HCEO WORKING PAPER SERIES

The Non-Market Benefits of Education and Ability

James J. Heckman
John Eric Humphries
Gregory Veramendi

Working Paper

2017-072

09/2017

HUMAN CAPITAL AND
ECONOMIC OPPORTUNITY
GLOBAL WORKING GROUP

The University of Chicago
1126 E. 59th Street Box 107
Chicago IL 60637

www.hceconomics.org

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

September 20, 2017

*James J. Heckman: Department of Economics, University of Chicago, 1126 East 59th Street, Chicago, IL 60637; phone: 773-702-0634; fax: 773-702-8490; email: jjh@uchicago.edu. John Eric Humphries: Department of Economics, Yale University, 28 Hillhouse Ave, New Haven, CT 06511; phone: 773-980-6575; email: john.humphries@yale.edu. Gregory Veramendi: Arizona State University, 501 East Orange Street, CPCOM 412A, Tempe, AZ 85287-9801; phone: 480-965-0894; email: gregory.veramendi@asu.edu. This paper was presented at the Becker Friedman Institute conference in honor of Gary Becker, Friday, October 16, 2015. We thank the editor, Isaac Ehrlich, the associate editor who handled this paper, Edward Vytlacil, and two anonymous referees for valuable comments. This research was supported in part by: the American Bar Foundation; the Pritzker Children's Initiative; the Buffett Early Childhood Fund; NIH grants NICHD R37HD065072 and NICHD R01HD054702; an anonymous funder; Successful Pathways from School to Work, an initiative of the University of Chicago's Committee on Education funded by the Hymen Milgrom Supporting Organization; and the Human Capital and Economic Opportunity Global Working Group, an initiative of the Center for the Economics of Human Development, affiliated with the Becker Friedman Institute for Research in Economics, and funded by the Institute for New Economic Thinking. Humphries acknowledges the support of a National Science Foundation Graduate Research Fellowship. The views expressed in this paper are solely those of the authors and do not necessarily represent those of the funders or the official views of the National Science Foundation or the National Institutes of Health. The Web Appendix for this paper is located at https://cehd.uchicago.edu/appendix_non-market_benefits.

Abstract

This paper analyzes the non-market benefits of education and ability. Using a dynamic model of educational choice we estimate returns to education that account for selection bias and sorting on gains. We investigate a range of non-market outcomes including incarceration, mental health, voter participation, trust, and participation in welfare. We find distinct patterns of returns that depend on the levels of schooling and ability. Unlike the monetary benefits of education, the benefits to education for many non-market outcomes are greater for low-ability persons. College graduation decreases welfare use, lowers depression, and raises self-esteem more for less-able individuals.

Keywords: Education and Inequality, Returns to Education, Government Policy, Health and Inequality, Household Behavior and Family Economics
JEL codes: I24, I26, I28, I14, D1

James J. Heckman
Department of Economics
University of Chicago
1126 East 59th Street
Chicago, IL 60637
Phone: 773-702-0634
Email: jjh@uchicago.edu

John Eric Humphries
Department of Economics
Yale University
37 Hillhouse Ave
New Haven, CT 06511
Phone: 773-980-6575
Email: john.humphries@yale.edu

Gregory Veramendi
Department of Economics
Arizona State University
501 East Orange Street, CPCOM 412A
Tempe, AZ 85287-9801
Phone: 480-965-0894
Email: gregory.veramendi@asu.edu

Contents

1 Introduction

In his pioneering analysis, Gary Becker (1964) discussed both the market and non-market benefits of education. Nonetheless, most of the subsequent literature on the returns to education has focused on the returns to wages and income.[1] However, education is strongly correlated with a variety of non-market outcomes. For example, a positive correlation between education and health has been well documented, with evidence that a considerate portion of this correlation is causal.[2] Education is positively correlated with better mental health, higher levels of civic engagement, lower welfare use and a lower propensity to commit crime. Outcomes like civic engagement are valued in democratic societies despite having little direct economic impact. For other outcomes like reduced incarceration or receipt of welfare, there are direct cost-savings for the government as well as private benefits. Ignoring these outcomes could lead policymakers to greatly underestimate the benefits of supporting education.

There are many challenges in identifying and interpreting the causal effects of education. Observed returns are subject to both selection bias and sorting on gains.[3] In addition, schooling choices have a dynamic character. Each year of schooling opens up options for additional schooling.

This paper investigates the non-market benefits of abilities and education while addressing these challenges. We analyze a range of non-market outcomes including crime, mental health, civic engagement, self-esteem, trust, and participation in welfare. We consider a model where the returns to education vary across educational decisions and across individuals. Unlike the early literature, we do not restrict the returns to a year of schooling to be the same for each individual, or the same across levels of schooling for any individual, thereby allowing for rich

[1] See, e.g., the review in Heckman, Lochner, and Todd (2006).

[2] The positive correlation between schooling and health is a well-established finding (Grossman, 1972, 2000, 2006). See also Adams (2002); Arendt (2005); Lleras-Muney (2005); Silles (2009); Spasojevic (2003); Arkes (2003); Auld and Sidhu (2005); Grossman (2008); Grossman and Kaestner (1997); Cutler and Lleras-Muney (2010); Conti, Heckman, and Urzúa (2010) and HHV (2017).

[3] In the context of a Mincer model $ln\ Y_i = \alpha_i + \rho_i S_i$, where $ln\ Y_i$ is log earnings for person i, S_i is years of schooling for person i, α_i is an individual-specific intercept, ρ_i is the rate of return for person i, selection bias arises if $\text{Cov}(\alpha_i, S_i) \neq 0$, and sorting on gains arises if $\text{Cov}(\rho_i, S_i) \neq 0$.

variation in the returns to schooling. For example, estimates of the model show that the nonmarket returns to high school are very similar across persons of different ability, but the returns to college are, in general, higher for low-ability people. We address both selection bias and sorting on gains by allowing observed covariates as well as a vector of unobserved (but proxied) skill endowments to affect baseline outcomes as well as the gains from education.

Our model accounts for the dynamics of educational decisions. At any stage of the life cycle, educational choices shape the possibility sets for future choices. Rather than estimating only pairwise comparisons between two final schooling levels, we identify dynamic treatment effects, which estimate the returns from making different educational decisions, accounting for the fact that these decisions shape the possibilities for future educational choices. Dynamic treatment effects are policy relevant because they estimate the benefits at each educational level, which assess policies that affect educational participation.

We find that the nonmarket returns to education are larger for low-ability individuals compared to high ability persons. For example, we find that college graduation decreases welfare use, lowers depression, and raises self-esteem more for low-ability individuals than high-ability individuals. This evidence is in contrast with that of Heckman, Humphries, and Veramendi (2017) (henceforth HHV, 2017), who show that market returns to education are typically larger for high-ability people, particularly when making post-secondary decisions.

This paper builds on a large literature that studies the impact of education on non-pecuniary benefits and the impact of cognitive and non-cognitive skills on non-pecunary outcomes. See Lochner (2011) for a comprehensive review of the literature.[4] Vila (2000) presents another review of the literature on the non-wage benefits of education, including improvements in family planning, health benefits, outcomes for children, job-search, and savings. Oreopoulos and Salvanes (2011) present additional evidence on the non-market

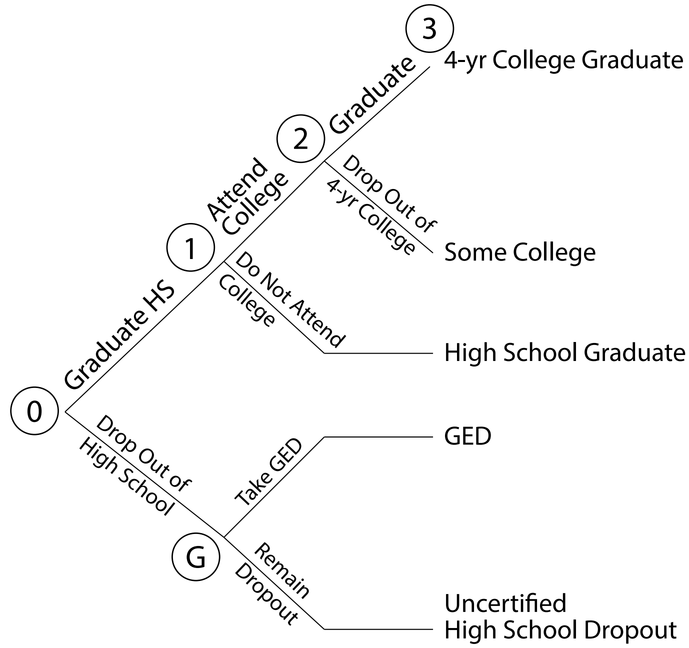
[4] A positive association between education and labor market outcomes has long been noted in the literature (Mincer, 1958, 1974). For surveys, see Card (1999) and Heckman, Lochner, and Todd (2006) and the references they cite. Lochner and Moretti (2004) review some of the early evidence on the non-market benefits of education. Ehrlich (1973) is a pioneering analysis of the economics of crime. His 1975 paper presents an early analysis of the correlation between education and crime.

benefits of education. They report that schooling improves happiness, leads to higher job-satisfaction and occupational prestige, less disability, improved health, decreased smoking, lowered probabilities of arrest, more schooling, leads to sorting with more highly educated spouses, lowers the probability of divorce, and improves trust. Almlund, Duckworth, Heckman, and Kautz (2011) review the literature on the importance of cognitive and non-cognitive skills. Section A of the Web Appendix presents a detailed review of the literature on each of the six non-market outcomes we consider in this paper.

2 Dynamic Discrete Choice Model of Education

This paper estimates the non-market benefits of education using the multistage sequential model of educational choice developed in HHV (2017) with transitions and decision nodes shown in Figure 1. We estimate returns to education at each node. People begin in high school and choose if they wish to graduate or not. Importantly, the set of future choices available to them depends on their earlier educational choices. If people choose to graduate from high school, they have the choice to enroll in college, and, if they enroll, they have the choice to graduate from college. If people choose to drop out of high school, they then have the option to take the GED to earn a high school equivalency certificate.

Figure 1: Unordered Dynamic Model



More generally, people are assumed to traverse a tree of educational decisions (Figure 1) where current choices affect future choice sets. We assume that at each choice node people make a binary decision. A person's final schooling level is $s \in \{G, 0, 1, 2, 3\}$, where G denotes earning a GED, 0 denotes high school dropout, 1 denotes high school graduate, 2 denotes some college, and 3 denotes college graduate. Individuals can only reach particular educational decisions (and particular final levels of schooling) if they have previously made specific choices. We define Q_j to be an indicator of whether an individual attains j and faces a choice between educational levels j and $j + 1$. Q_G indicates whether the person dropped out of high school and had the choice to get a GED or not.

Each person has a vector of characteristics $\boldsymbol{X}$ that affect educational decisions and outcomes at different schooling levels. These characteristics can include background variables such as parent's education or information on if one grew up in an urban environment. They also include contemporaneous variables such as the unemployment rate at the time decisions are taken. We also have access to instrument vector $\boldsymbol{Z}$, including the presence of a local college or local price of tuition, that affect educational decisions, but do not affect outcomes.[5] We assume $\boldsymbol{Z}$ and $\boldsymbol{X}$ are known by both the agent and the observing economist, but we allow agents to act on a multidimensional set of unobservables $\boldsymbol{\theta}$ that can affect educational decisions and outcomes, but are not directly observed by the economist. While $\boldsymbol{\theta}$ is not directly observed, it is proxied by measurements. These measurements render the random effects interpretable. In our previous research (HHV, 2017) we fit models with two unobserved factors which we estimate by a multivariate mixture of normals.[6] We test, and do not find evidence for, additional factors. The measurements of $\boldsymbol{\theta}$ are interpretable and correspond to cognitive and socio-emotional abilities.

Educational Decisions Educational decisions are modeled using latent variables crossing a threshold as described in Cameron and Heckman (2001). Educational decisions depend on functions of $\boldsymbol{X}$, $\boldsymbol{Z}$, and $\boldsymbol{\theta}$ as well as additive idiosyncratic shocks. For decision j the indices generating choices can be written as:

$$I_j = \boldsymbol{\phi}_j \underbrace{(\boldsymbol{X}, \boldsymbol{Z})}_{\substack{\text{Observed} \\ \text{by analyst}}} + \underbrace{\boldsymbol{\theta}'\boldsymbol{\alpha}_j - \nu_j}_{\substack{\text{Unbserved} \\ \text{by analyst}}}, \; j \in \{G, 0, 1, 2\}. \tag{1}$$

We do not impose rationality or rational expectations but Equation (1) is consistent with that. The unobserved components in $\boldsymbol{\theta}$ or observed components in $\boldsymbol{Z}$ need not necessarily affect all educational decisions in the same way.

Outcomes Outcomes $\boldsymbol{Y}$ are assumed to depend on a function of directly observed characteristics $\boldsymbol{X}$, the unobserved components $\boldsymbol{\theta}$, and an idiosyncratic shock. Importantly, we allow the intercepts, the observed characteristics $\boldsymbol{X}$, and the unobserved components $\boldsymbol{\theta}$ to affect outcomes differently depending on the final level of schooling. This can be thought of as a hedonic equation where all observed and unobserved (but proxied) variables are interacted

[5]The estimated model includes such instruments, but does not necessarily require the instruments for identification.

[6]See Section B in the Web Appendix for details on identification and estimation.

with schooling. The persistent (and known by the agent) proxied unobservable $\boldsymbol{\theta}$ enters the model in a fashion similar to a random effect, but the impact of the random effect can vary across schooling levels. For example, cognitive ability plays a more important role in explaining outcomes for those with a college degree compared to high school dropouts. The equation for outcome k for a person with final schooling level s is:

$$\tilde{Y}_s^k = \underbrace{\tau_s^k(\boldsymbol{X})}_{\substack{\text{Observed}\\ \text{by analyst}}} + \underbrace{\boldsymbol{\theta}'\boldsymbol{\alpha}_s^k + \omega_s^k}_{\substack{\text{Unobserved}\\ \text{by analyst}}}, \quad k \in \mathcal{K}, \quad s \in \{G, 0, 1, 2, 3\}, \tag{2}$$

where $\tilde{Y}_s^k$ is the outcome or a latent index for discrete outcomes.

The Measurement System The unobserved but proxied vector $\boldsymbol{\theta}$ plays an important role in our analysis. Along with the observed variables, they generate the dependence between choices and outcomes. Conditional on $\boldsymbol{X}$, $\boldsymbol{\theta}$ is assumed to capture the dependence between the unobserved components of the outcomes and the unobserved components of the educational decisions. If $\boldsymbol{\theta}$ were observed, we could condition on $(\boldsymbol{\theta}, \boldsymbol{X}, \boldsymbol{Z})$ and identify selection-bias-free estimates of a variety of causal effects. While $\boldsymbol{\theta}$ is not observed, it is proxied by N_M measures that facilitate identification of the distribution of $\boldsymbol{\theta}$. To accomplish this, we use a measurement system $\boldsymbol{M}$ consisting of early-life tests or outcomes which are thought to be generated by $\boldsymbol{\theta}$. Similar to the educational decisions and outcomes, we assume that they depend on observed characteristics $\boldsymbol{X}$, the proxied vector of factors $\boldsymbol{\theta}$, and idiosyncratic shocks. The $\boldsymbol{X}$ and $\boldsymbol{\theta}$ account for dependence across measures as well as the dependence among the measures, outcomes, and schooling decisions. Our measurement system includes subtests from the ASVAB achievement test, GPA in core subjects in 9th grade, and measures of whether persons committed risky or reckless behaviors when they were young. The system of measurement equations can be written as

$$\boldsymbol{M} = \begin{pmatrix} M_1 \\ \vdots \\ M_{N_M} \end{pmatrix} = \begin{pmatrix} \Phi_1(\boldsymbol{X}, \boldsymbol{\theta}, e_1) \\ \vdots \\ \Phi_M(\boldsymbol{X}, \boldsymbol{\theta}, e_{N_M}) \end{pmatrix}.$$

Using measurements, we can identify and interpret the estimated factors. We find that two factors are sufficient to explain our data (HHV, 2017). ASVAB scores, GPA, reckless or risky behavior, and educational decisions all determine the first factor. GPA, reckless behavior, and educational decisions determine the second factor. By excluding the second factor from the ASVAB tests, we are able to interpret the first factor as residual cognitive ability and the second factor as the residual ability that determines grades, educational decisions, and early behaviors that is not captured by cognitive ability—which we call socio-emotional ability.[7]

Identification The treatment effects developed in this paper can be identified using two different approaches. One approach relies on conditional independence. The other relies on instrumental variable exclusion restrictions. Heckman, Humphries, and Veramendi (2016) present a formal proof of model identification. Here we provide an intuitive explanation justifying each approach.

Conditioning on $\boldsymbol{\theta}$, $\boldsymbol{X}$, $\boldsymbol{Z}$, outcomes and choices are statistically independent. If we, in fact, could observe the unobserved factors $\boldsymbol{\theta}$, this would be equivalent to the assumptions made in matching. Since $\boldsymbol{\theta}$ is not observed, we treat it as a random effect. Unlike the usual random effect approach, we proxy $\boldsymbol{\theta}$ using a set of measurements described in Section 2. Under our assumptions, we can non-parametrically identify the distribution of $\boldsymbol{\theta}$. Effectively, we match on $\boldsymbol{X}$, $\boldsymbol{Z}$, and proxies for $\boldsymbol{\theta}$, accounting for measurement error in the proxies. This is a version of matching on mis-measured variables where, because of access to multiple measures of the latent true variable, we can correct for measurement error. As shown in Heckman, Humphries, and Veramendi (2016), the model can also be identified using instrumental variables. Different combinations of assumptions about IV and proxy matching also identify our model.

[7] See Williams (2017).

The estimates presented in this paper are based on parametric assumptions to facilitate the computation. The precise parametrization and the likelihood function for the model we estimate are presented in Web Appendix B.[8]

2.1 Defining Treatment Effects

This section discusses the various returns to education considered in this paper. Specifically we focus on two types of returns: (1) differences across final schooling levels, and (2) dynamic treatment effects that capture the benefits of choices, inclusive of continuation values associated with future educational decisions.[9]

Differences Across Final Schooling Levels The traditional approach in the literature on returns to education is to define them as the gains from choosing between a base and a terminal schooling level. Let $Y^k_{s'}$ be outcome k at schooling level s' and Y^k_s be outcome k at schooling level s. Conditioning on $\boldsymbol{X} = \boldsymbol{x}$, $\boldsymbol{Z} = \boldsymbol{z}$, and $\boldsymbol{\theta} = \overline{\boldsymbol{\theta}}$, the average treatment effect of s compared to s' is $E(Y^k_s - Y^k_{s'} | \boldsymbol{X} = \boldsymbol{x}, \boldsymbol{Z} = \boldsymbol{z}, \boldsymbol{\theta} = \overline{\boldsymbol{\theta}})$. Thus, we fix $S = s$ and condition on $\boldsymbol{X} = \boldsymbol{x}$, $\boldsymbol{Z} = \boldsymbol{z}$, and $\boldsymbol{\theta} = \bar{\boldsymbol{\theta}}$.[10] ATE ($ATE^k_{s,s'}$) is calculated by integrating over $\boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{\theta}$ for the subset of the population that completes one of the two final schooling levels ($S \in \{s, s'\}$), although in principle we could condition on one or the other element in this set or on other conditioning variables. This type of conditioning recognizes that the characteristics of people not making either final choice could be far away from the population making one of those choices, and hence, might not have any empirical or policy relevance. One can also compute parameters for other subpopulations.

Dynamic Treatment Effects Dynamic treatment effects take into account the direct effect of transiting to the next node in a decision tree, plus the benefits associated with the

[8] See the Web Appendix for HHV (2017) for results on how well this model fits the data.

[9] See HHV (2017) for formal definitions of these treatment effects.

[10] Fixing versus conditioning is the key notion in causal models since Haavelmo (1943). For a recent discussion, see Heckman and Pinto (2015).

options opened up by the additional choices made possible by such transitions. So, instead of fixing final schooling levels, dynamic treatment effects fix a decision ($Q_{j+1} = 1$ versus $Q_{j+1} = 0$) conditioning on the set of persons who visit the decision node ($Q_j = 1$). Those who go on to the next node ($Q_{j+1} = 1$) may choose to go on even further. So the outcome for fixing $Q_{j+1} = 1$ will include the probabilities and returns of future decisions. The mean gain (ATE_j^k) of fixing decision j is calculated by integrating over $\boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{\theta}$ for the subset of the population that reaches node j. Proceeding in a similar fashion, we can define conventional treatment effects, e.g., ATE, TT, TUT, AMTE for different subpopulations.[11]

To clarify the distinctions between the two types of treatment effects, consider a simplified example where there are five levels of schooling: high school dropout, GED recipient, high school graduate, some college, and college graduate ($s \in \{0, G, 1, 2, 3\}$) as shown in Figure 1. Consider outcome Y, where:

$$Y_i = \beta_{i,0} + \beta_{i,G}Q_{i,G} + \beta_{i,1}Q_{i,1} + \beta_{i,2}Q_{i,2} + \beta_{i,3}Q_{i,3} + U_i, \tag{3}$$

where $Q_{i,j} = 1$ indicates that the individual i made it to decision node j. $Q_{i,G} = 1$ if a person drops out of high school, but then chooses to earn a GED, $Q_{i,1} = 1$ if a person graduates from high school, $Q_{i,2} = 1$ is equal to one if a person enrolls in college, and $Q_{i,3} = 1$ is equal to one if a person graduates from college. For this simplified model, assume that U_i is an independent (across states) shock.[12] The $\beta_{i,j}$ are the marginal benefits of being at state j. Note that the β have i subscripts as we allow them to vary by both the individual's observable and unobservable characteristics.

For this model, the mean treatment effect associated with differences across adjacent final

[11] These are the average treatment effect (ATE), treatment on the treated (TT), treatment on the untreated (TUT) and the average marginal treatment effect (AMTE). AMTE is the treatment effect for those who are indifferent between treatment and non-treatment. See HHV (2017).

[12] With the factor structure used in this paper, U_i can be correlated with educational choices and outcomes.

schooling levels (s, s') are given by:

$$ATE_{s,s'} = E\left[\beta_{s'} | S \in \{s, s'\}\right]$$

where we have conditioned the expectation for populations at one of two final schooling levels s and s' and assume that the individual is deciding between staying at s or continuing their schooling to s'.[13]

In order to construct dynamic treatment effects, we need to account for the decision rule for how educational choices are made. Let $\mathbf{1}(Q_j = 1)$ be an indicator that denotes whether or not person i makes it to node j. The individual level dynamic treatment effect of graduating from high school is then:

$$T_{i,1} = \beta_{i,1} + \mathbf{1}(Q_{i,2} = 1)\beta_{i,2} + \mathbf{1}(Q_{i,2} = 1)\mathbf{1}(Q_{i,3} = 1)\beta_{i,3} - \mathbf{1}(Q_{i,G} = 1)\beta_{i,G}, \tag{4}$$

which is the direct benefit of graduating ($\beta_{i,1}$), plus the continuation value associated with the fact that the person may then choose to enroll in college or enroll in and graduate from college. [14] Given that agents who drop out can earn a GED, we must also subtract the continuation value associated with getting a GED from the treatment effect. Thus, decisions affect future choice sets (such as creating the possibility of college enrollment) while closing down other future decisions (like earning a GED).[15]

Note that even if we assumed that the direct gains are identical across individuals (i.e., $\beta_{i,j} = \beta_j, \forall\ i$), if decisions to continue on differ across people, then the dynamic returns to schooling will be also be heterogeneous. The other three dynamic treatment effects for

[13]Treatment effects without this conditioning can also be constructed.

[14]We abstract from discounting. The $\beta_{i,j}$ coefficients can be interpreted as incorporating the discounted value.

[15]Note that these treatment effects are similar to those in Heckman, Urzúa, and Vytlacil (2008), where in a multinomial model they consider the treatment effect of choosing between two choice sets. For an individual could choose between $A = \{\text{Dropout}, \text{GED}\}$ or $B = \{\text{HS Grad}, \text{Some College}, \text{College Grad}\}$. This would be similar to the treatment effects of graduating from high school or not in our model. The key difference is that here we assume that agents only know the idiosyncratic shock associated with their current choice, allowing us to break down the returns into direct effects and continuation values.

individual i are:

$$T_{i,2} = \beta_{i,2} + \mathbf{1}(Q_{i,3} = 1)\beta_{i,3} \tag{5}$$

$$T_{i,3} = \beta_{i,3}, \tag{6}$$

$$T_{i,G} = \beta_{i,G}, \tag{7}$$

where the returns to graduating college and the returns to earning a GED are equal to the direct returns in our example as they are final levels of education. To construct the population average dynamic treatment effects, we take the expectation of these individual returns over the full population:[16]

$$ATE_j^* = E\left[T_{i,j}\right]. \tag{8}$$

Equation (8) is the average treatment effect for the entire population. This likely entails using out-of-sample estimates far from the samples of people making specific choices such as the earnings of dropouts as college graduates. The full population ATE is less empirically and policy relevant than ATE for those who reach a particular decision node. In this paper we report:

$$ATE_j = E\left[T_{i,j} | Q_{i,j} = 1\right], \tag{9}$$

which is to say we estimate the average dynamic treatment effect for those that reach particular decision j.[17] Web Appendix B gives more detail on the specification of the model.

[16] This entails taking the expectation over the full population (and thus integrates over the full support of $\boldsymbol{X}$, $\boldsymbol{Z}$, $\boldsymbol{\theta}$, and $\boldsymbol{U}$).

[17] In particular, this can be thought as integrating over a particular support of $\boldsymbol{X}$, $\boldsymbol{Z}$, $\boldsymbol{\theta}$, and $\boldsymbol{U}$ that results in $Q_j = 1$.

3 The Data

This paper considers six non-market outcomes constructed from the National Longitudinal Survey of Youth 1979. These six outcomes are depression, self-esteem, incarceration, voting, welfare receipt, and trust. Table 1 provides a brief description of how each of the six outcomes are constructed. Section C of the Web Appendix provides details on the construction of each outcome.[18]

Table 1: Description of Outcomes

Outcome	**Description**
Depression	CES-D depression scale at age 40.
Self-Esteem	Rosenberg Scale administered in 2006.
Incarceration	Individual was incarcerated at time of interview between 1990 and 2010.
Voting	Individual reported voting in the 2006 election.
Welfare Receipt	Individual was ever on welfare between 1996 and 2006.
Trust	Individual reported “usually” or “always” trusting people in 2008.

Notes: All outcomes are from the National Longitudinal Survey of Youth 1979. See Section D of the Web Appendix for additional health outcomes from HHV (2017). See Section C of the Web Appendix for additional details on the construction of the outcomes.

Before discussing estimates from our model, it is informative to present adjusted and unadjusted associations between the outcomes and schooling. Figure 2 presents estimated regression relationships between different levels of schooling (relative to high school dropouts) and the six outcomes analyzed in this paper.[19] The black bars in each panel show the unadjusted mean differences in outcomes for persons at the indicated levels of educational attainment compared to those for high school dropouts.

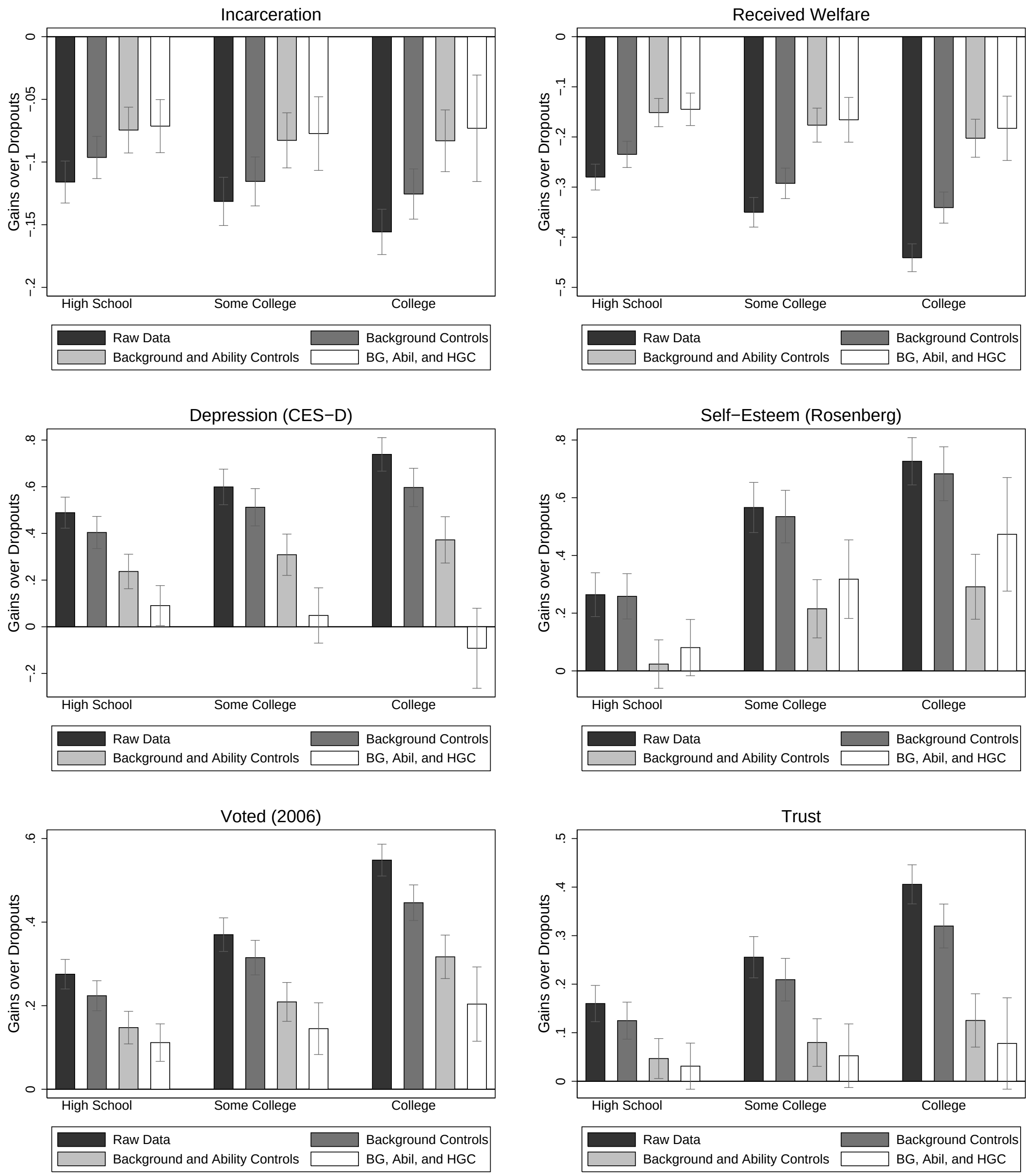
Higher ability is associated with better outcomes and more schooling. However, as shown

[18]In addition, the Web Appendix contains additional information and results on smoking at age 30 and if health ever limits work, two health outcomes considered in HHV (2017).

[19]Adjustments are made through linear regression.

by the grey bars in Figure 2, adjusting for family background and adolescent measures of ability attenuates, but does not eliminate, the estimated effects of education.

Figure 2: 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 9th 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. *Source:* NLSY79 data.

4 Results

4.1 The Effects of Endowments on Education and Outcomes

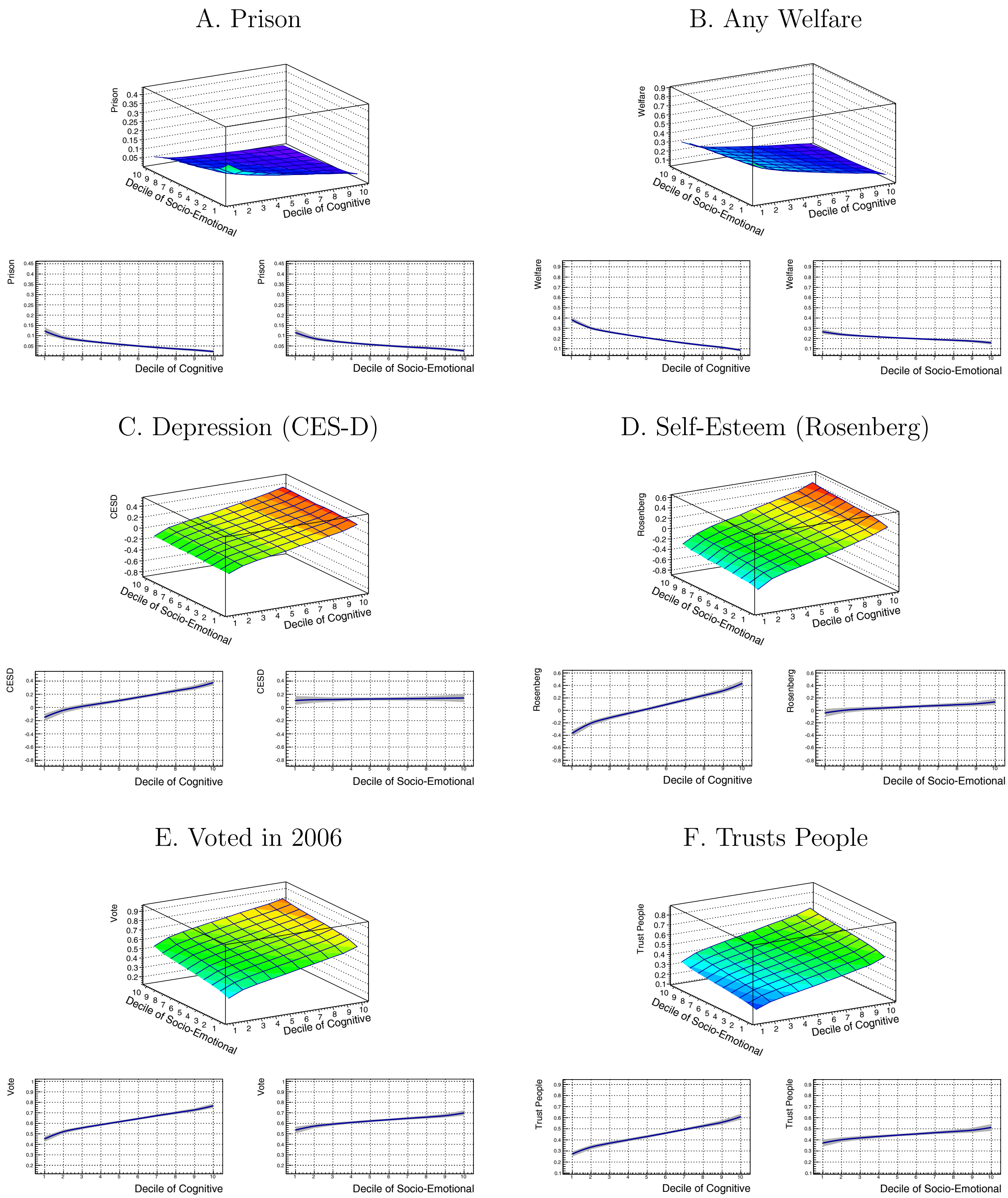
We first demonstrate the relationships between the latent endowment $\boldsymbol{\theta}$ and non-market outcomes and educational choices. Figure 3 plots the effects of the latent endowments on incarceration, use of welfare, depression, self-esteem, electoral participation, and trusting others. In the Web Appendix, we additionally reproduce results on health limits work and smoking reported in HHV (2017). Figure 3 shows the expected outcome based on deciles of the cognitive and socio-emotional endowment. The two lower figures in each panel show the marginal impact of cognitive (left) and socio-emotional (right) endowments. The cognitive endowment affects all six outcomes, while the effect of the socio-emotional endowment is statistically significant for the prison, welfare, voting, and trust outcomes.[20]

Moving persons from the lowest decile to the highest decile in both cognitive and socio-emotional ability lowers the probability of being incarcerated by 20.5%, the probability of receiving welfare by 40%, the depression score by 0.55 standard deviations, and the self-esteem score by 0.97 standard deviations. It increases the probability that they voted in the 2006 election by 46%, and the probability that they trust others by 46%.[21]

[20]While the option to get a GED is included in our model, we omit the treatment effects in our results for clarity and because we find that the GED has few causal benefits.

[21]Someone in the highest decile in both cognitive and socio-emotional ability has the following probabilities for each of the following outcomes: 0.9% to be incarcerated, 6.1% to use welfare, 83% to vote in the 2006 election, and 67% to trust others.

Figure 3: The Effect of Cognitive and Socio-Emotional Endowments



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. It marginalizes over $\boldsymbol{X}, \boldsymbol{Z}$. 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, marginalizing over the non-cognitive endowment as well as $\boldsymbol{X}, \boldsymbol{Z}$. The bottom right figure mimics the structure of the left-hand side figure but now for the socio-emotional endowment, marginalizing over the cognitive endowment, as well as $\boldsymbol{X}, \boldsymbol{Z}$.

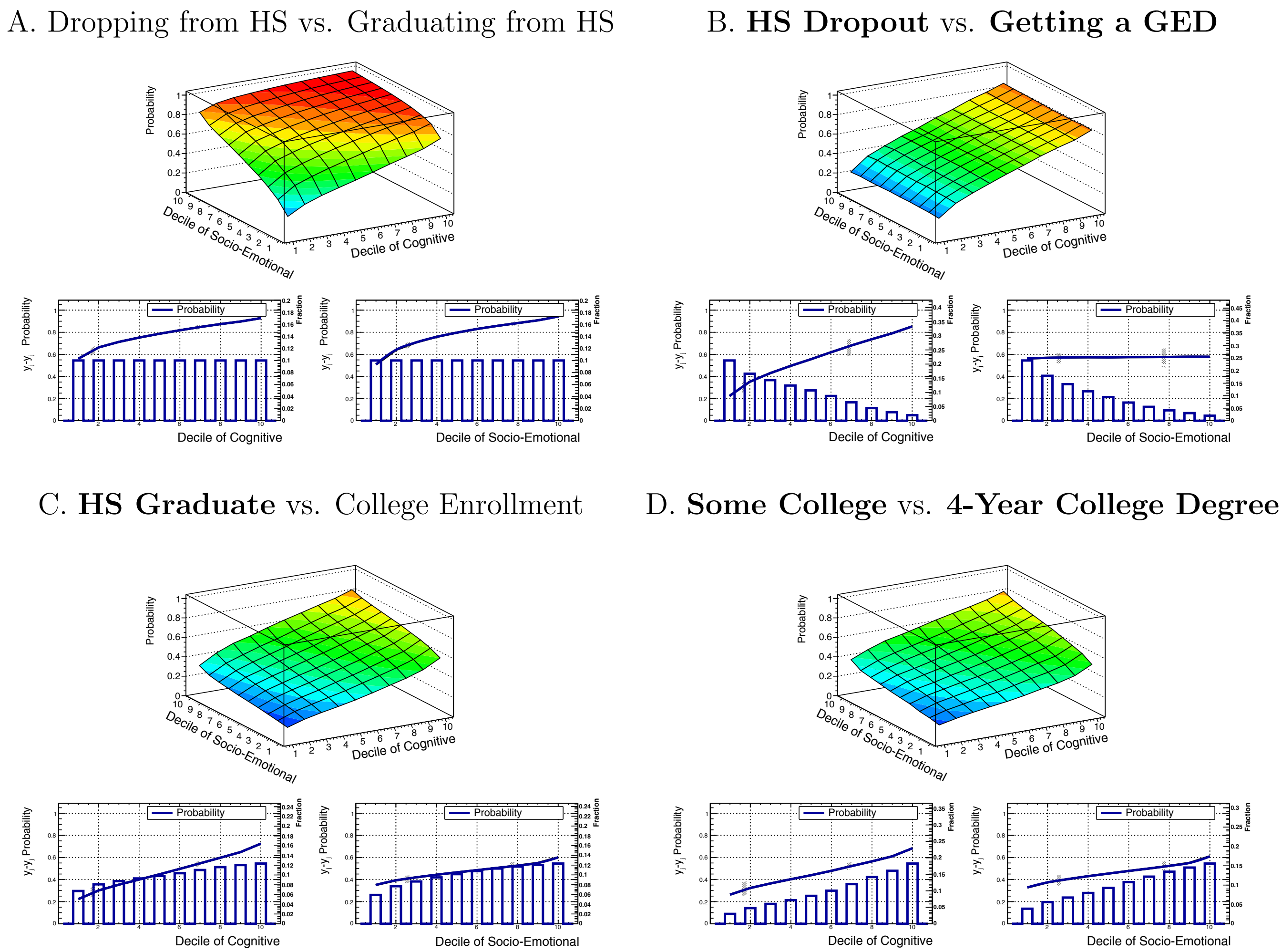
Figure 3 shows the total effect of ability on outcomes. A substantial portion of its impact comes through educational decisions. Figure 4 presents the probabilities of making the indicated educational choice at various levels of agent latent endowments. The top figure shows the probability of making the transition by decile of both cognitive and socio-emotional endowments, marginalizing over $\boldsymbol{X}, \boldsymbol{Z}$. The figures below the graphs of the joint distribution display the marginal effects on outcomes for cognitive and socio-emotional endowments respectively, marginalizing over $\boldsymbol{X}, \boldsymbol{Z}$, and the other endowment.[22] The estimates reveal clear evidence of sorting into education by both cognitive and socio-emotional endowments. For example, moving someone from the lowest decile to the highest decile in both cognitive and socio-emotional ability increases the probability of graduating high school by 75.7%, increases the probability of enrolling in college by 68%, and increases the probability of graduating from college by 63%.[23]

At the same time, these endowments have important impacts on adult outcomes. Together, these results imply strong selection biases in observed differences in outcomes by education level. This highlights the importance of accounting for observed and latent traits when estimating the causal impact of education.

[22]The bars in the subfigures show how the population reaching the decision fits into the deciles of the skill distribution for the whole population. Since everyone reaches the high school graduation decision, these are flat at 0.10 per decile. We see that the GED decision has few high endowment individuals, while we sort towards high endowment individuals as we work through the college enrollment and graduation decisions.

[23]Someone in the highest decile in both cognitive and socio-emotional ability has the following probabilities for each of the following outcomes: 99.1% to graduate high school, 83% to enroll in college, 81% to graduate with a four-year college degree.

Figure 4: The Probability of Educational Decisions, by Endowment Levels (Final Schooling Levels Are Highlighted Using Bold Letters)



Notes: For each of the four educational choices, we present three figures that present the probability of that specific educational choice. Final schooling levels do not allow for further options. For each pair of schooling levels j and $j+1$, the first subfigure (top) presents $Prob(Q_{j+1} = 1|d^C, d^{SE}, Q_j = 1)$, where d^C and d^{SE} denote the cognitive and socio-emotional deciles computed from the marginal distributions of cognitive and socio-emotional endowments. $Prob(Q_{j+1} = 1|d^C, d^{SE}, Q_j = 1)$ is computed for those who reach the decision node involving a decision between levels j and $j+1$. It marginalizes over $\boldsymbol{X}, \boldsymbol{Z}$ conditional on $Q_j = 1$. The bottom left subfigures present $Prob(Q_{j+1} = 1|d^C, Q_j = 1)$, where the socio-emotional factor and $\boldsymbol{X}, \boldsymbol{Z}$ are integrated out. The bars in these figures display, for a given decile of cognitive endowment, the fraction of individuals visiting the node leading to the educational decision involving levels j and $j+1$. The bottom right subfigures present $Prob(Q_{j+1} = 1|d^{SE}, Q_j = 1)$ for a given decile of socio-emotional endowment, as well as the fraction of individuals visiting the node leading to the educational decision involving levels j and $j+1$, marginalizing on $\boldsymbol{X}, \boldsymbol{Z}$, and the other endowment.

4.2 Estimated Causal Effects

We next present estimates of the main treatment effects for our model. Since our model is nonlinear and multidimensional, we report interpretable functions derived from it. We randomly draw sets of regressors from our sample and a vector of factors from the estimated factor distribution to simulate the treatment effects.[24] We first present the traditional treatment effects across final schooling levels. We then consider how dynamic treatment effects vary across decisions and endowment levels.

4.2.1 The Estimated Causal Effects of Final Educational Levels

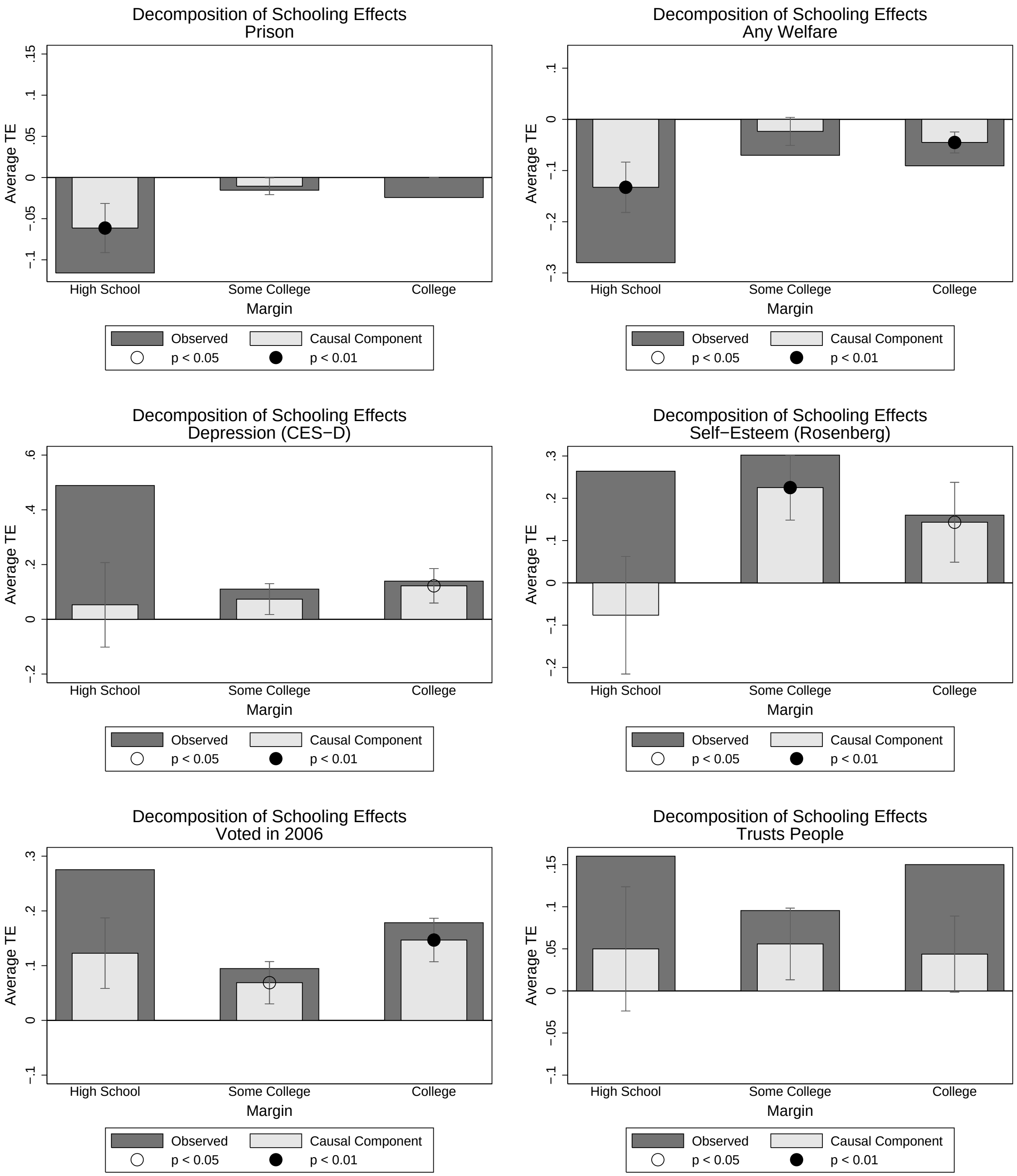
For certain outcomes almost none of the observed difference is causal. For example, for self-esteem and depression, we find that the observed differences between high school graduates and dropouts are large, but the estimated average treatment effect is almost zero. We first compare the outcomes from final schooling level s with those from $s-1$. The estimated treatment effects of education on the six outcomes are shown in Figure 5.[25] For each outcome, the bars labeled "Observed" display the unadjusted raw differences in the data. The bars labeled "Causal Component" display the average treatment effect obtained from comparing the outcomes associated with a particular schooling level s relative to $s-1$. These are defined for individuals at s or $s-1$.[26] There are substantial causal effects on incarceration, welfare use, self-esteem, and voting. At the same time, the causal average treatment effect is only one third to two thirds of the observed difference.

[24] We randomly draw an individual and use their full set of regressors.

[25] These are calculated by simulating the mean outcomes for the designated state and comparing it with the mean-simulated outcome for the state directly below it for the subpopulation of persons who are in either of the states.

[26] The circles and black circles indicate that the ATE is statistically significant at the 0.05 and 0.01 level, respectively.

Figure 5: Causal Versus Observed Differences by Final Schooling Level (Compared to Next Lowest Level)



Notes: Each bar compares the outcomes from a particular schooling level j and the next lowest level $j-1$. The "Observed" bar displays the observed differences in the data. The "Causal Component" bar displays the estimated average treatment effects (ATE). The difference between the observed and causal treatment effect is attributed to the effect of selection and ability. The error bars and significance levels for the estimated ATE are calculated using 200 bootstrap samples. Error bars correspond to the 16th and 84th 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.

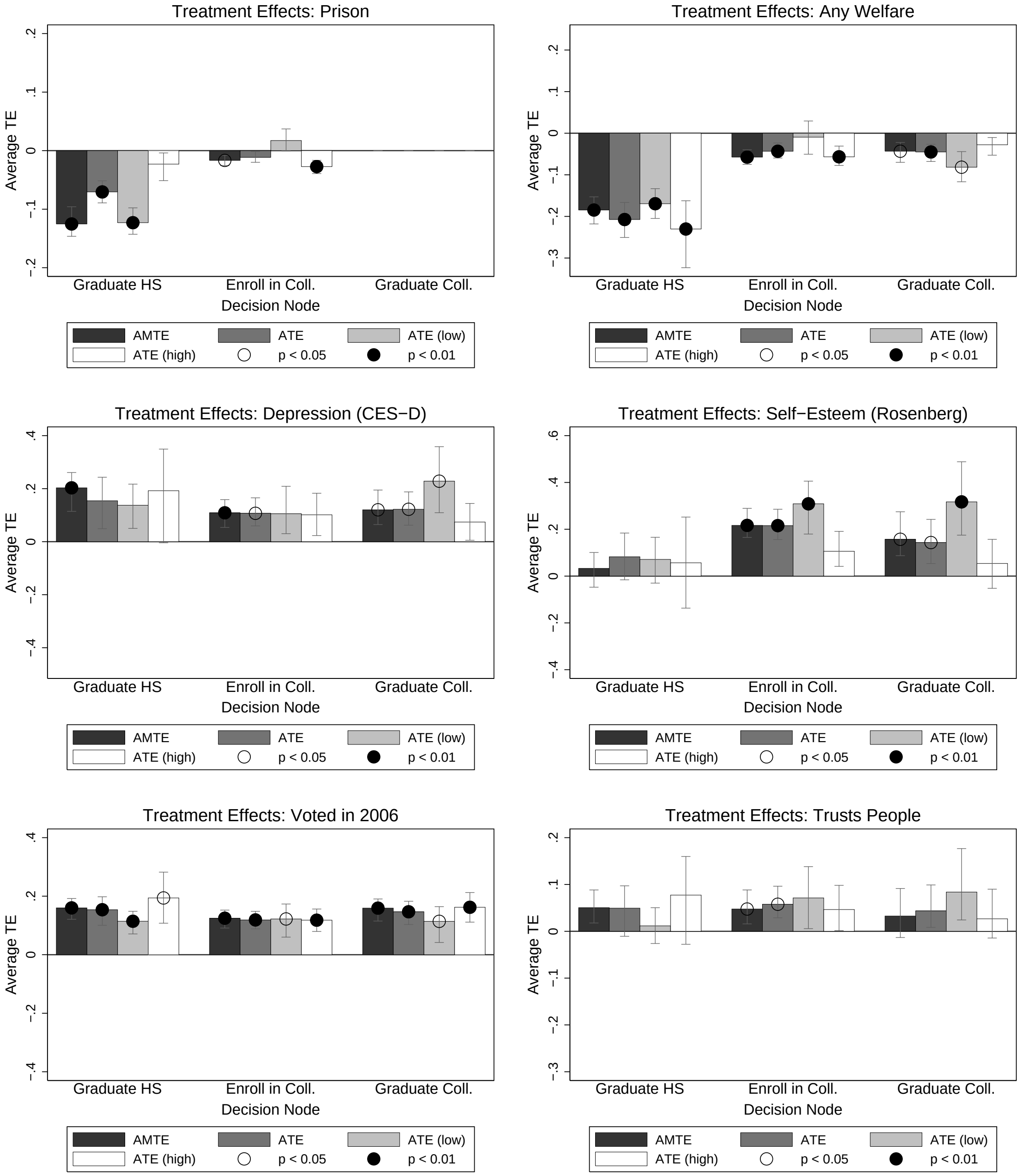
4.2.2 Dynamic Treatment Effects

We next report treatment effects by decision node (see Figure 6). We compute the gains to achieving (and possibly exceeding) the designated level of schooling (including continuation values) and compare them to the outcomes associated with not achieving that level. The Average Marginal Treatment Effect, AMTE, is the average treatment effect for those at or near the margin of indifference for that decision node.[27] AMTE is introduced in Carneiro, Heckman, and Vytlacil (2011) and formally developed in the context of our model in HHV (2017).

Each box of Figure 6 presents this array of educational treatment effects for each education level for the outcomes studied. The effects are presented as the height of different bars in each figure. They are defined as the differences in the outcomes associated with being at the designated level, compared to the one preceding it (not necessarily final or terminal schooling levels). ATE is calculated for the population that reaches the decision node. At each node j, the treatment effect is $E(Y^k|Fix\ Q_{j+1} = 1, Q_j = 1) - E(Y^k|Fix\ Q_{j+1} = 0, Q_j = 1)$ for those who reach node j. ATE (high) and ATE (low) are the ATEs for different ability groups. The high- (low-) ability group is defined for individuals with both cognitive and socio-emotional endowment above (below) the median of the full population. The whiskers show standard errors while the hollow and black circles indicate statistical significance at the 0.05 and 0.01 levels, respectively. The table below the figure displays the fraction of individuals at each educational choice who are in the high- or low-ability group.

[27] We define the margin of indifference to be $|| I_j/\sigma_j || \leq .01$, where σ_j is the standard deviation of I_j.

Figure 6: Treatment Effects of Outcomes by Decision Node



Sorting on Ability		
	Low Ability	High Ability
$j = 1$: Dropping from HS vs. Graduating from HS	0.31	0.31
$j = 2$: **HS Graduate** vs. College Enrollment	0.22	0.38
$j = 3$: **Some College** vs. **Four-Year College Degree**	0.13	0.51

Notes: Each schooling level might provide the option to pursuing higher schooling levels. Only final schooling levels do not provide an option value. The error bars and significance levels for the estimated ATE are calculated using 200 bootstrap samples. Error bars correspond to the 16th and 84th 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. The affect of graduating from college on prison was not estimated as very few individuals who enroll or graduate from college are ever incarcerated. The table below the figure shows the proportion of individuals at each decision node 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.

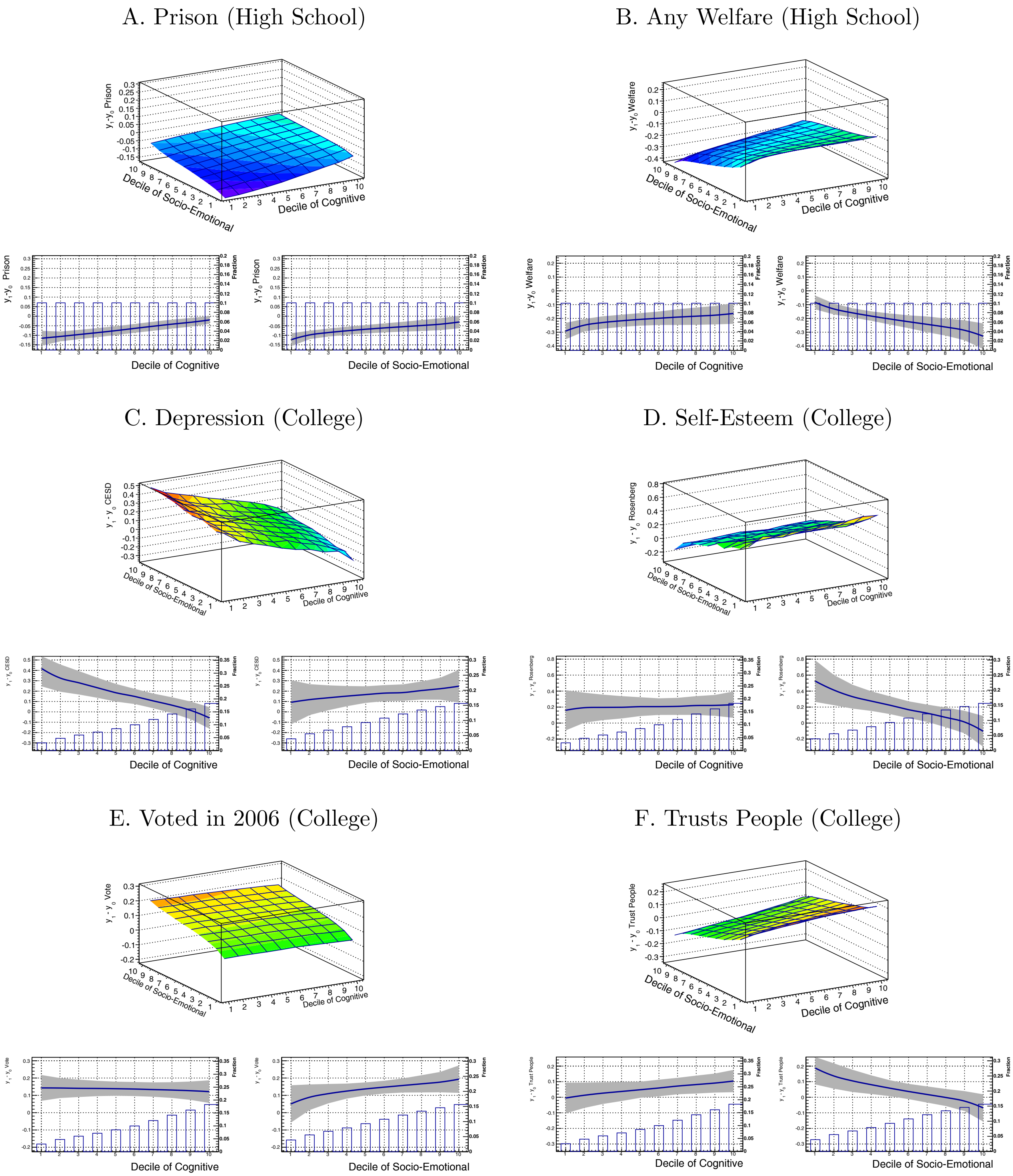
4.2.3 The Effects on Cognitive and Non-Cognitive Endowments on Treatment Effects

The disaggregation of the treatment effects for "high" and "low" endowment individuals in Figure 6 is coarse. One benefit of our approach is that we can determine the contribution of cognitive and non-cognitive endowments ($\boldsymbol{\theta}$) to explaining estimated treatment effects. We can decompose the overall effects of $\boldsymbol{\theta}$ into their contribution to the causal effects at each node and the contribution of endowments to attaining that node. We find substantial contributions of $\boldsymbol{\theta}$ to each component at each node.

To illustrate, the panels in Figure 7 display the estimated average treatment effect of either getting a four-year college degree (compared to stopping with some college) or graduating from high school (compared to dropping out) for each decile pair of cognitive and non-cognitive endowments.[28] Treatment effects in general depend on both measures of ability. Moreover, different outcomes depend in different ways on the two dimensions of ability.

[28]The deciles are calculated using the full population, rather than the population that reaches each node.

Figure 7: Average Treatment Effect of Graduating from High School or a Four-Year College by Outcome



Notes: Each panel in this figure studies the average effects of high school or college graduation 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 $Y_{four-yr\ degree}$ denote the outcomes associated with attaining some college and graduating with a four-year degree, respectively. For each outcome, the first figure (top) presents $E(Y_{four-yr\ degree} - Y_{some\ coll}|d^C, d^{SE})$ 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 $E(Y_{four-yr\ degree} - Y_{some\ coll}|d^C)$ 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 $E(Y_{four-yr\ degree} - Y_{some\ coll}|d^{SE})$ 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.

4.3 A Summary of Treatment Effects by Outcome

In this subsection we summarize the estimates of our model, outcome by outcome.

Prison We find that both the cognitive and socio-emotional endowments affect incarceration. An individual with both endowments in the top decile have only 0.98% probability of being incarcerated, while someone in the bottom decile in both endowments has a 21.7% probability. We find the largest observed difference in outcomes between high school graduates and high school dropouts, for which we also find the largest average causal impact (see Figure 5). For the population of high school graduates and high school dropouts, graduating from high school on average decreases the probability of being incarcerated by about 6% points. Turning to the dynamic treatment effects, Figure 6 shows that almost all of the benefits come from high school graduation. Interestingly, the average marginal treatment effect is larger than the average treatment effect and the average treatment effect for low-endowment individuals is much larger than the average treatment effect for high-endowment individuals. This suggests that low-skill individuals who graduate from high school are much less likely to be incarcerated (about 12% points lower)—an important benefit from high school graduation not fully captured by earnings. Figure 7 shows how the average treatment effects for graduating high school depend on the abilities of the individuals, where high-ability individuals receive almost no return.

Any Welfare Figure 3 shows that both cognitive and non-cognitive endowments affect welfare, though there is a stronger dependence on cognitive endowments. The probabilities of using welfare are 6% and 48.5% for those with both endowments in the top and bottom deciles, respectively. Like prison, we find that the largest observed difference in outcomes is between dropouts and high school graduates. As shown in Figure 5, we find that the average treatment effect between dropouts and high school graduates is about half of the observed difference (13% point decrease). Though smaller, we also find the average pairwise treatment effect between college and some college to be statistically significant. As shown in Figure 6, the average dynamic treatment effects are large for graduating from high school for both high- and low- endowment individuals. There are also statistically significant benefits to enrolling in and graduating from college, though they are much smaller. As we see with many of the non-market outcomes, low-endowment individuals gain most from a four-year degree.

Depression Lower levels of depression are associated with higher levels of the cognitive endowment. The relationship with the socio-emotional endowment is much weaker. An individual in the top and bottom deciles of both endowments has, respectively, depression levels that are 0.48 and -0.30 standard deviations from the population mean. For depression, the largest observed difference shown in Figure 5 is between high school graduates and dropouts, but we find that the average treatment effect is small and insignificant. The only pairwise average treatment effect we find to be statistically significant is graduating from college, which we find to be almost equal to the observed difference between the two groups. For the dynamic treatment effects we find that the average *marginal* treatment effect is statistically significant across the three educational nodes. Many of the treatment effects reported are not statistically significant for the high school graduation or college enrollment decisions, but we find that the ATE, AMTE, and ATE for low-ability individuals is statistically significant for college graduation. Figure 7 further illustrates how treatment effects vary with ability for the college graduation decison. We find that those with lower cognitive ability benefit much more from college graduation while the benefits are flat or slightly increasing in socio-emotional ability.

Self-Esteem Similar to depression, higher cognitive ability is associated with higher self-esteem, but the role of socio-emotional ability is small and statistically insignificant. An individual in the top and bottom deciles of both endowments have self-esteem levels that are 0.60 and -0.67 standard deviations from the population mean. Similar to depression, we find that the observed difference between dropouts and high school graduates is large, but that the average pairwise treatment effect is small and imprecisely estimated. Unlike depression, we find that there are also large observed differences between high school graduates and those with some college as well as between those with some college and those with college degrees. In addition, we find that the average pairwise treatment effect accounts for more than 75% of these observed differences. As shown in Figure 6, there are no benefits for self-esteem from graduating from high school, but the ATE, AMTE, and ATE (low) are large and statistically significant for college enrollment and college graduation. Low-ability individuals increase their self-esteem level by 0.31 standard deviations by enrolling in college and also by earning a four-year college degree. Interestingly, the effects are small and statistically insignificant for high-ability individuals at both margins. Figure 7 further demonstrates how returns differ by endowments. For the decision to graduate college, we find that the gains are larger for those with lower levels of socio-emotional skill, while the returns are more or less flat with respect to the cognitive endowment.

Voting As shown in Figure 3, voting depends strongly on both endowments. The probability of voting for an individual in the top and bottom deciles of both endowments is 83% and 31%, respectively. We find that the average pairwise effect accounts for half or more of the observed difference in outcomes. Figure 6 shows that we find returns to schooling for voting across all margins: graduating from college, enrolling in college, and graduating from high school. Unlike the other outcomes, we find that the returns are statistically significant across educational decisions and endowment levels. Graduating from high school, enrolling in college and graduating from high school increases the probability that an individual will vote by 15% points, 12% points, and 15% points, respectively.

Trust Trust also strongly depends on both cognitive and socio-emotional endowments. The probability that an individual in the top and bottom deciles of both endowments reports trusting others is 62% and 23%, respectively. We find large observed differences between adjacent schooling levels, but do not find any of the pairwise average treatment effects to be statistically significant. Looking at dynamic treatment effects, we find statistically significant returns for enrolling in college for the AMTE and ATE but find all other reported treatment effects to be statistically insignificant. Thus, trust serves as an example where there are large observed differences between schooling levels, but we find little evidence of causal effects.

In Section D of the Web Appendix we reproduce results from HHV (2017) for smoking and health limits work. We briefly review these outcomes as well.

Daily Smoking In our previous work, we find that the observed differences in smoking across education levels are largely causal. While the observed difference is somewhat larger than the estimated pairwise ATE for high school graduation, they are close to identical for college enrollment and graduation. For the dynamic treatment effects we find that education reduces smoking across the educational decisions we consider as well as across the various average treatment effects we calculate.

Health Limits Work For health limits work, we find that the estimated pairwise average treatment effects account for roughly 75% of the observed difference between high school graduates and dropouts as well as college graduates and those with some college. Considering dynamic treatment effects, we find large benefits to college graduation across endowment levels, but little benefit to enrolling in college except for a small but statistically significant ATE. Finally, for college graduation, we find that there are gains on average but that low-ability individuals do not benefit. This contrasts results with many of our other non-market outcomes where the low-ability individuals are the primary beneficiaries.

5 Summary and Conclusion

Becker (1964) emphasized both the market and non-market benefits of education. This paper demonstrates the wisdom of his insights and finds that there are substantial and diverse non-market benefits to education. Estimating a robust dynamic model of educational choices and their consequences, we account for selection bias and sorting on gains in a dynamic sequential model of educational decisions. We identify and estimate dynamic treatment effects that account for the options opened up from each educational choice. The dynamic treatment effects are policy relevant because most policies only affect choices at one margin of education, but do not restrict future choices.

Using our model, we investigate a range of non-market outcomes including incarceration, mental health, voter participation, trust, and participation in welfare. We find that the returns to education for many non-market outcomes appear to be larger for low-ability individuals—a feature of the returns to education that is missed if only market returns are analyzed. We also demonstrate the important effects of cognitive and non-cognitive abilities on educational choices and their effects on outcomes fixing educational levels.

References

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

ALMLUND, M., A. L. DUCKWORTH, J. J. HECKMAN, AND T. KAUTZ (2011): "Personality Psychology and Economics," in *Handbook of the Economics of Education*, ed. by E. A. Hanushek, S. Machin, and L. Wößmann, vol. 4, chap. 1, pp. 1–181. Elsevier B. V., Amsterdam.

ARENDT, J. N. (2005): "Does Education Cause Better Health? A Panel Data Analysis Using School Reforms for Identification," *Economics of Education Review*, 24(2), 149–160.

ARKES, J. (2003): "Does Schooling Improve Adult Health?," Unrestricted Draft DRU-3051, RAND, Santa Monica, CA.

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

BECKER, G. S. (1964): *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education*. University of Chicago Press for the National Bureau of Economic Research, Chicago, 1 edn.

CAMERON, S. V., AND J. J. HECKMAN (2001): "The Dynamics of Educational Attainment for Black, Hispanic, and White Males," *Journal of Political Economy*, 109(3), 455–499.

CARD, D. (1999): "The Causal Effect of Education on Earnings," in *Handbook of Labor Economics*, ed. by O. C. Ashenfelter, and D. Card, vol. 3A, chap. 30, pp. 1801–1863. Elsevier Science B.V., Amsterdam.

CARNEIRO, P., J. J. HECKMAN, AND E. J. VYTLACIL (2011): "Estimating Marginal Returns to Education," *American Economic Review*, 101(6), 2754–2781.

Conti, G., J. J. Heckman, and S. Urzúa (2010): "The Education-Health Gradient," *American Economic Review: Papers and Proceedings*, 100(2), 1–5.

Cutler, D. M., and A. Lleras-Muney (2010): "Understanding Differences in Health Behaviors by Education," *Journal of Health Economics*, 29(1), 1–28.

Ehrlich, I. (1973): "Participation in Illegitimate Activities: A Theoretical and Empirical Investigation," *Journal of Political Economy*, 81(3), 521–565.

——— (1975): "On the Relation between Education and Crime," in *Education, Income, and Human Behavior*, ed. by F. T. Juster, chap. 12, pp. 313–338. National Bureau of Economic Research, Cambridge, MA.

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

——— (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.

——— (2006): "Education and Nonmarket Outcomes," in *Handbook of the Economics of Education*, ed. by E. Hanushek, and F. Welch, vol. 1, chap. 10, pp. 577–633. Elsevier, Amsterdam.

——— (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 *The Social Benefits of Education*, ed. by J. R. Behrman, and N. Stacey, pp. 69–124. University of Michigan Press, Ann Arbor, MI.

Haavelmo, T. (1943): "The Statistical Implications of a System of Simultaneous Equations," *Econometrica*, 11(1), 1–12.

Heckman, J. J., J. E. Humphries, and G. Veramendi (2016): "Dynamic Treatment Effects," *Journal of Econometrics*, 191(2), 276–292.

——— (2017): "Returns to Education: The Causal Effects of Education on Earnings, Health and Smoking," Forthcoming, *Journal of Political Economy*.

Heckman, J. J., L. J. Lochner, and P. E. Todd (2006): "Earnings Functions, Rates of Return and Treatment Effects: The Mincer Equation and Beyond," in *Handbook of the Economics of Education*, ed. by E. A. Hanushek, and F. Welch, vol. 1, chap. 7, pp. 307–458. Elsevier, Amsterdam.

Heckman, J. J., and R. Pinto (2015): "Causal Analysis after Haavelmo," *Econometric Theory*, 31(1), 115–151.

Heckman, J. J., S. Urzúa, and E. J. Vytlacil (2008): "Instrumental Variables in Models with Multiple Outcomes: The General Unordered Case," *Annales d'Economie et de Statistique*, 91–92, 151–174.

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. (2011): "Nonproduction Benefits of Education: Crime, Health, and Good Citizenship," in *Handbook of the Economics of Education*, ed. by E. A. Hanushek, S. Machin, and L. Wößmann, vol. 4, chap. 2, pp. 183–282. Elsevier B. V., Amsterdam.

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.

Mincer, J. (1958): "Investment in Human Capital and Personal Income Distribution," *Journal of Political Economy*, 66(4), 281–302.

——— (1974): *Schooling, Experience, and Earnings*. Columbia University Press for National Bureau of Economic Research, New York.

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

Silles, M. A. (2009): "The Causal Effect of Education on Health: Evidence from the United Kingdom," *Economics of Education Review*, 28(1), 122–128.

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

Vila, L. E. (2000): "The Non-Monetary Benefits of Education," *European Journal of Education*, pp. 21–32.

Williams, B. (2017): "Identification of the Linear Factor Model," George Washington University, Department of Economics. Unpublished manuscript.