The Non-Market Benefits of Education and Ability

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

Gregory Veramendi
Arizona State University

April 13, 2016

*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, University of Chicago, 1126 East 59th Street, Chicago, IL 60637; phone: 773-980-6575; email: johneric@uchicago.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. 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. Unlike previous evidence on the monetary benefits of education, the benefits to education for many non-market outcomes appear to be larger for low-ability individuals. College graduation decreases welfare use, lowers depression, and raises self-esteem more for less-able individuals. Accounting for the non-market benefits of education is an important component of any analysis of educational policy.

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
University of Chicago
1126 East 59th Street
Chicago, IL 60637
Phone: 773-980-6575
Email: johneric@uchicago.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 .......................... 4

2 A Brief Survey of the Literature ........................................................................... 6

3 Dynamic Discrete Choice Model of Education ......................................................... 7

   3.1 Defining Treatment Effects ........................................................................... 12

4 The Data .................................................................................................................. 15

5 Results ..................................................................................................................... 18

   5.1 The Effects of Endowments on Education and Outcomes .......................... 18

   5.2 Estimated Causal Effects ............................................................................. 21

      5.2.1 The Estimated Causal Effects of Final Educational Levels .............. 22

      5.2.2 Dynamic Treatment Effects ............................................................... 24

      5.2.3 The Effects on Cognitive and Non-Cognitive Endowments on Treatment Effects ........................................................................... 26

   5.3 A Summary of Treatment Effects by Outcome .......................................... 28

6 Summary and Conclusion ....................................................................................... 31
1 Introduction

In his pioneering analysis, Gary Becker (1964) discusses both the market and non-market benefits of education. Despite this, 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 some evidence that a portion of this correlation is causal.\(^2\) Education is positively correlated with better mental health, more 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 support for 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.

Two general approaches have been developed to address these issues. One approach is to estimate treatment effect models using instrumental variables methods.\(^4\) Although the identifying assumptions are transparent, treatment effects estimated using instrumental variables often do not identify policy-relevant questions unless the instruments correspond very closely to the policy of interest. In addition, instrumental variables for education are often weak, requiring large datasets where information on non-market outcomes is often not

\(^1\)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); Heckman, Humphries, and Veramendi (2016b).

\(^3\)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 \(i\), \(\alpha_i\) is an individual-specific intercept, \(\rho_i\) 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 \(\text{Cov}(\rho_i, S_i) \neq 0\).

The second approach is the estimation of structural models. By taking a stand on the decision making process and information set of the agents, structural models can be used to identify well-defined treatment effects, even for policies that have never previously been implemented. While the structural approach provides interpretable treatment effects, it requires the researcher to take a stand on the decision making processes and expectations of agents. In addition, estimation typically requires parametric assumptions for tractability. While the results are interpretable, other researchers have called into question the robustness of structural estimates to variations in maintained assumptions.

This paper investigates the non-market benefits of abilities and education using a third approach. 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. 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 variation in the returns to schooling. For example, the model shows that the returns to high school are relatively homogenous across persons of different ability, but the returns to college are higher for low-ability people. We address both selection bias and sorting on gains by allowing for observed covariates as well as a multidimensional vector of unobserved skill endowments to affect baseline outcomes as well as the gains from education.

Our model accounts for the dynamics of the educational decisions. Early educational choices open up future choices. Rather than estimating only pairwise comparisons between two final schooling levels, we identify dynamic treatment effects, which estimate the return from making a particular educational decision, accounting for the fact that the decision opens up additional educational choices. Dynamic treatment effects are particularly policy relevant as they focus on changing behavior at a particular educational choice, which more closely

\footnote{See, e.g., Keane and Wolpin (1997); Eisenhauer, Heckman, and Mosso (2015).}

\footnote{See Angrist and Pischke (2010).}
correspond to potential policies affecting educational decisions.

We find that the returns to education for many non-market outcomes appear to be larger for low-ability individuals. 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 (2016b), who show that market returns to education are typically larger for high-ability people, particularly when making post-secondary decisions. The returns to education are substantial for low-ability people.

Our paper proceeds in the following way. Section 2 reviews the literature on education and non-market outcomes. Section 3 presents our model. Section 3.1 presents economically interpretable treatment effects (rates of return) that can be derived from it. Section 4 discusses the data analyzed and presents unadjusted associations and regression adjusted associations between different levels of education and the outcomes analyzed in this paper. Section 5 reports our estimated treatment effects and interprets them. Section 6 concludes.

2 A Brief Survey of the Literature

Lochner (2011) provides a comprehensive review of the literature. Lochner and Moretti (2004) review some of the early evidence on the non-market benefits of education. They present economic models for how individuals may optimally choose crime and health-related behaviors over the life-cycle and how these may be changed by education. 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. He considers a number of social benefits such as the impact on the level of inequality in society and the stability of social structures. Oreopoulos and Salvanes (2011) present

---

7A 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.

8Ehrlich (1973) is a pioneering analysis of the economics of crime. His 1975 paper presents an early analysis of the correlation between education and crime.
additional evidence on the non-market benefits of education. They report evidence that schooling improves happiness, leads to higher job-satisfaction and occupational prestige, leads to less disability, improves health, leads to more highly educated spouses, lowers the probability of divorce, and improves trust. They also find some evidence that education lowers the probability of being a smoker, being arrested, or reporting that the individual “lives for today”—outcomes they attribute to changing preferences.

Other papers consider the role of cognitive and non-cognitive abilities in educational attainment and later life outcomes. Heckman, Stixrud, and Urzúa (2006) analyze the role of cognitive and non-cognitive ability in educational decisions, economic outcomes such as wages and work experience, as well as some non-market outcomes such as incarceration by age 30. Almlund, Duckworth, Heckman, and Kautz (2011) review the literature. Roberts, Kuncel, Shiner, Caspi, and Goldberg (2007) compare the predictive power of early-life measures of cognitive ability, personality, and socio-economic status on mortality, divorce, and occupational attainment. Using three studies, they show that personality is as predictive as cognition and socio-economic status for the outcomes they study. Other work has evaluated the correlations between cognitive ability and crime (Frisell, Pawitan, and Långström, 2012), mental health (Hatch, Jones, Kuh, Hardy, Wadsworth, and Richards, 2007), and political engagement (Denny and Doyle, 2008). Section A of the Web Appendix provides a detailed literature review on each of the six non-market outcomes we consider in this paper.

3 Dynamic Discrete Choice Model of Education

This paper estimates the non-market benefits of education using the multistage sequential model of educational choice from Heckman, Humphries, and Veramendi (2016b) with transitions and decision nodes shown in Figure 1. Unlike the previous literature, which is often unclear about the margins of schooling choice identified by its instruments and other sources of exogenous variation, we estimate returns to education at each node. People begin in high school and must 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, we assume that people traverse a tree of educational decisions where their choice affects which choices are available in the future. As shown in Figure 1, we assume that at each choice node people make a binary decision. In our model, we assume a person’s final schooling level is \( s \in \{G, 0, 1, 2, 3\} \), where \( G \) denotes earning a GED, and high school dropout, high school graduate, some college, and college graduate are denoted as 0, 1, 2, and 3, respectively. 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.

We assume each person has a set of characteristics $X$ that affects educational decisions and outcomes at different schooling levels. The characteristics include background variables such as parent’s education or if one grew up in an urban environment, but they also include contemporaneous variables such as current unemployment rate. We assume that there are some variables $Z$, such as presence of local college or local price of tuition, that affect educational decisions, but do not affect outcomes.\footnote{The estimated model includes such instruments, but does not necessarily require the instruments for identification.} We assume $Z$ and $X$ are known by both the agent and the economist, but we allow agents to act on a multidimensional set of unobservables $\theta$ that can affect their educational decisions and outcomes, but are not directly observed by the economist. They can be thought of as a vector of random effects that potentially affect decisions and outcomes. These measurements make our random effects interpretable. While $\theta$ is not directly observed, it is proxied by measurements. We assume that there are two unobserved factors which we estimate by a multivariate mixture of normals.\footnote{See Section C.1 in the Web Appendix for details on identification and estimation.} In this paper, we use measurements of $\theta$ corresponding to cognitive and socio-emotional abilities. These unobserved factors are assumed to capture the correlation between schooling decisions and outcomes that generate selection bias as well as sorting on gains. Any remaining variation in decisions and outcomes that is not captured by $Z, X,$ and $\theta$ is assumed to be explained by independent shocks known when the agent reaches the outcome or decision, but otherwise unknown to the agent.\footnote{We allow components of $\theta$ to affect some transitions but not others, and thereby build in the possibility of shocks across states, a form of serial correlation across states.}

**Educational Decisions** Educational decisions are modeled using latent variables crossing a threshold as described in Cameron and Heckman (2001) and Heckman, Humphries, and Veramendi (2016b). We assume that educational decisions depend on functions of $X, Z,$ and $\theta$ as well as additive idiosyncratic shocks. This formulation does not assume that agents are maximizers. Our approach can approximate such a model if it is valid (Heckman, Humphries,
and Veramendi, 2016a). This model also does not require that agents make their educational decisions based on expected gains, or that the unobserved components in $\theta$ or observed components in $Z$ affect all educational decisions in the same way.

**Outcomes** Outcomes are assumed to depend on a function of directly observed characteristics $X$, the unobserved components $\theta$, and an idiosyncratic shock. Importantly, we assume that the intercept, the observed characteristics $X$, and the unobserved components $\theta$ 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 with schooling. The persistent (and known by the agent) proxied variable $\theta$ enters the model similar to a random effect, but the impact of the random effect can vary across schooling levels. For example, cognitive ability can play a more important role for those with a college degree compared to high school dropouts.

**The Measurement System** The unobserved but proxied variable ($\theta$) plays an important role in our analysis. Along with the observed variables, it generates the dependence between choices and outcomes. Conditional on $X$, $\theta$ is assumed to capture the correlation between the unobserved components of the outcomes and the unobserved components of the educational decisions. If $\theta$ were observed, we could condition on ($\theta, X, Z$) and identify selection-bias-free estimates of a variety of causal effects. While the $\theta$ is not observed, it is proxied. In order to identify and estimate the distribution of $\theta$, we use multiple measures of it. To accomplish this, we use a measurement system consisting of early-life tests or outcomes which are thought to be generated by $\theta$. Similar to the educational decisions and outcomes, we assume that these depend on observed characteristics $X$, the proxied vector of factors $\theta$, and idiosyncratic shocks. The $X$ and $\theta$ account for the correlations across measures as well as the correlations between the measures, outcomes, and schooling decisions. In this paper, the 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.

By using measurements, we can identify and interpret the individual factors. We find that two factors are sufficient to explain our data (Heckman, Humphries, and Veramendi, 2016b). ASVAB scores, GPA, reckless or risky behavior, educational decisions, and outcomes all determine the first factor. GPA, reckless behavior, educational decisions, and outcomes 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. By identifying and interpreting the two random effects, we are able to better interpret the sources of flourishing lives.

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

Under conditional independence, we assume that conditional on $\theta, X, Z$, outcomes and choices are statistically independent. If we, in fact, could observe the unobserved factors $\theta$, this would be equivalent to the assumptions made in matching. Since $\theta$ is not observed, we treat it as a random effect. Unlike the usual random effect approach, we proxy $\theta$ using a set of measurements as described in Section 3. Under these assumptions, we can identify the distribution of $\theta$ non-parametrically. Effectively, we match on $X, Z$, and proxies for $\theta$, 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 errors. The model can also be identified using instrumental variables. See Heckman, Humphries, and Veramendi (2016a).

While the model is identified non-parametrically (Heckman, Humphries, and Veramendi, 2016a), we make parametric assumptions in this paper to facilitate the computation. The
The Non-Market Benefits of Education and Ability

precise parametrization and the likelihood function for the model we estimate are presented in Web Appendix C.1.

3.1 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 a particular choice, inclusive of continuation values associated with future educational decisions.\(^{12}\)

**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 \(Y_{s'}^k\) be outcome \(k\) at schooling level \(s'\) and \(Y_s^k\) be outcome \(k\) at schooling level \(s\). Conditioning on \(X = x, Z = z, \text{and } \theta = \overline{\theta}\), the average treatment effect of \(s\) compared to \(s'\) is \(E(Y_s^k - Y_{s'}^k | X = x, Z = z, \theta = \overline{\theta})\). Thus, we fix \(S = s\) and condition on \(X = x, Z = z, \text{and } \theta = \overline{\theta}\).\(^{13}\) ATE \((ATE_{s,s'}^k)\) is calculated by integrating over \(X, Z, \theta\) for the subset of the population that completes one of the two final schooling levels \((S \in \{s, s'\})\). Conditioning in this fashion 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 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 \((Q_{j+1} = 1 \text{ versus } Q_{j+1} = 0)\) conditioning on the set of persons who visit the decision node \((Q_j = 1)\). Those

\(^{12}\)See Heckman, Humphries, and Veramendi (2016b) for formal definitions of these treatment effects.

\(^{13}\)Fixing versus conditioning is the key notion in causal models since Haavelmo (1943). For a recent discussion, see Heckman and Pinto (2015).
The Non-Market Benefits of Education and Ability April 13, 2016

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 $X, Z, \theta$ 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 defined for different populations.\(^{14}\)

To clarify the distinctions between the two types of treatment effects, consider a simplified example where there are four levels of schooling: high school dropout, high school graduate, some college, and college graduate ($s \in \{0, 1, 2, 3\}$). We focus on the upper branch of Figure 1. Consider outcome $Y$, where:

$$Y_i = \beta_{i,0} + \beta_{i,1}Q_{i,1} + \beta_{i,2}Q_{i,2} + \beta_{i,3}Q_{i,3} + U_i, \quad (1)$$

where $Q_{i,j} = 1$ indicates that the individual $i$ made it to decision node $j$. $Q_{i,1} = 1$ 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.\(^{15}\) The $\beta_{i,j}$ are the marginal benefits of being at state $j$. Note that the $\beta$ have $i$ subscripts as we allow them to vary by both the individual’s observable and unobservable characteristics.

Using this simplified model, the mean treatment effects associated with differences across final schooling levels are given by:

$$ATE_{s,s-1} = E[\beta_s | S \in \{s, s-1\}]$$

$$ATE_{s,0} = E \left[ \sum_{j=1}^{s} \beta_j | S \in \{s, 0\} \right],$$

where we have conditioned the expectation for populations at one of the two final schooling levels.

\(^{14}\)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 Heckman, Humphries, and Veramendi (2016b).

\(^{15}\)With the factor structure used in this paper, $U_i$ can be correlated with educational choices and outcomes.
levels considered.\footnote{16}{Treatment effects without this conditioning can also be constructed.}

In order to construct the dynamic treatment effects in our simplified model, we need to include a decision rule for how educational choices are made. We do not need to assume that agents are rational. Our model is consistent with irrationality or ad-hoc rules-of-thumb.

Let $1(Q_j = 1)$ 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:

$$T_{i,1} = \beta_{i,1} + 1(Q_{i,2} = 1)\beta_{i,2} + 1(Q_{i,2} = 1)1(Q_{i,3} = 1)\beta_{i,3},$$  \hspace{1cm} (2)

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.\footnote{17}{We abstract from discounting. The $\beta$ coefficients can be interpreted as incorporating the discounted value.} Note that even if we had 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 two dynamic treatment effects for individual $i$ are:

$$T_{i,2} = \beta_{i,2} + 1(Q_{i,3} = 1)\beta_{i,3}$$ \hspace{1cm} (3)

$$T_{i,3} = \beta_{i,3},$$ \hspace{1cm} (4)

where the returns to graduating college are equal to the direct returns in our example as it is the final level of education. To construct the population average dynamic treatment effects, we take the expectation of these individual returns over the full population:\footnote{18}{This entails taking the expectation over the full population (and thus integrates over the full support of $X$, $Z$, $\theta$, and $U$).}

$$ATE_j^* = E[T_{i,j}].$$ \hspace{1cm} (5)
Equation (5) is the average treatment effect for the entire population. This likely involves far out-of-sample estimates such as the earnings of dropouts as college graduates. The full population ATE is less empirically and policy relevant than considering only those who reach a particular decision node. In this paper we will focus on:

\[ ATE_j = E [T_{i,j}|Q_{i,j} = 1], \]  

which is to say we estimate the average dynamic treatment effect for those that reach a particular decision.\(^{19}\)

We estimate various treatment effects derived from estimates using the econometric model derived in Heckman, Humphries, and Veramendi (2016a,b). Web Appendix B gives details that are more complete on the model and the treatment effects derived from it.

4 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.\(^{20}\)

\(^{19}\)In particular, this can be thought as integrating over a particular support of \(X, Z, \theta,\) and \(U\) that results in \(Q_j = 1.\)

\(^{20}\)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 Heckman, Humphries, and Veramendi (2016b).
Table 1: Description of Outcomes

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

Notes: All outcomes are from the National Longitudinal Survey of Youth 1979. See Section D of the Web Appendix for additional health outcomes from Heckman, Humphries, and Veramendi (2016b). 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 we study 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.\(^{21}\) 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 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.

\(^{21}\)Adjustments are made through linear regression.
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.
5 Results

5.1 The Effects of Endowments on Education and Outcomes

To begin, we consider how the latent endowment $\theta$ affects 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 Heckman, Humphries, and Veramendi (2016b). 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.$^{22}$

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%.$^{23}$

\(^{22}\)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.

\(^{23}\)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

A. Prison

B. Any Welfare

C. Depression (CES-D)

D. Self-Esteem (Rosenberg)

E. Voted in 2006

F. Trusts People

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. 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 X, 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 X, 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 $X, Z$. The figures below the graphs of the joint distribution display the marginal effects on outcomes for cognitive and socio-emotional endowments respectively, marginalizing over $X, Z$, and the other endowment. 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%.

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.

---

24 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.

25 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.
The Non-Market Benefits of Education and Ability April 13, 2016

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

A. Dropping from HS vs. Graduating from HS

B. HS Dropout vs. Getting a GED

C. HS Graduate vs. College Enrollment

D. Some College vs. 4-Year College Degree

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 $\text{Prob}(Q_{j+1} = 1|d^C, d^{SE}, Q_j = 1)$, where $d^C$ and $d^{SE}$ denote the cognitive and socio-emotional endowments. $\text{Prob}(Q_{j+1} = 1|d^C, d^{SE}, Q_j = 1)$ is computed for those who reach the decision node involving a decision between levels $j$ and $j+1$. It marginalizes over $X, Z$ conditional on $Q_j = 1$. The bottom left subfigures present $\text{Prob}(Q_{j+1} = 1|d^C, Q_j = 1)$, where the socio-emotional factor and $X, Z$ 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 $\text{Prob}(Q_{j+1} = 1|d^{SE}, Q_j = 1)$ 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 $X, Z$, and the other endowment.

5.2 Estimated Causal Effects

We next present the main treatment effects estimated from our model. Since our model is nonlinear and multidimensional, in the main body of the paper 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.\footnote{We randomly draw an individual and use their full set of regressors.} We first present the main treatment effects across final schooling levels. We then consider how dynamic treatment effects vary across decisions and endowment levels.

### 5.2.1 The Estimated Causal Effects of Final Educational Levels

For some specific outcomes almost none of the observed difference is causal. For example, we find that the observed differences between high school graduates and dropouts are large for self-esteem and depression, but that 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.\footnote{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.} 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 $s - 1$.\footnote{The circles and black circles indicate that the ATE is statistically significant at the 0.05 and 0.01 level, respectively.} 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 between outcome.
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 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.
5.2.2 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 between the two options of the choice studied. \(^{29}\) AMTE is introduced in Carneiro, Heckman, and Vytlacil (2011) and formally developed in the context of our model in Heckman, Humphries, and Veramendi (2016b).

Each box of Figure 6 presents a number of educational treatment effects for each education level for a specified outcome. 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). The 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.

\[^{29}\text{We define the margin of indifference to be } || I_j/\sigma_j || \leq 0.01, \text{ where } \sigma_j \text{ is the standard deviation of } I_j.\]

24
Figure 6: Treatment Effects of Outcomes by Decision Node

<table>
<thead>
<tr>
<th>Treatment Effects: Prison</th>
<th>Treatment Effects: Any Welfare</th>
</tr>
</thead>
<tbody>
<tr>
<td>AMTE ATE ATE (low) ATE (high) p &lt; 0.05 p &lt; 0.01</td>
<td>AMTE ATE ATE (low) ATE (high) p &lt; 0.05 p &lt; 0.01</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Average TE</th>
<th>Graduate HS</th>
<th>Enroll in Coll.</th>
<th>Graduate Coll.</th>
</tr>
</thead>
<tbody>
<tr>
<td>-0.2</td>
<td>-1.1</td>
<td>-1.1</td>
<td>-1.1</td>
</tr>
<tr>
<td>-0.1</td>
<td>-0.9</td>
<td>-0.9</td>
<td>-0.9</td>
</tr>
<tr>
<td>0</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
</tr>
<tr>
<td>0.1</td>
<td>0.9</td>
<td>0.9</td>
<td>0.9</td>
</tr>
<tr>
<td>0.2</td>
<td>1.1</td>
<td>1.1</td>
<td>1.1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Treatment Effects: Depression (CES−D)</th>
<th>Treatment Effects: Self−Esteem (Rosenberg)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AMTE ATE ATE (low) ATE (high) p &lt; 0.05 p &lt; 0.01</td>
<td>AMTE ATE ATE (low) ATE (high) p &lt; 0.05 p &lt; 0.01</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Average TE</th>
<th>Graduate HS</th>
<th>Enroll in Coll.</th>
<th>Graduate Coll.</th>
</tr>
</thead>
<tbody>
<tr>
<td>-0.3</td>
<td>-2.2</td>
<td>-2.2</td>
<td>-2.2</td>
</tr>
<tr>
<td>-0.2</td>
<td>-2.1</td>
<td>-2.1</td>
<td>-2.1</td>
</tr>
<tr>
<td>-0.1</td>
<td>-1.9</td>
<td>-1.9</td>
<td>-1.9</td>
</tr>
<tr>
<td>0</td>
<td>-1.1</td>
<td>-1.1</td>
<td>-1.1</td>
</tr>
<tr>
<td>0.1</td>
<td>-0.9</td>
<td>-0.9</td>
<td>-0.9</td>
</tr>
<tr>
<td>0.2</td>
<td>-0.7</td>
<td>-0.7</td>
<td>-0.7</td>
</tr>
<tr>
<td>0.3</td>
<td>-0.5</td>
<td>-0.5</td>
<td>-0.5</td>
</tr>
<tr>
<td>0.4</td>
<td>-0.3</td>
<td>-0.3</td>
<td>-0.3</td>
</tr>
<tr>
<td>0.5</td>
<td>-0.1</td>
<td>-0.1</td>
<td>-0.1</td>
</tr>
<tr>
<td>0.6</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Treatment Effects: Voted in 2006</th>
<th>Treatment Effects: Trusts People</th>
</tr>
</thead>
<tbody>
<tr>
<td>AMTE ATE ATE (low) ATE (high) p &lt; 0.05 p &lt; 0.01</td>
<td>AMTE ATE ATE (low) ATE (high) p &lt; 0.05 p &lt; 0.01</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Average TE</th>
<th>Graduate HS</th>
<th>Enroll in Coll.</th>
<th>Graduate Coll.</th>
</tr>
</thead>
<tbody>
<tr>
<td>-0.4</td>
<td>-3.3</td>
<td>-3.3</td>
<td>-3.3</td>
</tr>
<tr>
<td>-0.2</td>
<td>-3.1</td>
<td>-3.1</td>
<td>-3.1</td>
</tr>
<tr>
<td>0</td>
<td>-2.9</td>
<td>-2.9</td>
<td>-2.9</td>
</tr>
<tr>
<td>0.2</td>
<td>-2.7</td>
<td>-2.7</td>
<td>-2.7</td>
</tr>
<tr>
<td>0.4</td>
<td>-2.5</td>
<td>-2.5</td>
<td>-2.5</td>
</tr>
<tr>
<td>0.6</td>
<td>-2.3</td>
<td>-2.3</td>
<td>-2.3</td>
</tr>
<tr>
<td>0.8</td>
<td>-2.1</td>
<td>-2.1</td>
<td>-2.1</td>
</tr>
<tr>
<td>1.0</td>
<td>-1.9</td>
<td>-1.9</td>
<td>-1.9</td>
</tr>
<tr>
<td>1.2</td>
<td>-1.7</td>
<td>-1.7</td>
<td>-1.7</td>
</tr>
<tr>
<td>1.4</td>
<td>-1.5</td>
<td>-1.5</td>
<td>-1.5</td>
</tr>
<tr>
<td>1.6</td>
<td>-1.3</td>
<td>-1.3</td>
<td>-1.3</td>
</tr>
<tr>
<td>1.8</td>
<td>-1.1</td>
<td>-1.1</td>
<td>-1.1</td>
</tr>
<tr>
<td>2.0</td>
<td>-0.9</td>
<td>-0.9</td>
<td>-0.9</td>
</tr>
<tr>
<td>2.2</td>
<td>-0.7</td>
<td>-0.7</td>
<td>-0.7</td>
</tr>
<tr>
<td>2.4</td>
<td>-0.5</td>
<td>-0.5</td>
<td>-0.5</td>
</tr>
<tr>
<td>2.6</td>
<td>-0.3</td>
<td>-0.3</td>
<td>-0.3</td>
</tr>
<tr>
<td>2.8</td>
<td>-0.1</td>
<td>-0.1</td>
<td>-0.1</td>
</tr>
<tr>
<td>3.0</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
</tr>
</tbody>
</table>

Sorting on Ability

<table>
<thead>
<tr>
<th>j = 1: Dropping from HS vs. Graduating from HS</th>
<th>Low Ability</th>
<th>High Ability</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.31</td>
<td>0.31</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>j = 2: HS Graduate vs. College Enrollment</th>
<th>Low Ability</th>
<th>High Ability</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.22</td>
<td>0.38</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>j = 3: Some College vs. Four-Year College Degree</th>
<th>Low Ability</th>
<th>High Ability</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.13</td>
<td>0.51</td>
<td></td>
</tr>
</tbody>
</table>

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 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. The table below the figure shows the proportion of individuals at each decision node that are high or 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.
5.2.3 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. A byproduct of our approach is that we can determine the contribution of cognitive and non-cognitive endowments ($\theta$) to explaining estimated treatment effects. We can decompose the overall effects of $\theta$ into their contribution to the causal effects at each node and the contribution of endowments to attaining that node. We find substantial contributions of $\theta$ 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.$^{30}$ Treatment effects in general depend on both measures of ability. Moreover, different outcomes depend in different ways on the two dimensions of ability.

---

$^{30}$The deciles are calculated using the full population, rather than the population that reaches each node.
Figure 7: Average Treatment Effect of Graduating from a Four-Year College by Outcome

A. Prison (High School)  

B. Any Welfare (High School)  

C. Depression (College)  

D. Self-Esteem (College)  

E. Voted in 2006 (College)  

F. Trusts People (College)

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 \( Y_{\text{some coll}} \) and \( Y_{\text{four-yr degree}} \) denote the outcomes associated with attaining some college and graduating with a four-year degree, respectively. For each outcome, the first figure (top) presents \( E(Y_{\text{four-yr degree}} - Y_{\text{some coll}} | d^C, d^{SE}) \) 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 \( E(Y_{\text{four-yr degree}} - Y_{\text{some coll}} | d^C) \) 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 \( E(Y_{\text{four-yr degree}} - Y_{\text{some coll}} | d^{SE}) \) 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.
5.3 A Summary of Treatment Effects by Outcome

This subsection provides a brief summary of the empirical results by outcome.

Prison  We find that both the cognitive and socio-emotional endowment affect incarceration, but that the effect is close to zero except for low-endowment individuals. Similarly, we find the largest observed difference in outcomes between high school graduates and high school dropouts, for which we find the largest average causal impact (see Figure 5). 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—a potentially important benefit to education not fully captured through earnings.

Any Welfare  Figure 3 shows that both cognitive and non-cognitive endowments affect welfare, though the cognitive endowment has a somewhat large impact. 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. 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 these are much smaller.

Depression  Lower levels of depression are causally associated with higher levels of the cognitive endowment, but not the socio-emotional endowment. 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 three educational decisions. 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 decision. We find that those with lower socio-emotional ability benefit much more from college graduation while the benefits are flat or slightly increasing in cognitive 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. 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 self-esteem benefits from graduating from high school, but the ATE, AMTE, and ATE (low) are large and statistically significant for college enrollment and college graduation. 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 college graduation decision, 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 3E, voting depends on both the cognitive and socio-emotional endowments. We find that the average pairwise effect accounts for half or more of the observed difference in outcomes. Figure 6 shows that we find persistent returns to schooling for voting across graduating from college, enrolling in college, and graduating from high school. Unlike the other outcomes, we find that the returns are persistent across educational decision and endowment level.

**Trust**  Trust depends on both the cognitive endowment and the socio-emotional endowment, though somewhat more strongly on cognition. 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 the 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 of where there are large observed differences between schooling levels, but we find very little evidence of causal effects.

In Section D of the Web Appendix we reproduce results from Heckman, Humphries, and Veramendi (2016b) 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. Looking at 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 such as self-esteem where it was
low-ability individuals who benefited the most.

6 Summary and Conclusion

Becker (1964) emphasized both the market and non-market benefits of education. This paper
demonstrates the wisdom of his insights and demonstrates that there are substantial and
diverse non-market benefits to education. Using a model of educational choice, we have
implemented a framework that accounts for selection bias, sorting on gains, and dynamic
sequential educational decisions. We construct dynamic treatment effects that account for
the additional options opened up by a particular 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.

Our approach offers a middle ground between reduced-form estimates of the returns to
schooling that are often difficult to interpret, and fully structural models that often make
strong assumptions about behavior. 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 often
missed if only individual market returns are considered. We also demonstrate the important
effects of cognitive and non-cognitive abilities on educational choices and their effects on
outcomes fixing educational levels.
References


