Web Appendix for The Non-Market Benefits of Education and Ability[*]

James J. Heckman
University of Chicago
and the American Bar Foundation

John Eric Humphries
Yale University

Gregory Veramendi
Arizona State University

February 5, 2018

1

Contents

A   Full Literature Review

A.1   Crime

Education may lower the probability that an individual commits crime. As highlighted in Lochner and Moretti (2004), education may raise the opportunity-cost of crime by raising the value of not being incarcerated. Education may also have a direct impact on individuals' risk preferences and ethical considerations of crime.

A number of papers use changes in compulsory schooling to look at the effects of education on crime. Lochner and Moretti (2004) estimate the effect of education on participation in criminal activity using changes in state compulsory schooling laws. The paper uses Census and FBI data and finds that additional schooling substantially reduces the probability of incarceration. The paper finds that graduating from high school leads to a 0.61-0.97 percent reduction in the probability of incarceration for whites and a 7.23-11.40 percent reduction for blacks. Machin, Marie, and Vujic (2010) similarly use changes in compulsory schooling law in England and Wales to get at the causal returns to schooling. Using a regression discontinuity design, they find that the increased education associated with policy changes resulted in a substantial drop in property crime. Meghir, Palme, and Schnabel (2012) use educational reforms in Sweden to estimate the impact of increased education on crime rates. The paper finds that the educational reform reduced criminal activity of individuals and that it also reduced criminal activity of their children as well. The paper is unique in showing that education may improve outcomes for parents and their children. They associate the gains for children with increased resources and better parenting. Also using Swedish data, Hjalmarsson, Holmlund, and Lindquist (2015) use changes in compulsory schooling laws to estimate the role of education in reducing criminal convictions and incarceration. The paper finds that an additional year of schooling reduces convictions by 6.7% and incarceration by 15.5% for men, but has no effects for women.

In other work, Buonanno and Leonida (2009) use panel data from 20 Italian regions

from 1980 to 1995 to look at the relationship between crime and education. Controlling for regional time trends and other covariates, the authors find that education leads to reduced crime that is not fully explained by increased employment or wages. Groot and van den Brink (2010) use data from the Netherlands Survey on Criminality and Law Enforcement, finding that there are positive returns to education. This study uses self-reported crime rather than incarcerations or arrests, suggesting that the reduction in arrests is not caused by lower probabilities that educated individuals are caught.

Other work has found a strong correlation between cognitive ability and crime. Frisell, Pawitan, and Långström (2012) find that adolescent cognitive ability is negatively correlated with being convicted of a violent crime even after controlling for socio-economic status. Beaver and Wright (2011) review the literature on the correlation between IQ and delinquent behavior and find a state-level correlation between average IQ scores and crime rates after controlling for the demographic composition of the state. Bartels, Ryan, Urban, and Glass (2010) similarly demonstrate a negative correlation between state-level IQ and various crime rates.

A.2 Mental Health

One possible benefit of education is improved mental health. This may operate through mechanisms such as education increasing income and prestige, or there may be direct mental health benefits, potentially through better understanding or lower costs of acquiring health information.

Cutler and Lleras-Muney (2006) provide evidence on the health benefits of education, including results on depression, anxiety, and if depression hindered the individuals life in the past month. When controlling for a broad set of controls, the authors find that a year of schooling lowers the probability that depression has hindered life in the past month by 0.6%. Chevalier and Feinstein (2006) use the National Childhood Development Study to consider the impact of education on mental health. The NCDS provides longitudinal data on

health, allowing the authors to better control for pre-existing mental illness or other mental health factors. Overall, the authors find that education improves mental health, though this is less true for advanced levels of education. They also find that the impacts are larger for women or individuals with previous mental health issues. The authors estimate a number of models that aim to control for pre-existing mental health. They also provide estimates that instrument years of schooling with (1) teacher's expectation of future schooling for the individual, and (2) proxies for the student's discount rate. They find that a post-secondary degree decreases the risk of adult depression by 5 to 7 percentage points. Across the board, their IV estimates are larger than OLS estimates. In more recent work, Lee (2011) uses Korean data to explore the education-depression gradient in Koreans over the age of 45. They find that a number of mediating factors, such as cognitive ability, wealth, social status, and health-related behaviors, explain all of the gradient. The authors interpret these as mediators that are changed by education, and do not rigorously consider if differences in cognitive ability may be pre-existing.

One concern with outcomes such as mental health is reverse causality. It may be that early mental health directly affects educational decisions and later mental health outcomes. For example, Fletcher (2008, 2010) estimates the role of adolescent mental health on educational attainment.

Hatch, Jones, Kuh, Hardy, Wadsworth, and Richards (2007) explore how childhood IQ is associated with mental health outcomes as an adult using the Medical Research Council National Survey of Health and Development. The paper finds that adolescent IQ is negatively associated with anxiety and depression, as recorded through the General Health Questionnaire, but were more prone to alcoholism.

A.3   Self-Esteem

Self-esteem is seen as an important life outcome, though little work exists studying the causal effect of education on self-esteem. More commonly, early measures of self-esteem are used

as predictors of educational success. Papers such as Heckman, Stixrud, and Urzúa (2006) use self-esteem as a measure of early non-cognitive skills, while papers such as Trzesniewski, Donnellan, Moffitt, Robins, Poulton, and Caspi (2006), Flouri (2006), and Ross and Broh (2000) use early measures of self-esteem directly. Bachman and O'Malley (1977) provide early evidence that education improves self-esteem and occupational attainment, but find that post-secondary educational attainment has little impact on self-esteem.

A.4   Trust

Trust is less studied than many of our other outcomes. A more trusting society may be valued, but it is not clear that it is always better for an individual to be more trusting. Less work exists evaluating the role of education on trust, but some papers provide evidence of a relationship between education and trust.

Helliwell and Putnam (1999) provide evidence that raising the average level of education in a region improves trust, but does not increase political participation. Glaeser, Laibson, Scheinkman, and Soutter (1999) use undergraduates at Harvard to study trust. They find that survey questions about trust may actually measure trustworthiness rather than trust. They find that students who are persistently more trusting do not report being more trusting in surveys. This suggests that our measure of self-reported trust may rather be a measure of trustworthiness. Alesina and La Ferrara (2002) use individual data across many localities to study the determinants of trust. They find that income and education are both positively correlated with trust. For example, individuals with less than 12 years of education were 32% less likely to report being trusting than individuals with more than 16 years of education.

A.5   Voting

Many believe that education is an essential component for functional democracy and several papers have explored the relationship between civic participation and education. Dee (2004) uses instrumental variable techniques to test the role of education on civic participation,

such as voting, believing in free speech, and reading newspapers. Using the High School and Beyond longitudinal study, he uses proximity of a junior or community college to instrument years of schooling and finds large LATE estimates for registering to vote and voted in the last twelve months. Using the General Social Survey, the paper also estimates the civic returns to schooling instrumenting with exposure to child labor laws as an instrument. The paper finds large LATE estimates for voting in the last presidential election, reading newspapers, and a number of outcomes related to believing in free speech. The author also provides results from OLS and probit models without instruments, and finds that they produce smaller estimates, but still finds that additional schooling is highly associated with increased civic participation. Milligan, Moretti, and Oreopoulos (2004) provide additional evidence that education increases political involvement in both the US and the UK, but that it only increases voting in the US. The paper estimates OLS models of participation, but also estimates IV models using cross-state variation in mandatory schooling laws as instruments. Beyond voting, the paper finds that extra schooling increases a number of other civic related activities, such as following campaigns on TV or newspapers, following public affairs, and being interested in elections.

In earlier work, Hauser (2000) explores the hypothesis that education is only a proxy for cognitive ability, and that previous correlation studies may falsely attribute increased education to increased political participation due to confounding correlations with ability. The author tests this hypothesis with three data sets in the US and finds that pre-existing cognitive ability does not explain the strong correlation between education and civic participation.

In more recent work, Sondheimer and Green (2010) use two experimental interventions and one natural experiment that affected educational attainment to evaluate the link between education and voter turnout. Using the Perry and STAR early-life intervention experiments, the authors find that the gains in education coming from the intervention notably increased voting. Using the "I Have A Dream" program—a program aimed at helping low SES students have increased access to education and a number of other services—the authors build control groups of students the year below or the year above the treatment group. Across all studies,

they find that the intervention increased educational attainment and voting, though it is not clear if the authors isolate the effect of education and not the larger effects of early-life interventions.

Denny and Doyle (2008) investigate the correlations between voter turnout in a British general election and cognition and personality. Using the National Child Development Study, the authors find that both cognitive ability and personality are both significantly correlated with voter turnout.

A.6 Welfare

Welfare is closely linked to the economic returns to education, but focuses more heavily on the wage and employment benefits for low SES individuals. Reducing the use of welfare through increased education has a direct cost-saving component. Oreopoulos and Salvanes (2011), as discussed above, provide evidence that education reduces welfare receipt. Using compulsory schooling laws, they find LATE estimates of -0.015 for being on welfare. Coelli, Green, and Warburton (2007) study the impact of high school graduation on welfare receipt of students who come from families on welfare in Canada. The paper uses linked administrative data and includes performance measures in school, such as GPA and test scores. The paper finds that graduating from high school greatly reduces the probability of being on welfare (a 50-70 percent reduction). The paper tries to estimate the causal benefit through a number of ways: they use test scores to control for ability, instrument high school graduation using principle fixed-effects, and account for unobserved latent heterogeneity, as proposed in Heckman and Singer (1986).

A.7 Additional Evidence From the Literature on the Causal Effects of Education on Health

The literature finds that health and education are positively correlated both at the individual level (for example Grossman, 1975) and at the country level (see Jayachandran and Lleras-

Muney, 2009 for an overview). More recently, the literature has focused on the causal relationship between education and health. While some papers have looked at how health affects educational attainment (Madsen, 2012), most papers focus on the causal impact of education on later life health outcomes and health-related behaviors.[1] Most of the literature uses natural experiments to evaluate the "causal" impact of additional education on health. Using instrumental variables or regression discontinuity, these papers require a valid instrument and typically need large samples. Many of these papers exploit changes in mandatory schooling requirements in the early 20th century. Lleras-Muney (2005) uses the US Census to construct synthetic panels, evaluates the impact of compulsory education laws in 1915 and 1939, and finds that increases in schooling lead to a significant decrease in mortality. Similarly, van Kippersluis, O'Donnell, and van Doorslaer (2011) find that changes in compulsory schooling laws in the Netherlands decreased mortality rates for elderly men. Kemptner, Jürges, and Reinhold (2011) consider compulsory school changes in Germany in 1949 and 1969 and find that additional schooling lowers the prevalence of long-term illness, and find weak evidence that it reduced obesity, but no evidence that it reduced smoking.

Not all of this literature finds a statistically significant role of education on health. Albouy and Lequien (2009) find no statistically significant evidence of education decreasing mortality using French schooling reforms. Arendt (2005) finds no statistically significant impact of education on self-reported health and BMI using Danish data.

The instrumental variables literature suggests that changes in compulsory schooling causally increased health, but this literature is limited in a number of ways. First, these natural experiments focus on changes in compulsory schooling levels that are well below current requirements, and the health gains may not extrapolate to current educational decisions, such as graduating from high school or college. Second, the natural experiments focus on changes predominantly at the beginning of the twentieth century—which may not extrapolate to current conditions. Third, these natural experiments may demonstrate causal

---

[1] Conti and Heckman (2010) model both the causal role of education on health and the role of early health on education and later health.

benefits, but they do not help us understand the causal pathways. Oreopoulos and Salvanes (2011) overview the potential ways schooling may lead to non-pecuniary benefits, such as health.[2] Cutler and Lleras-Muney (2010) review the literature on the causal benefits of education on health and similarly discuss the pathways though which education may act, and when possible, try to test these hypotheses. The authors summarize their results as finding that resources such as income or health insurance can explain 11-32% of the education-health behavior gradient, cognitive ability accounts for 30% (though they note that this is just their best guess), and social integration accounts for 11% of the gradient. Notably, Cutler and Lleras-Muney (2010) find that discount rate, risk aversion, value of the future, as well as personality factors, do not account for any of the education-health gradient. While this is a featured result of their paper, the proxies for these parameters used by the authors are quite weak. For example, discount rates were proxied using a question from the MIDUS survey, which asks how strongly the respondent agrees with the statement, "I live one day at a time and don't really think about the future," and the value of the future is measured with the respondent's answer to the question, "Looking ahead ten years into the future, what do you expect your life will overall be like at that time?" Similarly, measures of personality are restricted to measures of impulsiveness or lack of self-control. We believe that being limited to rough proxies of economic parameters and personality should temper the conclusions reached by the authors.

While not directly comparable to Cutler and Lleras-Muney (2010), Conti and Heckman (2010) find that non-cognitive ability plays an important role in health outcomes. It finds that the importance of cognitive ability is overstated when non-cognitive ability is not included, and that education has a causal impact on health. Prior to Conti and Heckman (2010), Auld and Sidhu (2005) attempt to disentangle the causal roles of cognition and education on health. The paper sets up an extended Grossman model where schooling, cognitive ability, and material inputs enter the production of health. The paper then estimates the model

---

[2]The paper also provides an overview of different approaches taken by the literature to establish causal relationships.

using the NLSY79 and an instrumental variables technique (with local unemployment rates and parent's education as instruments). The paper questions if health limits the amount of work an individual can perform as their later-life measure of health. Once accounting for cognition, the paper finds that there is little statistically significant evidence that education affects health outcomes. The paper finds that benefits are the largest for those with low educational attainment.

Many other papers develop frameworks for integrating health into economic models. The most broadly known is the Grossman model (see Grossman, 2000). Many extensions or adaption of the Grossman model have been considered (for example, Galama, 2011). Similarly, Cervellati and Sunde (2013) extend the Ben-Porath model to consider how increased life expectancy changes education and labor supply decisions. Ehrlich and Chuma (1990) and Ehrlich (2000) develop a model that introduces the uncertainty of mortality into a lifecycle framework. Ehrlich and Yin (2005) build on this work and calibrate a dynamic life-cycle model where mortality risk is endogenous. The authors find that a non-trivial portion of the observed educational differences in life expectancy can be explained by endogeneous choices on "life protection" (*i.e.* health spending).

For additional overviews of the associations between education and health activities, see Cawley and Ruhm (2012) for an overview of the causal impacts of education on risky behaviors, Chaloupka and Warner (2000) for smoking, and Cook and Moore (2000) for drinking.

B   Parameterization of the Model and the Likelihood

This section presents more details on how the model is parameterized and estimated.

B.1   Parameterization of the Model

Each educational decision D_j is modeled as a binary decision

$$D_j = \left\{ \begin{array}{ll} 0 & \text{if } I_j \geq 0, \quad j \in \mathcal{J} = \{G, 0, 1, 2\} \\ 1 & \text{otherwise,} \end{array} \right\} \quad j \in \{G, 0, \ldots, \bar{s} - 1\}, \tag{1}$$

where I_j is the *perceived* value at node j of going on to $j + 1$ for a person at node j.[3] We approximate I_j using a linear-in-the-parameters model:

$$I_j = \boldsymbol{Z}_j' \boldsymbol{\gamma}_j + \boldsymbol{X}_j' \boldsymbol{\beta}_j + \boldsymbol{\theta}' \boldsymbol{\alpha}_j - \nu_j, \ j \in \{0, \ldots, \bar{s} - 1\}, \tag{2}$$

where $\boldsymbol{Z}_j$ is a vector of instruments and $\boldsymbol{X}$ is a vector of variable observed by the economist that determine the schooling transition decision of the agent with schooling level j and $\boldsymbol{\theta}$ is a vector of unobserved (by the economist) endowments. This approximation is a starting point for a more general analysis of dynamic discrete choice models. Endowments $\boldsymbol{\theta}$ are not directly observed by the econometrician but are proxied by measurements. $\boldsymbol{\theta}$ plays an important role in our model. Along with the observed variables, it generates dependence among schooling choices and outcomes. ν_j represents an idiosyncratic error term assumed to be independent across agents and states: $\nu_j \perp\!\!\!\perp (\boldsymbol{X}_j, \boldsymbol{Z}_j, \boldsymbol{\theta})$, where "$\perp\!\!\!\perp$" denotes statistical independence.

Outcomes can be either continuous or binary and are estimated separately by final schooling level:

$$Y_s^k = \left\{ \begin{array}{ll} \tilde{Y}_s^k & \text{if } Y_s^k \text{ is continuous,} \\ \boldsymbol{1}(\tilde{Y}_s^k \geq 0) & \text{if } Y_s^k \text{ is a binary outcome,} \end{array} \right\} k \in \mathcal{K}, \ \ s \in \mathcal{S}. \tag{3}$$

---

[3]This formulation follows a well-established tradition in the literature. See Heckman (1981), Cameron and Heckman (1987, 2001), Eckstein and Wolpin (1989), Geweke and Keane (2001), and Arcidiacono and Ellickson (2011).

We approximate $\tilde{Y}_s^k$ by the linear-in-the-parameters model

$$\tilde{Y}_s^k = \boldsymbol{X}_s^k \boldsymbol{\beta}_s^k + \boldsymbol{\theta}' \boldsymbol{\alpha}_s^k + \omega_s^k, \tag{4}$$

where $\boldsymbol{X}_s^k$ is a vector of observed controls relevant for outcome k and $\boldsymbol{\theta}$ is the vector of unobserved endowments. ω_s^k represents an idiosyncratic error term that satisfies $\omega_s^k \perp\!\!\!\perp (\boldsymbol{X}_s^k, \boldsymbol{\theta})$.

B.2   Measurement System for Unobserved Endowments $\boldsymbol{\theta}$

Most of the literature estimating the causal effect of schooling develops strategies for eliminating the effect of $\boldsymbol{\theta}$ in producing spurious relationships between schooling and outcomes.[4] Our approach is different. We proxy $\boldsymbol{\theta}$ to identify the interpretable sources of omitted variable bias and to determine how the unobservables mediate the causal effects of education. We follow a recent literature documenting the importance of both cognitive and non-cognitive endowments in shaping schooling choices and mediating the effects of schooling on outcomes.

Given $\boldsymbol{\theta}$, and conditional on $\boldsymbol{X}$, all educational choices and outcomes are assumed to be statistically independent. If $\boldsymbol{\theta}$ were observed, we could condition on $(\boldsymbol{\theta}, \boldsymbol{X})$ and identify selection bias-free estimates of causal effects and model parameters. We do not directly measure $\boldsymbol{\theta}$ and instead, we proxy it and correct for the effects of measurement error on the proxy. We test the robustness of our approach by allowing for an additional unproxied unobservable that accounts for dependence between schooling and economic outcomes not captured by our proxies. These additional sources of dependence can be identified without proxy measurements under the conditions stated in Heckman and Navarro (2007).

Let θ^C and θ^{SE} denote the levels of cognitive and socio-emotional endowments and suppose $\boldsymbol{\theta} = (\theta^C, \theta^{SE})$. We allow θ^C and θ^{SE} to be correlated. Let $t_{m,s}^C$ be the mth cognitive test score and $t_{m,s}^{C,SE}$ the mth measure influenced by both cognitive and socio-emotional endowments, all measured at schooling level s. Parallel to the treatment of the index and outcome equations,

---

[4]See Heckman (2008).

we assume linear measurement systems for these variables:

$$t_{m,s}^C \quad = \quad \boldsymbol{X}_{m,s}^C \boldsymbol{\beta}_{m,s}^C + \theta^C \alpha_{m,s}^C + e_{m,s}^C, \tag{5}$$

$$t_{m,s}^{C,SE} \quad = \quad \boldsymbol{X}_{m,s}^{C,SE} \boldsymbol{\beta}_{m,s}^{C,SE} + \theta^C \tilde{\alpha}_{m,s}^C + \theta^{SE} \tilde{\alpha}_{m,s}^{SE} + e_{m,s}^{C,SE}. \tag{6}$$

The structure assumed in Equations (5) and (6) is identified, even when allowing for correlated factors, if we have one measure that is a determinant of cognitive endowments $(t_{m,s}^C)$ and at least four measures that load on both cognitive ability and socio-emotional ability, and conventional normalizations are assumed.[5] In the main text, we report results from models that use measurements to proxy $\boldsymbol{\theta}$. Let $H_{i,s}^m$ be an indicator for if an individual i took test t at schooling level s.

Specification of the measurement system   When estimating the factor model, we must make normalizations and exclusion restrictions. There is no "best" method for determining these restrictions. As presented below, we use a variety of measures joined with the theory for determining our measurement system.

Factors have no natural scale. To address this, we normalize one loading for each factor to unity. This normalization does not affect the relative loadings of the two factors, but rather determines the units in which the factors are measured. We normalize the measure that has the largest correlation with the other measures. In the case of our paper, we normalize the cognitive loading to one for the arithmetic reasoning ASVAB measure and we normalize the socio-emotional loading to one for the language arts grade measure. Switching the normalization to the loadings on other measures has no substantive effect on the results.

Following Heckman, Stixrud, and Urzúa (2006), the model imposes that the ASVAB measures do not load on socio-emotional factors. If any particular ASVAB score is excluded, it does not substantively change the analysis. Course grades are assumed to load on both

---

[5]See, e.g., the discussion in Williams (2011) and Anderson and Rubin (1956). One of the factor loadings for θ^C and θ^{SE} has to be normalized to set the scale of the factors. Non-parametric identification of the distribution of $\boldsymbol{\theta}$ is justified by an appeal to the results in Cunha, Heckman, and Schennach (2010).

the cognitive and socio-emotional factors. As discussed in the main paper, this assumption is supported by Duckworth and Seligman (2005) and Borghans, Golsteyn, Heckman, and Humphries (2011), who find that grades are largely determined by endowments other than cognitive ability.

As discussed above, the identification strategy used in the paper requires one measure that loads exclusively on cognitive ability. We assume ASVAB tests only measure cognition. Subject-specific 9th grade GPA, educational choices, and early risky behavior are assumed to depend on both factors.

We include violent behavior, smoking regularly by age 15, drinking regularly by age 15, ever smoking marijuana by age 15, and sexual intercourse by age 15, as early "outcomes" in our model. These do not inform the cognitive or socio-emotional factor, but provide a robustness check of our interpretation of our factors and aid in interpretation.

B.3   Likelihood

We estimate our model in two stages using maximum likelihood. The measurement system, the distribution of latent endowments, and the model of schooling decisions are estimated in the first stage. The outcome equations are estimated in the second stage using estimates from the first stage. We follow Hansen, Heckman, and Mullen (2004) and correct estimated factor distributions for the causal effect of choices on measurements by jointly estimating the choice and measurement equations in the first stage. The distribution of the latent factors is estimated only using data on educational choices and measurements. This allows us to interpret the factors as cognitive and socio-emotional endowments. It links our estimates to an emerging literature on the economics of personality and psychological traits, but the link is not strictly required if we seek only to control for selection in schooling choices and do not seek to identify the system of measurement equations presented in the text. We do not use the final outcome system to estimate the distribution of factors, thus avoiding tautologically strong predictions of outcomes from the system of estimated factors.

For convenience, we repeat the definitions from Section **??**. Let $\mathcal{J}$ denote the set of possible terminal states. Let $D_j \in \mathcal{D}$ be the set of possible transition decisions that can be taken by the individual over the decision horizon. Let $\mathcal{S}$ denote the finite and bounded set of stopping states with $S = s$ if the agent stops at $s \in \mathcal{S}$. Define $\bar{s}$ as the highest attainable element in $\mathcal{S}$. $Q_j = 1$ indicates that an agent *gets to* decision node j. $Q_j = 0$ if the person never gets there. The history of nodes visited by an agent can be described by the collection of the Q_j, such that $Q_j = 1$. To ensure consistent notation, we define $Q_0 := 1$. $\boldsymbol{Y}_i$, $\boldsymbol{D}_i$, and $\boldsymbol{M}_i$ are vectors of individual i's outcomes, educational decisions and measurements of endowments, respectively. $\boldsymbol{Z}$ is a vector of observed determinants of decisions, $\boldsymbol{X}$ is a vector of observed determinants of outcomes, and $\boldsymbol{\theta}$ is the vector of unobserved endowments. The $\boldsymbol{Z}$ can include all variables in $\boldsymbol{X}$. When instrumental variable methods are used to identify components of the model, it is assumed that there are some variables in $\boldsymbol{Z}$ not in $\boldsymbol{X}$.

Assuming independence across individuals (denoted by i), the likelihood is:

$$
\begin{aligned}
\mathcal{L} &= \prod_i f(\boldsymbol{Y}_i, \boldsymbol{D}_i, \boldsymbol{M}_i | \boldsymbol{X}_i, \boldsymbol{Z}_i) \\
&= \prod_i \int f(\boldsymbol{Y}_i | \boldsymbol{D}_i, \boldsymbol{X}_i, \boldsymbol{Z}_i, \boldsymbol{\theta}) f(\boldsymbol{D}_i, \boldsymbol{M}_i | \boldsymbol{X}_i, \boldsymbol{Z}_i, \boldsymbol{\theta}) f(\boldsymbol{\theta}) d\boldsymbol{\theta},
\end{aligned}
$$

where $f(\cdot)$ denotes a probability density function. The last step is justified from the assumptions in Heckman, Humphries, and Veramendi (2017).

For the first stage, the sample likelihood is

$$
\begin{aligned}
\mathcal{L}^1 &= \prod_i \int_{\bar{\theta} \in \Theta} f(\boldsymbol{D}_i, \boldsymbol{M}_i | \boldsymbol{X}_i, \boldsymbol{Z}_i, \boldsymbol{\theta} = \overline{\boldsymbol{\theta}}) f_\theta(\overline{\boldsymbol{\theta}}) \, d\overline{\boldsymbol{\theta}} \\
&= \prod_i \int_{\bar{\theta} \in \Theta} \left[\prod_{j \in \mathcal{S} \setminus \{\bar{s}\}} f(\boldsymbol{D}_{i,j} | \boldsymbol{Z}_{i,j}, \boldsymbol{X}_{i,j}, \boldsymbol{\theta} = \overline{\boldsymbol{\theta}}; \boldsymbol{\gamma}_j)^{Q_{i,j}} \right] \\
&\quad \times \left[\prod_{m=1}^{N_M} \prod_{s \in \mathcal{S}^M} f(\boldsymbol{M}_{i,m,s} | \boldsymbol{X}_{i,m,s}, \boldsymbol{\theta} = \overline{\boldsymbol{\theta}}; \boldsymbol{\gamma}_{m,s})^{H_{i,s}^m} \right] f_\theta(\overline{\boldsymbol{\theta}}; \boldsymbol{\gamma}_\theta) \, d\overline{\boldsymbol{\theta}},
\end{aligned}
$$

where we integrate over the distributions of the latent factors. H_s^m is an indicator for the

level of the choice variable at the time the measurement m is taken, and is equal to one if the individual had attained s at the time of the measurement, and zero otherwise. Let $\mathcal{S}^M$ denote the set of possible states at the time of the measurement. The goal of the first stage is to secure estimates of $\boldsymbol{\gamma}_j$, $\boldsymbol{\gamma}_{m,s}$, and $\boldsymbol{\gamma}_\theta$, where $\boldsymbol{\gamma}_j$, $\boldsymbol{\gamma}_{m,s}$, and $\boldsymbol{\gamma}_\theta$ are the parameters for the educational decision models, the measurement models and the factor distribution, respectively. We assume that the idiosyncratic shocks are mean zero normal variates.

We approximate the factor distribution using a mixture of normals.[6] We define the index ℓ for each mixture, where $f_\theta(\boldsymbol{\theta}; \boldsymbol{\gamma}_\theta) = \sum_\ell \rho_\ell f_\theta^\ell(\boldsymbol{\theta}; \boldsymbol{\gamma}_\theta^\ell)$. The weights for each mixture are ρ_ℓ and they must satisfy $\sum_\ell \rho_\ell = 1$. $f_\theta^\ell(\boldsymbol{\theta}; \boldsymbol{\gamma}_\theta^\ell)$ is the PDF for mixture ℓ. Since the mean of the overall factor distribution is not identified, we also require that $E[\boldsymbol{\theta}] = \mathbf{0}$, which places constraints on the mixture parameters $\boldsymbol{\gamma}_\theta^\ell$. The log-likelihood can be rewritten as

$$
\begin{aligned}
\log \mathcal{L}^1 &= \sum_i \log \int_{\overline{\theta} \in \Theta} \left[\prod_{j \in \mathcal{S} \backslash \overline{s}} f(\boldsymbol{D}_{i,j} | \boldsymbol{Z}_{i,j}, \boldsymbol{X}_{i,j}, \boldsymbol{\theta} = \overline{\boldsymbol{\theta}}; \boldsymbol{\gamma}_j)^{Q_{i,j}} \right] \\
&\quad \times \left[\prod_{m=1}^{N_M} \prod_{s \in \mathcal{S}^M} f(\boldsymbol{M}_{i,m,s} | \boldsymbol{X}_{i,m,s}, \boldsymbol{\theta} = \overline{\boldsymbol{\theta}}; \boldsymbol{\gamma}_{m,s})^{H_{i,s}^m} \right] \times \left[\sum_\ell \rho_\ell f_\theta^\ell(\overline{\boldsymbol{\theta}}; \boldsymbol{\gamma}_\theta^\ell) \right] d\overline{\boldsymbol{\theta}} \\
&= \sum_i \log \left\{ \sum_\ell \rho_\ell \int_{\overline{\theta} \in \Theta} \left[\prod_{j \in \mathcal{S} \backslash \overline{s}} f(\boldsymbol{D}_{i,j} | \boldsymbol{Z}_{i,j}, \boldsymbol{X}_{i,j}, \boldsymbol{\theta} = \overline{\boldsymbol{\theta}}; \boldsymbol{\gamma}_j)^{Q_{i,j}} \right] \right. \\
&\quad \left. \left[\prod_{m=1}^{N_M} \prod_{s \in \mathcal{S}^M} f(\boldsymbol{M}_{i,m,s} | \boldsymbol{X}_{i,m,s}, \boldsymbol{\theta} = \overline{\boldsymbol{\theta}}; \boldsymbol{\gamma}_{m,s})^{H_{i,s}^m} \right] f_\theta^\ell(\overline{\boldsymbol{\theta}}; \boldsymbol{\gamma}_\theta^\ell) d\overline{\boldsymbol{\theta}} \right\}.
\end{aligned}
$$

We use Gauss-Hermite quadrature to numerically evaluate the integral. Although there are a number of ways to numerically evaluate an integral, one advantage of Gaussian quadrature is that it gives analytical expressions for the integral. Analytical expressions for the gradient and hessian can then be calculated, which allows for the use of efficient second-order optimization routines. Since the models are very smooth, a second-order optimization strategy leads to faster convergence. Given that we are using a mixture of normals, $f_\theta^\ell(\boldsymbol{\theta}; \boldsymbol{\gamma}_\theta^\ell) = \phi(\boldsymbol{\theta}; \boldsymbol{\mu}_\theta^\ell, \boldsymbol{\sigma}_\theta^\ell)$

---

[6]Mixtures of normals can be used to identify the true density non-parametrically, where the number of mixtures can be increased based on the size of the sample. For a discussion of sieve estimators, see Chen (2007).

is a multivariate normal, where we assume, for now, that the components are independent. This assumption can easily be relaxed, but keeping it simplifies notation. The Gauss-Hermite quadrature rule is $\int f(v)e^{-v^2}dv = \sum_n \lambda_n f(v_n)$, where the weights λ_n and nodes v_n are defined by the quadrature rule depending on the number of points used (Judd, 1998).[7] Applying the Gauss-Hermite rule and making a change of variables $(\overline{\boldsymbol{\theta}} = \sqrt{2}\boldsymbol{\sigma}_\theta^\ell \circ \boldsymbol{v}_n + \boldsymbol{\mu}_\theta^\ell)$[8], we can rewrite the likelihood as

$$\log \mathcal{L}^1 = \sum_i \log \left\{ \sum_\ell \rho_\ell \sum_{n1} \lambda_{n1} \sum_{n2} \lambda_{n2} \left[\prod_{j \in \mathcal{S} \backslash \overline{s}} f(\boldsymbol{D}_{i,j} | \boldsymbol{Z}_{i,j}, \boldsymbol{X}_{i,j}, \boldsymbol{\theta} = \sqrt{2}\boldsymbol{\sigma}_\theta^\ell \circ \boldsymbol{v}_n + \boldsymbol{\mu}_\theta^\ell; \boldsymbol{\gamma}_j)^{Q_{i,j}} \right] \right.$$
$$\left. \times \left[\prod_{m=1}^{N_M} \prod_{s \in \mathcal{S}^M} f(\boldsymbol{M}_{i,m,s} | \boldsymbol{X}_{i,m,s}, \boldsymbol{\theta} = \sqrt{2}\boldsymbol{\sigma}_\theta^\ell \circ \boldsymbol{v}_n + \boldsymbol{\mu}_\theta^\ell; \boldsymbol{\gamma}_{m,s})^{H_{i,s}^m} \right] \right\},$$

where $\boldsymbol{v_n} = \begin{pmatrix} v_{n1}, & v_{n2} \end{pmatrix}$ represents the vector of nodes. Multivariate normal variables with correlated components can be rewritten as the sum of independent standard normal variables and then one can use the same procedure.

The goal of the first stage is then to maximize $\log \mathcal{L}^1$ and obtain estimates $\hat{\boldsymbol{\gamma}}_j$, $\hat{\boldsymbol{\gamma}}_{m,s}$, $\hat{\boldsymbol{\sigma}}_\theta^\ell$, $\hat{\boldsymbol{\mu}}_\theta^\ell$, and $\hat{\rho}_\ell$ for $j \in J^{MS}$. If a density $f(\cdot)$ cannot be calculated either because of missing data or because that model does not apply to individual i[9], then $f(\cdot) = 1$.

One can think of the inner brackets as the PDF of $\boldsymbol{\theta}$ for each individual i. This is useful in two respects. First, we can now predict the factor scores $(\hat{\boldsymbol{\theta}}_i)$ via maximum likelihood, where the likelihood for each individual i is

$$\mathcal{L}_i^\theta = \left[\prod_{j \in \mathcal{S} \backslash \overline{s}} f(\boldsymbol{D}_{i,j} | \boldsymbol{Z}_{i,j}, \boldsymbol{X}_{i,j}, \boldsymbol{\theta}_i; \hat{\boldsymbol{\gamma}}_j)^{Q_{i,j}} \right] \times \left[\prod_{m=1}^{N_M} \prod_{s \in \mathcal{S}^M} f(\boldsymbol{M}_{i,m,s} | \boldsymbol{X}_{i,m,s}, \boldsymbol{\theta}_i; \hat{\boldsymbol{\gamma}}_{m,s})^{H_{i,s}^m} \right].$$

Secondly, we can correct for measurement error in the outcome equations by integrating

---

[7]We use 16 quadrature points. Using 32 point did not substantively change any of our results.

[8]$\circ$ is the Hadamard, or entrywise, product.

[9]For example, the individual i is a high school dropout and the model corresponds to the graduate college decision.

over the PDF of the latent factor. The likelihood for the outcome equations is

$$\log \mathcal{L}_k^2 = \sum_i \log \left\{ \sum_\ell \rho_\ell \sum_{n1} \lambda_{n1} \sum_{n2} \lambda_{n2} \left[\prod_{j \in \mathcal{S} \setminus \bar{s}} f(\boldsymbol{D}_{i,j} | \boldsymbol{Z}_{i,j}, \boldsymbol{X}_{i,j}, \boldsymbol{\theta} = \sqrt{2} \hat{\boldsymbol{\sigma}}_\theta^\ell \circ \boldsymbol{v}_n + \hat{\boldsymbol{\mu}}_\theta^\ell; \hat{\boldsymbol{\gamma}}_j)^{Q_{i,j}} \right] \right.$$

$$\times \left[\prod_{m=1}^{N_M} \prod_{s \in \mathcal{S}^M} f(\boldsymbol{M}_{i,m,s} | \boldsymbol{X}_{i,m,s}, \boldsymbol{\theta} = \sqrt{2} \hat{\boldsymbol{\sigma}}_\theta^\ell \circ \boldsymbol{v}_n + \hat{\boldsymbol{\mu}}_\theta^\ell; \hat{\boldsymbol{\gamma}}_{m,s})^{H_{i,s}^m} \right]$$

$$\left. \times \left[\prod_{s \in \mathcal{S}} f(\boldsymbol{Y}_{i,s}^k | \boldsymbol{X}_{i,k,s}, \boldsymbol{\theta} = \sqrt{2} \hat{\boldsymbol{\sigma}}_\theta^\ell \circ \boldsymbol{v}_n + \hat{\boldsymbol{\mu}}_\theta^\ell; \boldsymbol{\gamma}_{s,k})^{H_{i,s}} \right] \right\},$$

where $H_{i,s}$ is an indicator for the highest level of schooling attained by individual i. The goal of the second stage is to maximize $\log \mathcal{L}_k^2$ and obtain estimates $\hat{\boldsymbol{\gamma}}_{s,k}$. Since outcomes $(\boldsymbol{Y}_s^k)$ are independent from the first stage outcomes conditional on $\boldsymbol{X}, \boldsymbol{\theta}$, and we impose no cross-equation restrictions, we obtain consistent estimates of the parameters for the adult outcomes. Standard errors and confidence intervals are calculated by estimating two hundred bootstrap samples for the combined stages.

C   Detailed Description of the Data

Depression

Depression is measured using the Center for Epidemiologic Studies Depression Scale (CES-D), which was administered once individuals were 40 years of age. The CES-D is one of the most common screening tests for helping an individual determine his or her depression quotient. The scale measures symptoms of depression, discriminates between clinically depressed individuals and others, and is highly correlated with other depression rating scales (see Radloff, 1977 and Devins, Orme, Costello, Binik, Frizzell, Stam, and Pullin, 1988 for details). We form the scale summing the scores from the items: (a) "I did not feel like eating; my appetite was poor," (b) "I had trouble keeping my mind on what I was doing," (c) "I felt depressed," (d) "I felt that everything I did was an effort," (e) "My sleep was restless," (f) "I felt sad," and (g) "I could not get going." For each item the potential answers are:

(i) "0 Rarely/None of the time/1 Day," (ii) "1 Some/A little of the time/1-2 Days," (iii) "2 Occasionally/Moderate amount of the time/3-4 Days," and (iv) "3 Most/All of the time/5-7 Days." We standardized the scores to have mean 0 and variance 1 in the overall population where higher scores imply less depression.

Self-Esteem

Individuals were administered the Rosenberg Self-Esteem Scale in 2006. Rosenberg's Self-Esteem Scale consists of 11 items which are answered on a 4-point scale (4 strongly agree, 3 agree, 2 disagree, 1 strongly disagree).[10] We form the scale summing the scores from the items, and standardizing the scores to have mean 0 and variance 1 in the overall population. See Rosenberg and Pearlin (1978) and Robins, Hendin, and Trzeniewski (2001) for details on the Rosenberg Self-Esteem Scale.

Incarceration

The NLSY does not include detailed questions on illegal activity or incarceration as an adult, but records if the residence the respondent lives in is a jail at the time of interview. Using these reports, we construct an indicator for if the individual resided in jail during any interview between 1990 and 2010. This variable will not capture criminal activity that does not involve incarceration and may miss short spells of incarceration.

Voting

The NLSY asks a single question about voting. In 2008, respondents are asked if they voted in the 2006 national elections. Individuals could respond, "I did not vote in the November 2006 elections," "I thought about voting in 2006, but didn't," "I usually vote, but didn't

---

[10]The items are (a) "I feel that I'm a person of worth, at least on equal basis with others," (b) "I feel that I have a number of good qualities," (c) "All in all, I am inclined to feel that I am a failure," (d) "I am able to do things as well as most other people," (e) "I feel I do not have much to be proud of," (f) "I take a positive attitude toward myself," (g) "On the whole, I am satisfied with myself," (h) "I wish I could have more respect for myself," (i) "I certainly feel useless at times," and (j) "At times I think I am no good at all."

in 2006," and "I am sure I voted." We construct a binary variable that is equal to 1 if the individual responded, "I am sure I voted," and zero otherwise.

Welfare Receipt

The NLSY constructs variables indicating if the respondent received any AFDC, food stamps, SSI, welfare, or other public assistance during the calendar year. We construct an indicator showing whether or not the respondent reports ever receiving welfare between 1996 and 2006.
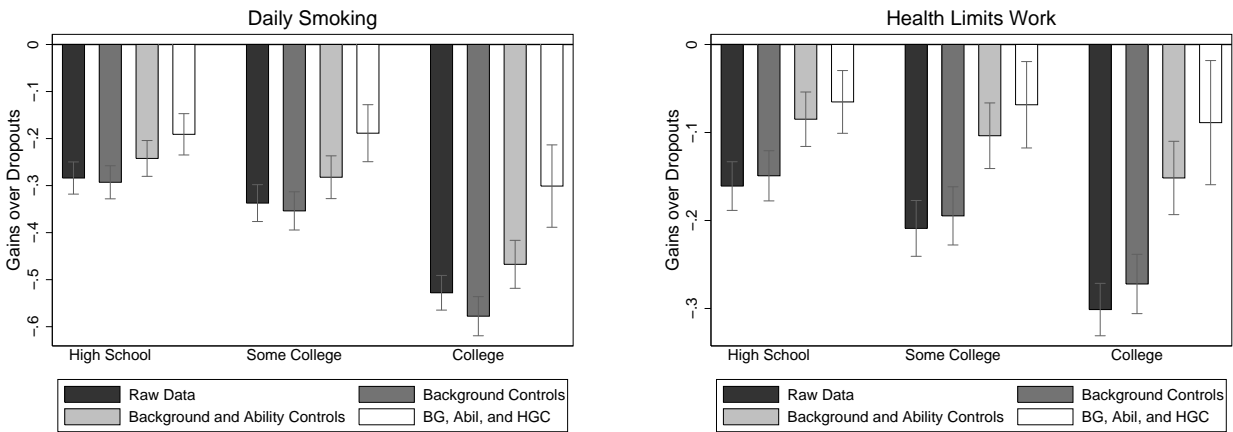
Trust

In 2008, the NLSY79 asks respondents, "Generally speaking, how often can you trust other people?" to which they can respond: (1) always, (2) most of the time, (3) about half the time, (4) once in a while, or (5) never. From this variable, we create an indicator that is equal to one if the individual reports trusting other people "always" or "most of the time".
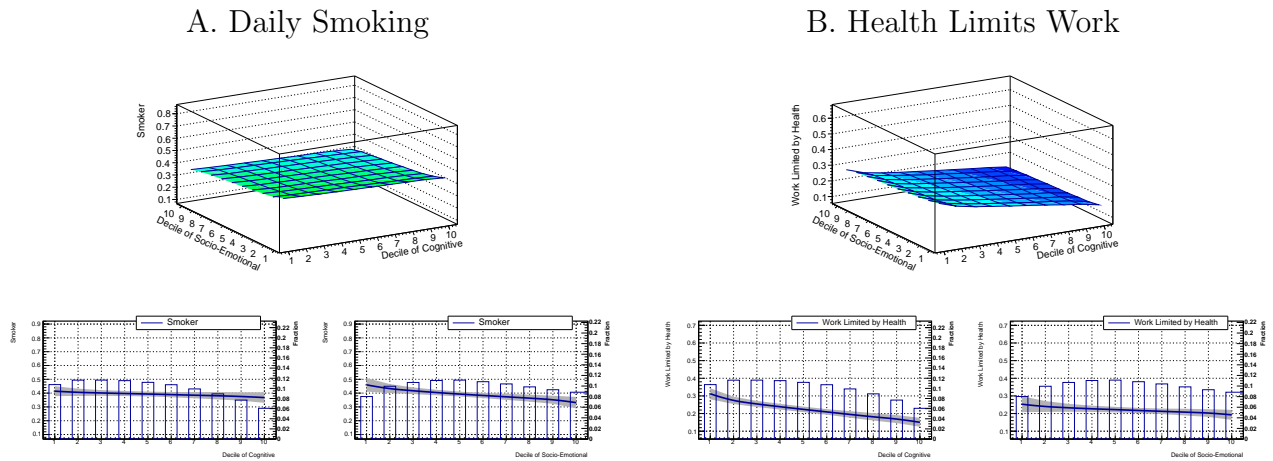
D   Results on Health

In this section, we repeat the results on health outcomes from Heckman, Humphries, and Veramendi (2017).

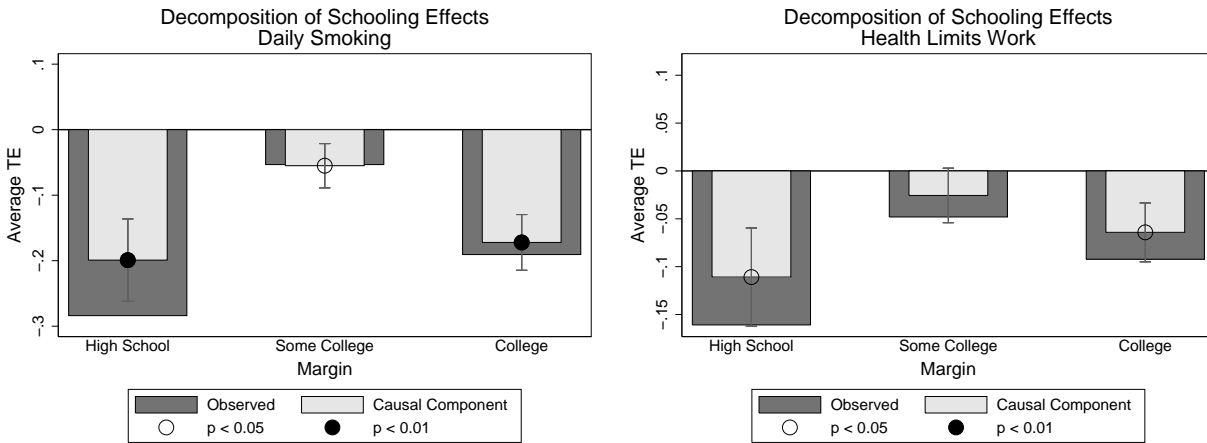## Figure 1: Raw and Adjusted Benefits from Education

Notes: The bars represent the coefficients from a regression of the designated outcome on dummy variables for educational attainment, where the omitted category is high school dropout. Regressions are run adding successive controls for background and proxies for ability. Background controls include race, region of residence in 1979, urban status in 1979, broken home status, number of siblings, mother's education, father's education, and family income in 1979. Proxies for ability are average score on the ASVAB tests and ninth grade GPA in core subjects (language, math, science, and social science). "Some College" includes anyone who enrolled in college, but did not receive a four-year college degree. The white bars additionally controls for highest grade completed (HGC). *Source:* NLSY79 data.

Figure 2: The Effect of Cognitive and Socio-Emotional Endowments

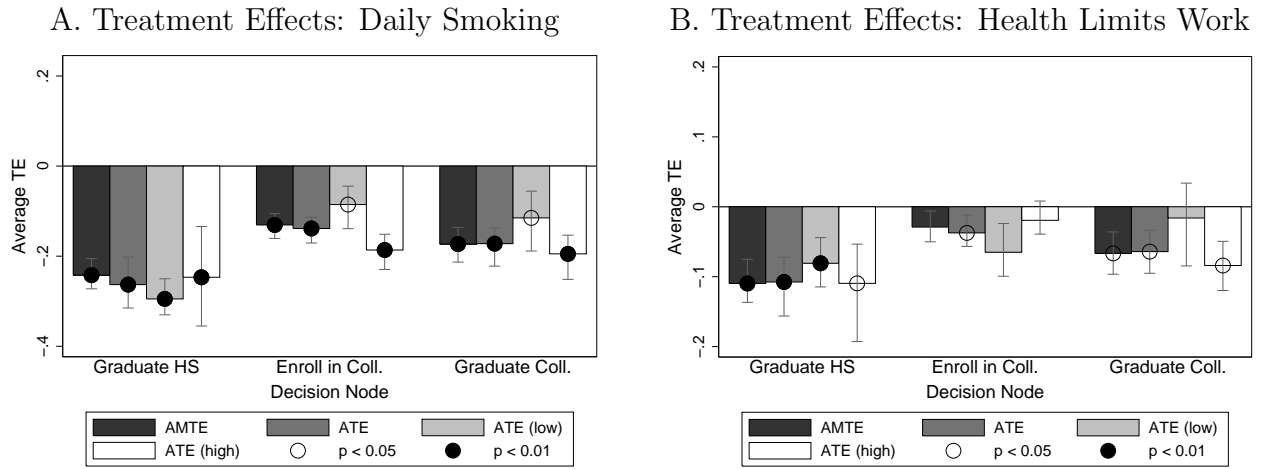### A. Daily Smoking   B. Health Limits Work

Notes: For each of the six outcomes, we present three figures that study the impact of cognitive and socio-emotional endowments. The top figure in each panel displays the levels of the outcome as a function of cognitive and socio-emotional endowments. In particular, we present the average level of outcomes for different deciles of cognitive and socio-emotional endowments. Notice that we define as "decile 1" the decile with the lowest values of endowments and "decile 10" as the decile with the highest levels of endowments. The bottom left figure displays the average levels of endowment across deciles of cognitive endowments. The bottom right figure mimics the structure of the left-hand side figure but now for the socio-emotional endowment.

Figure 3: Causal Versus Observed Differences by Final Schooling Level
(Compared to Next Lowest Level)

Notes: These figures report pairwise treatment effect for the indicated schooling nodes. Each bar compares the mean outcomes from a particular schooling level j and the next lowest level $j-1$ defined for the set of persons who complete schooling at $j-1$ or j. The "Observed" bar displays the observed differences in the data. The "Causal Component" bar displays the estimated average treatment effect to those who get treated (ATE) for the indicated group. The difference between the observed and causal treatment effect is attributed to the effect of selection and ability. Selection includes sorting on gains. The error bars and significance levels for the estimated ATE are calculated using 200 bootstrap samples. Error bars show one standard deviation and correspond to the 15.87th and 84.13th percentiles of the bootstrapped estimates, allowing for asymmetry. Significance at the 5% and 1% levels is shown by open and filled circles on the plots, respectively.
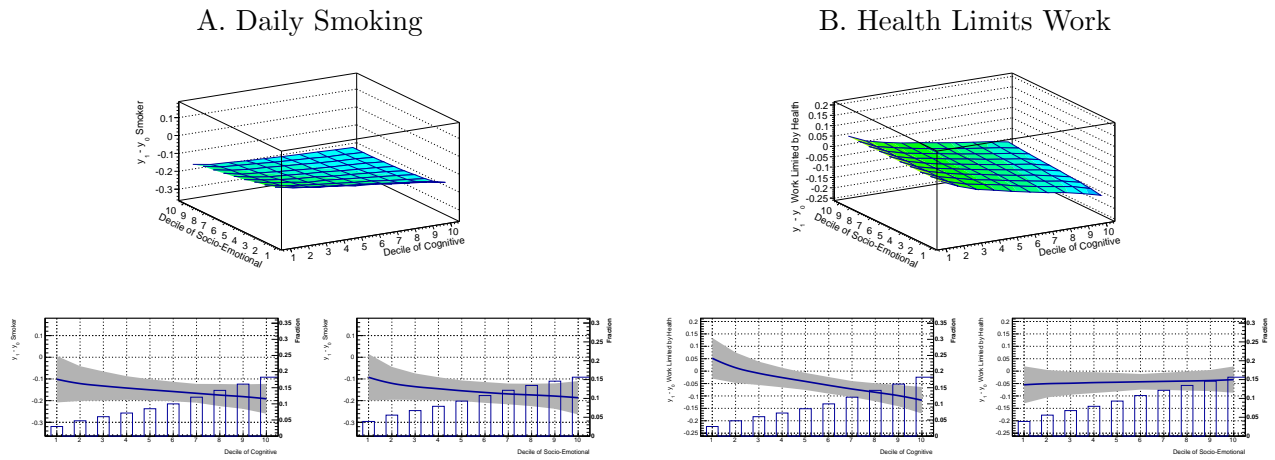
Figure 4: Treatment Effects of Outcomes by Decision Node

A. Treatment Effects: Daily Smoking

B. Treatment Effects: Health Limits Work

Sorting on Ability		
	Low Ability	High Ability
D_1: Dropping from HS vs. Graduating from HS	0.31	0.31
D_2: **HS Graduate** vs. College Enrollment	0.22	0.38
D_3: **Some College** vs. **Four-Year College Degree**	0.13	0.51

Notes: Each schooling level might provide the option to pursuing higher schooling levels. The error bars and significance levels for the estimated ATE are calculated using 200 bootstrap samples. Error bars show one standard deviation and correspond to the 15.87th and 84.13th percentiles of the bootstrapped estimates, allowing for asymmetry. Significance at the 5% and 1% level are shown by hollow and black circles on the plots respectively. The figure reports various treatment effects for those who reach the decision node, including the estimated ATE conditional on endowment levels. The high- (low-) ability group is defined as those individuals with cognitive and socio-emotional endowments above (below) the median in the overall population. These categories are not mutually exclusive, as some people may be high-ability in one dimension but low-ability in another. The table below the figure shows the proportion of individuals at each decision ($Q_j = 1$) that are high- and low-ability. The larger proportion of the individuals are high-ability and a smaller proportion are low-ability in later educational decisions. In this table, final schooling levels are highlighted using bold letters.

Figure 5: Average Treatment Effect of Graduating from a Four-Year College by Outcome

A. Daily Smoking

B. Health Limits Work

Notes: Each panel in this figure studies the average effects of graduating with a four-year college degree on the outcome of interest. The effect is defined as the differences in the outcome between those with a four-year college degree and those with some college. For each panel, let $Y_{some\ coll}$ and $Y_{four-year\ degree}$ denote the outcomes associated with attaining some college and graduating with a four-year degree, respectively. For each outcome, the first figure (top) presents $E(Y_{four-year\ degree} - Y_{some\ coll}|d^C, d^{SE})$ where d^C and d^{SE} denote the cognitive and socio-emotional deciles computed from the marginal distributions of cognitive and socio-emotional endowments. The second figure (bottom left) presents $E(Y_{four-year\ degree} - Y_{some\ coll}|d^C)$ so that the socio-emotional factor is integrated out. The bars in this figure display, for a given decile of cognitive endowment, the fraction of individuals visiting the node leading to the educational decision involving graduating from a four-year college. The last figure (bottom right) presents $E(Y_{four-year\ degree} - Y_{some\ coll}|d^{SE})$ and the fraction of individuals visiting the node leading to the educational decision involving graduating from a four-year college for a given decile of socio-emotional endowment.

References

ALBOUY, V., AND L. LEQUIEN (2009): "Does Compulsory Education Lower Mortality?," *Journal of Health Economics*, 28(1), 155–168.

ALESINA, A., AND E. LA FERRARA (2002): "Who Trusts Others?," *Journal of Public Economics*, 85(2), 207–234.

ANDERSON, T. W., AND H. RUBIN (1956): "Statistical Inference in Factor Analysis," in *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, ed. by J. Neyman, vol. 5, pp. 111–150, Berkeley, CA. University of California Press.

ARCIDIACONO, P., AND P. B. ELLICKSON (2011): "Practical Methods for Estimation of Dynamic Discrete Choice Models," *Annual Review of Economics*, 3(1), 363–394.

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.

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

BACHMAN, J. G., AND P. M. O'MALLEY (1977): "Self-Esteem in Young Men: A Longitudinal Analysis of the Impact of Educational and Occupational Attainment," *Journal of Personality and Social Psychology*, 35(6), 365–380.

BARTELS, J. M., J. J. RYAN, L. S. URBAN, AND L. A. GLASS (2010): "Correlations Between Estimates of State IQ and FBI Crime Statistics," *Personality and Individual Differences*, 48(5), 579–583.

BEAVER, K. M., AND J. P. WRIGHT (2011): "The Association Between County-Level IQ and County-Level Crime Rates," *Intelligence*, 39(1), 22–26.

BORGHANS, L., B. H. H. GOLSTEYN, J. J. HECKMAN, AND J. E. HUMPHRIES (2011): "Identification Problems in Personality Psychology," *Personality and Individual Differences*, 51(3: Special Issue on Personality and Economics), 315–320.

BUONANNO, P., AND L. LEONIDA (2009): "Non-Market Effects of Education on Crime: Evidence from Italian Regions," *Economics of Education Review*, 28(1), 11–17.

CAMERON, S. V., AND J. J. HECKMAN (1987): "Son of CTM: The DCPA Approach Based on Discrete and Factor Structure Models," Unpublished working paper, University of Chicago, Department of Economics.

——— (2001): "The Dynamics of Educational Attainment for Black, Hispanic, and White Males," *Journal of Political Economy*, 109(3), 455–499.

CAWLEY, J., AND C. J. RUHM (2012): "The Economics of Risky Health Behaviors," in *Handbook of Health Economics*, ed. by M. V. Pauly, T. G. Mcguire, and P. P. Barros, vol. 2 of *Handbooks in Economics*, chap. 3, pp. 95–199. Elsevier, Amsterdam.

CERVELLATI, M., AND U. SUNDE (2013): "Life Expectancy, Schooling, and Lifetime Labor Supply: Theory and Evidence Revisited," *Econometrica*, 81(5), 2055–2086.

CHALOUPKA, F. J., AND K. E. WARNER (2000): "The Economics of Smoking," in *Handbook of Health Economics*, ed. by A. J. Cuyler, and J. P. Newhouse, vol. 1B, chap. 29, pp. 1539–1627. Elsevier B. V., Amsterdam.

CHEN, X. (2007): "Large Sample Sieve Estimation of Semi-Nonparametric Models," in *Handbook of Econometrics*, ed. by J. J. Heckman, and E. E. Leamer, vol. 6B, chap. 76, pp. 4777–5752. Elsevier Science B. V., Amsterdam.

CHEVALIER, A., AND L. FEINSTEIN (2006): "Sheepskin or Prozac: The Causal Effect of Education on Mental Health," SSRN Scholarly Paper ID 923530, Social Science Research Network, Rochester, NY.

COELLI, M. B., D. A. GREEN, AND W. P. WARBURTON (2007): "Breaking the Cycle? The Effect of Education on Welfare Receipt among Children of Welfare Recipients," *Journal of Public Economics*, 91(7–8), 1369–1398.

CONTI, G., AND J. J. HECKMAN (2010): "Understanding the Early Origins of the Education-Health Gradient: A Framework That Can Also Be Applied to Analyze Gene-Environment Interactions," *Perspectives on Psychological Science*, 5(5), 585–605.

COOK, P. J., AND M. J. MOORE (2000): "Alcohol," in *Handbook of Health Economics*, ed. by A. J. Cuyler, and J. P. Newhouse, vol. 1B, chap. 30, pp. 1629–1673. Elsevier B. V., Amsterdam.

CUNHA, F., J. J. HECKMAN, AND S. M. SCHENNACH (2010): "Estimating the Technology of Cognitive and Noncognitive Skill Formation," *Econometrica*, 78(3), 883–931.

CUTLER, D. M., AND A. LLERAS-MUNEY (2006): "Education and Health: Evaluating Theories and Evidence," Working Paper 12352, National Bureau of Economic Research.

——— (2010): "Understanding Differences in Health Behaviors by Education," *Journal of Health Economics*, 29(1), 1–28.

DEE, T. S. (2004): "Are There Civic Returns to Education?," *Journal of Public Economics*, 88(9–10), 1697–1720.

DENNY, K., AND O. DOYLE (2008): "Political Interest, Cognitive Ability and Personality: Determinants of Voter Turnout in Britain," *British Journal of Political Science*, 38(02), 291–310.

DEVINS, G. M., C. M. ORME, C. G. COSTELLO, Y. M. BINIK, B. FRIZZELL, H. J. STAM, AND W. M. PULLIN (1988): "Measuring Depressive Symptoms in Illness Populations: Psychometric Properties of the Center for Epidemiologic Studies Depression (CES-D) Scale," *Psychology and Health*, 2(2), 139–156.

DUCKWORTH, A. L., AND M. E. P. SELIGMAN (2005): "Self-Discipline Outdoes IQ in Predicting Academic Performance of Adolescents," *Psychological Science*, 16(12), 939–944.

ECKSTEIN, Z., AND K. I. WOLPIN (1989): "The Specification and Estimation of Dynamic Stochastic Discrete Choice Models: A Survey," *Journal of Human Resources*, 24(4), 562–598.

EHRLICH, I. (2000): "Uncertain lifetime, life protection, and the value of life saving," *Journal of Health Economics*, 19(3), 341–367.

EHRLICH, I., AND H. CHUMA (1990): "A Model of the Demand for Longevity and the Value of Life Extension," *Journal of Political Economy*, 98(4), 761–782.

EHRLICH, I., AND Y. YIN (2005): "Explaining Diversities in Age-Specific Life Expectancies and Values of Life Saving: A Numerical Analysis," *Journal of Risk and Uncertainty*, 31(2), 129–162.

FLETCHER, J. M. (2008): "Adolescent Depression: Diagnosis, Treatment, and Educational Attainment," *Health Economics*, 17(11), 1215–1235.

——— (2010): "Adolescent Depression and Educational Attainment: Results Using Sibling Fixed Effects," *Health Economics*, 19(7), 855–871.

FLOURI, E. (2006): "Parental Interest in Children's Education, Children's Self-Esteem and Locus of Control, and Later Educational Attainment: Twenty-Six Year Follow-Up of the 1970 British Birth Cohort," *British Journal of Educational Psychology*, 76(1), 41–55.

FRISELL, T., Y. PAWITAN, AND N. LÅNGSTRÖM (2012): "Is the Association between General Cognitive Ability and Violent Crime Caused by Family-Level Confounders?," *PLoS ONE*, 7(7), e41783.

GALAMA, T. (2011): "A Contribution to Health Capital Theory," RAND Working Paper WR-831, RAND Corporation.

GEWEKE, J., AND M. KEANE (2001): "Computationally Intensive Methods for Integration in Econometrics," in *Handbook of Econometrics*, ed. by J. J. Heckman, and E. E. Leamer, vol. 5, chap. 56, pp. 3463–3568. Elsevier Science B. V., Amsterdam.

GLAESER, E. L., D. LAIBSON, J. A. SCHEINKMAN, AND C. L. SOUTTER (1999): "What Is Social Capital? The Determinants of Trust and Trustworthiness," Working Paper 7216, National Bureau of Economic Research.

GROOT, W., AND H. M. VAN DEN BRINK (2010): "The Effects of Education on Crime," *Applied Economics*, 42(3), 279–289.

GROSSMAN, M. (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, chap. 7, pp. 347–408. Elsevier Science B. V., Amsterdam.

HANSEN, K. T., J. J. HECKMAN, AND K. J. MULLEN (2004): "The Effect of Schooling and Ability on Achievement Test Scores," *Journal of Econometrics*, 121(1–2), 39–98.

HATCH, S. L., P. B. JONES, D. KUH, R. HARDY, M. E. J. WADSWORTH, AND M. RICHARDS (2007): "Childhood Cognitive Ability and Adult Mental Health in the British 1946 Birth Cohort," *Social Science & Medicine*, 64(11), 2285–2296.

HAUSER, S. M. (2000): "Education, Ability, and Civic Engagement in the Contemporary United States," *Social Science Research*, 29(4), 556–582.

HECKMAN, J. J. (1981): "The Empirical Content of Alternative Models of Labor Earnings," Unpublished manuscript, University of Chicago, Department of Economics.

——— (2008): "The Principles Underlying Evaluation Estimators with an Application to Matching," *Annales d'Economie et de Statistiques*, 91–92, 9–73.

HECKMAN, J. J., J. E. HUMPHRIES, AND G. VERAMENDI (2017): "Returns to Education: The Causal Effects of Education on Earnings, Health and Smoking," Forthcoming, *Journal of Political Economy*.

HECKMAN, J. J., AND S. NAVARRO (2007): "Dynamic Discrete Choice and Dynamic Treatment Effects," *Journal of Econometrics*, 136(2), 341–396.

HECKMAN, J. J., AND B. S. SINGER (1986): "Econometric Analysis of Longitudinal Data," in *Handbook of Econometrics*, ed. by Z. Griliches, and M. D. Intriligator, vol. 3, chap. 29, pp. 1690–1763. North-Holland, Amsterdam.

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

HELLIWELL, J. F., AND R. D. PUTNAM (1999): "Education and Social Capital," Working Paper 7121, National Bureau of Economic Research.

HJALMARSSON, R., H. HOLMLUND, AND M. J. LINDQUIST (2015): "The Effect of Education on Criminal Convictions and Incarceration: Causal Evidence from Micro-data," *Economic Journal*, 125(587), 1290–1326.

JAYACHANDRAN, S., AND A. LLERAS-MUNEY (2009): "Life Expectancy and Human Capital Investments: Evidence from Maternal Mortality Declines," *Quarterly Journal of Economics*, 124(1), 349–397.

JUDD, K. L. (1998): *Numerical Methods in Economics*. MIT Press, Cambridge, MA.

KEMPTNER, D., H. JÜRGES, AND S. REINHOLD (2011): "Changes in Compulsory Schooling

and the Causal Effect of Education on Health: Evidence from Germany," *Journal of Health Economics*, 30(2), 340–354.

LEE, J. (2011): "Pathways from Education to Depression," *Journal of Cross-Cultural Gerontology*, 26(2), 121–135.

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

LOCHNER, L. J., AND E. MORETTI (2004): "The Effect of Education on Crime: Evidence from Prison Inmates, Arrests, and Self-Reports," *American Economic Review*, 94(1), 155–189.

MACHIN, S., O. MARIE, AND S. VUJIC (2010): "The Crime Reducing Effect of Education," CEP Discussion Paper 979, Centre for Economic Performance.

MADSEN, J. (2012): "Health, Human Capital Formation, and Knowledge Production: Two Centuries of International Evidence," Working Paper 18461, National Bureau of Economic Research.

MEGHIR, C., M. PALME, AND M. SCHNABEL (2012): "The Effect of Education Policy on Crime: An Intergenerational Perspective," Working Paper 18145, National Bureau of Economic Research.

MILLIGAN, K., E. MORETTI, AND P. OREOPOULOS (2004): "Does Education Improve Citizenship? Evidence from the United States and the United Kingdom," *Journal of Public Economics*, 88(9–10), 1667–1695.

OREOPOULOS, P., AND K. G. SALVANES (2011): "Priceless: The Nonpecuniary Benefits of Schooling," *Journal of Economic Perspectives*, 25(1), 159–184.

RADLOFF, L. S. (1977): "The CES-D Scale: A Self-Report Depression Scale for Research in the General Population," *Applied Psychological Measurement*, 1(3), 385–401.

ROBINS, R. W., H. M. HENDIN, AND K. H. TRZENIEWSKI (2001): "Measuring Global Self-Esteem: Construct Validation of a Single-Item Measure and the Rosenberg Self-Esteem Scale," *Personality and Social Psychology Bulletin*, 27(2), 151–161.

ROSENBERG, M., AND L. I. PEARLIN (1978): "Social Class and Self-Esteem Among Children and Adults," *American Journal of Sociology*, 84(1), 53–77.

ROSS, C. E., AND B. A. BROH (2000): "The Roles of Self-Esteem and the Sense of Personal Control in the Academic Achievement Process," *Sociology of Education*, 73(4), 270–284.

SONDHEIMER, R. M., AND D. P. GREEN (2010): "Using Experiments to Estimate the Effects of Education on Voter Turnout," *American Journal of Political Science*, 54(1), 174–189.

TRZESNIEWSKI, K. H., M. B. DONNELLAN, T. E. MOFFITT, R. W. ROBINS, R. POULTON, AND A. CASPI (2006): "Low Self-Esteem During Adolescence Predicts Poor Health, Criminal Behavior, and Limited Economic Prospects During Adulthood," *Developmental Psychology*, 42(2), 381–390.

VAN KIPPERSLUIS, H., O. O'DONNELL, AND E. VAN DOORSLAER (2011): "Long-Run Returns to Education: Does Schooling Lead to an Extended Old Age?," *Journal of Human Resources*, 46(4), 695–721.

WILLIAMS, B. (2011): "Notes on Identification of Factor Models," Unpublished manuscript, University of Chicago, Department of Economics.