	Socio-emotional	Health	Cognitive	Socio-emotional	Health
	Labor Market Outcomes					
Log of hourly wage ft	0.109	0.110	0.051	0.159	0.040	0.035
	(4.950)	(4.400)	(1.889)	(5.300)	(1.290)	(0.946)
Manager	0.102	0.024	0.046	0.090	0.040	-0.049
	(4.044)	(0.871)	(1.497)	(3.065)	(1.243)	(-1.261)
	Healthy Behaviors					
Food fried in fat	-0.048	-0.021	0.005	-0.029	0.017	-0.068
	(-2.890)	(-1.195)	(0.222)	(-1.482)	(1.201)	(-2.692)
Eating fish	0.074	0.099	0.035	0.110	0.158	0.040
	(1.480)	(1.800)	(0.583)	(1.774)	(2.431)	(0.488)
Cannabis use	0.178	-0.122	-0.034	0.151	-0.029	-0.038
	(6.640)	(-4.075)	(-0.994)	(4.674)	(-0.837)	(-0.887)
Smoking	0.046	-0.181	-0.032	0.007	-0.104	-0.021
	(1.587)	(-5.962)	(-0.923)	(0.233)	(-3.379)	(-0.523)
Exercise	-0.013	0.075	0.020	-0.054	0.089	-0.004
	(-0.608)	(3.185)	(0.795)	(-2.328)	(3.405)	(-0.113)
	Health and Marriage					
Poor Health	0.022	-0.127	-0.040	-0.004	-0.050	-0.023
	(1.124)	(-5.719)	(-1.739)	(-0.191)	(-2.018)	(-0.739)
High Blood Pressure	0.005	-0.026	0.017	0.016	-0.005	0.009
	(0.436)	-(1.970)	(1.068)	(1.272)	(-0.398)	(0.447)
BMI	-0.480	0.175	0.078	0.408	-0.467	-0.110
	(2.341)	(0.795)	(0.305)	(1.872)	(1.996)	(0.387)
Married	-0.006	0.117	0.019	-0.010	0.057	0.109
	(-0.223)	(4.384)	(0.614)	(-0.321)	(1.738)	(2.498)

Numbers in parentheses below the estimated marginal effects are *t*-statistics.

Table 6: Treatment Effect of Education on Outcomes

Outcome	Observed	ATE	TT	TUT	AMTE	Baseline
Labor Market Outcomes						
Log of hourly wage	0.250***	0.082***	0.086***	0.080***	0.091***	1.909***
Manager	0.185***	0.064***	0.044***	0.077***	0.056***	0.241***
Healthy Behaviors						
Food fried in fat	-0.033***	-0.011***	-0.004***	-0.016***	-0.003	0.130***
Eating Fish	0.217***	0.039***	0.072***	0.019***	0.023	-0.037***
Cannabis use	0.061***	-0.028***	-0.018***	-0.034***	-0.046**	0.607***
Smoking	-0.189***	-0.141***	-0.143***	-0.140***	-0.171***	0.372***
Exercise	0.087***	0.067***	0.076***	0.061***	0.089***	0.777***
Health and Marriage						
BMI	-0.434***	-0.123***	-0.104***	-0.135***	0.050	25.525***
Poor Health	-0.070***	-0.036***	-0.040***	-0.033***	-0.047***	0.164***
High Blood Pressure	-0.014***	-0.017***	-0.010***	-0.021***	-0.005	0.061***
Married	-0.024***	-0.051***	-0.066***	-0.041***	-0.068***	0.398***

Note: The parameters in this table are obtained using simulated data from our structural model. "Observed" represents the observed difference in outcomes across schooling groups. TT represents the treatment effect on those treated, i.e., $E(Y_1 - Y_0|D = 1)$. ATE represents the average effect of the treatment in the population, i.e., $E(Y_1 - Y_0)$. TUT represents the treatment effect on those untreated, i.e., $E(Y_1 - Y_0|D = 0)$. Finally, $AMTE$ represents the average marginal treatment effect. This parameter is obtained as $E(Y_1 - Y_0|m(Z, V) < \epsilon)$ where $m(Z, V) = |Z\gamma + V|$ (see Carneiro, Heckman, and Vytlacil, 2008). $AMTE$ is computed for $\epsilon = 0.0005 \times StDev(Z\gamma + V)$. *** denotes statistical significance at 1%, ** denotes statistical significance at 5%, * denotes statistical significance at 10%.

Table 7: Treatment Effect of Education on Outcomes - Individuals with Endowment Levels Below the Median

Outcome	Observed	ATE	TT	TUT	AMTE	Baseline
Labor Market Outcomes						
Log of hourly wage	0.241***	0.098***	0.102***	0.096***	0.110***	1.771***
Manager	0.181***	0.092***	0.076***	0.098***	0.040	0.159***
Healthy Behaviors						
Food fried in fat	-0.033***	-0.016***	-0.012***	-0.018***	-0.046*	0.167 ***
Eating Fish	0.146***	-0.017***	0.017***	-0.029***	0.058	-0.142***
Cannabis use	0.015***	-0.062***	-0.051***	-0.065***	0.000	0.588***
Smoking	-0.206***	-0.164***	-0.171***	-0.162***	-0.238***	0.457***
Exercise	0.118***	0.093***	0.115***	0.084***	0.066*	0.736***
Health and Marriage						
BMI	-0.545***	-0.201***	-0.260***	-0.180***	-0.73**	25.668***
Poor Health	-0.105***	-0.070***	-0.080***	-0.066***	-0.073**	0.238***
High Blood Pressure	-0.028***	-0.029***	-0.026***	-0.031***	-0.046**	0.065***
Married	-0.031***	-0.048***	-0.061***	-0.043***	-0.079*	0.331***

Note: The parameters in this table are obtained using simulated data from our structural model. "Observed" represents the observed difference in outcomes across schooling groups. Let $\theta = (\theta^C, \theta^N, \theta^H)$. The numbers in this table are obtained for those individuals $\theta \in \Omega_{<M}$ where $\Omega_{<M}$ is the region in which endowments are below the median. TT represents the treatment effect on those treated, i.e., $E(Y_1 - Y_0 | D = 1, \theta \in \Omega_{<M})$. ATE represents the average effect of the treatment in the population, i.e., $E(Y_1 - Y_0 | \theta \in \Omega_{<M})$. TUT represents the treatment effect on those untreated, i.e., $E(Y_1 - Y_0 | D = 0, \theta \in \Omega_{<M})$. Finally, $AMTE$ represents the average marginal treatment effect. This parameter is obtained as $E(Y_1 - Y_0 | m(Z, V) < \epsilon, \theta \in \Omega_{<M})$ where $m(Z, V) = |Z\gamma + V|$ (see Carneiro, Heckman, and Vytlacil, 2008). $AMTE$ is computed for $\epsilon = 0.0005 \times StDev(Z\gamma + V)$. *** denotes statistical significance at 1%, ** denotes statistical significance at 5%, * denotes statistical significance at 10%.

Table 8: Treatment Effect of Education on Outcomes - Individuals with Endowment Levels Above the Median

Outcome	Observed	ATE	TT	TUT	AMTE	Baseline
Labor Market Outcomes						
Log of hourly wage	0.195***	0.067***	0.070***	0.065***	0.050*	2.048***
Manager	0.136***	0.024***	0.005***	0.046***	0.047	0.334***
Healthy Behaviors						
Food fried in fat	-0.019***	-0.007***	-0.005***	-0.008***	-0.023	0.096***
Eating Fish	0.227***	0.093***	0.109***	0.075***	0.095	0.070***
Cannabis use	0.067***	0.006***	0.009***	0.003**	-0.039	0.629***
Smoking	-0.147***	-0.116***	-0.117***	-0.114***	-0.155***	0.287***
Exercise	0.058***	0.045***	0.054***	0.035***	0.132**	0.816***
Health and Marriage						
BMI	-0.339***	-0.048***	-0.084***	-0.009	0.252	25.388***
Poor Health	-0.022***	-0.005***	-0.011***	0.001	0.023	0.098***
High Blood Pressure	0.000	-0.002***	0.002***	-0.006***	0.016	0.056***
Married	-0.034***	-0.053***	-0.062***	-0.043***	-0.023	0.467***

Note: The parameters in this table are obtained using simulated data from our structural model. "Observed" represents the observed difference in outcomes across schooling groups. Let $\theta = (\theta^C, \theta^N, \theta^H)$. The numbers in this table are obtained for those individuals $\theta \in \Omega_{<M}$ where $\Omega_{>M}$ is the region in which endowments are below the median. TT represents the treatment effect on those treated, i.e., $E(Y_1 - Y_0|D = 1, \theta \in \Omega_{>M})$. ATE represents the average effect of the treatment in the population, i.e., $E(Y_1 - Y_0|\theta \in \Omega_{>M})$. TUT represents the treatment effect on those untreated, i.e., $E(Y_1 - Y_0|D = 0, \theta \in \Omega_{>M})$. Finally, $AMTE$ represents the average marginal treatment effect. This parameter is obtained as $E(Y_1 - Y_0|m(Z,V) < \epsilon, \theta \in \Omega_{>M})$ where $m(Z,V) = |Z\gamma + V|$ (see Carneiro, Heckman, and Vytlacil, 2008). $AMTE$ is computed for $\epsilon = 0.0005 \times StDev(Z\gamma + V)$. *** denotes statistical significance at 1%, ** denotes statistical significance at 5%, * denotes statistical significance at 10%.

Table 9: Test of Equality of Parameters Across Schooling Levels

Outcome	Test	Intercept	High SC at Birth	Total Gross at 10	Mother's Age at Birth	Mother's Education at Birth	Parity	Cognitive Endowment	Socio-Emotional Endowment	Health Endowment
					Labor Market Outcomes					
Log of hourly	Distribution (KS)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
wage	Mean	0.36	0.42	0.26	0.21	0.45	0.47	0.09	0.04	0.36
Manager	Distribution (KS)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Mean	0.01	0.10	0.13	0.16	0.35	0.30	0.24	0.40	0.03
					Healthy Behaviors					
Food fried	Distribution (KS)	0.00	0.00	0.00	0.00	0.14	0.00	0.00	0.00	0.00
in fat	Mean	0.10	0.33	0.39	0.03	0.49	0.36	0.27	0.09	0.01
Eating Fish	Distribution (KS)	0.00	0.00	0.00	0.00	0.05	0.00	0.00	0.00	0.00
	Mean	0.49	0.14	0.42	0.43	0.47	0.26	0.32	0.24	0.48
Cannabis use	Distribution (KS)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Mean	0.08	0.27	0.01	0.29	0.32	0.04	0.26	0.02	0.47
Smoking	Distribution (KS)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Mean	0.04	0.12	0.06	0.33	0.45	0.11	0.20	0.14	0.47
Exercise	Distribution (KS)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Mean	0.15	0.02	0.06	0.49	0.01	0.37	0.07	0.17	0.31
					Health and Marriage					
Poor Health	Distribution (KS)	0.00	0.12	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Mean	0.40	0.50	0.08	0.05	0.17	0.04	0.20	0.02	0.38
High Blood	Distribution (KS)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Pressure	Mean	0.44	0.16	0.41	0.32	0.05	0.44	0.21	0.19	0.41
BMI	Distribution (KS)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Mean	0.11	0.31	0.28	0.14	0.04	0.40	0.00	0.02	0.31
Married	Distribution (KS)	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00
	Mean	0.21	0.46	0.35	0.35	0.47	0.11	0.46	0.09	0.04

Note: The numbers represent the *p-values* from two different tests. Specifically, for each control variable and outcome, we test whether the distributions of the parameters are statistically the same across schooling levels (Distribution), and whether the means of the parameters' distributions are the same across schooling levels. For the first test we utilize Kolgomorov-Smirnov. We report the *p-value*. For the second test, we construct conventional t-test of equal means. We also report the *p-values* in this case.

PRELIMINARY DRAFT: do not cite

Table 10: Decomposition of "Selection Bias"

Outcome	Observed	TT	Selection Bias	Contribution of Unobservables (Selection Bias$_\theta$)	Contribution of Observables (Selection Bias$_X$)
Labor Market Outcomes					
Manager	0.185	0.044	0.141	0.089	0.089
Healthy Behaviors					
Food fried in fat	-0.033	-0.004	-0.029	-0.031	-0.008
Cannabis use	0.061	-0.018	0.079	0.060	0.048
Smoking	-0.189	-0.143	-0.046	-0.042	-0.018
Exercise	0.087	0.076	0.011	0.013	0.0005
Health and Marriage					
Poor Health	-0.07	-0.04	-0.03	-0.030	-0.009
High Blood Pressure	-0.014	-0.01	-0.004	-0.008	0.002
Married	-0.024	-0.066	0.042	0.038	0.016

Note: Observed=$E[Y_1|D=1] - E[Y_0|D=0]$, TT=$E[Y_1|D=1] - E[Y_0|D=1]$, Selection Bias=$E[Y_0|D=1] - E[Y_0|D=1]$. Formally, Observed=$OLS = \Pr(Y_1 = 1|D=1) - \Pr(D_0 = 1|D=0)$ and we decompose OLS as $OLS = TT +$ Selection Bias where

$$
\begin{aligned}
TT &= \Pr(Y_1 = 1|D=1) - \Pr(Y_0 = 1|D=1) \\
\text{Selection Bias} &= \Pr(Y_0 = 1|D=1) - \Pr(Y_0 = 1|D=0).
\end{aligned}
$$

In this context, we investigate the role of observable characteristics by computing:

$$
\text{Selection Bias}_X = \widetilde{\Pr}(Y_0 = 1|D=1) - \Pr(Y_0 = 1|D=0)
$$

where

$$
\widetilde{\Pr}(Y_0 = 1|D=1) = \int \int_{(X,\theta)\in\Omega_1} \Pr(Y_0 = 1|D=1, x, t) \frac{\Pr(D=1|X=x, \theta=t) f_X(x) f_\theta(t)}{\Pr(D=1) f_{\theta|D=1}(t)} f_{\theta|D=0}(t) \, dxdt.
$$

so the conditional distribution of unobserved factors in schooling level 0 is utilized when integrating out the unobserved components. The formula analyzing the effect of the unobserved characteristics is analogous to this last expression (Selection Bias$_\theta$).