Early Endowments, Education, and Health

Gabriella Conti
James J. Heckman
Sergio Urzua

October, 2011

EARLY ENDOWMENTS, EDUCATION AND HEALTH [*]

Gabriella Conti	James J. Heckman	Sergio Urzúa
University of Chicago	University of Chicago	Northwestern University
	and University College Dublin	and IZA

January 24, 2011

Abstract

This paper examines the early origins of observed health disparities by education. We determine the role played by cognitive, noncognitive and early health endowments, and we identify the causal effect of education on health and health-related behaviors. We show that family background characteristics, cognitive, noncognitive and health endowments developed as early as age 10 are important determinants of health disparities at age 30. We also show that not properly accounting for personality traits overestimates the importance of cognitive ability in determining later health. We show that selection explains more than half of the observed difference in poor health, depression and obesity, while education has an important causal effect in explaining differences in smoking rates. We also uncover significant gender differences. We then go beyond the current literature which usually estimates mean effects to compute distributions of treatment effects. We show how the health returns to education can vary also among individuals who are similar in their observed characteristics, and how a mean effect can hide gains and losses for different individuals. This analysis highlights the crucial role played by the early years in promoting health and the importance of prevention in the reduction of health disparities, and refocuses the role of education policy as health policy.

Keywords: health, education, cognitive ability, personality traits, health endowments, factor models, treatment effects.
JEL Codes: I12, I21, C31.

1 Background

Much of the policy debate on reducing health disparities has focused in the past on improving health insurance coverage and access to health care. However, in the recent years increasing attention is being paid to the social determinants of health (Commission on Social Determinants of Health, 2008; Marmot, 2010), with great emphasis placed on early childhood interventions (Currie, 2009b).

A growing literature establishes strong relationships between early childhood conditions and adult outcomes (Almond and Currie, 2010). Gaps in both cognitive and noncognitive abilities across different families emerge at an early age (Cunha et al., 2006). So do gaps in health status (Case et al., 2002). Various studies suggest it appears possible to partially compensate children damaged by adverse environments (Heckman et al., 2009). Still, very little of this research has focused on the role of these early factors on later health, and there is still much to know. Our research aims to fill this gap. The concept of developmental health, comprising physical, cognitive and psychosocial dimensions of child development, has been influential in life course epidemiology (Kuh and Ben-Shlomo, 1997), but has not yet been fully accepted into the mainstream economic or medical literature (McCormick, 2008).

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 (Grossman, 1972, 2008), that it runs from health to schooling (Currie, 2009a), and that both are determined by a third factor, such as time or risk preferences (Fuchs, 1982). Understanding the relative importance of each of these mechanisms in generating observed differences in health by education is relevant to designing policy to promote health.

Health gaps between education groups are rising (Meara et al., 2008). Many authors have noted that better health early in life is associated with higher educational attainment (Grossman, 1975; Perri, 1984; Wolfe, 1985; Currie, 2009a), 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). However, the exact mechanisms that produce this relationship remain to be identified. Education may merely proxy capabilities developed in the early years. Much of the

---

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

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 factors that are also determinants of health behaviors. The literature in economics addresses this problem largely using instrumental variables (see, e.g., Currie and Moretti, 2003; Lleras-Muney, 2005).

As Dow et al. (2010) notice, "relatively few out of the thousands of SES health gradient studies are even able to convincingly tease out what portion of that observed relationship reflects causal pathways from SES to health as opposed to the adverse effects of ill health, or third variable explanations." We address this concern in our research. This paper examines the origins of health disparities by education in the context of a general framework to analyze the effect of interventions and to disentangle causality from selection effects. The paper is organized as follows. We provide a brief overview of the relevant literature in the next section 2. In section 3 we outline our econometric model. Data and empirical implementation are discussed in section 4, while in section 5 we show how to estimate the causal effects of education. Estimates are reported in section 6. Section 7 develops a simple decomposition to disentangle the role of observables and unobservables in explaining selection bias. Section 8 compares our results with conventional propensity score matching. Section 9 concludes.

2   Literature Review

This paper joins together different strands of the literature in economics, epidemiology and psychology. The first strand refers to the relationship between health and cognitive ability. While the importance of the 'ability bias' has long been recognized in labor economics (see, e.g., Griliches, 1977), the effect of cognitive ability on health has received relatively less attention.[2] However, this topic has recently received considerable attention in the field of cognitive epidemiology: large epidemiological studies have found that intelligence in childhood predicts substantial differences in adult morbidity and mortality (Whalley and Deary, 2001; Gottfredson and Deary, 2004; Batty et al., 2007).

The second strand refers to the relationship between personality traits and health. While there is already an established tradition in psychology on their importance (see, for example, Roberts

---

[2]Grossman (1975); Shakotko et al. (1982); Hartog and Oosterbeek (1998); Elias (2005); Auld and Sidhu (2005); Cutler and Lleras-Muney (2007); Kaestner (2009) are the only exceptions.

et al. (2006, 2007); Hampson and Friedman (2008)), economists have just started to explore the effects of personality traits on health (Kaestner, 2009) and health-related behaviors (Heckman et al., 2006; Cutler and Lleras-Muney, 2007).

Our work also relates to the literature on biological programming (Gluckman and Hanson, 2006) and on the role of early-life conditions on adult outcomes (Kaestner, 2009; Case et al., 2005), and to life-course epidemiology (Kuh and Ben-Shlomo, 1997). We go beyond the current literature which looks at the effect of a single health indicator (e.g. height in adolescence) on later outcomes. We model health as a latent factor to fully capture its multiple indicators of and the possibility that each is measured with error (for a recent example of this approach, see also Dahly et al. (2008)).

The final strand of literature we refer to is that on the non-market returns to education. The positive correlation between education and health has long been recognized in the economic, epidemiologic and medical literature, and several attempts at disentangling correlation from causality have been made.[3] Our methodology allows us to disentangle the fraction of the health gap by education that can be explained by these early factors from that what can be attributed to the causal effect of education.

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

This paper develops a semiparametric structural model of schooling choice in which individuals sort across schooling levels on the basis of their gains in terms of health and labor market outcomes. Specifically, in our model, different schooling levels have associated different health and labor market outcomes. These differences arise not only because of the effects of observed variables on labor market productivity and health behaviors, but also because of unobserved factors, that we model and interpret as cognitive ability, personality traits, and health status.

---

[3]In an extensive review of the literature, Grossman (2006) concludes that there seems to be evidence of a causal effect of education on health.

3.1 Schooling Choice Model

This paper studies the schooling decision of whether or not to stay-on in schooling beyond the minimum compulsory school-leaving age and its causal effects on health and labor market outcomes.[4] We model the schooling decision using a binary decision model with a latent index structure. Let D_i^* denote the net utility of an individual from staying-on, and D_i a binary variable indicating individual's decision ($D_i = 1$ if the individual stays on, and $D_i = 0$ otherwise). Thus, we assume:

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

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

$$D_i^* = \mu_D(Z_i) + U_{Di}$$

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

Once the individual has decided his schooling level, all future outcomes (labor market outcomes, health status and healthy behaviors) are potentially causally related to this decision. Importantly, as described in detail below, this model allows individuals to select their schooling level taking into account the potential health and labor market outcomes in the two possible educational states. This feature of our model is extremely important. To the extent that individuals make their schooling choices anticipating 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 observed and unobserved variables (unobserved from the point of view of the researcher but known to the agent) to be correlated across schooling levels and outcomes. Finally, we link the unobserved variables in our schooling and outcome models to individual's early cognitive, noncognitive, and health endowments.

---

[4]This decision is particularly important in the United Kingdom (the country we study), where the dropout rate is particularly high.

This last feature of our approach represents an important contribution because not only does it allow for a simultaneous role for cognitive, noncognitive and health endowments as determinants of schooling choices and outcomes, but also recognizes that some of these endowments are unobserved by the researchers but are known to the agents.[5] Our model includes both continuous and discrete outcomes. We now turn to the discussion of how we model each of them in turn.

3.2 Continuous Outcomes

Let (Y_{i0}, Y_{i1}) 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 an individual's observed and unobserved characteristics. Specifically, we write the potential outcome associated with post-compulsory education as:

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

and the potential outcome obtained if a person stops at compulsory education as:

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

where X_i is a vector of observed characteristics and (U_{i1}, U_{i0}) denote the unobserved components. It is not strictly required that X_i is statistically independent of U_{i1}, U_{i0}, and U_{Di}. We condition on X throughout.[6] An additively 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 $\mu_1 (X_i, U_{i1}) = \beta_1 X_i + U_{i1}$. We do not impose any assumptions on the correlations among U_{i1}, U_{i0}, and U_{Di}. We allow the unobserved components from outcomes and schooling choices to be correlated, and as previously explained, any comparison of outcomes across schooling groups should take into account the potential selection problem. Notice that in

---

[5]For example, we allow for the possibility that individuals with better noncognitive skills (e.g. more willpower) are more successful at school and also less likely to engage in unhealthy behaviors.

[6]For purposes of estimation, it is convenient to assure that X_i is independent of U_{i1}, U_{i0}, and U_{D_i}, but this is not strictly required.

this setup, the observed outcome Y_i is produced by potential outcomes and schooling decisions:[7]

$$Y_i = D_i Y_{1i} + (1 - D_i) Y_{0i}. \tag{4}$$

3.3 Discrete Outcomes

Our general approach allows for the presence of dichotomous outcomes. 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:

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

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

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

As in the case of continuous outcomes, in our empirical implementation of the model we assume linear-in-parameters and additive specifications for the functions $\kappa_0 (Q_i, \epsilon_{i0})$ and $\kappa_1 (Q_i, \epsilon_{i1})$, i.e., $\kappa_0 (Q_i, \epsilon_{i0}) = \lambda_0 Q_i + \epsilon_{i0}$ and $\kappa_1 (Q_i, \epsilon_{i1}) = \lambda_1 Q_i + \epsilon_{i1}$. We also allow for correlations among $\epsilon_{i1}, \epsilon_{i0}, U_{i1}, U_{i0}$, and V_i. In this context, the observed outcome B_i can be written as:

$$B_i = B_{i1} D_i + B_{i0} (1 - D_i). \tag{5}$$

---

[7]Equations 2 - 3 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).

3.4 Unobserved Endowments

Our model allows for general correlations among the unobserved components, namely U_{Di}, U_{i1}, U_{i0}, ϵ_{i0}, ϵ_{i1}. Formally, we allow:

$$U_{Di} \not\perp U_{i1} \not\perp U_{i0} \not\perp \epsilon_{i0} \not\perp \epsilon_{i1} | (X_i, Z_i, Q_i)$$

where $A \not\perp B|C$ denotes "A and B are not statistically independent conditional on C." We model these correlations by assuming that the error terms are governed by a factor structure which we interpret as cognitive, noncognitive and health endowments. Specifically, and suppressing the sub-index i to simplify the exposition, if we let θ denote a vector of unobserved factors, with $\theta = (\theta_C, \theta_N, \theta_H)$, where θ_C, θ_N and θ_H represent the cognitive, noncognitive and health endowments, respectively, we assume:[8]

$$
\begin{aligned}
U_D &= \alpha_{U_D}\theta + \upsilon_{U_D} \\
U_1 &= \alpha_{U_1}\theta + \upsilon_{U_1} \\
U_0 &= \alpha_{U_0}\theta + \upsilon_{U_0} \\
\epsilon_0 &= \alpha_{\epsilon_0}\theta + \upsilon_{\epsilon_0} \\
\epsilon_1 &= \alpha_{\epsilon_1}\theta + \upsilon_{\epsilon_1}
\end{aligned}
$$

where $(\upsilon_{U_D}, \upsilon_{U_1}, \upsilon_{U_0}, \upsilon_{\epsilon_0}, \upsilon_{\epsilon_1} \perp\!\!\!\perp \theta)$ and $(\upsilon_{U_D}, \upsilon_{U_1}, \upsilon_{U_0}, \upsilon_{\epsilon_0}, \upsilon_{\epsilon_1})$ are mutually independent. Using this structure, we can analyze the effect of each of the components of θ (cognitive, noncognitive and health factors) on each of the outcomes controlling for the endogeneity of the schooling choice.[9]

However, without further structure the model is not identified. Up to this point, there is nothing in our model that allows us to identify the levels (and distributions) of the components of θ. Schooling decisions are endogenous, and the outcomes are conditional on observed schooling. We must then supplement our model 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]Here we posit the existence of three endowments. In ongoing work Conti et al. (2011), we relax this assumption about the dimensionality of the factor structure and we estimate the number of factors simultaneously with the educational choice and the outcomes.

[9]We can relax the additive separability and most of the independence assumptions using the nonparametric factor identification analysis of Cunha et al. (2010).

3.5 The Measurement System

Following Carneiro et al. (2003), we posit a linear measurement system to identify the joint distribution of the unobserved endowments θ. Specifically, we supplement the model introduced above with a set of equations linking early cognitive (M_C), noncognitive (M_{N_C}) and health measures (M_H) with the unobserved cognitive (θ_C), noncognitive (θ_N) and health (θ_H) factors, so that we can give them a meaningful interpretation. Specifically, denoting by $\{M_{Cl}\}_{l=1}^{N_C}$, $\{M_{Nj}\}_{j=1}^{N_N}$, $\{M_{Hk}\}_{k=1}^{N_H}$ the set of early cognitive, noncognitive and health variables, with N_C, N_N and N_H denoting the number of cognitive, noncognitive and health measurements available, respectively, and assuming they are "dedicated", we have:

$$
\begin{aligned}
M_{C1} &= \delta_{C1}X + \alpha_{C1}\theta_C + \upsilon_{C1} \\
&\vdots \\
M_{CN_C} &= \delta_{CN_C}X + \alpha_{CN_C}\theta_C + \upsilon_{CN_C} \\
M_{N1} &= \delta_{N1}X + \alpha_{N1}\theta_N + \upsilon_{N1} \\
&\vdots \\
M_{NN_C} &= \delta_{NN_N}X + \alpha_{NN_N}\theta_N + \upsilon_{NN_N} \\
M_{H1} &= \delta_{H1}X + \alpha_{H1}\theta_H + \upsilon_{H1} \\
&\vdots \\
M_{HN_H} &= \delta_{HN_H}X + \alpha_{HN_H}\theta_H + \upsilon_{HN_H}
\end{aligned}
$$

where X denotes the set of observed variables determining the measures, and we assume that $\upsilon_{C1} \perp\!\!\!\perp \ldots \perp\!\!\!\perp \upsilon_{CN_C} \perp\!\!\!\perp \upsilon_{N1} \perp\!\!\!\perp \ldots \perp\!\!\!\perp \upsilon_{NN_N} \perp\!\!\!\perp \upsilon_{H1} \perp\!\!\!\perp \ldots \perp\!\!\!\perp \upsilon_{HN_H}$. Our assumption of dedicated measurements implies, for example, that intelligence tests are solely a measure of cognitive ability. Note this is not saying that there is no interaction between abilities, as we allow the factors to be correlated, as we explain in section 4.4. below. A sketch of identification for the correlated factors case is provided in Appendix A.[10]

3.6 Identification Strategy

Our identification strategy is based on the following conditional independence assumption:

$$(Y_0, Y_1) \perp\!\!\!\perp D | X, Z, \theta$$

---

[10] We make usual required normalizations to set the scale of the factors, as detailed in section 4.2.

We notice that, if we did observe θ, we could do matching; since we observe θ only imperfectly, we account for imperfect measurement in our estimation (see Heckman et al. (2010) for a formal justification): hence, our method can be interpreted as a form of matching on imperfectly measured observables. The support of the estimated probability of schooling is essentially the full unit interval (see Figure 1) so that identification of the model is over the full support of the unobservables in the choice equation. We implement the conditional independence assumption in two ways: one way uses a "quasi-structural" factor model (as detailed in section 4.4), another way does direct matching on the factor scores (see section 8). We find that the results from both approaches are very similar.

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.[11] There have been seven follow-ups so far to trace all members of the birth cohort: in 1975, 1980, 1986, 1996, 2000, 2004, and 2008. We draw information from the birth survey, the second sweep (age 10) and the fifth sweep (age 30).[12]

After removing children born with congenital abnormalities and non-whites (or those with missing information on ethnicity), and deleting responses with missing information on the covariates, we are left with a sample of 3,777 men and 3,620 women.

4.1 Schooling and Post-Schooling Outcomes

The outcomes considered in our model are:

- Schooling. Our schooling measure is a dummy variable indicating 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.[13]

---

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

[12]We select the fifth sweep in order to secure the comparability of our results to those in the literature (Heckman et al., 2006).

[13]The decision to stay on in school after age 16 is crucial in the British educational system, given the big proportion

- Labor Market Outcomes. We analyze two labor market outcomes: (log) hourly wages and full-time employment status. Both are measured at age 30.

- Healthy Behaviors. We analyze three healthy behaviors, all measured at age 30: ever used cannabis, daily smoking and regular exercise.[14]

- Health. We include three variables characterizing individual's health status by age 30. These are: self-reported poor health, obesity and depression.[15]

Summary statistics for our outcome measures are displayed in Table 1. Figure 2 displays more directly the educational differentials in the outcome measures we consider. It is interesting to notice that the magnitude of the differential varies depending on the outcome, but for many of them a sizeable educational disparity is already present by age 30.

4.2 Measurement System

As indicators of cognitive ability we use the following seven test scores administered to the children at age ten: the Picture Language Comprehension Test,[16] the Friendly Math Test,[17] the Shortened Edinburgh Reading Test,[18] and the four British Ability Scales.[19] We performed a preliminary factor analysis of these measurements. Both Velicer (1976) minimum average partial correlation criterion and Kaiser (1960) eigenvalue rule suggested to retain one component, which we interpret as Spearman (1904)'s 'g'. As measurements of noncognitive ability we use six scales, one administered

---

of pupils who drop out after having reached the minimum school-leaving age (see, for example, Pissarides (1981) and Micklewright (1989)).

[14]The variable "smoking" takes the value 1 if the individual smokes cigarettes every day. The variable "exercise" takes the value 1 if the individual does any regular exercise. The variable "cannabis" takes the value 1 if the individual reports having ever used cannabis by age 30.

[15]The variable "poor health" takes the value 1 if the individual reports his/her health to be generally "fair" or "poor". The variable "obesity" is constructed in the standard way as having a BMI>25 (for females) or a BMI>30 (for males), where the BMI is weight in kilograms divided by height in meters squared. Note we use a different threshold for males and females as the difference between high- and low-educated females is barely statistically significant if we use as threshold BMI>30. The variable "depression" takes the value 1 if the individual is categorized as depressed; it is measured using the Malaise Inventory (Rutter et al., 1970), which includes 24 'yes-no' items which cover emotional disturbances and associated physical symptoms.

[16]This is a new test specifically developed for the BCS70 on the basis of the American Peabody Picture Vocabulary Test and the English Picture Vocabulary Test; it covers vocabulary, sequence and sentence comprehension.

[17]This is a new test specifically designed for the BCS70; it covers arithmetic, fractions, algebra, geometry and statistics.

[18]This is a shortened version of the Edinburgh Reading Test, which is a test of word recognition particularly designed to capture poor readers; it covers vocabulary, syntax, sequencing, comprehension, and retention.

[19]They measure a construct similar to IQ, and include two verbal scales (Word Definition and Word Similarities) and two non-verbal scales (Recall Digits and Matrices).

11

to the child (the locus of control scale),[20] and five to the teacher (perseverance,[21] cooperativeness,[22] completeness,[23] attentiveness[24] and persistence)[25]. We performed a preliminary factor analysis of these measurements. Both Velicer (1976) minimum average partial correlation criterion and Kaiser (1960) eigenvalue rule suggested to retain one component.[26] Following Rothbart (1981) and Rothbart (1989) self-regulative model of temperament, we define this noncognitive trait "self-regulation".[27] As measures of the health endowment we use the height and the head circumference of the child at age 10, and the height of the mother and of the father (also measured when the child was aged 10). We performed a preliminary factor analysis of these measurements. Both Velicer (1976) minimum average partial correlation criterion and Kaiser (1960) eigenvalue rule suggested to retain one component. Summary statistics for the measurements are presented in Table 2.

4.3 Observed Characteristics

We include the following set of covariates in both the measurement system and in the outcome equations: mother's age at birth, mother's education at birth (a dichotomous variable for whether or not the mother continued education beyond the minimum school-leaving age), father's high social class at birth,[28] total gross family income at age 10,[29] whether the child lived with both

---

[20]This is administered to the child and includes sixteen items which measure whether an individual's locus of control is external or internal.

[21]This is administered to the teacher, who answers to the question "How much perseverance does the child show in face of difficult tasks?" on a scale from 1 to 47.

[22]This is administered to the teacher, who makes an estimate of how cooperative is the child with his peers, on a scale from 1 to 47.

[23]This is administered to the teacher, who assesses "The child completes tasks which are started", on a scale from 1 to 47.

[24]This is administered to the teacher, who assesses "Child pays attention to what is being explained in class", on a scale from 1 to 47.

[25]This is administered to the teacher, who assesses "Child shows perseverance, persists with difficult or routine work", on a scale from 1 to 47.

[26]In ongoing work Conti et al. (2011) we fully exploit the richness of the BCS data and we use all available measurements of child behavioral traits without making a priori assumptions on the underlying latent structure.

[27]Rothbart (1989) defines this trait "effortful control", and argues that it is related to the executive system in the frontal lobe structure (which provides a rationale for the existence of the high correlation that we report with cognition), and that it is developmentally related to a major dimension of adult personality, namely conscientiousness (Costa and McCrae, 1988).

[28]A dichotomous variable for father belonging to Social Class I, II or III Non Manual. The BCS70 uses the Registrar General's classification for measuring social class (SC). Social class I includes professionals, such as lawyers, architects and doctors; Social Class II includes intermediate workers, such as shopkeepers, farmers and teachers; Social Class III Non Manual includes skilled non-manual workers, such as shop assistants and clerical workers in offices.

[29]A categorical variable: 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.

parents since birth until age 10, parity, and the number of children in the family at age 10.[30] The

schooling choice model also includes as covariate the gender-specific seasonally-adjusted rate of

unemployment-related benefit claims (the claimant count) as observed in January 1986.

Summary statistics for the covariates included in our model are presented in Table 3.

4.4 Distributional Assumption and Estimation Strategy

To avoid dependence of estimates on distributional assumptions, we use mixtures of multivariate

normals to characterize the distributions of the latent capabilities. Specifically, we assume:

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

where μ_1 and μ_2 are vectors of dimension 3×1, and Σ_1 and Σ_2 are matrices of dimension 3×3.[31] 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 choice models $(\upsilon_V, \upsilon_{\epsilon_0}, \upsilon_{\epsilon_1})$ we assume

independent normal distributions 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.

The density of outcomes given observables is:

$$
f(Y, D, B, M_C, M_N, M_H | X, Z, Q)
$$

where $f(\cdot)$ is the joint density of continuous and discrete outcomes, schooling choices, cognitive

measures, noncognitive scales, and early health variables. Written in terms of unobservables, the

---

[30]We also include child's weight in the measurement equation for child's height and head circumference, and mother(father) weight in the measurement equations for maternal(paternal) height.

[31]We find that a two-point mixture provides the best fit.

density is:

$$\int \int \int_{(\theta_C, \theta_N, \theta_H) \in \Theta} f(Y, D, B, M_C, M_N, M_H | X, Z, Q, t_C, t_N, t_H) dF_\theta(t_C, t_N, t_H)$$

where $F_\theta(\cdot)$ denotes the joint cumulative density associated with unobserved cognitive, noncognitive and health endowments. Notice that conditional on unobserved factors (and observed characteristics) $(D_i, M_C, M_N,$ and $M_H)$ are independent, and the sample likelihood simplifies accordingly.[32] This demonstrates the empirical convenience of using latent factors to account for the correlation across outcomes, schooling decisions, and measurements. We use Bayesian MCMC (Markov Chain Monte Carlo) methods to compute the sample likelihood.

5   Defining the Causal Effects of Education

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. We illustrate how to use our framework to compute treatment parameters in the context of a single outcome. However, our discussion directly extends to the more general case of vectors of continuous and discrete outcomes.

Δ_i involves factual and counterfactual outcomes: 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 average 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:

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

---

[32] Y and B are not independent of D given θ. See equations (4) and (5). However, conditional on θ, any effect of D on Y and B is causal.

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, where $F_{X,\theta}(x, t)$ is the joint distribution of X and θ evaluated at x, t.

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:

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

where $F_{X,\theta|D=1}(x, t)$ is the conditional distribution of X, θ given $D = 1$ evaluated at x, t.

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 et al. (2005) and Carneiro et al. (2003). We now discuss our empirical results.

6   Empirical Results

We first document the importance of modeling early endowments as correlated factors. The estimated correlations between the cognitive and noncognitive endowments is 0.544, between cognitive and health 0.176, and between noncognitive and health is 0.093 for males (see Figure 3). We also find substantial evidence of measurement error, as shown in Figure 4, which presents for each measurement the fraction of its variance explained by the uniqueness (darker region) (vs in the measurement system). We also notice that the estimated model passes tests of goodness of fit (see Table 4 and Figure 5).

6.1   The Role of Early Endowments as Determinants of Adult Outcomes

Figure 6 presents the sorting of individuals across schooling levels in terms of the distributions of cognitive, noncognitive and health endowments. Panels (B1) and (B2) demonstrate the importance of not imposing normal distributions for the unobserved endowments. We observe a clear sorting of high cognitive and noncognitive individuals into post-compulsory level. This patten is observed for both males and females. The sorting on the health endowment is not as strong as the ones observed

in panels (A1)-(A2) (cognitive) and (B1)-(B2) (noncognitive) but it is statistically significant for females. To gain a better understanding of the overall impact of early life factors, including their effect through education, we have computed the predicted unconditional outcome and we have plotted it by percentile of the respective factors (see Figures 7-14). In each case, for a given outcome Y, endowment θ, and percentile P, we have computed $E(Y|\theta_P)$ by integrating out the observable characteristics and fixing the remaining two unobserved endowments at their overall mean, and we have normalized the predicted outcome to zero at the first percentile of the distribution of each factor, so that we can compare the relative magnitude of their effects for both genders. Cognitive ability mostly matters for education and labor market outcomes, and is positively associated with the probability of having ever used cannabis by age 30.[33] Our first striking result points to a much lesser role for cognitive ability than has been emphasized in the cognitive epidemiology literature. The result is especially strong for males: a shift from the bottom to the top of the cognitive ability distribution brings about no significant change in the probability of daily smoking, of having poor health or of being obese at age 30. The picture is only slightly different for females: cognitive ability also plays no role on the probability of being a daily smoker or of being obese, but it is an important determinant of the probability of having poor health. The opposite is true for noncognitive ability, which plays a much lesser role in determining labor market outcomes, but exerts a powerful role in reducing the probability of engaging in unhealthy behaviors such as smoking or using cannabis, and experiencing unhealthy conditions such as poor health or depression. In addition, we notice that cognition matters for health more for females than for males (in the case of depression, exercise, and self-reported health it plays no role for males), a finding that we confirm in our ongoing work Conti et al. (2011). We also uncover an interesting finding on the role played by early health conditions. For males, early health has no significant effect on the probability of staying on beyond the minimum compulsory level of education (but it has a direct effect on many health outcomes at age 30). For females, the effect of health conditions at age 10 seems to work mainly through the educational channel. Notice that, for both males and females, children with a better health endowment at 10 are less likely to be obese by age 30, which is consistent with our modeling of the health factor as a physical health endowment.

---

[33]Conti (2009) first documents the positive association between experimentation with cannabis and cognitive ability, and provides evidence from several data sources to understand the nature of it.

6.2 Education

Figure 15 presents the observed disparities in outcomes due to education. These disparities are decomposed into the causal effect of education (darker region) and the effect of selection. Notice that education has a causal effect on most outcomes for both males and females. We also notice that, while in the data the more educated are more likely to have experimented with cannabis, the causal effect of education is actually negative. This demonstrates the importance of properly accounting for confounding factors and going beyond simple associations in understanding the relationship between education and health. In order to gain a better understanding of the role played by education in reducing health disparities, Figure 16 displays the fraction of the observed differential which can be attributed to education. We see that education plays an important role in explaining differences in smoking behavior, but it accounts for less than half of the observed differential in self-reported health and depression. We also uncover significant gender differences: education plays a much more important role in accounting for the gap in obesity rates, exercise, and employment for males than for females (notice the difference in obesity by education is entirely due to selection for females). This emphasizes the importance of taking the gender dimension into account when studying health disparities.

6.3 Distribution of Treatment Effects

We move beyond the traditional literature which only considers mean effects and estimate distributions of treatment effects (see Figure 17-Figure 19). The knowledge of these distributions is fundamental if we want to uncover what lies behind a "zero" average treatment effect, and what is the proportion of the individuals who actually benefit from the treatment. Consider the case of smoking: the proportion of people who stop smoking is much bigger than the proportion of people who start, so the average treatment effect turns out to be negative (Figure 17, panels C1 and C2).

Compare these to the results for obesity for females (Figure 18, panel C2). Underlying a statistically insignificant average treatment effect of education there are gains and losses which balance each other out: the same proportion of women (almost 20%) lose and gain from the treatment. While usually overlooked in traditional studies on the impact of treatments on outcomes, knowledge of these distributional parameters is fundamental in order to understand if there is effectively a

fraction of individuals who benefit from a particular policy, beyond the average treatment effect.[34]

6.4 Treatment Effect Heterogeneity: the Role of Early Endowments

The average treatment effect of education varies with the level of endowment of cognitive and noncognitive skills, and early health. While there is a significant amount of heterogeneity in the effect of education across outcomes by levels of endowments, some patterns emerge. For males, the beneficial effect of education is much bigger at the bottom of the noncognitive ability distribution, and is greater at the top of the cognitive ability distribution[35] (see Figure 20-Figure 22). This last finding is consistent with the interpretation that the information content on the dangers of smoking provided by post-compulsory education needs to be combined with the capacity to process that information in order for it to be effective.

7   Understanding the Selection Mechanism

In this section we investigate the role of observable and unobservable variables in explaining selection bias. 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:

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

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

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

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

---

[34]See Abbring and Heckman (2007) for a discussion of distributional treatment effects.

[35]The only exception to this pattern occurs for the outcome exercise.

We then write the observed difference in outcomes as:

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

and we decompose it as:

$$\text{Treatment on the treated } (TT) + \text{Selection bias } (SB)$$

where:

$$
\begin{aligned}
TT &= \Pr\left(Y_1 = 1 | D = 1\right) - \Pr\left(Y_0 = 1 | D = 1\right) \\
SB &= \Pr\left(Y_0 = 1 | D = 1\right) - \Pr\left(Y_0 = 1 | D = 0\right).
\end{aligned}
$$

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

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

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

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

so finally:

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

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

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

where:

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

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.[36]

The decomposition results are presented in Figure 23. They clearly show that early endowments account for most of the selection bias, for all the outcomes.

8 Matching

Our "quasi-structural" method can be interpreted as a form of matching on both observables and unobservables, where the unobservables are proxied, and we account for the errors in the proxies in the unobservables. As mentioned in section 3.6, we estimate our model also using propensity score matching, where we match directly on X, Z, and on the estimated factor scores θ. We obtain results which are in agreement for all the parameters identified by both methods. They are presented in Figure 24.

9 Conclusions

This paper examines the early origins of health disparities across education groups. We have determined the role played by cognitive, noncognitive and early health endowments, and we have identified the causal effect of education on health and health-related behaviors. We develop an empirical model of schooling choice and post-schooling outcomes, where both dimensions are influenced by latent factors (cognitive, noncognitive and health). We show that family background characteristics, and cognitive, noncognitive, and health endowments developed as early as age 10, are important determinants of labor market and health disparities at age 30. We show that not properly accounting for personality traits overestimates the importance of cognitive ability in determining later health. We show that selection explains more than half of the observed difference by education in poor health, depression, and obesity. Education has an important causal effect in explaining differences in smoking rates. We also uncover significant gender differences. We go

---

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

beyond the current literature which usually estimates mean effects to compute distributions of treatment effects. We show how the health returns to education can vary also among individuals who are similar under their observed characteristics, and how a mean effect can hide gains and losses for different individuals. We decompose the sources of selection and show that early cognitive, noncognitive and health capabilities explain a significant part of the selection bias. This highlights the crucial role played by the early years in promoting health and the importance of prevention in the reduction of health disparities.

Figure 1: Distribution of $P(Z)$

Males

Females
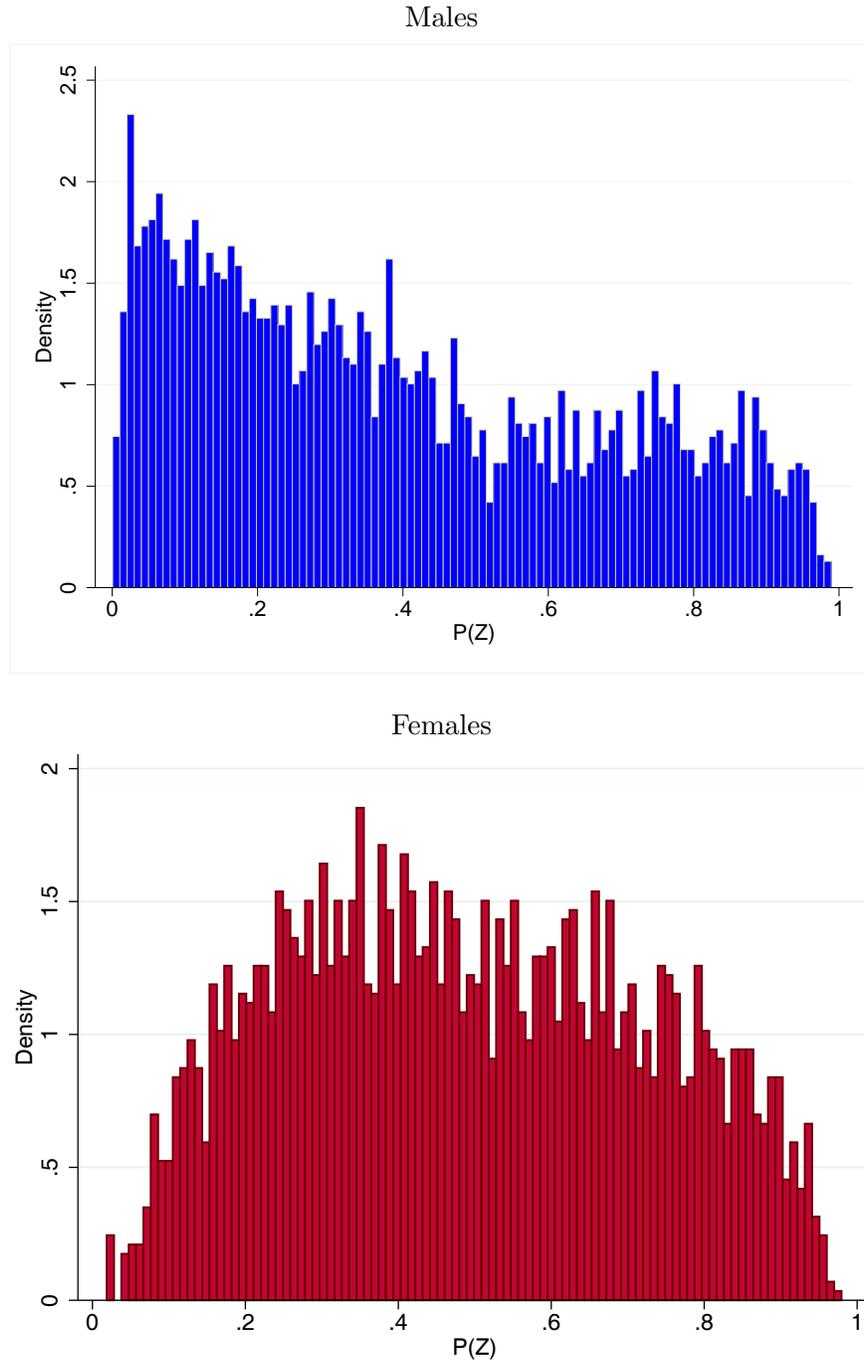
Note: The figure displays the estimated probability of schooling.

Figure 2: Disparities by Education
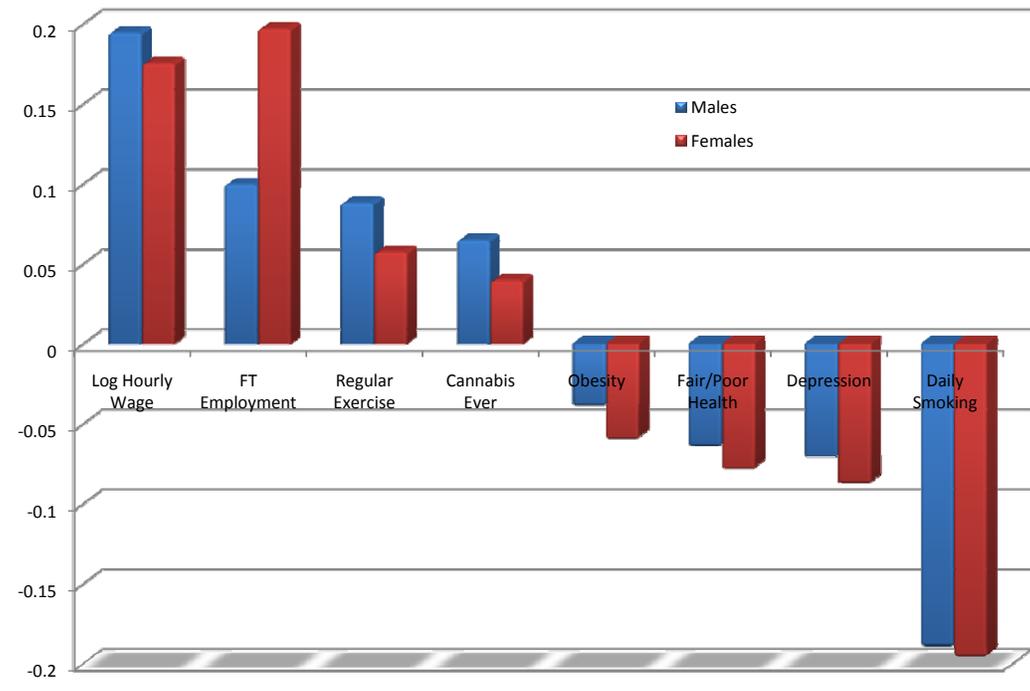
Note: The figure displays the differences in health, health behaviors and labor market outcomes by education, between individuals with educational level equal to compulsory education and individuals with some post-compulsory education. The differences are also presented by gender.

Figure 3: Joint Distributions of Endowments

(a) Males

(b) Females
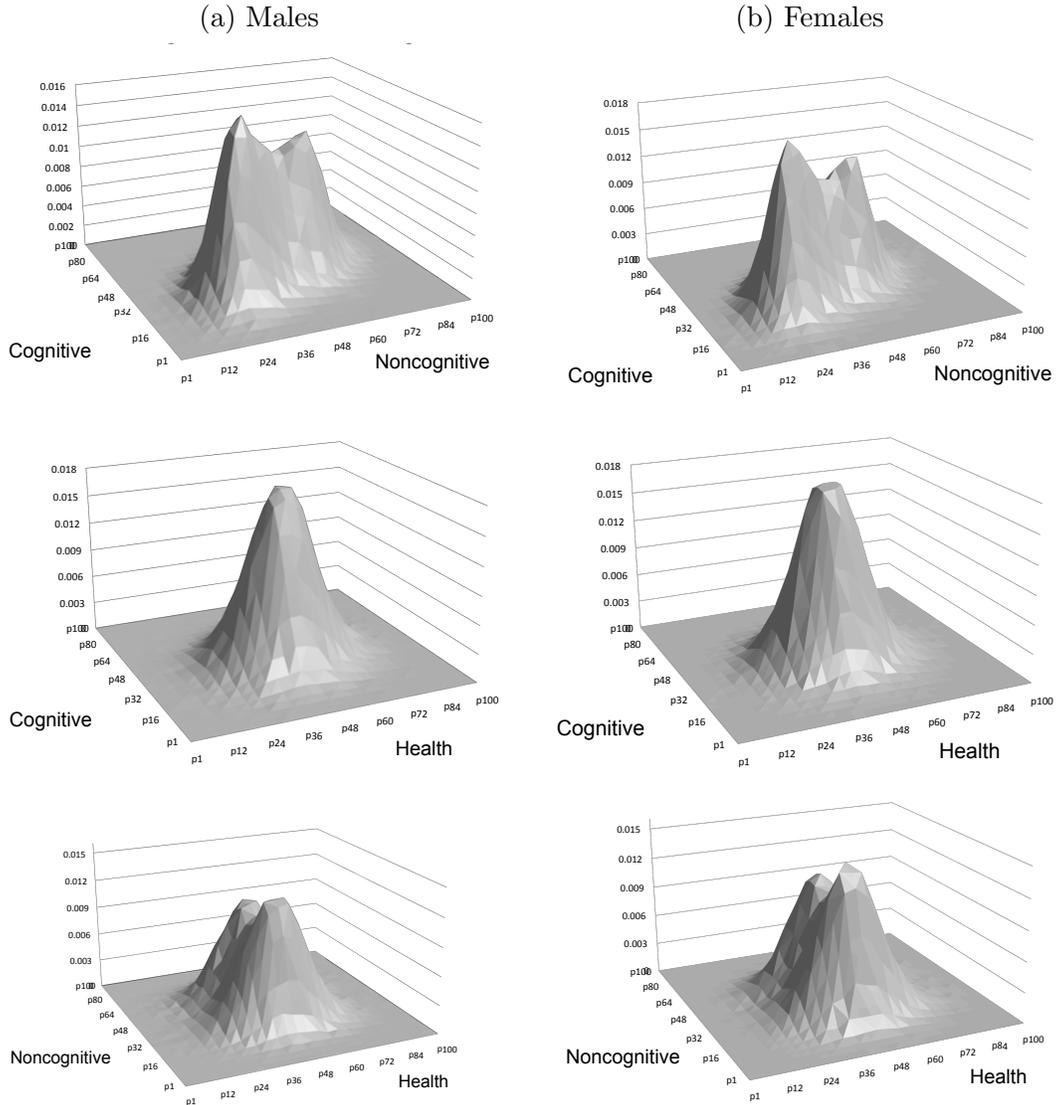
Note: The figure shows the joint distributions of cognitive, noncognitive, and health endowments, and are generated using simulated data from our model. The simulated data contains the same number of observations as the actual data. The estimated correlations are as follows: cognitive and noncognitive endowments = 0.544 for males and 0.541 for females, cognitive and health endowments = 0.176 for males and 0.153 for females, and noncognitive and health = 0.093 for males and 0.040 for females. Finally, for each endowment, the mean is standardized to be zero.

Figure 4: Share of Measurement Variance Explained by Uniqueness

Males

Father's Height, Mother's Height, Head Circumference, Height, Persistence, Attentiveness, Completeness, Cooperativeness, Perseverance, Locus of control, British Ability Scales - Word Definition, British Ability Scales - Similarities, British Ability Scales - Recall Digits, British Ability Scales - Math, Reading Test, Math Test, Picture Comprehension Test

■ Measurement Error

Females

Father's Height, Mother's Height, Head Circumference, Height, Persistence, Attentiveness, Completeness, Cooperativeness, Perseverance, Locus of control, British Ability Scales - Word Definition, British Ability Scales - Similarities, British Ability Scales - Recall Digits, British Ability Scales - Math, Reading Test, Math Test, Picture Comprehension Test
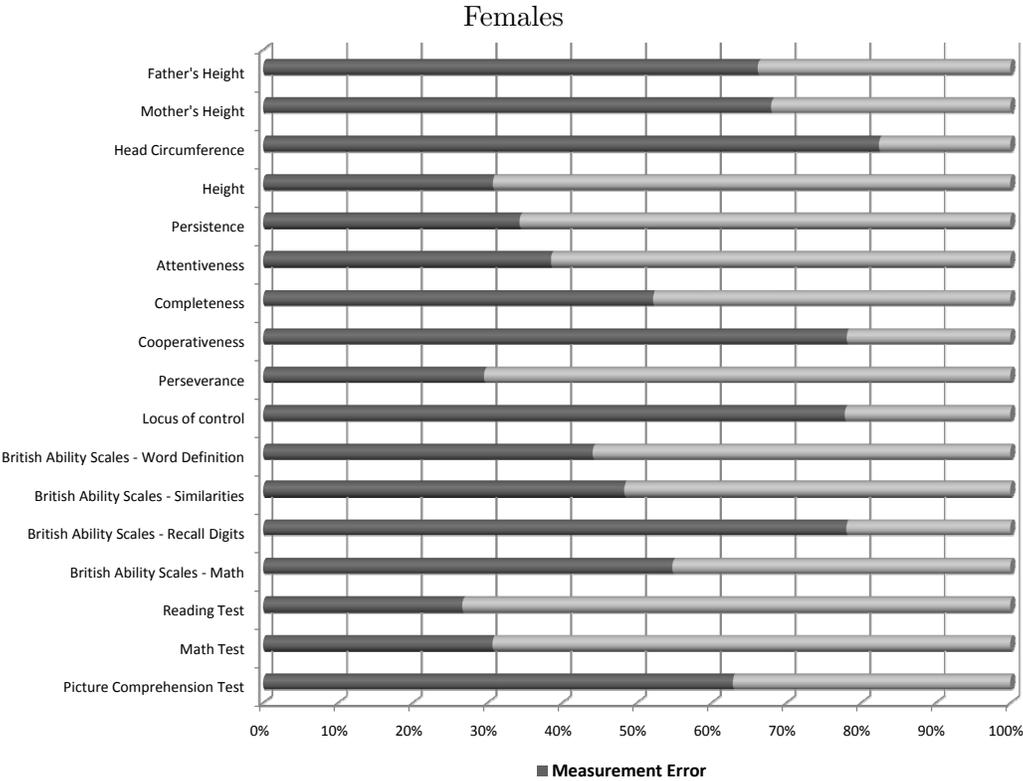
■ Measurement Error

Figure 5: Goodness of Fit - Wages

(A1) Compulsory Education [Males]

(A2) Compulsory Education [Females]

(B1) Post-Compulsory Education [Males]

(B2) Post-Compulsory Education [Females]

Note: The simulated data are generated from the model's estimates and contains the same number of observations as the actual data.

Figure 6: Marginal Distributions of Endowments by Schooling Level

Note: The endowments are simulated from the estimates of the model. The simulated data contains the same number of observations as the actual data.

Figure 7: Total Effects of Endowments: Ever Used Cannabis
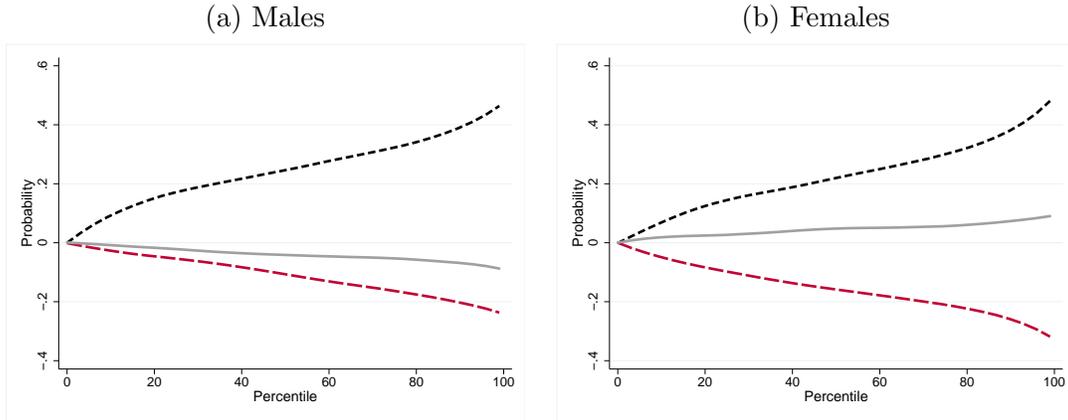
(a) Males

(b) Females

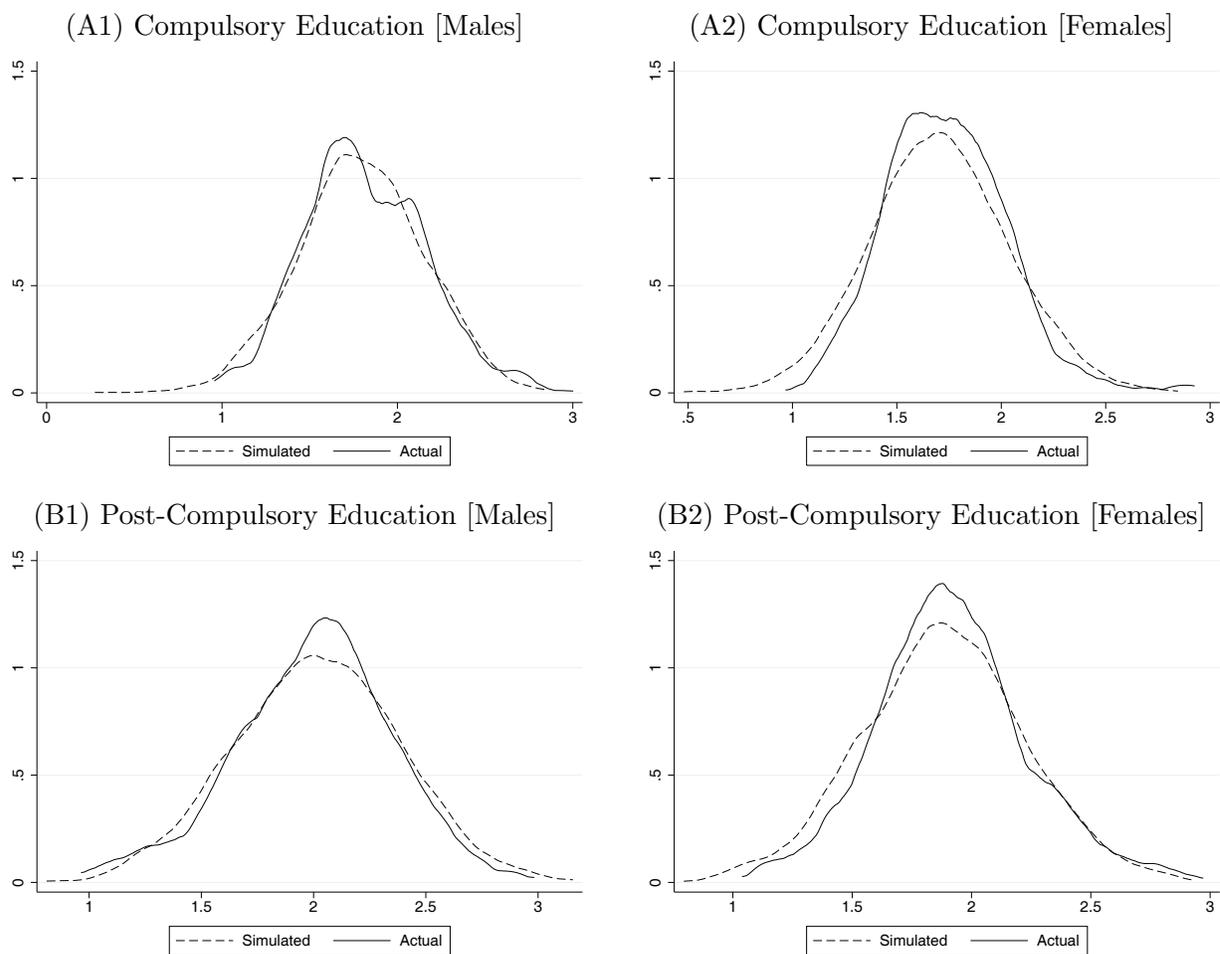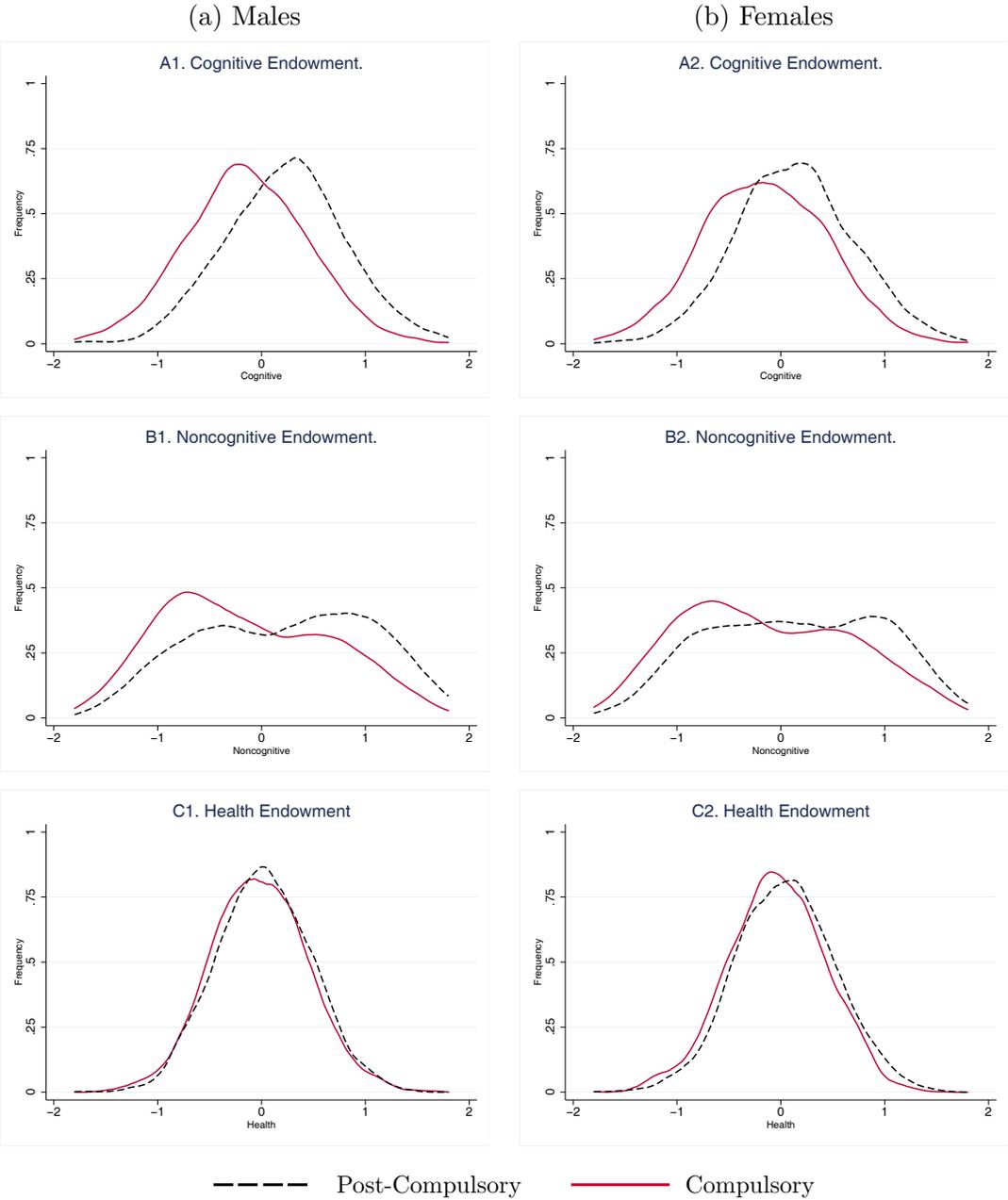Note: The endowments and the outcomes are simulated from the estimates of the model in each panel; we integrate out the observable and unobservable characteristics using the simulated distributions for each schooling group.

Figure 8: Total Effects of Endowments: Regular Exercise
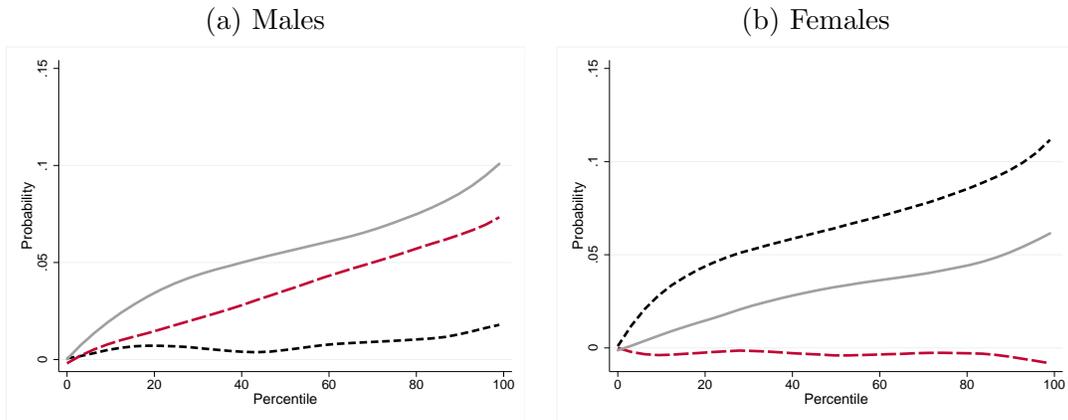
(a) Males

(b) Females

Note: The endowments and the outcomes are simulated from the estimates of the model in each panel; we integrate out the observable and unobservable characteristics using the simulated distributions for each schooling group.

Figure 9: Total Effects of Endowments: Daily Smoking

(a) Males (b) Females

Note: The endowments and the outcomes are simulated from the estimates of the model in each panel; we integrate out the observable and unobservable characteristics using the simulated distributions for each schooling group.

Figure 10: Total Effects of Endowments: Depression

(a) Males (b) Females

Note: The endowments and the outcomes are simulated from the estimates of the model in each panel; we integrate out the observable and unobservable characteristics using the simulated distributions for each schooling group.

Figure 11: Total Effects of Endowments: Fair/Poor Health

(a) Males

(b) Females

Note: The endowments and the outcomes are simulated from the estimates of the model in each panel; we integrate out the observable and unobservable characteristics using the simulated distributions for each schooling group.

Figure 12: Total Effects of Endowments: Obesity
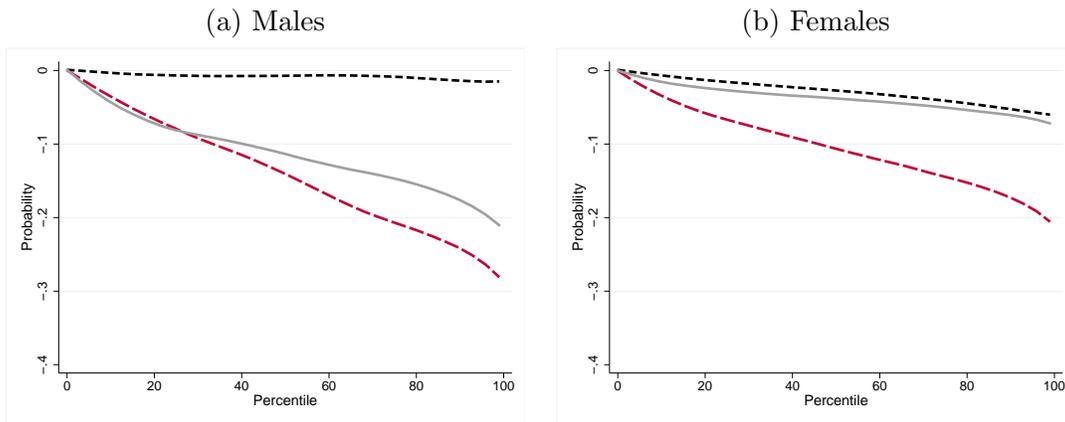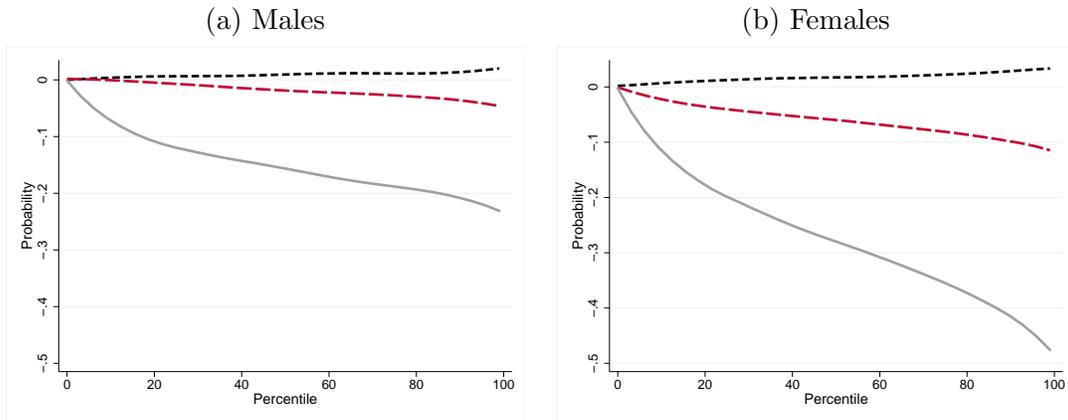
(a) Males

(b) Females

Note: The endowments and the outcomes are simulated from the estimates of the model in each panel; we integrate out the observable and unobservable characteristics using the simulated distributions for each schooling group.

Figure 13: Total Effects of Endowments: Full-Time Employment
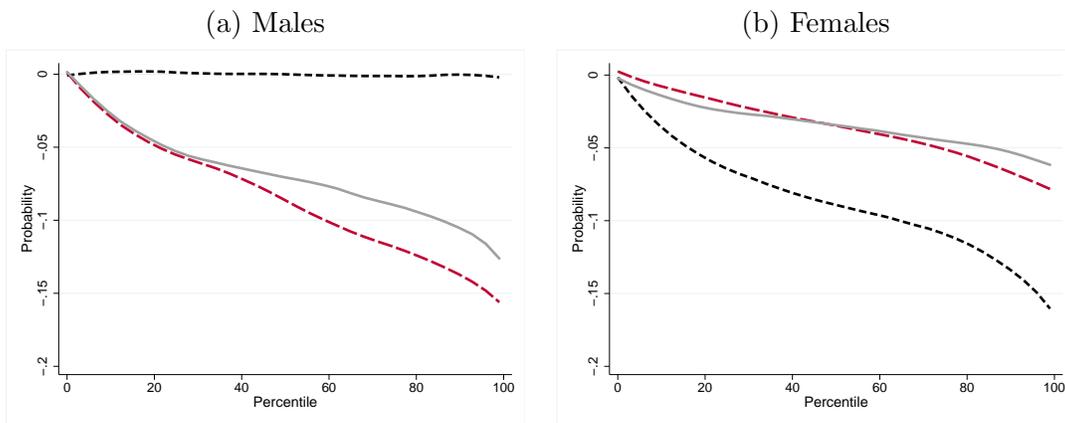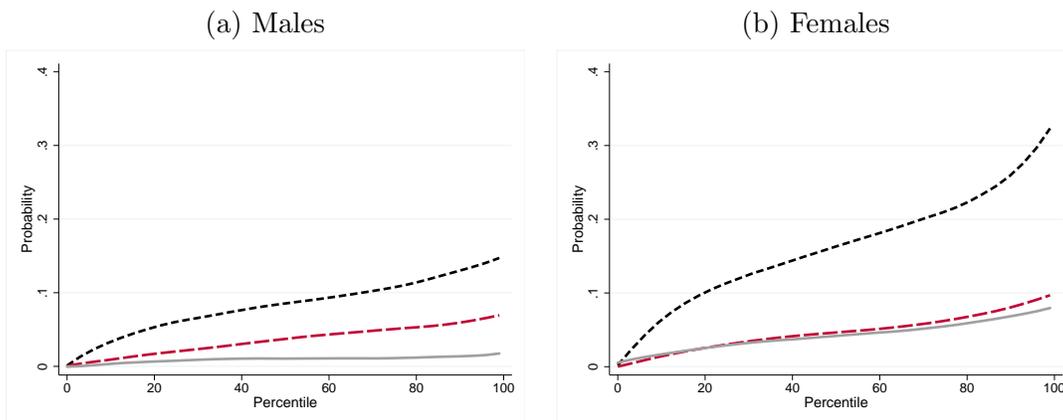
(a) Males	(b) Females

Note: The endowments and the outcomes are simulated from the estimates of the model in each panel; we integrate out the observable and unobservable characteristics using the simulated distributions for each schooling group.

Figure 14: Total Effects of Endowments: Log Wage
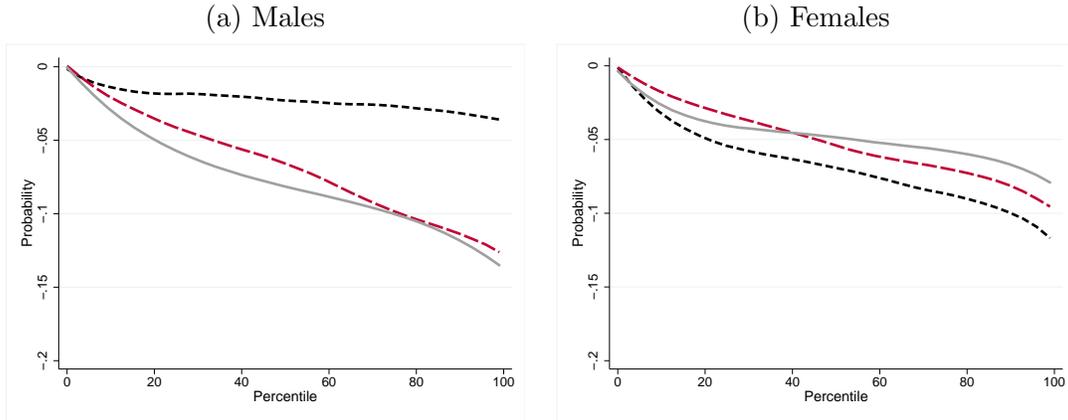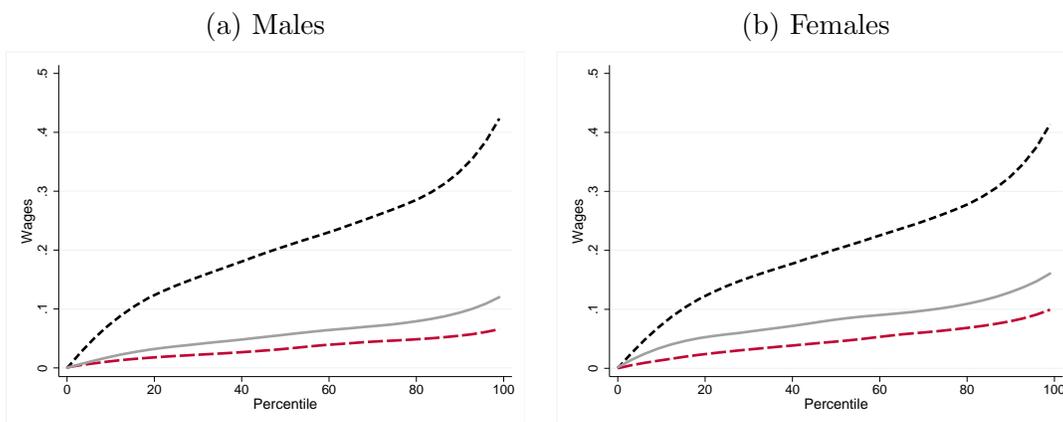
(a) Males	(b) Females

Note: The endowments and the outcomes are simulated from the estimates of the model in each panel; we integrate out the observable and unobservable characteristics using the simulated distributions for each schooling group.

31

Figure 15: Disparities in Outcomes by Education and Gender (outcomes measured at age 30)
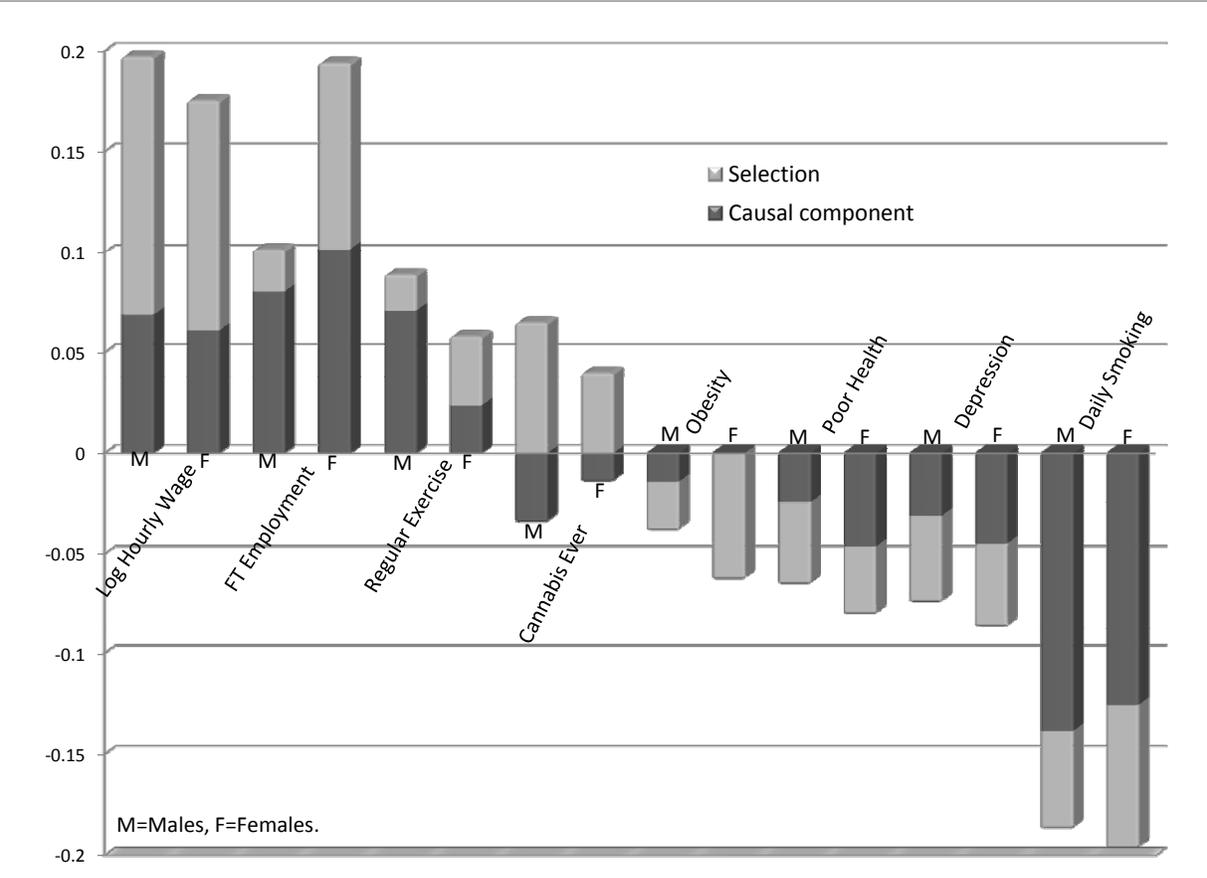
Note: The bars show the difference in outcomes by educational level (post-compulsory schooling level vs. compulsory schooling). The darker region within each bar shows the fraction of the raw gap arising from the causal contribution of education (ATE). The rest is associated with selection.

Figure 16: Fraction of the observed disparities in health and labor market outcomes due to Education
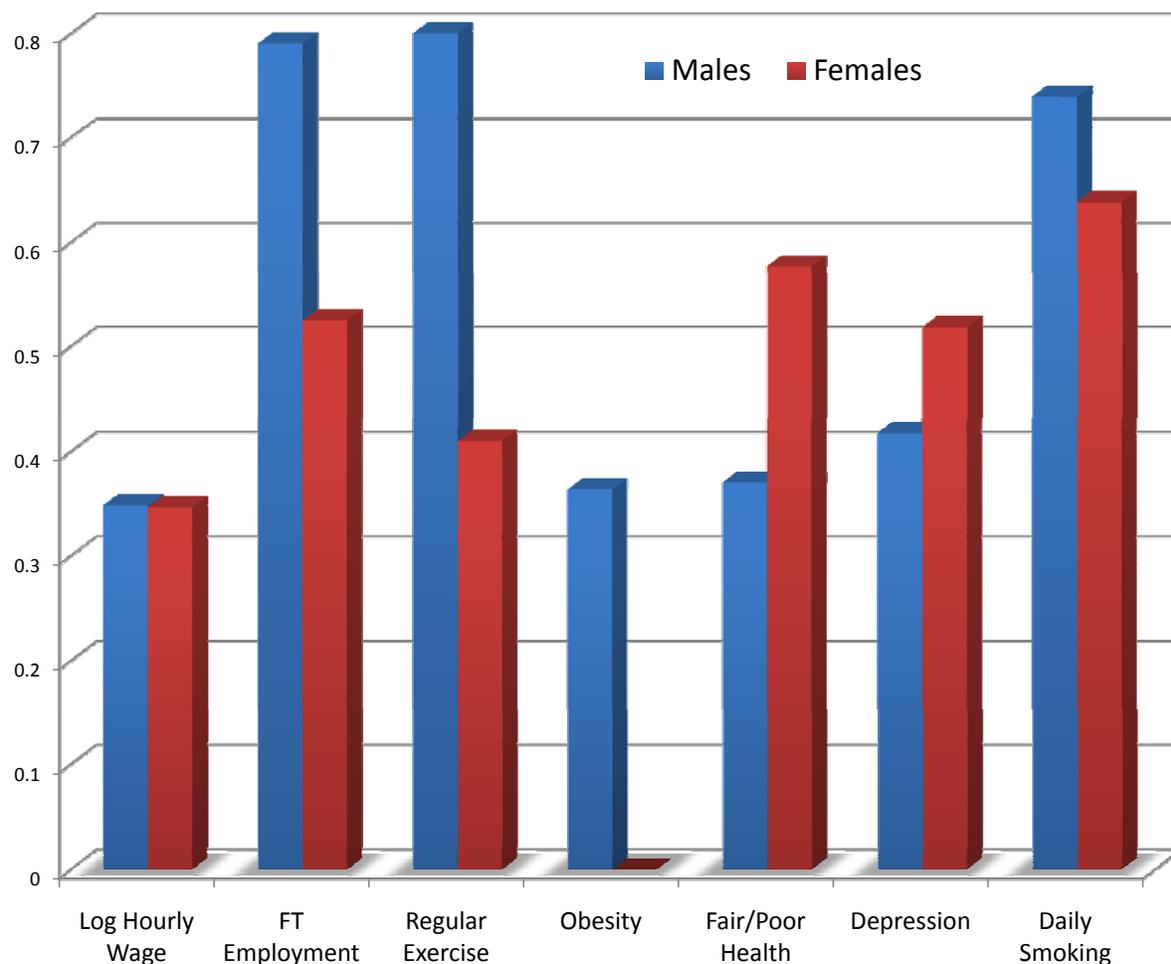
*Note for females the differential in obesity by education is entirely explained by selection (see the text for more details).

Note: The figure displays the fractions of the observed differentials which can be attributed to the effect of education. Specifically, if we denote by Δ the observed differences in outcome Y, i.e. $\Delta = E[Y_1|D=1] - E[Y_0|D=0]$, in this figure we present $\frac{E[Y_1-Y_0]}{E[Y_1|D=1]-E[Y_0|D=0]}$.

Figure 17: Population Distribution of the Average Treatment Effect - Health Behaviors
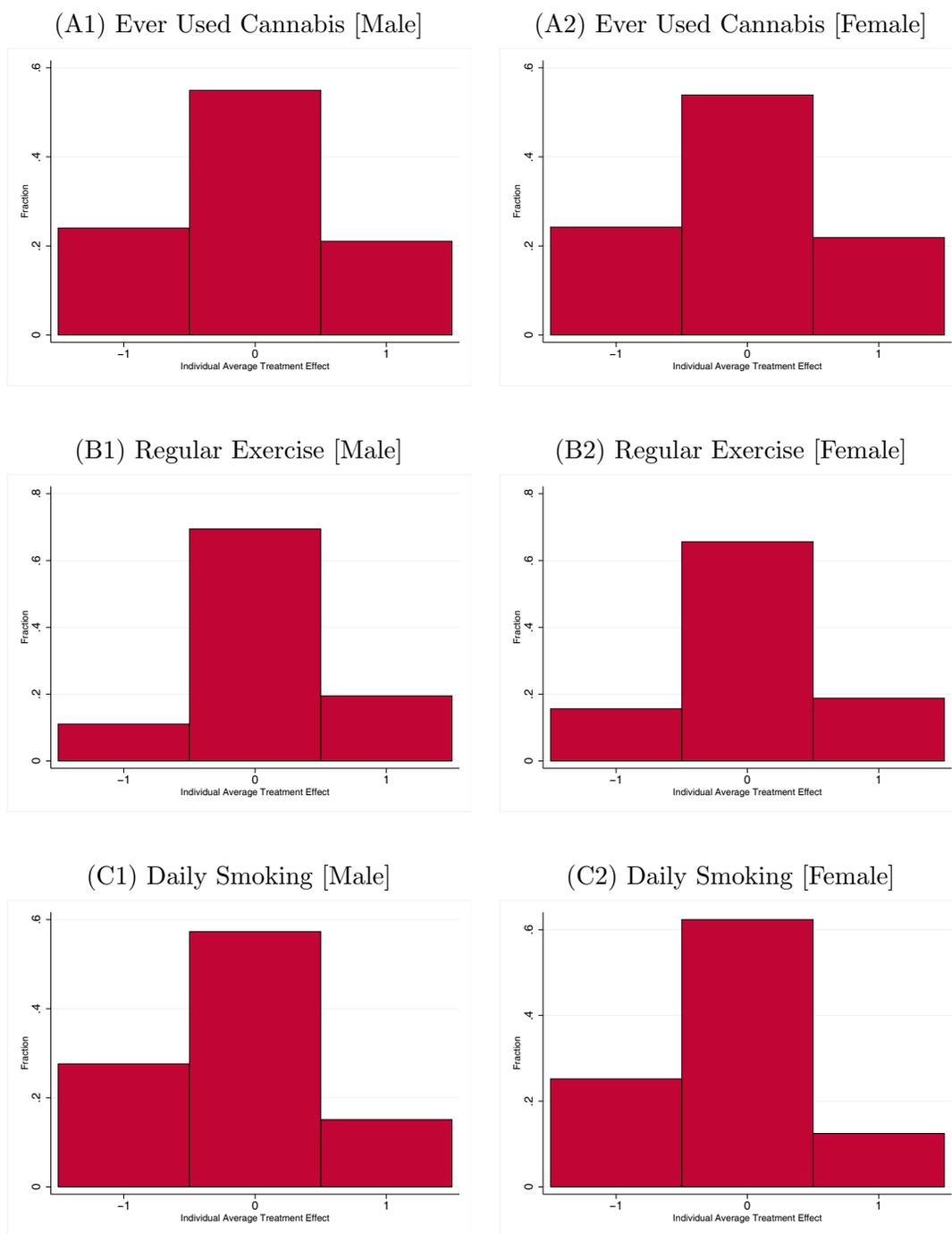
(A1) Ever Used Cannabis [Male]

(A2) Ever Used Cannabis [Female]

(B1) Regular Exercise [Male]

(B2) Regular Exercise [Female]

(C1) Daily Smoking [Male]

(C2) Daily Smoking [Female]

Note: The figures display the distribution of the average treatment effect by gender. The outcomes are simulated from the estimates of the model. The simulated data contains the same number of observations as the actual data.

Figure 18: Population Distribution of the Average Treatment Effect - Health Outcomes
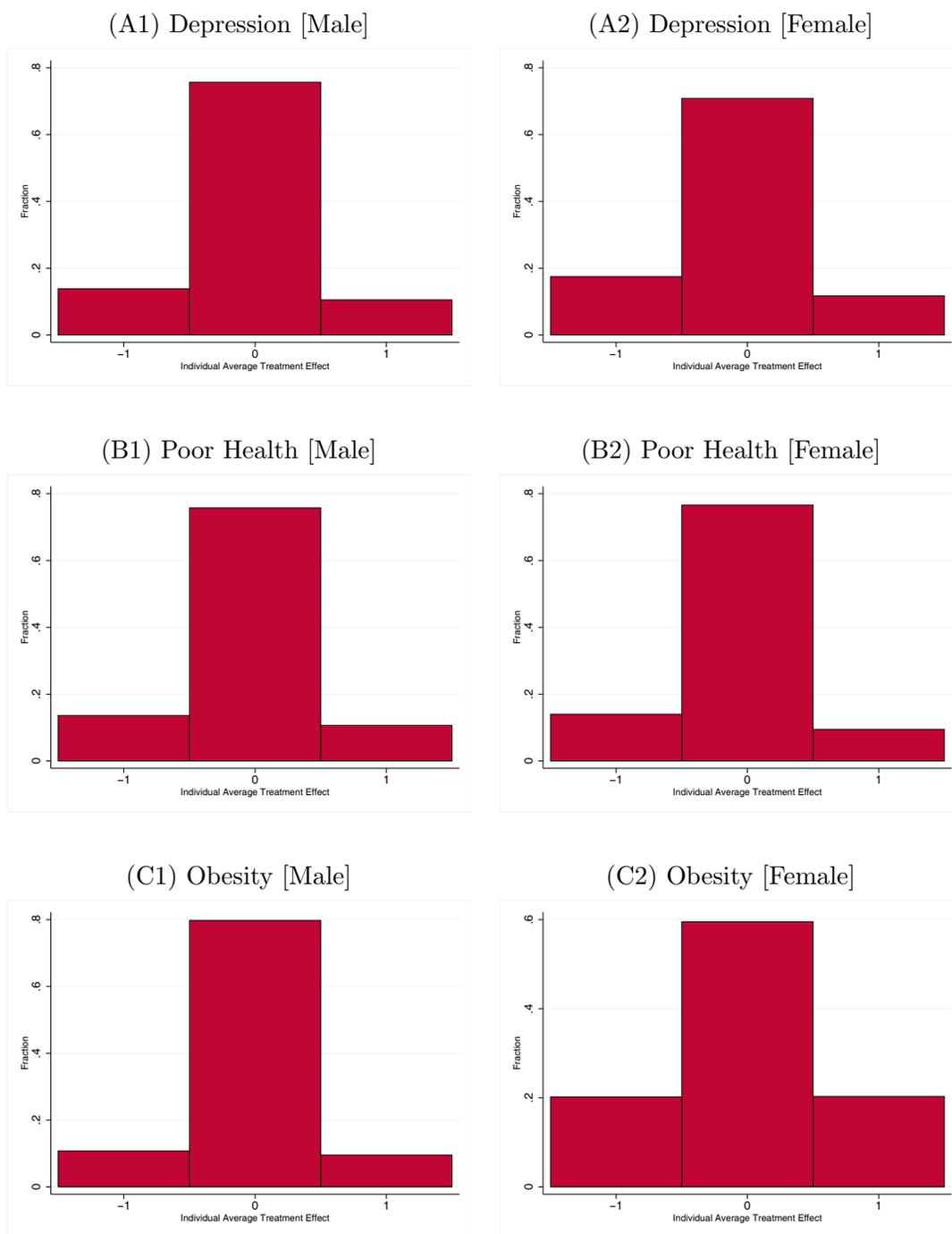
Note: The figures display the distribution of the average treatment effect by gender. The outcomes are simulated from the estimates of the model. The simulated data contains the same number of observations as the actual data.

Figure 19: Population Distribution of the Average Treatment Effect - Labor Market Outcomes
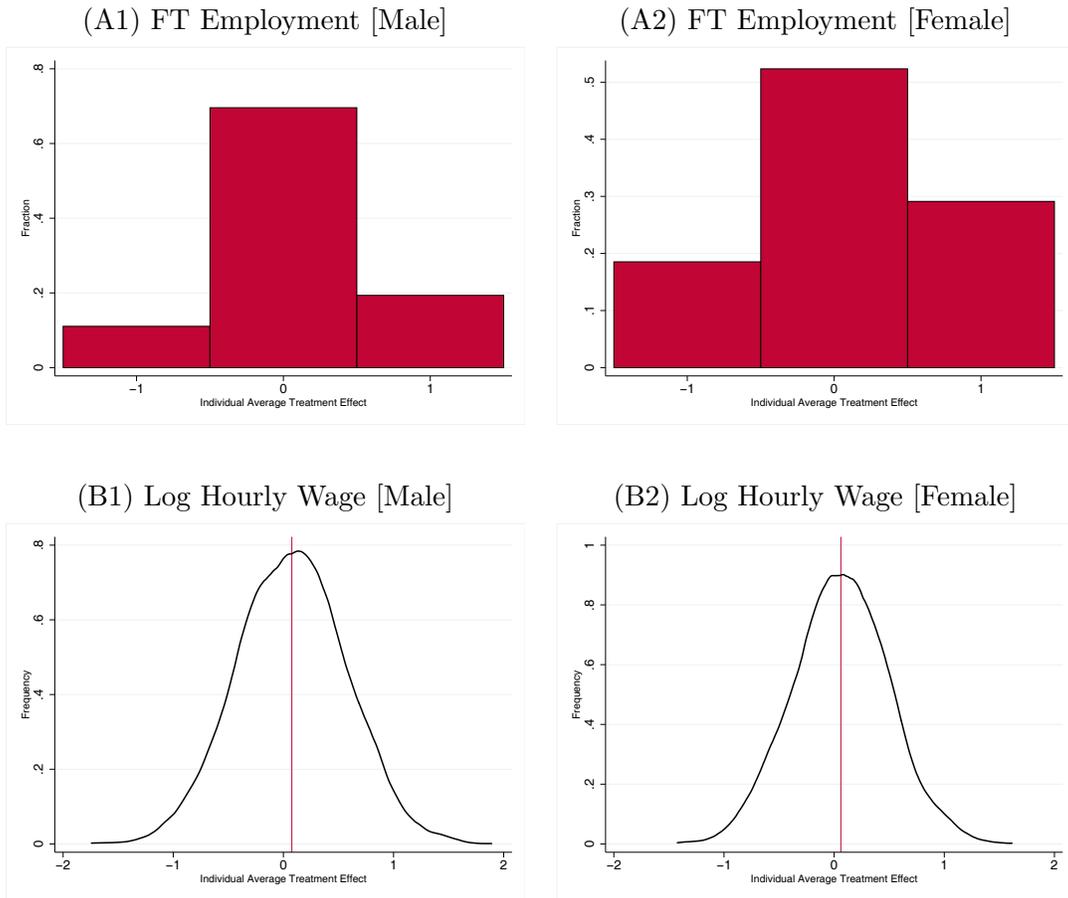
(A1) FT Employment [Male]

(A2) FT Employment [Female]

(B1) Log Hourly Wage [Male]

(B2) Log Hourly Wage [Female]

Note: The figures display the distribution of the average treatment effect by gender. The outcomes are simulated from the estimates of the model. The simulated data contains the same number of observations as the actual data.

Figure 20: Treatment Effect Heterogeneity: Health Outcomes
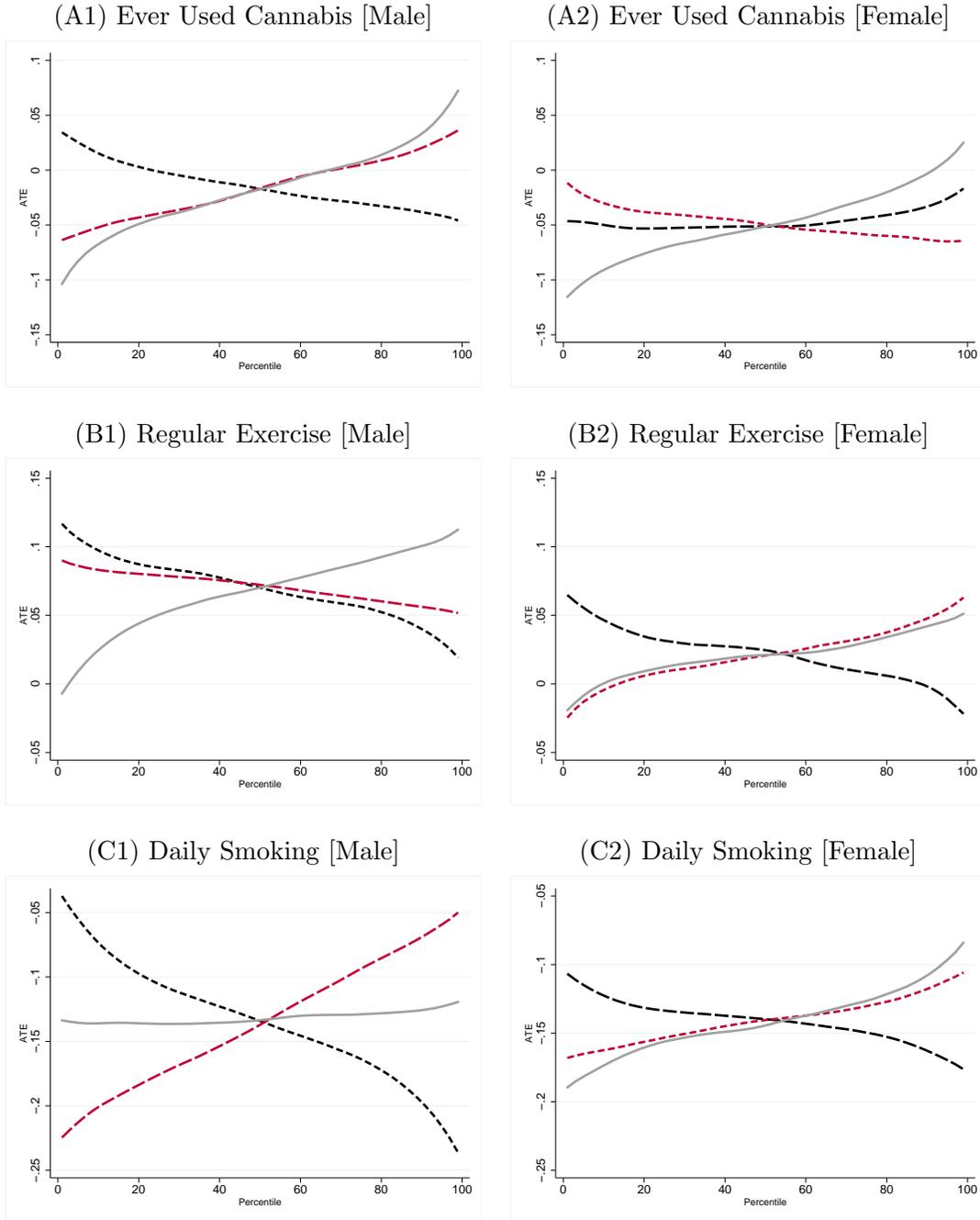
(A1) Ever Used Cannabis [Male]

(A2) Ever Used Cannabis [Female]

(B1) Regular Exercise [Male]

(B2) Regular Exercise [Female]

(C1) Daily Smoking [Male]

(C2) Daily Smoking [Female]

Note: The endowments and the outcomes are simulated from the estimates of the model in each panel; we integrate out the observable and unobservable characteristics using the simulated distributions for each schooling group.

Figure 21: Treatment Effect Heterogeneity: Health Outcomes
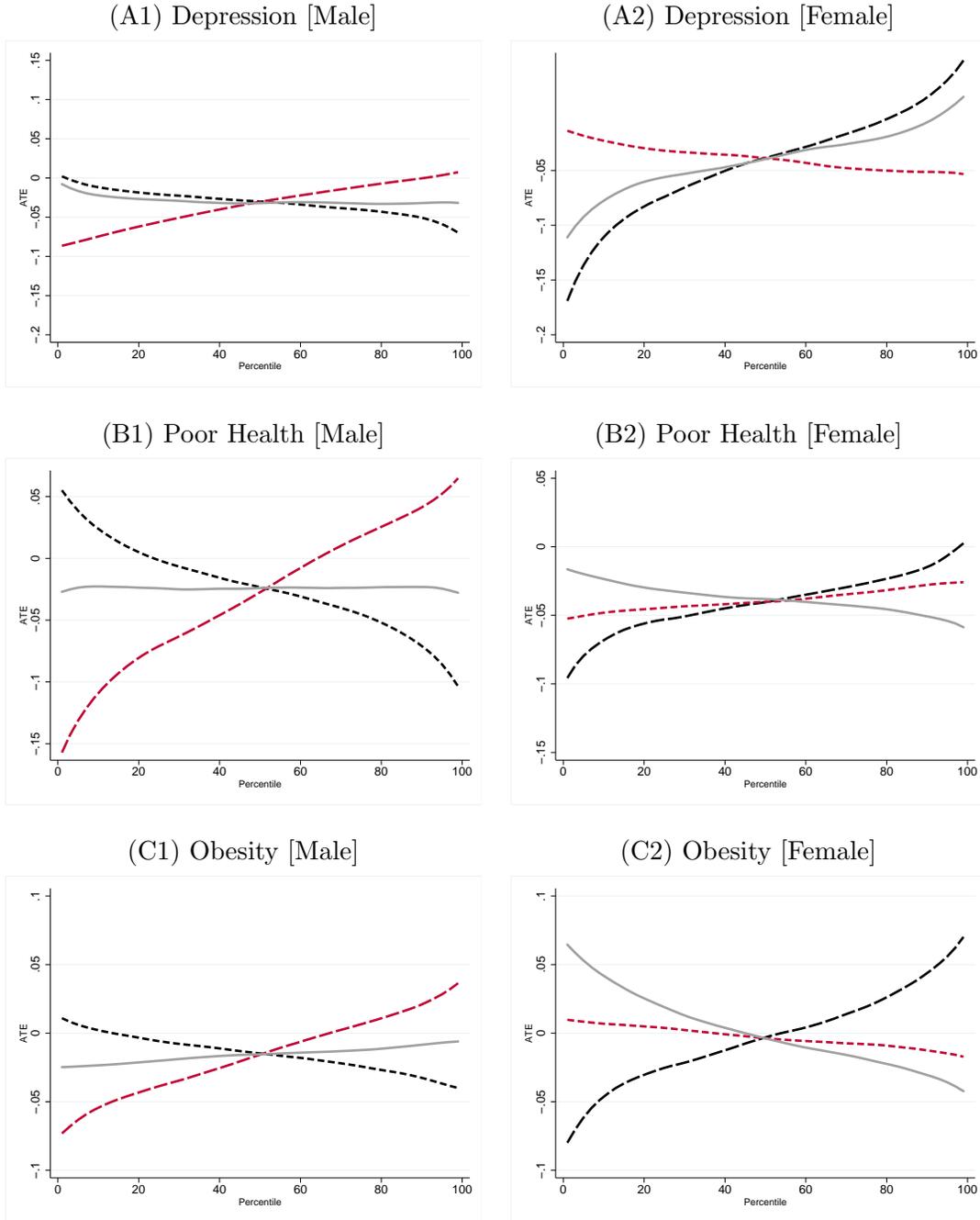
(A1) Depression [Male]　　　　(A2) Depression [Female]

(B1) Poor Health [Male]　　　　(B2) Poor Health [Female]

(C1) Obesity [Male]　　　　(C2) Obesity [Female]

Note: The endowments and the outcomes are simulated from the estimates of the model in each panel; we integrate out the observable and unobservable characteristics using the simulated distributions for each schooling group.

Figure 22: Treatment Effect Heterogeneity: Labor Market Outcomes
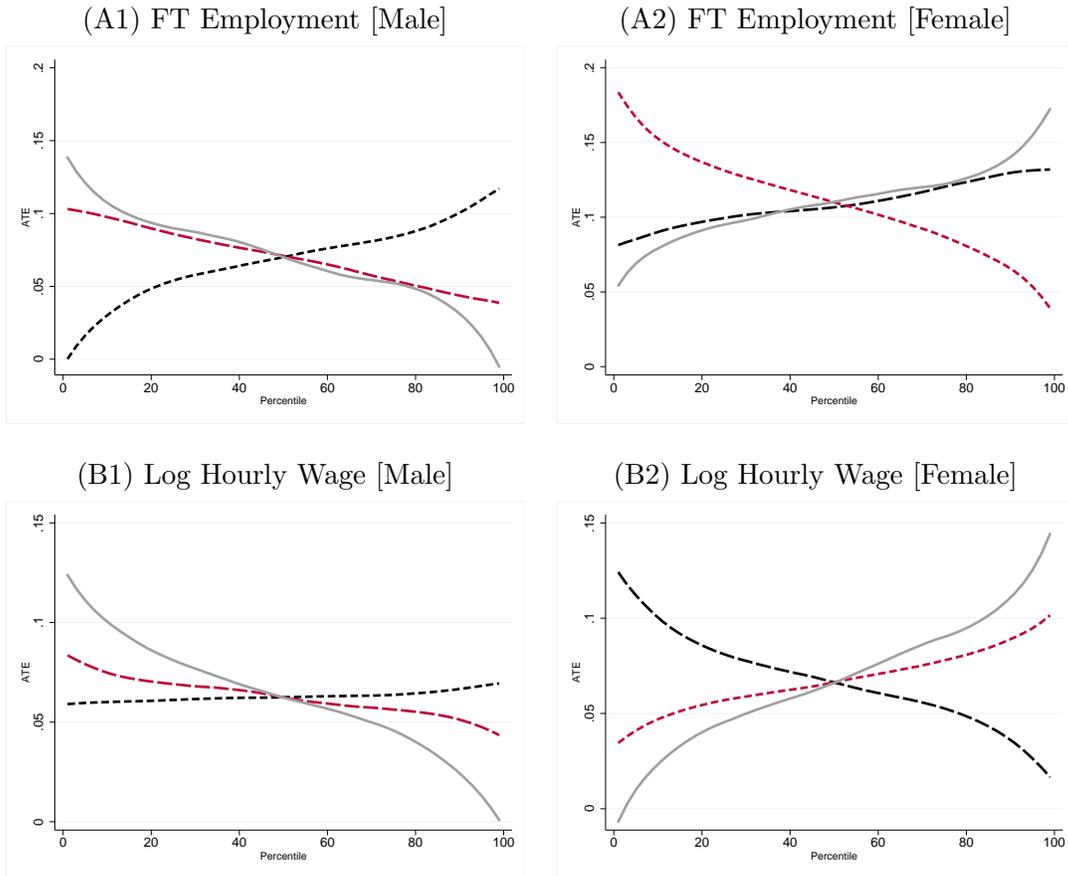
(A1) FT Employment [Male]

(A2) FT Employment [Female]

(B1) Log Hourly Wage [Male]

(B2) Log Hourly Wage [Female]

Note: The endowments and the outcomes are simulated from the estimates of the model in each panel; we integrate out the observable and unobservable characteristics using the simulated distributions for each schooling group.

Figure 23: Decomposition of Selection Bias
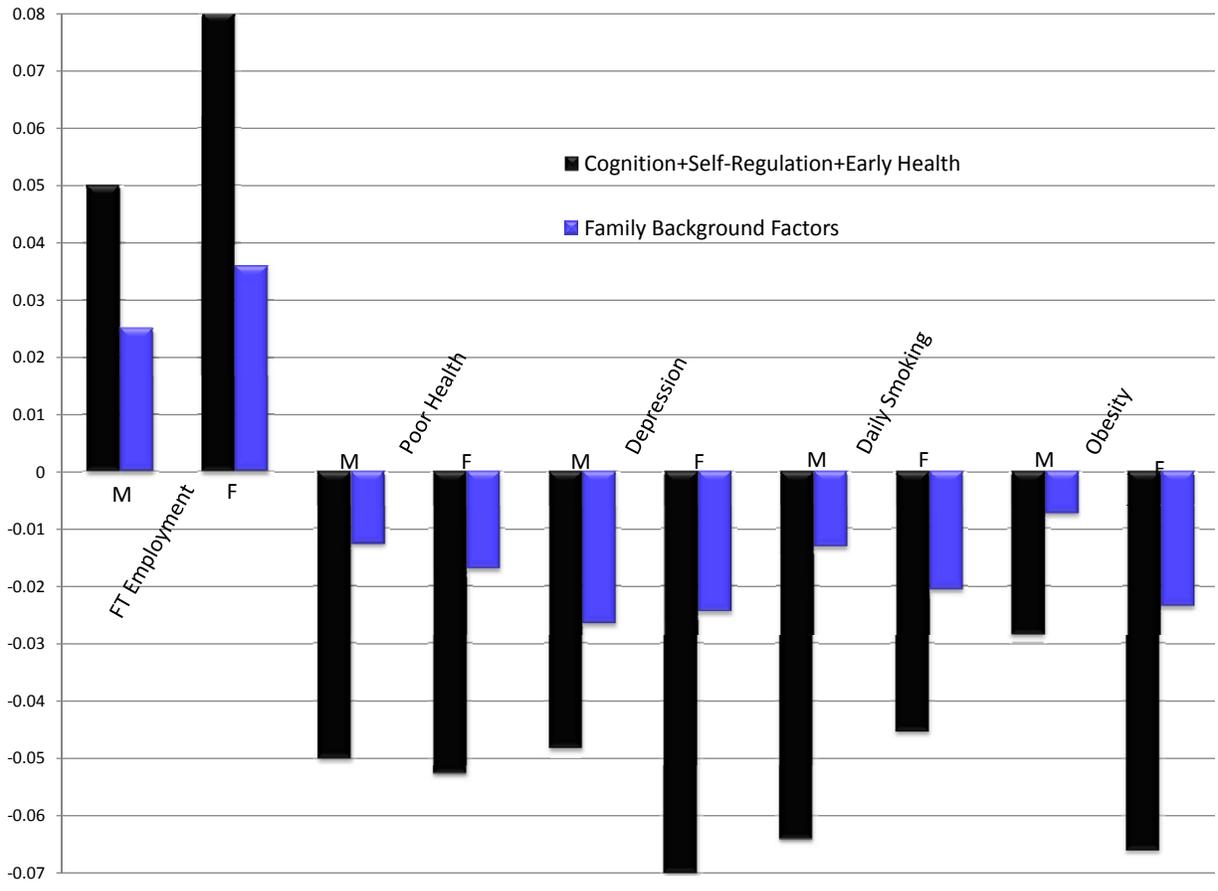
Note: The darker bars show the component of the selection bias arising from the contribution of the unobservables, the lighter bars from the contribution of observables.

Figure 24: Structural Matching and Propensity Score Matching Results
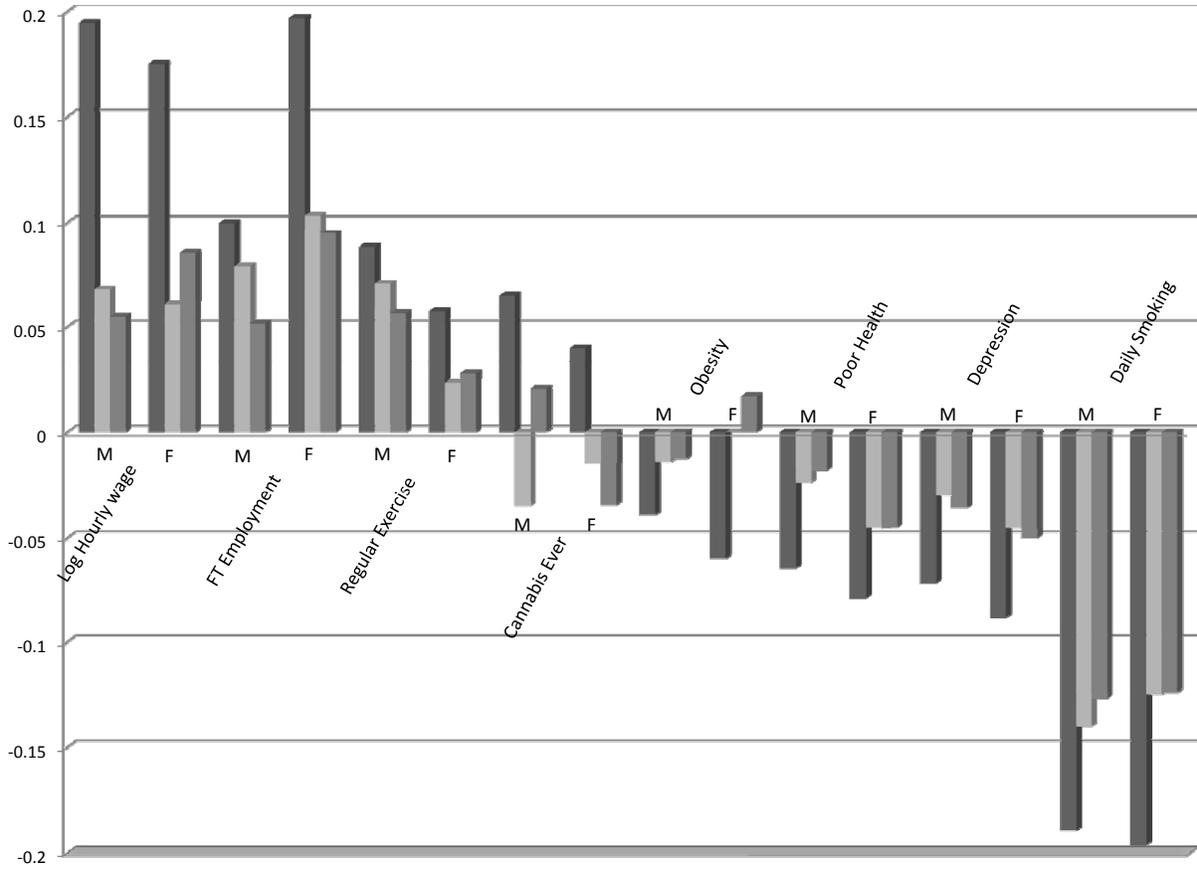
Note: The left bar displays the difference in outcomes by educational level; the central bar displays the average treatment effect of education obtained from our structural model; the right bar displays the same parameter obtained using propensity score matching. M=males, F=females.

Table 1: Summary Statistics: Outcomes

Males

	Compulsory (0.598)					Post-Compulsory (0.402)						
	Mean	S.D.	Min.	Max.	n	Mean	S.D.	Min.	Max.	n	Δ	p-value
Healthy Behaviors												
Ever Used Cannabis	0.575	0.494	0	1	1568	0.640	0.480	0	1	1064	0.065	0.001
Regular Exercise	0.772	0.419	0	1	1585	0.861	0.346	0	1	1070	0.088	0.000
Daily Smoking	0.380	0.485	0	1	1585	0.193	0.395	0	1	1069	-0.187	0.000
Health												
Depression	0.172	0.377	0	1	1565	0.098	0.297	0	1	1065	-0.074	0.000
Poor Health	0.173	0.378	0	1	1587	0.107	0.310	0	1	1071	-0.065	0.000
Obesity	0.128	0.334	0	1	1560	0.090	0.286	0	1	1048	-0.038	0.002
Labor Market Outcomes												
Full-Time Employed	0.765	0.424	0	1	1587	0.866	0.340	0	1	1071	0.101	0.000
Log of Hourly Wage	1.817	0.350	0.957	3.000	1182	2.014	0.363	0.959	3.029	890	0.197	0.000

Females

	Compulsory (0.502)					Post-Compulsory (0.498)						
	Mean	S.D.	Min.	Max.	n	Mean	S.D.	Min.	Max.	n	Δ	p-value
Healthy Behaviors												
Ever Used Cannabis	0.404	0.491	0	1	1421	0.444	0.497	0	1	1420	0.040	0.032
Regular Exercise	0.751	0.433	0	1	1430	0.809	0.393	0	1	1429	0.058	0.000
Daily Smoking	0.366	0.482	0	1	1431	0.169	0.375	0	1	1428	-0.197	0.000
Health												
Depression	0.228	0.420	0	1	1420	0.142	0.350	0	1	1418	-0.086	0.000
Poor Health	0.177	0.382	0	1	1431	0.097	0.295	0	1	1429	-0.080	0.000
Obesity	0.350	0.477	0	1	1383	0.287	0.452	0	1	1390	-0.063	0.000
Labor Market Outcomes												
Full-Time Employed	0.395	0.489	0	1	1431	0.589	0.492	0	1	1430	0.194	0.000
Log of Hourly Wage	1.734	0.317	0.968	2.944	548	1.910	0.321	1.016	2.970	822	0.175	0.000

The column Δ shows the difference in means for each outcomes between the individuals with compulsory and post-compulsory level of education. The last column shows the p-value of a two-sided test for the statistical significance of that difference.

Table 2: Summary Statistics: Measurements

	Males					Females				
	Mean	S.D.	Min.	Max.	n	Mean	S.D.	Min.	Max.	n
Cognitive Scores (Age 10)										
Picture Language Comprehension Test	41.83	7.29	18	71	3769	40.30	7.15	17	71	3614
Friendly Math Test	45.20	12.48	1	72	3770	43.98	11.23	4	72	3614
Shortened Edinburgh Reading Test	45.10	15.55	1	74	3772	47.36	14.46	0	74	3618
BAS Matrices	15.35	5.39	0	28	3777	15.88	5.24	1	28	3620
BAS Recall Digits	22.34	4.31	1	34	3763	22.62	4.22	8	34	3608
BAS Similarities	12.38	2.60	0	20	3746	12.00	2.42	0	20	3601
BAS Word Definition	10.89	5.19	0	30	3760	9.86	4.74	0	29	3611
Noncognitive Scales (Age 10)										
Locus of control	9.87	2.45	1.5	16	3477	9.66	2.38	1.5	16	3346
Perseverance	27.04	11.60	1	47	3777	30.08	10.80	1	47	3620
Cooperativeness	31.39	9.05	1	47	3748	32.33	8.818	1	47	3586
Completeness	30.98	13.72	1	47	3749	35.33	12.41	1	47	3596
Attentiveness	30.27	12.92	1	47	3752	33.87	12.16	1	47	3581
Persistence	25.84	13.78	1	47	3754	30.22	12.96	1	47	3607
Health Measurements (Age 10)										
Height (ms)	1.39	0.06	1.18	1.61	3777	1.38	0.06	1.17	1.63	3620
Head circumference (mms)	539.36	17.87	318	597	3690	530.64	18.05	305	597	3501
Mother height (cms)	161.15	6.56	124	183	3774	161.51	6.65	132	188	3618
Father height (cms)	175.28	7.49	147	211	3622	149.55	62.68	0	201	3616

BAS = British Ability Scales.

Table 3: Summary Statistics: Covariates

	Males					Females				
	Mean	S.D.	Min.	Max.	n	Mean	S.D.	Min.	Max.	n
Covariates										
Mother's age at birth	25.84	5.31	14	52	3777	25.78	5.28	15	46	3620
Mother's education a birth	0.33	0.47	0	1	3777	0.34	0.47	0	1	3620
# Children at age 10	2.55	1.03	1	13	3777	2.54	1.03	1	10	3620
Father high SC at birth	0.29	0.45	0	1	3777	0.29	0.45	0	1	3620
Broken family	0.12	0.32	0	1	3777	0.13	0.34	0	1	3620
Total gross family income at age 10[1]	4.03	1.23	1	7	3777	4.06	1.25	1	7	3620
Parity	1.12	1.27	0	11	3777	1.11	1.29	0	10	3620
Additional Covariates in the Measurement System										
Weight[2] (kgs)	32.24	4.90	23	48	3777	32.80	5.62	23	51	3620
Mother weight[2] (kgs)	60.82	10.34	35	164	3777	61.03	10.60	29	132	3620
Father weigh[2] (kgs)	75.14	10.16	41	167	3777	75.34	9.89	44	139	3620
Additional Covariate in the Schooling Choice Equation										
Claimant Count	12.89	3.42	7.8	19.8	3777	7.77	1.48	5.2	10.1	3620

Note: SC = Social Class. High Social Class comprises SCI, SCII and SCIIINM (Non-Manual) . The BCS70 uses the Registrar General's classification for measuring social class (SC). Social class I includes professionals, such as lawyers, architects and doctors; Social Class II includes intermediate workers, such as shopkeepers, farmers and teachers; Social Class III Non Manual includes skilled non-manual workers, such as shop assistants and clerical workers in offices.

[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] Only in the measurement system for health.

Table 4: Goodness of Fit

	Males			Females		
	Actual	Simulated	Goodness of Fit	Actual	Simulated	Goodness of Fit
Education	0.40	0.39	0.391	0.50	0.49	0.208
A. Healthy Behaviors						
Ever Cannabis [C]	0.57	0.58	0.771	0.40	0.39	0.259
Ever Cannabis [PC]	0.64	0.63	0.657	0.44	0.45	0.404
Regular Exercise [C]	0.77	0.76	0.271	0.75	0.74	0.321
Regular Exercise [PC]	0.86	0.87	0.454	0.81	0.82	0.268
Daily Smoking [C]	0.38	0.38	0.468	0.37	0.37	0.569
Daily Smoking [PC]	0.19	0.21	0.194	0.17	0.17	0.728
B. Health Status						
Depression [C]	0.17	0.18	0.232	0.23	0.24	0.260
Depression [PC]	0.10	0.10	0.826	0.14	0.13	0.161
Poor Health [C]	0.17	0.18	0.142	0.18	0.18	0.469
Poor Health [PC]	0.11	0.10	0.517	0.10	0.10	0.321
Obesity [C]	0.13	0.12	0.480	0.35	0.36	0.593
Obesity [PC]	0.09	0.09	0.917	0.29	0.29	0.819
C. Labor Market Outcomes						
FT Employment [C]	0.76	0.76	0.820	0.39	0.38	0.280
FT Employment [PC]	0.87	0.87	0.674	0.59	0.58	0.561

Note: The simulated data are generated from the model's estimates and contains the same number of observations as the actual data. Goodness of fit is tested using a χ^2 test where the Null Hypothesis is *Simulated=Actual* (*p*-values are reported). FT=full-time. C=compulsory; PC=post-compulsory.

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.

Abbring, J. H. and J. J. Heckman (2007). Econometric evaluation of social programs, part III: Distributional treatment effects, dynamic treatment effects, dynamic discrete choice, and general equilibrium policy evaluation. In J. Heckman and E. Leamer (Eds.), *Handbook of Econometrics*, Volume 6B, pp. 5145–5303. Amsterdam: Elsevier.

Almond, D. and J. Currie (2010). Human Capital Development Before Age Five. *NBER Working Papers*.

Auld, M. C. and N. Sidhu (2005, October). Schooling, cognitive ability and health. *Health Economics 14*(10), 1019–1034.

Batty, G. D., I. J. Deary, I. Schoon, and C. R. Gale (2007, November). 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, May). 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.

Case, A., A. Fertig, and C. Paxson (2005, March). The lasting impact of childhood health and circumstance. *Journal of Health Economics 24*(2), 365–389.

Case, A., D. Lubotsky, and C. Paxson (2002, December). Economic status and health in childhood: The origins of the gradient. *American Economic Review 92*(5), 1308–1334.

Commission on Social Determinants of Health (2008). Closing the gap in a generation: Health equity through action on the social determinants of health. Final report, World Health Organization, Geneva.

Conti, G. (2009). Cannabis, cognition and wages. Unpublished manuscript, University of Chicago, Department of Economics.

Conti, G., J. Heckman, H. Lopes, and R. Piatek (2011). Constructing Economically Justified Aggregates: An Application to the Early Origins of Health. *Unpublished manuscript, University of Chicago, Department of Economics*.

Costa, P. T. and R. R. McCrae (1988, August). From catalog to classification: Murray's needs and the five-factor model. *Journal of Personality and Social Psychology 55*(2), 258–265.

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

Cunha, F., J. J. Heckman, L. J. Lochner, and D. V. Masterov (2006). Interpreting the evidence on life cycle skill formation. In E. A. Hanushek and F. Welch (Eds.), *Handbook of the Economics of Education*, Chapter 12, pp. 697–812. Amsterdam: North-Holland.

Cunha, F., J. J. Heckman, and S. M. Schennach (2010). Estimating the technology of cognitive and noncognitive skill formation. Forthcoming, *Econometrica*.

Currie, J. (2009a). Healthy, wealthy, and wise: Socioeconomic status, poor health in childhood, and human capital development. *Journal of Economic Literature 47*(1), 87–122.

Currie, J. (2009b, November). Policy interventions to address child health disparities: moving beyond health insurance. *Pediatrics 124*(Supplement), S246–S254.

Currie, J. and E. Moretti (2003, November). Mother's education and the intergenerational transmission of human capital: Evidence from college openings. *Quarterly Journal of Economics 118*(4), 1495–1532.

Cutler, D. M. and A. Lleras-Muney (2007, August). Understanding differences in health behaviors by education. Unpublished manuscript, Harvard University, Department of Economics.

Dahly, D. L., L. S. Adair, and K. A. Bollen (2008, February). A structural equation model of the developmental origins of blood pressure. *International Journal of Epidemiology 37*(1), 1–11.

Dow, W., R. Schoeni, N. Adler, and J. Stewart (2010). Evaluating the evidence base: Policies and interventions to address socioeconomic status gradients in healtha. *Annals of the New York Academy of Sciences 1186*(1), 240–251.

Elias, J. J. (2005, December). *The Effects of Ability and Family Background on Non-Monetary Returns to Education.* Ph. D. thesis, University of Chicago, Chicago, IL.

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

Fuchs, V. R. (1982). Time preference and health: An exploratory study. In V. R. Fuchs (Ed.), *Economic Aspects of Health*, pp. 93–120. Chicago, IL: University of Chicago Press.

Gluckman, P. D. and M. A. Hanson (2006). *Developmental Origins of Health and Disease.* Cambridge, UK: Cambridge University Press.

Gottfredson, L. S. and I. J. Deary (2004, February). Intelligence predicts health and longevity, but why? *Current Directions in Psychological Science 13*(1), 1–4.

Griliches, Z. (1977, January). Estimating the returns to schooling: Some econometric problems. *Econometrica 45*(1), 1–22.

Grossman, M. (1972, March-April). On the concept of health capital and the demand for health. *Journal of Political Economy 80*(2), 223–255.

Grossman, M. (1975). The correlation between health and schooling. In N. E. Terleckyj (Ed.), *Household Production and Consumption*, pp. 147–211. New York: Columbia University Press.

Grossman, M. (2000). The human capital model. In A. J. Culyer and J. P. Newhouse (Eds.), *Handbook of Health Economics*, Volume 1, pp. 347–408. Amsterdam: Elsevier.

Grossman, M. (2006). Education and nonmarket outcomes. In E. Hanushek and F. Welch (Eds.), *Handbook of the Economics of Education*, Volume 1, Chapter 10, pp. 577–633. Amsterdam: Elsevier.

Grossman, M. (2008). The relationship between health and schooling: Presidential address. *Eastern Economic Journal 34*(3), 281–292.

Grossman, M. and R. Kaestner (1997). Effects of education on health. In J. R. Behrman and N. Stacey (Eds.), *The Social Benefits of Education*, pp. 69–124. Ann Arbor, MI: University of Michigan Press.

Hampson, S. E. and H. S. Friedman (2008). Personality and health: A lifespan perspective. In O. P. John, R. Robins, and L. Pervin (Eds.), *The Handbook of Personality: Theory and Research* (Third ed.)., pp. 770–794. New York: Guilford.

Hartog, J. and H. Oosterbeek (1998, June). Health, wealth and happiness: Why pursue a higher education? *Economics of Education Review 17*(3), 245–256.

Heckman, J., S. Schennach, and B. Williams (2010). Matching with Error-Ridden Covariates. *Unpublished manuscript, University of Chicago, Department of Economics*.

Heckman, J. J. and B. E. Honoré (1990, September). The empirical content of the Roy model. *Econometrica 58*(5), 1121–1149.

Heckman, J. J., S. H. Moon, R. Pinto, P. A. Savelyev, and A. Q. Yavitz (2009). A Reanalysis of the HighScope Perry Preschool Program. Unpublished manuscript, University of Chicago, Department of Economics. First draft, September, 2006.

Heckman, J. J., J. Stixrud, and S. Urzua (2006, July). The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. *Journal of Labor Economics 24*(3), 411–482.

Kaestner, R. (2009). Adolescent cognitive and non-cognitive correlates of adult health. Technical report, National Bureau of Economic Research.

Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement 20*(1), 141–151.

Kolata, G. (2007). A surprising secret to a long life: Stay in school. *New York Times January 3*(Health). http://www.nytimes.com/2007/01/03/health/03aging.html.

Kuh, D. and Y. Ben-Shlomo (1997). *A Lifecourse Approach to Adult Disease.* New York: Oxford University Press, Projected Date: 9710.

49

Lleras-Muney, A. (2005). The relationship between education and adult mortality in the United States. *Review of Economic Studies 72*(1), 189–221.

Marmot, M. (2010). *Fair society, healthy lives: the Marmot review; strategic review of health inequalities in England post-2010.* The Marmot Review.

McCormick, M. C. (2008, March-April). Issues in measuring child health. *Ambulatory Pediatrics 8*(2), 77–84.

Meara, E. R., S. Richards, and D. M. Cutler (2008, March-April). The gap gets bigger: Changes in mortality and life expectancy, by education, 1981-2000. *Health Affairs 27*(2), 350–360.

Micklewright, J. (1989). Choice at sixteen. *Economica 56*(221), 25–39.

Neyman, J. (1923). Statistical problems in agricultural experiments. *Journal of the Royal Statistical Society II (Supplement)*(2), 107–180.

Perri, T. J. (1984, January). Health status and schooling decisions of young men. *Economics of Education Review 3*(3), 207–213.

Pissarides, C. (1981). Staying-on at school in England and Wales. *Economica 48*(192), 345–363.

Quandt, R. E. (1972, June). A new approach to estimating switching regressions. *Journal of the American Statistical Association 67*(338), 306–310.

Roberts, B. W., P. Harms, J. L. Smith, D. Wood, and M. Webb (2006). Using multiple methods in personality psychology. In M. Eid and E. Diener (Eds.), *Handbook of Multimethod Measurement in Psychology*, pp. 321–335. Washington, D.C.: American Psychological Association.

Roberts, B. W., N. R. Kuncel, R. L. Shiner, A. Caspi, and L. R. Goldberg (2007, December). The power of personality: The comparative validity of personality traits, socioeconomic status, and cognitive ability for predicting important life outcomes. *Perspectives in Psychological Science 2*(4), 313–345.

Rothbart, M. (1981). Measurement of temperament in infancy. *Child Development 52*(2), 569–578.

Rothbart, M. (1989). Temperament and development.

Roy, A. (1951, June). Some thoughts on the distribution of earnings. *Oxford Economic Papers 3*(2), 135–146.

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

Rutter, M., J. Tizard, and K. Whitmore (1970). *Education, health and behaviour*. Longmans London.

Shakotko, R. A., L. N. Edwards, and M. Grossman (1982). An exploration of the dynamic relationship between health and cognitive development in adolescence. Technical report, National Bureau of Economic Research.

Spearman, C. (1904). "general intelligence," objectively determined and measured. *American Journal of Psychology 15*, 201–293.

Velicer, W. F. (1976). Determining the number of components from the matrix of partial correlations. *Psychometrika 41*(3), 321–327.

Whalley, L. J. and I. J. Deary (2001). Longitudinal cohort study of childhood IQ and survival up to age 76. *British Medical Journal 322*(7290), 819.

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

A   Identification of the correlated 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:

$$Y = A\theta + \epsilon \tag{6}$$

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 et al. (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 (6):

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

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:

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

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:

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