The Causal Effects of Education on Earnings and Health*

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

John Eric Humphries
University of Chicago

Gregory Veramendi
Arizona State University

January 14, 2016

*James 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, October 30, 2014. It was also presented as the Sandmo Lecture at the Norwegian School of Economics, January 13, 2015. We thank Chris Taber for insightful comments. We also thank Ariel Pakes and other participants at a Harvard Labor Economics Workshop April, 2014, for helpful comments on a previous draft. We thank Eleanor Dillon and Mathew Wiswall for comments on this draft received at a seminar at Arizona State University, February, 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 R01HD54702; 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 Institutes of Health. The Web Appendix for this paper is https://heckman.uchicago.edu/eff-ed-earn-health.

Abstract

This paper estimates a robust dynamic model of the causal effects of different levels of schooling on earnings and health. Our framework synthesizes approaches used in the dynamic discrete choice literature with approaches used in the reduced form treatment effect literature. We estimate economically interpretable and policy relevant treatment effects. Cognitive and noncognitive endowments play important roles in explaining observed differences in earnings and health across education levels. Nonetheless, after controlling for them, there are substantial causal effects of education at all stages of schooling. Continuation values associated with dynamic sequential schooling choices are empirically important components of estimated causal effects. There is considerable heterogeneity in the effects of schooling on outcomes at different schooling levels and in these effects across persons. We find strong sorting on gains consistent with comparative advantage, but only at higher levels of schooling. This result is not imposed in our estimation procedure. We find that the estimated causal effects of education vary with the level of cognitive and noncognitive endowments. Estimates of causal effects using standard instrumental variables are often quite different from the economically interpretable and policy relevant treatment effects derived from our model.

Keywords: education, earnings, health, rates of return, causal effects of education

JEL codes: C32, C38, I12, I14, I21

James 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 **5**

2 Model **11**

2.1 A Sequential Decision Model 12

2.2 Parameterizations of the Decision Rules and Potential Outcomes for Final States 13

2.3 Structure of the Unobservables 14

2.4 Measurement System for Unobserved Factors $\boldsymbol{\theta}$ 15

3 Defining Treatment Effects **16**

3.1 Dynamic Treatment Effects 18

3.2 Mean Differences Across Final Schooling Levels 19

3.3 Average Marginal Treatment Effects 20

3.4 Policy Relevant Treatment Effects 21

4 Identification and Model Likelihood **21**

5 Our Data, A Benchmark OLS Analysis of the Outcomes We Study and Our Exclusion Restrictions **23**

5.1 Exclusion Restrictions 24

6 Estimated Causal Effects **25**

6.1 The Estimated Causal Effects of Educational Choices 27

6.1.1 Dynamic Treatment Effects 28

6.1.2 Continuation Values 31

6.1.3 The Effects on Cognitive and Noncognitive Endowments on Treatment Effects 32

6.2 Distributions of Treatment Effects 33

6.3 Interpreting the Estimated Treatment Effects 34

7 Policy Simulations from our Model **36**

7.1 Policy Relevant Treatment Effects 37

7.2 Boosting Cognitive and Noncognitive Endowments 39

8 Robustness and Comparison of Our Estimates with those Obtained from Other Methods **40**

8.1 Testing the Two Factor Assumption 41

8.2 Comparisons with Alternative Treatment Effect Estimators 42

9 Summary and Conclusion **44**

1 Introduction

In his pioneering research on human capital, Gary Becker (1962; 1964) identified the rate of return to education as a central policy parameter. He launched an active industry on estimating rates of return.[1]

Becker focused on internal rates of return that equate the discounted values of the earnings streams associated with different levels of education. He noted that the full return to schooling also includes nonmarket benefits and nonpecuniary costs. Individuals should continue schooling as long as their marginal internal rate of return exceeds their marginal opportunity cost of funds. If the social return exceeds the social opportunity cost of funds, there is aggregate under-investment in education.

Formidable challenges are faced in estimating internal rates of return: (a) lifetime earnings profiles are required; (b) observed earnings profiles are subject to selection bias; and (c) quantifying nonmarket benefits and nonpecuniary costs is a difficult task. In a neglected paper, Becker and Chiswick (1966) addressed challenge (a) and developed a tractable framework for measuring rates of return to schooling that utilizes cross-section synthetic cohort data on earnings to approximate life cycle earnings data. Mincer (1974) improved on this model by adding work experience. The "Mincer Equation" has become the workhorse of the empirical literature on estimating rates of return:

$$\ln Y(S_i, X_i) = \alpha_i + \rho_i \underbrace{S_i}_{\substack{\text{years of}\\ \text{schooling}}} + \phi(\underbrace{X_i}_{\substack{\text{work}\\ \text{experience}}}) \tag{1}$$

where $Y(S_i, X_i)$ is the earnings of individual i with S_i years of schooling and work experience X_i, α_i is an "ability to earn" parameter that is common across all schooling levels and ρ_i is the "rate of return" to schooling for person i that is assumed to vary among individuals.

Equation (1) and its variants have become the standard framework for estimating rates

[1] For surveys of this literature, see, e.g., Card (1999, 2001); Heckman, Lochner, and Todd (2006); Oreopoulos and Salvanes (2011); McMahon (2009), and Oreopoulos and Petronijevic (2013).

of return.[2] While ρ_i is not, in general, an internal rate of return for individual i, it is the causal effect of an increase in one year of schooling on log earnings from any base state of schooling holding α_i and X_i fixed.[3]

ρ_i ignores the continuation values arising from the dynamic sequential nature of the schooling decision where information is updated and schooling at one stage opens up options for schooling at later stages.[4] Given the functional form imposed by the Mincer Model, ρ_i would be one piece needed to construct continuation values, but this has not been the focus of most previous empirical studies. The distribution of ρ_i and its correlation with S_i have become central targets of empirical studies of the causal effects of education. A positive correlation is consistent with a meritocratic society. People who benefit from schooling get more of it. A negative correlation indicates problems with access to schooling.

Two approaches have been developed to address challenge (b) and estimate rates of return in the general case where ρ_i is correlated with S_i (sorting bias) and S_i is correlated with α_i (ability bias). They are: (I) structural models that jointly analyze outcomes and schooling choices,[5] and (II) treatment effect models that use instrumental variables methods (including randomization and regression discontinuity methods) as well as matching on observed variables to identify "causal parameters."[6]

The structural approach explicitly models agent decision rules. It uses a variety of sources of identification including exclusion restrictions (instrumental variables), conditional independence assumptions on unobservables and functional form assumptions (see, e.g., Blevins, 2014). The final two sources of identification are often controversial. The structural approach identifies the margins of choice identified by instruments and can evaluate the

[2]See, e.g., Cutler and Lleras-Muney (2010) who apply model (1) to estimate the causal effect of education on health.

[3]The stringent conditions under which ρ_i is an internal rate of return and evidence that they are not satisfied in many commonly used samples are presented in Heckman, Lochner, and Todd (2006).

[4]Weisbrod (1962) first raised this point. There is later work by e.g.,Z Comay, Melnik, and Pollatschek (1973); Altonji (1993); Cameron and Heckman (1993), and Eisenhauer, Heckman, and Mosso (2015).

[5]See e.g., Willis and Rosen (1979); Keane and Wolpin (1997); Eisenhauer, Heckman, and Mosso (2015).

[6]See, e.g., Dothan and Williams (1981); Angrist and Imbens (1995); Angrist and Pischke (2009) for IV, and Heckman, Ichimura, and Todd (1998).

impacts of different policies never previously implemented.

The instrumental variable approach is agnostic about agent decision rules and relies on exclusion restrictions to identify its estimands.[7] This approach is often more transparent in securing identification than is the structural approach. However, the economic interpretation of its estimands is often obscure. In a model with multiple levels of schooling, LATE often does not identify the separate margins of choice traced out by instruments or the subpopulations affected by them. Its estimands are irrelevant for addressing policy questions except when the variation induced by the instruments corresponds closely to variation induced by the policies of interest.[8]

This paper develops and applies a methodology that offers a middle ground between the reduced form treatment approach and the fully structural dynamic discrete choice approach. Like the structural literature, we estimate causal effects at clearly identified margins of choice. Our methodology identifies which agents are affected by instruments as well as which agents would be affected by alternative policies never previously implemented. Like the treatment effect literature, we are agnostic about the precise rules used by agents to make decisions. Unlike that literature, we recognize the possibility that people somehow make decisions and account for the consequences of their choices. We approximate agent decision rules and do not impose the cross-equation restrictions that are the hallmark of the structural approach.[9]

Using a generalized Roy framework, we estimate a multistage sequential model of educational choices and their consequences.[10] An important feature of our model is that educational

[7]Instrumental variables still requires assumptions about the validity of the instrument. If there are heterogeneous treatment effects we need additional assumptions such as monotonicity to interpret IV estimates. See Angrist and Imbens (1995); Angrist and Pischke (2009) for details.

[8]See Heckman (2010).

[9]Such approximations are discussed in Heckman (1981), Eckstein and Wolpin (1989), Cameron and Heckman (2001), and Geweke and Keane (2001).

[10]Our approach is related to the analyses of Heckman and Vytlacil (1999, 2005, 2007a,b), Carneiro, Heckman, and Vytlacil (2010, 2011), and Eisenhauer, Heckman, and Vytlacil (2015), who introduce choice theory into the instrumental variables literature. They focus their analysis on binary choice models but also analyze ordered and unordered choice models with multiple outcomes to estimate economically interpretable treatment effects. Expanding on that body of research, we consider multiple sources of identification besides instrumental variables, and link our analysis more closely than they do to the dynamic discrete choice literature.

choices at one stage open up educational options at later stages. Each educational decision is approximated using a reduced form discrete choice model. While not necessary for identification, we assume correlations between decisions, measures and outcomes can be captured by a latent factor structure. The anticipated consequences of future choices and their costs are *implicitly* valued by individuals when deciding whether or not to continue their schooling. Our model approximates a dynamic discrete choice model without taking a stance on exactly what agents are maximizing or how their information sets are being updated.

Like structural models, our model is identified though multiple sources of variation. Drawing from the matching literature, we identify the causal effects of schooling at different stages of the life cycle by using a rich set of observed variables and by proxying unobserved endowments. Unlike previous work on matching, we correct our proxies for measurement error and the bias introduced into the measurements by family background. We can also use exclusion restrictions to identify our model as in the IV and control function literatures. Unlike many structural models, we provide explicit proofs of model identification.

Our framework allows for *ex-ante* valuations as in dynamic discrete choice models but does not explicitly identify them.[11] However, we can estimate *ex-post* returns to schooling, and model how they depend on both observed and unobserved variables. We decompose the *ex-post* treatment effects into (i) the direct benefits of going from one level of schooling to the next[12] and (ii) continuation values arising from access to additional education beyond the immediate next step.

Estimating our model on NLSY79 data, we investigate foundational issues in human capital theory. We report the following findings. (1) Ability bias accounts for a substantial portion (ranging between a third and two-thirds) of the raw differences in outcomes classified by education. At the same time, there are substantial causal effects of education on earnings

[11] See, e.g., Eisenhauer, Heckman, and Mosso (2015), where this is done.

[12] The human capital literature traditionally focused on the direct causal benefits of one final schooling level compared to another, but makes sequential comparisons from the lowest levels of schooling to the highest (Becker, 1964)

and health.[13] (2) Estimated causal effects differ by schooling level and depend on observed and unobserved characteristics of individuals. While the returns to high school are roughly the same across endowment levels, only high-endowment individuals benefit from college graduation. There is positive sorting on gains ("sorting bias" or "pursuit of comparative advantage") only at higher educational levels, but there is sorting into schooling based on observed and unobserved variables in earnings equations across all schooling levels ("ability bias").

(3) The early literature ignored the dynamics of schooling decisions. We find that continuation values arising from sequential choices are empirically important. Continuation values depend on cognitive and noncognitive endowments. Low endowment individuals gain mostly from the direct effect of high school graduation while high endowment individuals gain mostly in terms of continuation values. Low endowment individuals do not benefit from graduating college.

(4) Our schooling choice model is consistent with a variety of decision rules and allows for time inconsistency, regret and systematic mistakes due to cognitive failures. We use model estimates to test the assumptions of forward looking behavior and selection on gains often assumed in estimating dynamic discrete choice models.[14] We find that agents do not know, or act on, publicly available information on college tuition costs in making decisions about graduating high school. Nonetheless, agents sort into schooling on *ex-post* gains, especially at higher schooling levels. A core tenet of human capital theory is thus confirmed.

(5) Our paper contributes to an emerging literature on the importance of both cognitive and noncognitive endowments in shaping life outcomes.[15] Consistent with the recent literature, we find that both cognitive and noncognitive endowments are important predictors of educational attainment. Within schooling levels, cognitive and noncognitive endowments have additional

[13] This finding runs counter to a common interpretation in the literature based on comparing IV and OLS estimates of Equation (1). See, e.g., Griliches (1977) and Card (1999, 2001).

[14] See e.g., Rust (1994); Keane and Wolpin (1997); Blevins (2014).

[15] See, e.g., Borghans, Duckworth, Heckman, and ter Weel (2008); Heckman, Stixrud, and Urzúa (2006); Almlund, Duckworth, Heckman, and Kautz (2011).

impacts on most outcomes.[16]

(6) We meet challenge (c) and estimate substantial causal effects of education on health and healthy behaviors in addition to its large effects on wages.[17]

Using our estimated model, we conduct two policy experiments. In the first, we examine the impact of a tuition subsidy on college enrollment. We identify who is impacted by the policy, how their decisions change, and how much they benefit. Those induced to enroll benefit from the policy, and many go on to graduate from college. In a second experiment, we exploit the structural properties of our model. We analyze a policy that improves the endowments of those at the bottom of the distribution to see how this impacts educational choices and outcomes. Such improvements are produced by early intervention programs.[18] Increasing cognitive endowments has a positive impact on all outcomes, while increasing noncognitive endowments mostly impacts health outcomes.

This paper proceeds in the following way. Section 2 presents our model. Section 3 presents the economically interpretable treatment effects that can be derived from it. Section 4 discusses identification. Section 5 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 6 reports our estimated treatment effects and their implications. Section 7 uses the estimated model to address two policy-relevant questions. Section 8 considers the robustness of our estimates to alternative methodological approaches such as OLS and matching. Section 9 concludes.

[16] Our estimates of the causal effects of education do *not* require that we separately isolate the effects of individual cognitive and noncognitive endowments on outcomes, just that we control for them as a set.

[17] There is a small, but growing literature on this topic. See Grossman (2000); McMahon (2000); Lochner (2011); Oreopoulos and Salvanes (2011); Cutler and Lleras-Muney (2010). For a review of this literature see Web Appendix A.1.

[18] Heckman, Pinto, and Savelyev (2013).

2 Model

This paper estimates a multistage sequential model of educational choices with transitions and decision nodes shown in Figure 1. Let $\mathcal{J}$ denote a set of possible terminal states. At each node there are only two possible choices: remain at j or transit to the next node ($j+1$ if $j \in \{1, \ldots, \overline{s}-1\}$). $D_j = 0$ if a person at j does not stop there and goes on to the next node. $D_j = 1$ if the person stops at j for $j \neq 0$. $D_0 = 1$ opens an additional branch of the decision tree. For $D_0 = 1$, we define the attainable sets as $\{0, G\}$. Thus, a person may remain a dropout or may get the GED.[19] Thus, in the lower branch ($D_0 = 1$), agents can terminate as a dropout ($D_0 = 1, D_G = 1$) or as a dropout who exam certifies ($D_0 = 1, D_G = 0$). $D_j \in \mathcal{D}$ is the set of possible transition decisions that can be taken by the individual over the decision horizon. Let $\mathcal{S} = \{G, 0, \ldots, \overline{s}\}$ denote the finite and bounded set of stopping states with $S = s$ if the agent stops at $s \in \mathcal{S}$, so for example $D_s = 1$ for $s \in \mathcal{S}\backslash\{0, G\}$. Define $\overline{s}$ as the highest attainable element in $\mathcal{S}$. We assume that the environment is time-stationary and decisions are irreversible.[20]

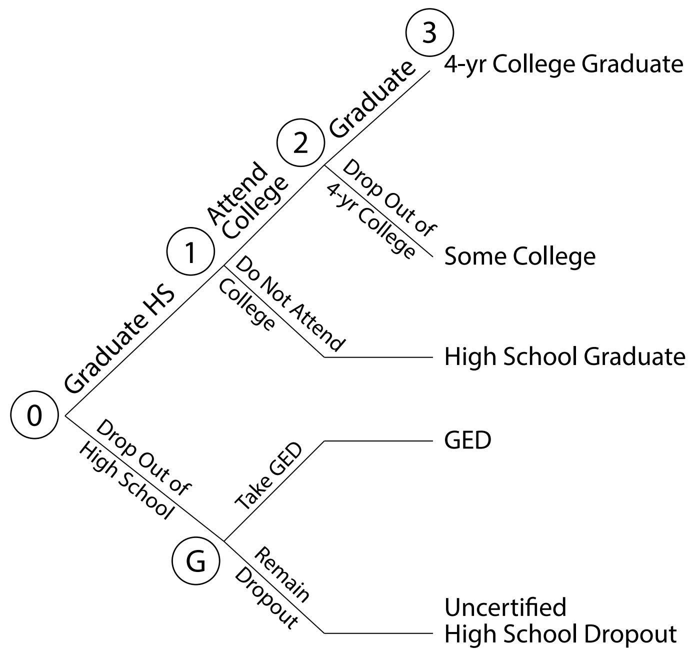
$Q_j = 1$ indicates that an agent *gets to* decision node j. $Q_j = 0$ if the person never gets there. $Q_G = 1$ if the agent drops out of high school and confronts the GED option. The history of nodes visited by an agent can be described by the collection of the Q_j such that $Q_j = 1$. Observe that $D_s = 1$ and $D_{s-1} = 0$ are equivalent to $S = s$ for $s \in \{1, \ldots, \overline{s}\}$ and $D_{\overline{s}} = 1$ if $D_j = 0, \forall j \in \mathcal{S}$.[21] Finally, $D_0 = 1$ and $D_G = 0$ is equivalent to $S = G$.

[19] The GED is an exam whose proponents claim that successful examinees are the equivalents of high school graduates. For strong evidence to the contrary, see Heckman, Humphries, and Kautz (2014b).

[20] This model is also analyzed in Cunha, Heckman, and Navarro (2007) and in Heckman and Navarro (2007).

[21] For notational convenience, we assign $D_j = 0$ for all $j > s$.

Figure 1: A Multistage Dynamic Decision Model



2.1 A Sequential Decision Model

The decision process at each node is characterized by an index threshold-crossing property:

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

where I_j is the *perceived* value at node j of going on to the next node for a person at node j. The requirement $Q_j = 1$ ensures that agents are able to make the transition at j.

Associated with each final state $s \in \mathcal{S} = \{G, 0, \ldots, \overline{s}\}$ is a set of K_s potential outcomes for each agent with indices $k \in \mathcal{K}_s$. We define $\tilde{Y}_s^k$ as latent variables that map into potential outcomes Y_s^k:

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

Using the switching regression framework of Quandt (1958, 1972), the observed outcome Y^k for a k common across transitions is

$$Y^k = \left(\sum_{\mathcal{S} \setminus \{0,G\}} D_s Y_s^k \right) (1 - D_0) + \left(Y_0^k D_G + Y_G^k (1 - D_G) \right) D_0. \tag{4}$$

2.2 Parameterizations of the Decision Rules and Potential Outcomes for Final States

Following a well-established tradition in the treatment effect and structural literatures, we approximate I_j using a separable model:

$$I_j = \boldsymbol{\phi}_j \underbrace{(\boldsymbol{Z})}_{\substack{\text{Observed} \\ \text{by analyst}}} - \underbrace{\eta_j}_{\substack{\text{Unobserved} \\ \text{by analyst}}}, \ j \in \{G, 0, \ldots, \overline{s} - 1\} \tag{5}$$

where $\boldsymbol{Z}$ is a vector of variables observed by the analyst, components of which determine the transition decisions of the agent at different stages and η_j is unobserved by the analyst. A separable representation of the choice rule is an essential feature of LATE (Vytlacil, 2002) and dynamic discrete choice models (Blevins, 2014).

Outcomes are also separable:

$$\tilde{Y}_s^k = \tau_s^k \underbrace{(\boldsymbol{X})}_{\substack{\text{Observed} \\ \text{by analyst}}} + \underbrace{U_s^k}_{\substack{\text{Unobserved} \\ \text{by analyst}}}, \quad k \in \mathcal{K}_s, \quad s \in \mathcal{S}, \tag{6}$$

where $\boldsymbol{X}$ is a vector of observed determinants of outcomes and U_s^k is unobserved by the analyst.[22] Separability of the unobserved variables in the outcome equations is often invoked in the structural literature but is not strictly required (see Blevins, 2014).[23] It is not required in the IV literature.

2.3 Structure of the Unobservables

Central to our main empirical strategy is the existence of a finite dimensional vector $\boldsymbol{\theta}$ of unobserved (by the economist) endowments that generate all of the dependence across the η_j and the U_s^k. We assume that

$$\eta_j = -(\boldsymbol{\theta}'\boldsymbol{\alpha_j} - \nu_j,) \quad j \in \{G, 0, \ldots, \overline{s} - 1\} \tag{7}$$

and

$$U_s^k = \boldsymbol{\theta}'\alpha_s^k + \omega_s^k, \quad k \in \mathcal{K}_s, s \in S, \tag{8}$$

where ν_j is an idiosyncratic error term for transition j.

Conditional on $\boldsymbol{\theta}, \boldsymbol{X}, \boldsymbol{Z}$, choices and outcomes are statistically independent. Thus controlling for this set of variables eliminates selection effects. If the analyst knew $\boldsymbol{\theta}$, he/she could use matching to identify the model.[24]

Standard "random effects" approaches in the structural literature integrate out $\boldsymbol{\theta}$ and do not interpret it. Our approach is different. We proxy $\boldsymbol{\theta}$ using multiple measurements of it and we identify, and correct for, errors in the proxy variables. The measurements facilitate the interpretation of $\boldsymbol{\theta}$. We develop this intuition further in Section 4, after presenting the rest of our model.

We array the $\nu_j, j \in \mathcal{J}$, into a vector $\boldsymbol{\nu} = (\nu_G, \nu_0, \nu_1, \ldots, \nu_{\overline{s}-1})$ and the η_j into $\boldsymbol{\eta} = (\eta_G, \eta_0, \ldots, \eta_{\overline{s}-1})$. ω_s^k represents an idiosyncratic error term for outcome k in state s. Array the ω_s^k into a vector $\boldsymbol{\omega}_s = (\omega_s^1, \ldots, \omega_s^{Ks})$. Array the U_s^k into vector $\boldsymbol{U_s} = (U_s^1, \ldots, U_s^{Ks})$ and

[22]In practice $\boldsymbol{X}$ and $\boldsymbol{Z}$ can vary by decision or outcome. See Table 1 for details.
[23]Moreover, we can condition on observable covariates $\boldsymbol{X}$.
[24]See Carneiro, Hansen, and Heckman (2003).

array the $\boldsymbol{U_s}$ into $\boldsymbol{U} = (\boldsymbol{U}_G, \boldsymbol{U}_0, \ldots, \boldsymbol{U}_{\bar{s}})$.

Letting "$\perp\!\!\!\perp$" denote statistical independence, we assume that *conditional on* $\boldsymbol{X}$

$$\nu_j \perp\!\!\!\perp \nu_l \quad \forall l \neq j \quad l, j \in \{G, 0, \ldots, \bar{s}-1\} \tag{A-1a}$$

$$\omega_s^k \perp\!\!\!\perp \omega_{s'}^k \quad \forall s \neq s' \quad \forall k \tag{A-1b}$$

$$\boldsymbol{\omega}_s \perp\!\!\!\perp \boldsymbol{\nu}, \quad \forall s \in \mathcal{S} \tag{A-1c}$$

$$\boldsymbol{\theta} \perp\!\!\!\perp \boldsymbol{Z} \tag{A-1d}$$

$$(\boldsymbol{\omega}_s, \boldsymbol{\nu}) \perp\!\!\!\perp (\boldsymbol{\theta}, \boldsymbol{Z}) \quad \forall s \in \mathcal{S} \tag{A-1e}$$

Assumption (A-1a) maintains independence of the shocks affecting transitions; (A-1b) independence of shocks across all states; (A-1c) independence of the shocks to transitions and the outcomes; (A-1d) independence of $\boldsymbol{\theta}$ with respect to the observables; and (A-1e) independence of the shocks and the factors with the observed variables. Versions of assumptions (A-1d) and (A-1e) play fundamental roles in the structural dynamic discrete choice literature. For example, the widely-used "types" assumption of Keane and Wolpin (1997) postulates conditional independence between choices and outcomes conditional on types ($\boldsymbol{\theta}$) that operate through the initial conditions of their model.

2.4 Measurement System for Unobserved Factors $\boldsymbol{\theta}$

We allow for the possibility that $\boldsymbol{\theta}$ cannot be measured precisely, but that it can be proxied with multiple measurements. We correct for the effects of measurement error in the proxy. The structural literature treats the $\boldsymbol{\theta}$ as nuisance variables, invokes conditional independence assumptions, and integrates $\boldsymbol{\theta}$ out using random effect procedures.[25] Instead, we link $\boldsymbol{\theta}$ to measurements, and adjoin measurement equations to choice and outcome equations, rendering $\boldsymbol{\theta}$ interpretable.

Let $\boldsymbol{M}$ be a vector of N_M measurements on $\boldsymbol{\theta}$. They may consist of lagged or future

[25] See e.g., Keane and Wolpin (1997); Rust (1994); Adda and Cooper (2003); Blevins (2014).

values of the outcome variables or additional measurements.[26] The system of equations determining $\boldsymbol{M}$ is:

$$\boldsymbol{M} = \Phi(\boldsymbol{X}, \boldsymbol{\theta}, \boldsymbol{e}), \tag{9}$$

where $\boldsymbol{X}$ are observed variables, $\boldsymbol{\theta}$ are the factors and

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

where we array the e_j into $\boldsymbol{e} = (e_1, \ldots, e_{N_M})$. We assume in addition to the previous assumptions that *conditional on* X

$$e_j \perp\!\!\!\perp e_l, \quad j \neq l, \quad j, l \in \{1, \ldots, N_M\} \tag{A-1g}$$

$$\text{and} \qquad \boldsymbol{e} \perp\!\!\!\perp (\boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{\theta}, \boldsymbol{\nu}, \boldsymbol{\omega}). \tag{A-1h}$$

For the purpose of identifying treatment effects, we do not need to identify each equation of system (9). We just have to identify the span of $\boldsymbol{\theta}$ that preserves the information on $\boldsymbol{\theta}$ in (9), and that is sufficient to produce conditional independence between choices and outcomes.[27] However, in this paper we estimate equation system (9).

3 Defining Treatment Effects

A variety of *ex post* counterfactual outcomes and associated treatment effects can be generated from our model. They can be used to predict the effects of manipulating education levels through different policies for people of different backgrounds and abilities. They allow us to understand the effectiveness of policies for different identifiable segments of the population, and the benefits to people at different margins of choice.

[26] See, e.g., Abbring and Heckman (2007); Schennach, White, and Chalak (2012).

[27] See e.g., Heckman, Schennach, and Williams (2013).

In principle, we can define and estimate a variety of treatment effects, many of which are implausible. For example, many empirical economists would not find estimates of the effect of fixing (manipulating) $D_j = 0$ if $Q_j = 0$ to be credible (i.e., the person for whom we fix $D_j = 0$ is not at the decision node to take the transition).[28] In the spirit of credible econometrics, we define treatment effects associated with fixing $D_j = 0$ conditioning on $Q_j = 1$. This approach blends structural and treatment effect approaches. Our causal parameters recognize agent heterogeneity and are allowed to differ across populations, contrary to standard approaches in structural econometrics.[29]

The person-specific treatment effect T_j^k for outcome k for an individual selected from the population $Q_j = 1$ with characteristics $\boldsymbol{X} = \boldsymbol{x}, \boldsymbol{Z} = \boldsymbol{z}, \boldsymbol{\theta} = \overline{\boldsymbol{\theta}}$, making a decision at node j between going on to the next node or stopping at j is the difference between the individual's outcomes under the two actions. This can be written as

$$\begin{aligned} T_j^k[Y^k|\boldsymbol{X} = \boldsymbol{x}, \boldsymbol{Z} = \boldsymbol{z}, \boldsymbol{\theta} = \overline{\boldsymbol{\theta}}] :=&(Y^k|\boldsymbol{X} = \boldsymbol{x}, \boldsymbol{Z} = \boldsymbol{z}, \boldsymbol{\theta} = \overline{\boldsymbol{\theta}}, Q_j = 1, Fix\ D_j = 0) \\ &- (Y^k|\boldsymbol{X} = \boldsymbol{x}, \boldsymbol{Z} = \boldsymbol{z}, \boldsymbol{\theta} = \overline{\boldsymbol{\theta}}, Q_j = 1, Fix\ D_j = 1). \end{aligned} \tag{10}$$

The random variable $(Y^k|\boldsymbol{X} = \boldsymbol{x}, \boldsymbol{Z} = \boldsymbol{z}, \boldsymbol{\theta} = \overline{\boldsymbol{\theta}}, Fix\ D_j = 0, Q_j = 1)$ is the outcome variable Y^k at node j for a person with characteristics $\boldsymbol{X} = \boldsymbol{x}, \boldsymbol{Z} = \boldsymbol{z}, \boldsymbol{\theta} = \overline{\boldsymbol{\theta}}$ from the population who attain node j (or higher), $Q_j = 1$, and for whom we fix $D_j = 0$ so they go on to the next node. Random variable $(Y^k|\boldsymbol{X} = \boldsymbol{x}, \boldsymbol{Z} = \boldsymbol{z}, \boldsymbol{\theta} = \overline{\boldsymbol{\theta}}, Fix\ D_j = 1, Q_j = 1)$ is defined for the same individual but forces the person with these characteristics not to transit to the next node.

We next present population level treatment effects based on (10). We focus our discussion on means but we can also formulate distributional counterparts for all of the treatment effects considered in this paper.

[28] The distinction between *fixing* and *conditioning* traces back to Haavelmo (1943). White and Chalak (2009) use the terminology "setting" for the same notion. For a recent analysis of this crucial distinction, see Heckman and Pinto (2015).

[29] See, e.g., Hansen and Sargent (1980).

3.1 Dynamic Treatment Effects

A main contribution of this paper is to define and estimate treatment effects that take into account the direct effect of moving to the next node of a decision tree, plus the benefits associated with the further schooling that such movement opens up. This treatment effect is the difference in expected outcomes arising from changing a single educational decision in a sequential schooling model and tracing through its consequences, accounting for the dynamic sequential nature of schooling.

The person-specific treatment effect can be decomposed into two components: the *Direct Effect* of going from j to $j+1$: $DE_j^k = Y_{j+1}^k - Y_j^k$, the effect often featured in the literature on the returns to schooling, and the *Continuation Value* of going *beyond* $j+1$:

$$C_{j+1}^k = \sum_{r=1}^{\overline{s}-(j+1)} \left[\prod_{l=1}^{r} (1 - D_{j+l}) \right] (Y_{j+r+1}^k - Y_{j+r}^k).$$ [30]

The continuation value for the lower branch ($D_0 = 1$) is defined for the attainable set $\{0, G\}$. Essentially, G is the only option available to a high school dropout in our model.

Thus, at the individual level, the *Total Effect* of fixing $D_j = 0$ on Y^k is decomposed into

$$T_j^k = DE_j^k + C_{j+1}^k. \tag{11}$$

The associated population level average treatment effect conditional on $Q_j = 1$ is

$$ATE_j^k := \int \dots \int E[T_j^k(Y^k | \boldsymbol{X} = \boldsymbol{x}, \boldsymbol{Z} = \boldsymbol{z}, \boldsymbol{\theta} = \overline{\boldsymbol{\theta}})] \, dF_{\boldsymbol{X},\boldsymbol{Z},\boldsymbol{\theta}}(\boldsymbol{x}, \boldsymbol{z}, \overline{\boldsymbol{\theta}} \,|\, Q_j = 1) \tag{12}$$

which can be decomposed into direct and continuation value components.

Integrating over $\boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{\theta}$, conditioning on $Q_j = 1$, the population continuation value at

[30] The relationship between this notion of continuation value and the definition in the dynamic discrete choice literature is explored in Web Appendix A.3.

$j+1$ is

$$E_{\boldsymbol{X},\boldsymbol{Z},\boldsymbol{\theta}}(C^k_{j+1}) = E_{\boldsymbol{X},\boldsymbol{Z},\boldsymbol{\theta}}\Bigg[\sum_{l=j+1}^{\overline{s}-1}\Big\{E(Y^k_{l+1} - Y^k_l|\boldsymbol{X}=\boldsymbol{x},\boldsymbol{Z}=\boldsymbol{z},\boldsymbol{\theta}=\overline{\boldsymbol{\theta}},Q_{l+1}=1,\ Fix\ Q_{j+1}) \cdot Pr(Q_{l+1}=1|Fix\ Q_{j+1}=1,\boldsymbol{X}=\boldsymbol{x},\boldsymbol{Z}=\boldsymbol{z},\boldsymbol{\theta}=\overline{\boldsymbol{\theta}},Q_j=1)\Big\}|Q_j=1\Bigg]. \quad (13)$$

where $Q_{\overline{s}} = 1$ if $S = \overline{s}$.

We can also specify population distributions of total effects as in Heckman, Smith, and Clements (1997):[31]

$$Pr(T^k_j < t^k_j|\boldsymbol{X}=\boldsymbol{x},\boldsymbol{Z}=\boldsymbol{z},\boldsymbol{\theta}=\overline{\boldsymbol{\theta}},Q_j=1) \quad (14)$$

with population distribution counterpart

$$E_{\boldsymbol{X},\boldsymbol{Z},\boldsymbol{\theta}}\left[Pr(T^k_j < t^k_j|Q_j=1)\right] \quad (15)$$

which can be decomposed into the distribution of direct effects and continuation values. (The modifications for the unordered case require that we define these terms over the admissible options available for $D_0 = 1$ or $D_0 = 0$.)

Because we do not specify or attempt to identify choice-node-specific agent information sets, we can only identify *ex-post* treatment effects. Hence, we can identify continuation values associated with choices, but cannot identify option values. However, a benefit of this more agnostic approach is that it does not impose specific decision rules. Our model allows for irrationality, regret, and mistakes in decision-making associated with agent maturation and information acquisition.

3.2 Mean Differences Across Final Schooling Levels

Becker's original approach (1964) can be interpreted to define returns to education as the gains from choosing between a base and a terminal schooling level. Let $Y^k_{s'}$ be outcome k at

[31] See Abbring and Heckman (2007) for a review of the literature.

schooling level s' and Y_s^k be outcome k at schooling level s. Conditioning on $\boldsymbol{X} = \boldsymbol{x}$ and $\boldsymbol{\theta} = \overline{\boldsymbol{\theta}}$, the average treatment effect of s compared to s' is $E(Y_s^k - Y_{s'}^k | \boldsymbol{X} = \boldsymbol{x}, \boldsymbol{Z} = \boldsymbol{z}, \boldsymbol{\theta} = \overline{\boldsymbol{\theta}})$. Integrating out $\boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{\theta}$ produces a pairwise ATE parameter over the available supports of these variables.

A more empirically credible version, and the one we report here, calculates the mean gain for the subset of the population that completes one of the two final schooling levels:

$$ATE_{s,s'}^k \equiv \iiint E(Y_s^k - Y_{s'}^k | \boldsymbol{X} = \boldsymbol{x}, \boldsymbol{Z} = \boldsymbol{z}, \boldsymbol{\theta} = \overline{\boldsymbol{\theta}}) \, dF_{\boldsymbol{X},\boldsymbol{Z},\boldsymbol{\theta}}(\boldsymbol{x}, \boldsymbol{z}, \overline{\boldsymbol{\theta}} \,|\, S \in \{s, s'\}). \tag{16}$$

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 be far away from having any empirical or policy relevance.[32]

3.3 Average Marginal Treatment Effects

In order to understand treatment effects for persons at the margin of indifference at each node of the decision tree of Figure 1, we estimate the Average Marginal Treatment Effect (AMTE).[33] It is the average effect of transiting to the next node for individuals at the margin of indifference between the two nodes:

$$AMTE_j^k := \iiint E\left[T_j^k\left(Y^k | \boldsymbol{X} = \boldsymbol{x}, \boldsymbol{Z} = \boldsymbol{z}, \boldsymbol{\theta} = \overline{\boldsymbol{\theta}}\right)\right] dF_{\boldsymbol{X},\boldsymbol{Z},\boldsymbol{\theta}}(\boldsymbol{x}, \boldsymbol{z}, \bar{\boldsymbol{\theta}} \mid Q_j = 1, |I_j| \leq \varepsilon), \tag{17}$$

where ε is an arbitrarily small neighborhood around the margin of indifference. These effects are inclusive of all consequences of taking the transition at j, including the possibility of attaining final schooling levels well beyond j. AMTE defines causal effects at well-defined and empirically identified margins of choice. It is the proper measure of the marginal gross benefit for evaluating the gains from moving from one stage of the decision tree to the next

[32] The estimated differences in treatment effects for the conditional and unconditional population are not large for outcomes associated with the decision to enroll in college, but is substantial for the choice to graduate from college. See Tables A56, A58, A60, and A62 in the Web Appendix.

[33] See Carneiro, Heckman, and Vytlacil (2010, 2011).

for those at that margin of choice. In general it is distinct from LATE, which is not defined for any specific margin of choice.[34] Since we identify the distribution of I_j, we can identify the characteristics of agents in the indifference set, something not possible using IV or matching.

The population distribution counterpart of AMTE is defined over the set of agents for whom $| I_j | \leq \varepsilon$, which can be generated from our model: $Pr(T_j^k < t_j^k | Q_j = 1, |I_j| \leq \varepsilon)$. Distributional versions can be defined for all of the treatment effects considered in this section.

3.4 Policy Relevant Treatment Effects

The policy relevant treatment effect (PRTE) is the average treatment effect for those induced to change their choices in response to a particular policy intervention. Let $Y^k(p)$ be the aggregate outcome under policy p for outcome k. Let $S(p)$ be the final state selected by an agent under policy p. The policy relevant treatment effect from implementing policy p compared to policy p' for outcome k is:

$$PRTE_{p,p'}^k := \iiint E(Y^k(p) - Y^k(p') | \boldsymbol{X} = \boldsymbol{x}, \boldsymbol{Z} = \boldsymbol{z}, \boldsymbol{\theta} = \overline{\boldsymbol{\theta}}) dF_{\boldsymbol{X},\boldsymbol{Z},\boldsymbol{\theta}}(\boldsymbol{x}, \boldsymbol{z}, \overline{\boldsymbol{\theta}} | S(p) \neq S(p')), \quad (18)$$

where $S(p) \neq S(p')$ denotes the set of the characteristics of people for whom attained states differ under the two policies. In general, it is different from AMTE because the agents affected by a policy can be at multiple margins of choice. PRTE is often confused with LATE. In general, they are different unless the proposed policy change coincides with the instrument used to define LATE.[35]

4 Identification and Model Likelihood

The treatment effects defined in Section 3 can be identified using alternative empirical approaches. The main approach used in this paper exploits the fact that conditional on

[34] See Heckman and Vytlacil (2007a) and Carneiro, Heckman, and Vytlacil (2010). The LATE can correspond to people at multiple margins. See Angrist and Imbens (1995).

[35] See Carneiro, Heckman, and Vytlacil (2011) for an empirical example. The differences between the two parameters can be substantial as we show in Web Appendix A.5.2.

$\boldsymbol{\theta}, \boldsymbol{X}, \boldsymbol{Z}$, outcomes and choices are statistically independent. $\boldsymbol{X}$ and $\boldsymbol{Z}$ are observed. $\boldsymbol{\theta}$ is not. If $\boldsymbol{\theta}$ were observed, one could condition on $\boldsymbol{\theta}, \boldsymbol{X}, \boldsymbol{Z}$ and identify the model of Equations (2) - (8) and the treatment effects that can be generated from it. We use nonlinear factor model (9) to proxy $\boldsymbol{\theta}$.

Under the conditions presented in Web Appendix A.4, we can nonparametrically identify the model of Equations (2) - (8) including the distribution of $\boldsymbol{\theta}$, as well as the Φ functions and the distribution of $\boldsymbol{e}$ (which can be interpreted as measurement errors). Effectively, we match on proxies for $\boldsymbol{\theta}$ and correct for the effects of measurement error ($\boldsymbol{e}$) in creating the proxies. Such corrections are possible because with multiple measures on $\boldsymbol{\theta}$ we can identify the distribution of $\boldsymbol{e}$.

Under full linearity assumptions, one can directly estimate the $\boldsymbol{\theta}$ and use factor regression methods.[36] Full details of this approach are spelled out in Web Appendix A.4.[37] Another approach to identification uses instrumental variables which, if available, under the conditions presented in Web Appendix A.4 can be used to identify the structural model (2) - (8) without factor structure (7) and (8).

The precise parameterization and the likelihood function for the model we estimate is presented in Web Appendix A.6. While in principle it is possible to identify the model semi-parametrically, in this paper we make parametric assumptions in order facilitate computation. We subject the estimated model to rigorous goodness of fit tests which we pass.[38]

[36] See, e.g., Heckman, Pinto, and Savelyev (2013) and the references cited therein.

[37] As noted in Web Appendix A.4.1, and Heckman, Schennach, and Williams (2011), we do not need to solve classical identification problems associated with estimating equation system (9) in order to extract measure-preserving transformations of $\boldsymbol{\theta}$ on which we can condition in order to identify treatment effects. In the linear factor analysis literature these are rotation and normalization problems.

[38] See Web Appendix A.7.

5 Our Data, A Benchmark OLS Analysis of the Outcomes We Study and Our Exclusion Restrictions

We estimate our model on a sample of males extracted from the widely used National Longitudinal Sample of Youth (NLSY 79).[39] Before discussing estimates from our model, it is informative to set the stage and 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 four outcomes analyzed in this paper: wages, log present value of wages, health limitations, and smoking.[40]

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 higher earnings 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.

Figure 2 shows that controlling for proxies for ability substantially reduces the observed differences in earnings across educational groups. Nonetheless, there are still strong causal effects of education. It has been claimed that a model that is linear in years of schooling fits the data well.[41] The white bar in Figure 2 displays the estimated adjusted effect of schooling controlling for years of completed schooling as in Equation (1).[42] The white bars in all figures suggest that the linear-in-years-of-schooling Mincer specification (1) does not describe our data. There are effects of schooling beyond those captured by a linear years of schooling specification.

[39] Web Appendix A.8 presents a detailed discussion of the data we analyze and our exclusion restrictions.
[40] Adjustments are made through linear regression.
[41] See e.g., Card (1999, 2001). Heckman, Lochner, and Todd (2006) dispute this claim.
[42] Mis-measurement of schooling is less of a concern in our data as the survey asks numerous educational questions every year which we use to determine an individual's final schooling state.

5.1 Exclusion Restrictions

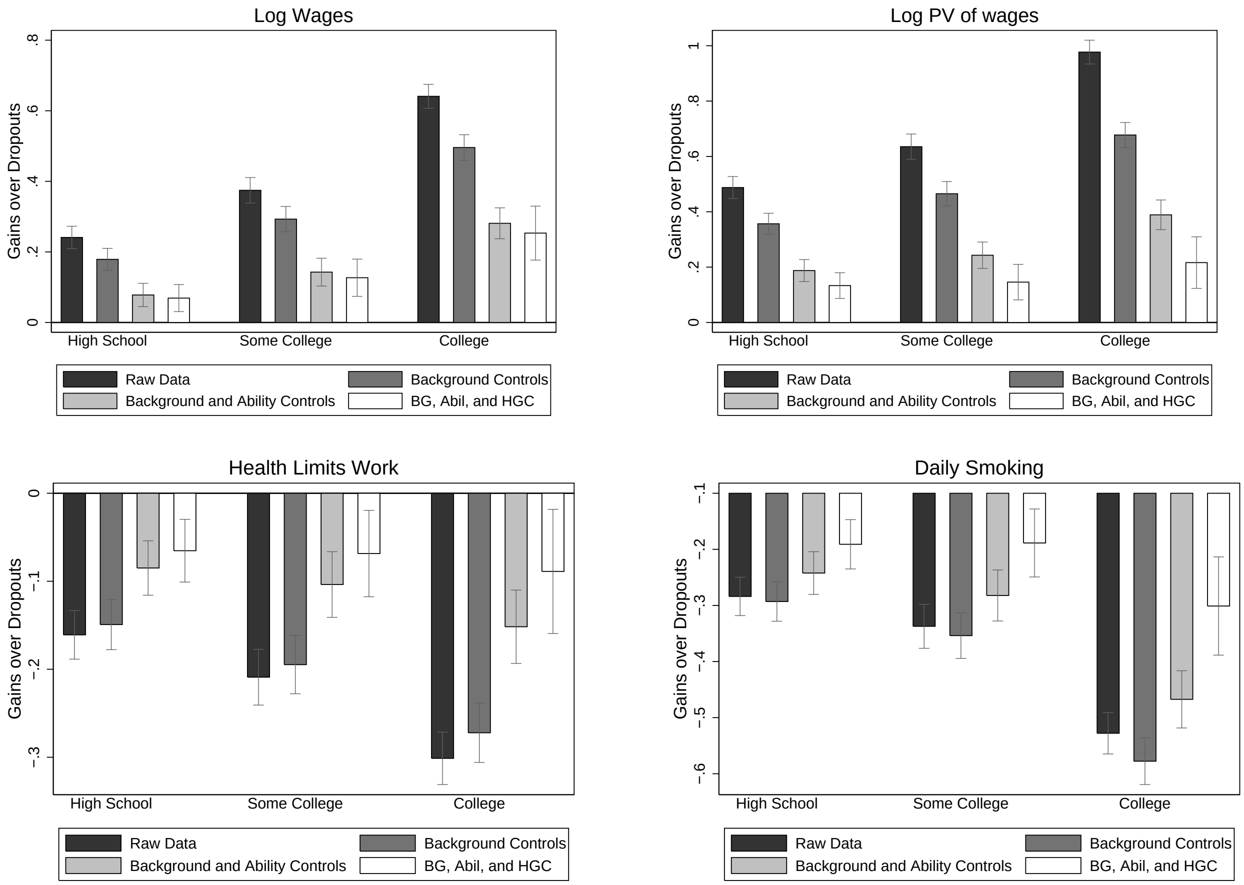
As noted in Section 4, identification does not depend exclusively on conditional independence assumptions associated with our factor model although they alone justify the identification of our model using matching on mismeasured variables.[43] Node-specific instruments can nonparametrically identify treatment effects without invoking the full set of conditional independence assumptions.[44] We have a variety of exclusion restrictions that affect choices but not outcomes. Table 1 documents the $\boldsymbol{X}$ and $\boldsymbol{Z}$ used in this paper. Our instruments are traditional in the literature that estimates the causal effects of education.[45]

[43] See Carneiro, Hansen, and Heckman (2003).

[44] See Web Appendix A.4.

[45] For example, presence of a nearby college or distance to college is used by Cameron and Taber (2004); Kling (2001); Carneiro, Meghir, and Parey (2013); Cawley, Conneely, Heckman, and Vytlacil (1997); Heckman, Carneiro, and Vytlacil (2011); and Eisenhauer, Heckman, and Vytlacil (2015). Local tuition at two or four year colleges is used as an instrument by Kane and Rouse (1993); Heckman, Carneiro, and Vytlacil (2011); Eisenhauer, Heckman, and Vytlacil (2015); and Cameron and Taber (2004). Local labor market shocks are used by Heckman, Carneiro, and Vytlacil (2011) and Eisenhauer, Heckman, and Vytlacil (2015).

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 ninth grade GPA in core subjects (language, math, science, and social science). "Some College" includes anyone who enrolled in college, but did not receive a four-year college degree. The white bar additionally controls for highest grade completed (HGC). *Source:* NLSY79 data.

6 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.[46] We randomly draw sets of regressors from our sample and a

[46]Parameter estimates for individual equations are reported in Web Appendix A.9.

Table 1: Control Variables and Instruments Used in the Analysis

Variables	Measurement Equations	Choice	Outcomes
Race	x	x	x
Broken Home	x	x	x
Number of Siblings	x	x	x
Parents' Education	x	x	x
Family Income (1979)	x	x	x
Region of Residence[a]	x	x	x
Urban Status[a]	x	x	x
Age[b]	x	x	x
Local Unemployment[c]			x
Local Long-Run Unemployment		x	
Instruments			
Local Unemployment at Age 17[d]		x	
Local Unemployment at Age 22[e]		x	
College Present in County 1977[f]		x	
Local College Tuition at Age 17[g]		x	
Local College Tuition at Age 22[h]		x	

Notes: [a]Region and urban dummies are specific to the age that the measurement, educational choice, or outcome occurred. [b]Age in 1979 is included as a cohort control. We also included individual cohort dummies which did not change the results. [c]For economic outcomes, local unemployment at the time the outcome is measured. [d]This is an instrument for choices at nodes 1 and 2. It represents opportunity costs at the time schooling decisions are made. [e]This is an instrument for the choice at node 3. [f]Presence of a 4-year college in the county in 1977 is constructed from Kling (2001) and enters the choice to enroll and the choice to graduate from college. [g]Local college tuition at age 17 only enters the college enrollment graduation decisions. [h]Local college tuition at age 22 only enters the college completion equation. The measurement system includes the arithmetic reasoning, coding speed, paragraph comprehension, word knowledge, mathematical knowledge, and numerical operations sub-tests of the AS VAB, 9th grade GPA in math, english, science, and social studies, and early risky and reckless behavior. We assume ASVAB only loads on the cognitive factor . See Appendix Section A.8 for details.

vector of factors from the estimated factor distribution to simulate the treatment effects.[47] We first present (Section 6.1) the main treatment effects across final schooling levels, by node, and their decomposition into direct and indirect effects. We discuss how endowments affect the treatment effects. We next (Section 6.2) discuss distributions of treatment effects. In Section 6.3 we interpret these estimates for each of the four outcomes studied.

Educational decisions at each node depend on both endowments. In addition we find sorting on gains (comparative advantage) for the college enrollment and college graduation decisions, but not the high school graduation decision.[48] Sorting on levels (ability bias) is

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

[48]This finding generalizes the analysis of Willis and Rosen (1979) to multiple schooling levels.

found across all educational levels. The Mincer model (Equation (1)) does not capture these types of sorting patterns. It overlooks the differences in the distributions of returns across schooling levels.

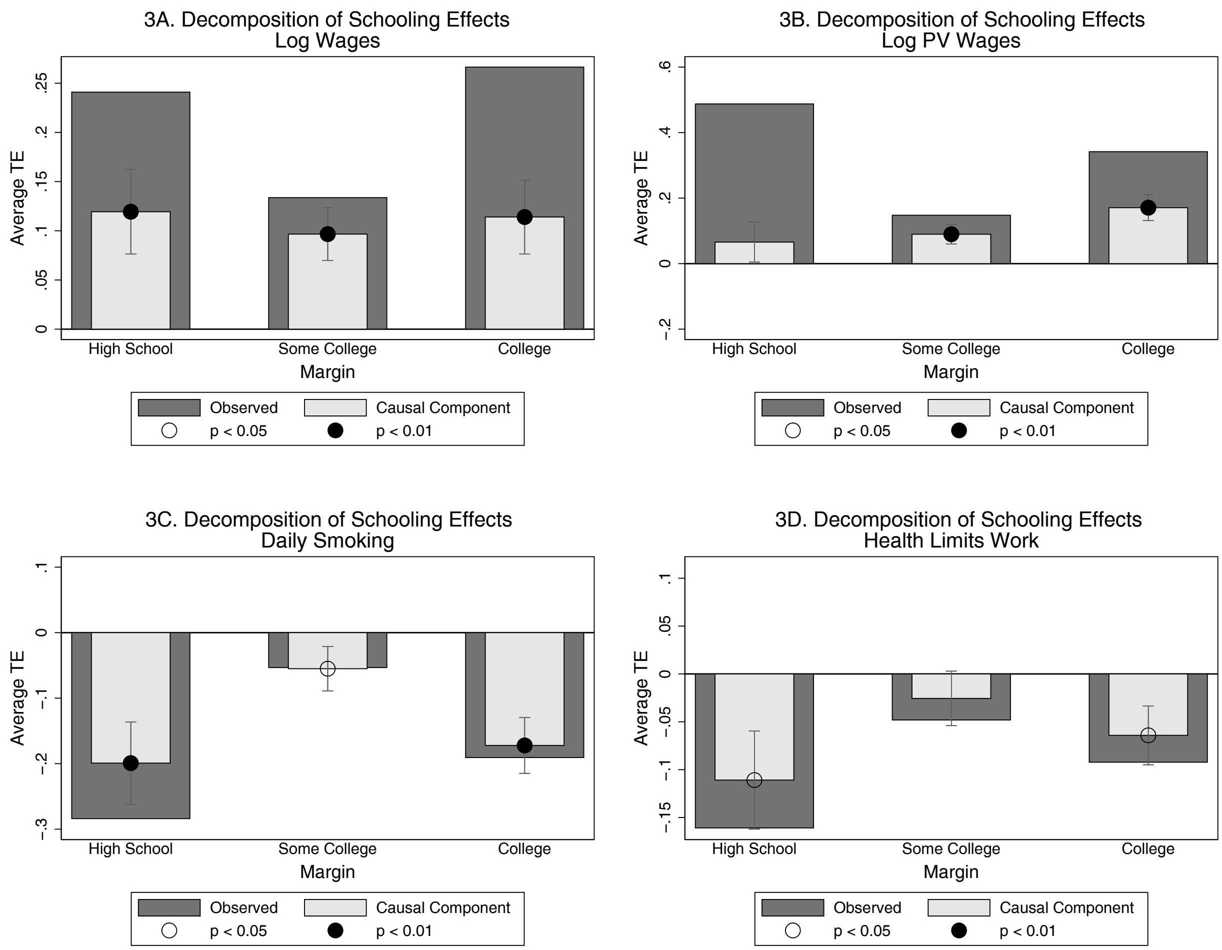
6.1 The Estimated Causal Effects of Educational Choices

We first compare the outcomes from final schooling level s with those from $s-1$.[49] The estimated treatment effects of education on log wages, log PV wage income, smoking, and health limits work are shown in Figure 3.[50] 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 $s-1$. There are substantial causal effects on earnings and health at each level of schooling. But at most levels there is also considerable ability bias.

[49] See expression (16) for the case $s' = s - 1$.

[50] 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.

Figure 3: 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.

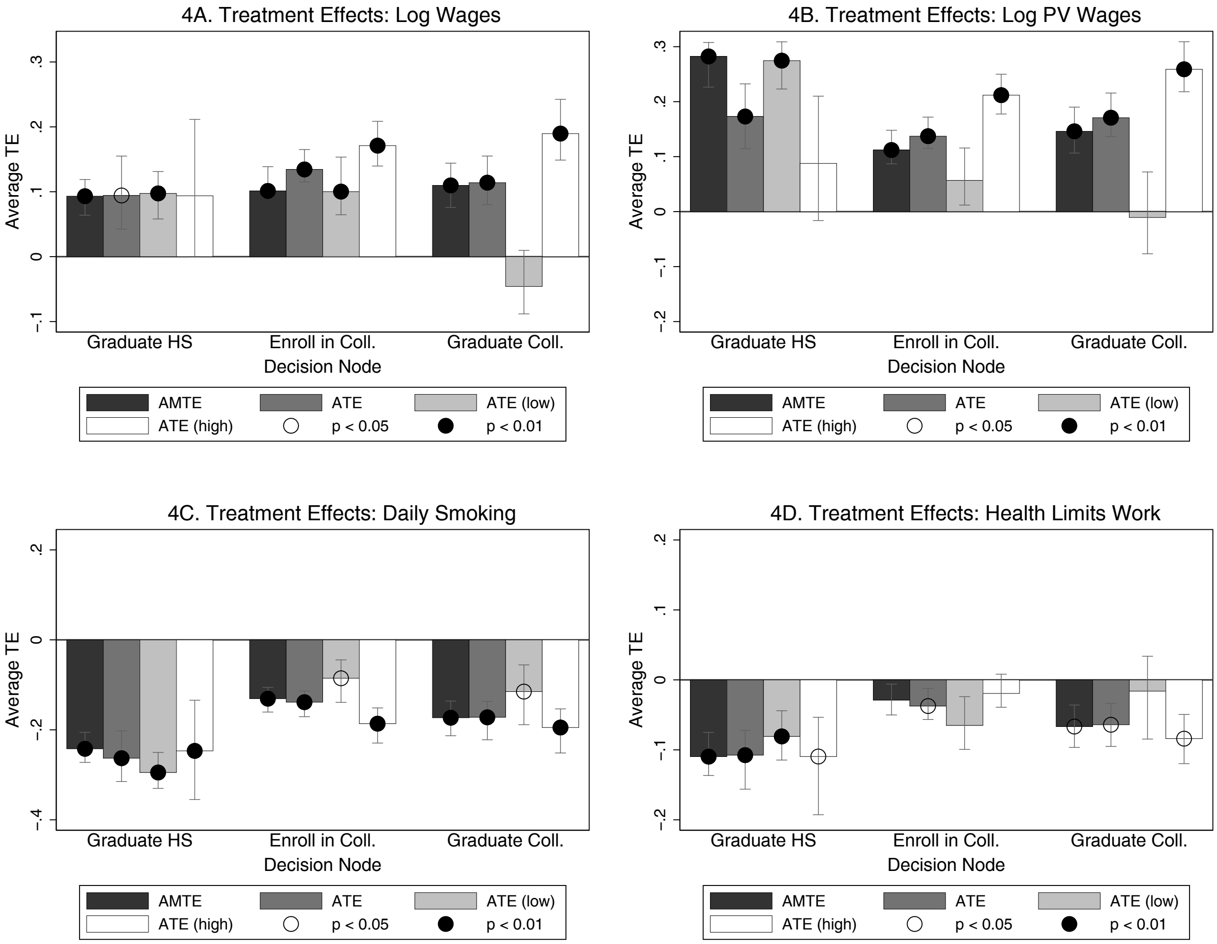
6.1.1 Dynamic Treatment Effects

We next report treatment effects by decision node (see Figure 4). 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 indifferent to the two options of the choice studied.[51]

Each box of Figure 4 presents the average effects of educational choices on the 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), for those for whom $Q_j = 1$. The ATE is calculated for the population that reaches the decision node. At each node j, the treatment effect is $E(Y^k|Q_j = 1, Fix\ D_j = 0) - E(Y^k|Q_j = 1, Fix\ D_j = 1)$. 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 socioemotional endowment above (below) the median of the full population. The table below the figure displays the fraction of individuals at each educational choice who are in the high- or low-ability group.

[51] 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 4: Treatment Effects of Outcomes by Decision Node
$E(Y^k|\textbf{Fix } D_j = 0, Q_j = 1) - E(Y^k|\textbf{Fix } D_j = 1, Q_j = 1)$



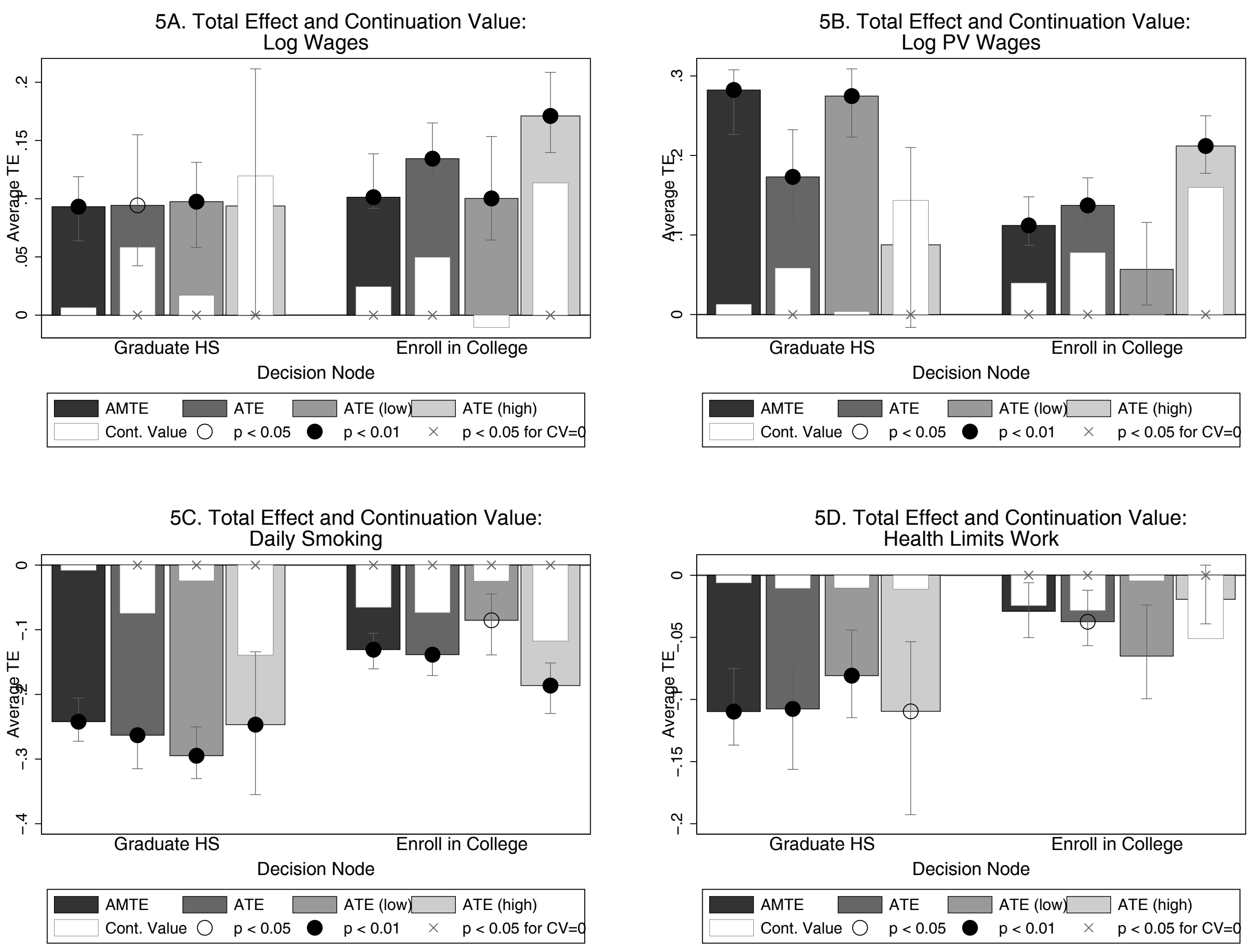
Sorting on Ability		
	Low Ability	High Ability
D_1: Dropping from HS vs. Graduating from HS	0.31	0.31
D_2: **HS Graduate** vs. College Enrollment	0.22	0.38
D_3: **Some College** vs. **4-year college degree**	0.13	0.51

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 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 socioemotional endowments above (below) the median in the overall population. The table below the figure shows the proportion of individuals at each decision ($Q_j = 1$) that are high and low ability. The larger proportion of the individuals are high ability and a smaller proportion are low ability in later educational decisions. In this table, final schooling levels are highlighted using bold letters.

6.1.2 Continuation Values

We next decompose the node-specific treatment effects reported in Table 4 into the total effect and the continuation value components. Figure 5 presents graphs of each causal effect in Figure 4 and shows the continuation value component (in white). Continuation values are important components of the dynamic treatment effects for all outcomes except health limits work.

Figure 5: Dynamic Treatment Effects:
Continuation Values and Total Treatment Effects by Node



Notes: High-ability individuals are those in the top 50% of the distributions of both cognitive and socioemotional endowments. Low-ability individuals are those in the bottom 50% of the distributions of both cognitive and socioemotional endowments. 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. Statistical significance for continuation values at the 5% level are shown by x. Section 3 provides details on how the continuation values and treatment effects are calculated.

6.1.3 The Effects on Cognitive and Noncognitive Endowments on Treatment Effects

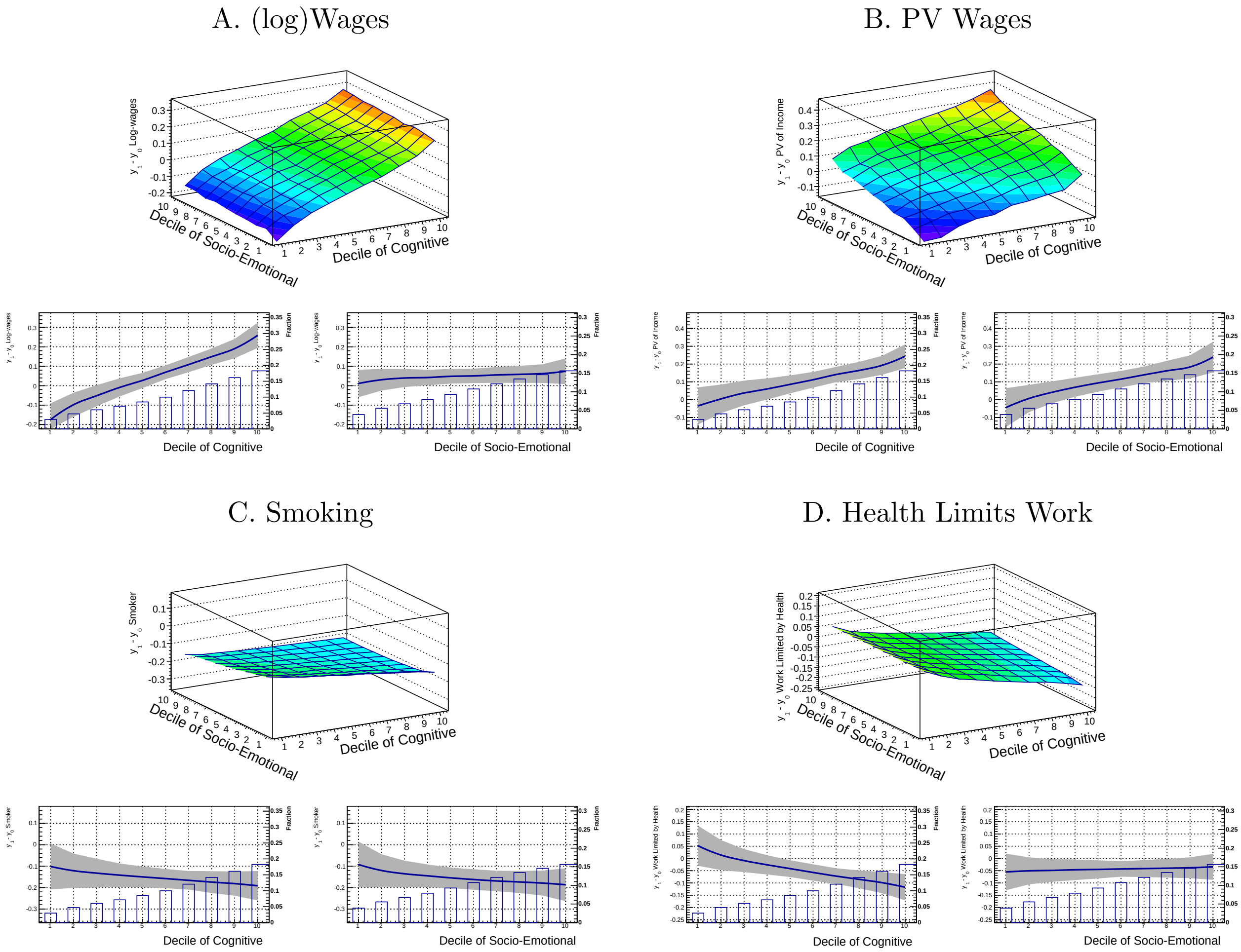
While we disaggregate the treatment effects for "high" and "low" endowment individuals in Figure 4, this division is coarse. A byproduct of our approach is that we can determine the contribution of cognitive and noncognitive endowments ($\boldsymbol{\theta}$) to explaining estimated treatment effects. 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 6 display the estimated average treatment effect of getting a four-year degree (compared to stopping with some college) for each decile pair of cognitive and noncognitive endowments.[52,53] Treatment effects in general depend on both measures of ability. Moreover, different outcomes depend in different ways on the two dimensions of ability. For example, the treatment effect of graduating college is increasing in both dimensions for present value of wages, but the reductions in health limitations with education depend mostly on cognitive endowments.

[52] Web Appendix A.10 reports a full set of results.

[53] They show average benefits by decile over the full population, rather than for the population that reaches each node.

Figure 6: Average Treatment Effect of Graduating from a Four-Year College by Outcome

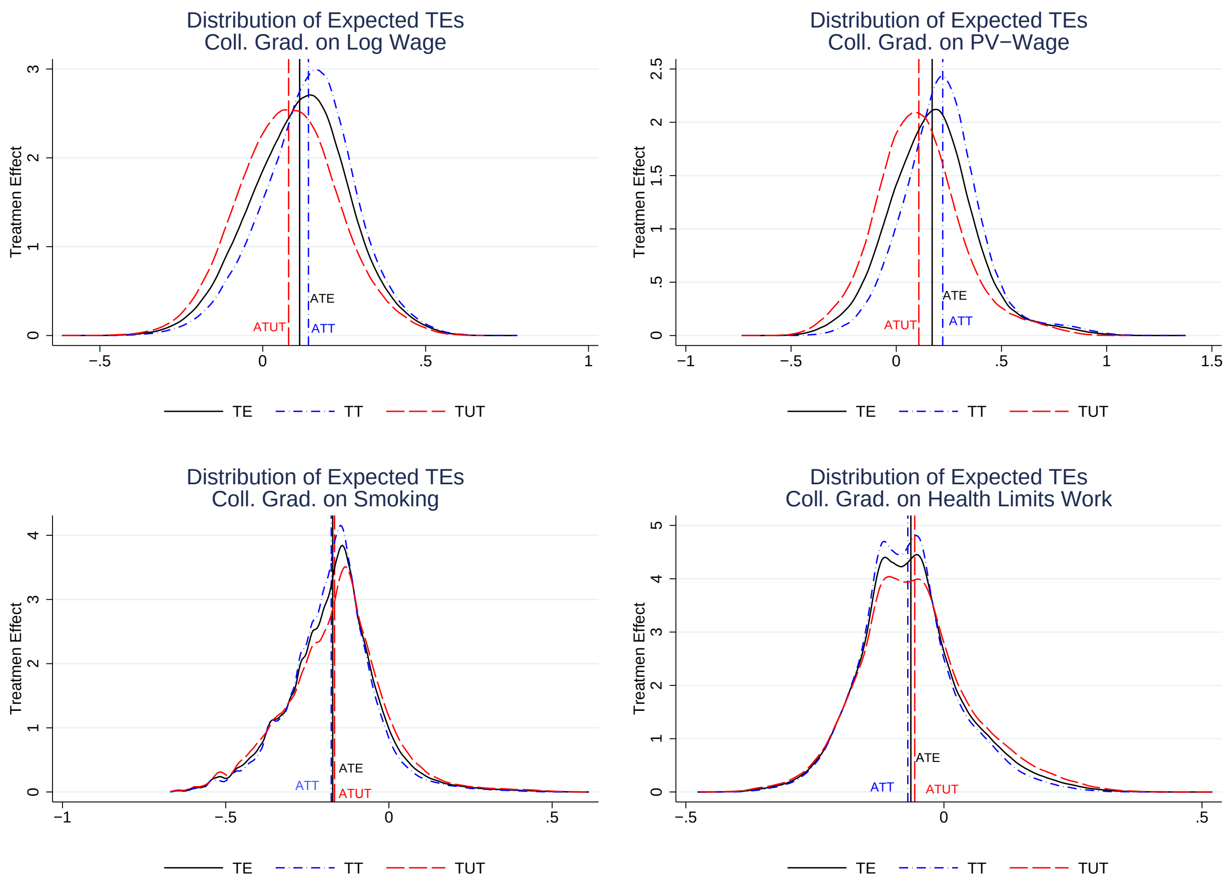


Notes: Each panel in this figure studies the average effects of graduating with a four-year college degree on the outcome of interest. The effect is defined as the differences in the outcome between those with a four-year college degree and those with some college. For each panel, let $Y_{somecoll}$ and $Y_{4-yr\ degree}$ denote the outcomes associated with attaining some college and graduating with a four-year degree, respectively. For each outcome, the first figure (top) presents $E(Y_{4-yr\ degree} - Y_{somecoll}|d^C, d^{SE})$ where d^C and d^{SE} denote the cognitive and socioemotional deciles computed from the marginal distributions of cognitive and socioemotional endowments. The second figure (bottom left) presents $E(Y_{4-yr\ degree} - Y_{somecoll}|d^C)$ so that the socioemotional 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 $E(Y_{4-yr\ degree} - Y_{somecoll}|d^{SE})$ 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 socioemotional endowment.

6.2 Distributions of Treatment Effects

One benefit of our approach over the standard IV approach is that we can identify the distributions of expected treatment effects. This feature is missing from the standard treatment effect literature. Figure 7 plots the distribution of gains for persons who graduate from college (compared to attending college but not attaining a four-year degree) along with the mean treatment effects.[54] Expectations are computed over the idiosyncratic error terms (ω_s^k).[55] Variation in the expected treatment effect comes from the variation in observed variables ($\boldsymbol{X}$) and the unobserved endowments ($\boldsymbol{\theta}$).

Figure 7: Distributions of Expected Treatment Effects: College Graduation



Notes: Distributions of treatment effects are for those who reach the educational choice.

6.3 Interpreting the Estimated Treatment Effects

Treatment Effects on Log Wages Comparing final educational levels, the average treatment effect is statistically significant for graduating from high school, attending college,

[54] Web Appendix A.11 reports a full set of distributions of treatment effects for all outcomes.
[55] The individual's expected treatment effect is $E_{\{\omega_s, \omega_s'\}}(Y_s^k - Y_{s'}^k | Q_s = 1)$.

and attaining a four-year college degree. About half of the observed difference in wages at age 30 is explained by the $\boldsymbol{X}, \boldsymbol{Z}$, and $\boldsymbol{\theta}$.

Estimates for node-specific treatment effects show that more schooling causally boosts wages although low-endowment individuals gain very little from getting a four-year college degree. Figure 4 shows that individuals with high cognitive ability capture most of the gains from a four-year degree. In fact, our estimates suggest those with very low cognitive and socioemotional endowments *lose* wage income at age 30 by graduating from college.[56] Figure 5 shows that continuation values are an important component of average treatment effects for high ability individuals. Figure 6 shows that most of the effect of abilities on the average treatment effect of college graduation comes from cognitive channels. Figure 7 shows the sorting pattern for college graduation. Even though it is not imposed by our estimation procedure, we find sorting on gains.

Treatment Effects on Present Value of Wage Income The pattern for the present value of wages is similar to that for wages with some interesting exceptions. Figures 4 and 5 show that low ability students appear to benefit substantially from graduating from high school, while only high ability individuals benefit from enrolling in and graduating from college. The treatment effect of college graduation is especially strong for high ability students. The benefits to low ability people and people at the margin of graduating high school come primarily from direct effects. The larger effects for present values than for wages comes from labor supply responses of high school graduates.[57] We find sorting on gains for the higher educational nodes. Figure 6 shows that noncognitive endowments play a much stronger role in generating the average treatment effect of college graduation on the PV of wages than they do for wages.

[56]See Web Appendix Section A.1.2 for a brief overview of the literature on the outcomes considered in this paper.

[57]See Heckman, Humphries, and Kautz (2014a, Chapter 5).

Treatment Effects on Smoking Controlling for unobserved endowments, education causally reduces smoking. The endowments and observables account for about one-third of the observed effect of education. The effects are especially strong for high school graduation. Looking at the node-specific treatment effects, each level of education has a substantial causal effect in reducing smoking. For high-endowment individuals, more than half of the average treatment effect of graduating high school and enrolling in college is derived from continuation values. Almost all of the treatment effect comes from the direct effect for low-endowment individuals.

Treatment Effects on Health Limits Work There are strong treatment effects for graduating high school but much weaker, and less precisely determined, treatment effects at higher levels of education. Continuation values are small and generally statistically insignificant. As in the case of smoking, the treatment effects are especially strong for high ability individuals except in this case noncognitive endowments play a small role.

7 Policy Simulations from our Model

Using our model, it is possible to conduct a variety of counterfactual policy simulations, a feature not shared by standard treatment effect models. We achieve these results without imposing strong assumptions on the choice model. We consider two policy experiments: (1) a tuition subsidy; and (2) an increase in the cognitive or non-cognitive endowments of those at the bottom of the endowment distribution. The first policy experiment is similar to what is estimated by LATE only in the special case where the instrument corresponds to the exact policy experiment. The second policy experiment is of interest because early childhood programs boost these endowments (Heckman, Pinto, and Savelyev, 2013). The counterfactuals generated cannot be estimated by instrumental variable methods. We ignore general equilibrium effects in all of these simulations.

7.1 Policy Relevant Treatment Effects

Unless the instruments correspond to policies, IV does not identify policy relevant treatment effects. The PRTE allows us to identify who would be induced to change educational choices under specific policy changes, and how these individuals would benefit on average. As an example, we simulate the response to a policy intervention that provides a one standard deviation subsidy to early college tuition (approximately $850 dollars for the first year of college). Column 1 of Table 2 presents the average treatment effect (including continuation values) in our estimated model for those who are induced to change education levels by the tuition subsidy. Since tuition at age 17 only enters the choice to enroll in college, the subsidy only induces high school graduates to change their college enrollment decisions. Those induced to enroll may then go on to graduate with a four year degree.[58] Columns 2 and 3 of Table 2 decompose the PRTE into the average gains for those induced to enroll and then go on to earn 4 year degrees and the average gains for those who do not. For the most part, the PRTE is larger for those who go on to earn 4 year degrees.

Figure 8 shows which individuals are induced to enroll in college within the deciles of the distribution of the unobservable in the choice equation for node 2,[59] conditional on $Q_2 = 1$ (the node determining college enrollment). These are the unobserved components of heterogeneity acted upon by the agent but unobserved by the economist.

The policy induces some individuals at every decile to switch, but places more weight on those in the middle deciles of the distribution. The figure further decomposes the effect of those induced to switch into the effect for those who go on to graduate with four year degrees and the effect for those who do not. Those induced to switch in the top deciles are more likely to go on to graduate.

[58]Models were estimated that include tuition as a determinant of the high school graduation decision. However, estimated effects of tuition on high school graduation are small and statistically insignificant. We do not impose the requirement that future values of costs affect current educational choices. This highlights the benefits of our more robust approach.

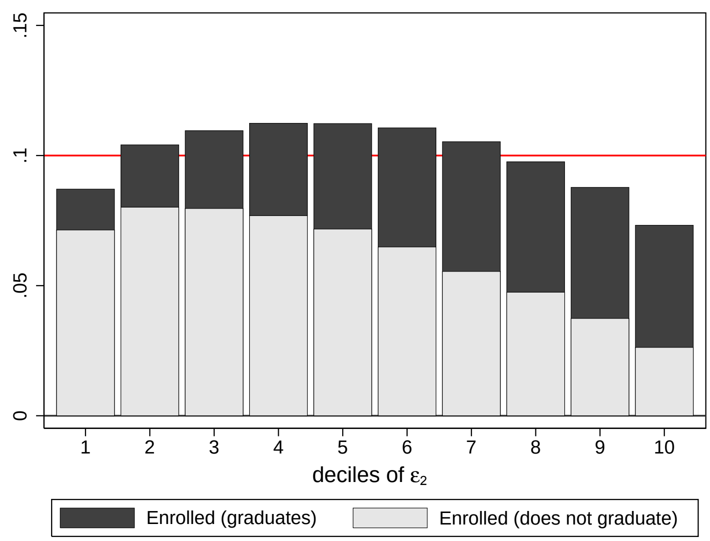
[59]$(\eta_2 = \boldsymbol{\theta}'\boldsymbol{\alpha}_2 - \nu_2)$

Table 2: PRTE: Standard Deviation Decrease in Tuition

	PRTE		4-year degree		no 4-year degree	
Log Wages	0.125	(0.023)	0.143	(0.027)	0.114	(0.027)
PV Log Wages	0.129	(0.03)	0.138	(0.033)	0.123	(0.028)
Health Limits Work	-0.036	(0.022)	-0.025	(0.021)	-0.043	(0.023)
Smoking	-0.131	(0.029)	-0.166	(0.030)	-0.108	(0.030)

Notes: Table shows the policy relevant treatment effect (PRTE) of reducing tuition for the first two years of college by a standard deviation (approx. $850). The PRTE is the average treatment effect of those induced to change educational choices as a result of the policy: $PRTE^k_{p,p'} := \iiint E(Y^k(p) - Y^k(p')|\boldsymbol{X} = \boldsymbol{x}, \boldsymbol{Z} = \boldsymbol{z}, \boldsymbol{\theta} = \overline{\boldsymbol{\theta}})dF_{\boldsymbol{X},\boldsymbol{Z},\boldsymbol{\theta}}(\boldsymbol{x}, \boldsymbol{z}, \overline{\boldsymbol{\theta}}|S(p) \neq S(p'))$. Column 1 shows the overall PRTE. Column 2 shows the PRTE for those induced to enroll by the policy who then go on to complete 4-year college degrees. Column 3 shows the PRTE for individuals induced to enroll but who do not complete 4-year degrees.

Figure 8: PRTE: Who is induced to switch



Notes: The figure plots the proportion of individuals induced to switch from the policy that lay in each decile of η_2, where $\eta_2 = \boldsymbol{\theta\alpha_2} - \nu_2$. η_2 is the unobserved component of the educational choice model. The deciles are conditional on $Q_2 = 1$, so η_2 for individuals who reach the college enrollment decision. The bars are further decomposed into those that are induced to switch that then go on to earn 4-year degrees and those that are induced to switch but do not go on to graduate.

The $850 subsidy induces 12.8% of high school graduates who previously did not attend college to enroll in college. Of those induced to enroll, more than a third go on to graduate with a 4 year degree. For outcomes such as smoking, the benefits are larger for those who graduate with a 4-year degree. The large gains for marginal individuals induced to enroll is consistent with the literature that finds large psychic costs are necessary to justify why more individuals do not attend college.

Using the estimated benefits, we can determine if the monetary gains in the present value of wages at age 18 is greater than the $850 subsidy.[60] Given a PRTE of 0.13 for log present value of wage income, the average gains for those induced to enroll is $36,401. If the subsidy is given for the first two years of college, then the policy clearly leads to monetary gains for those induced to enroll. If the subsidy is also offered to those already enrolled, the overall monetary costs of the subsidy is much larger because it is given to more than 8 students previously enrolled for each new student induced to enroll (dead weight).

7.2 Boosting Cognitive and Noncognitive Endowments

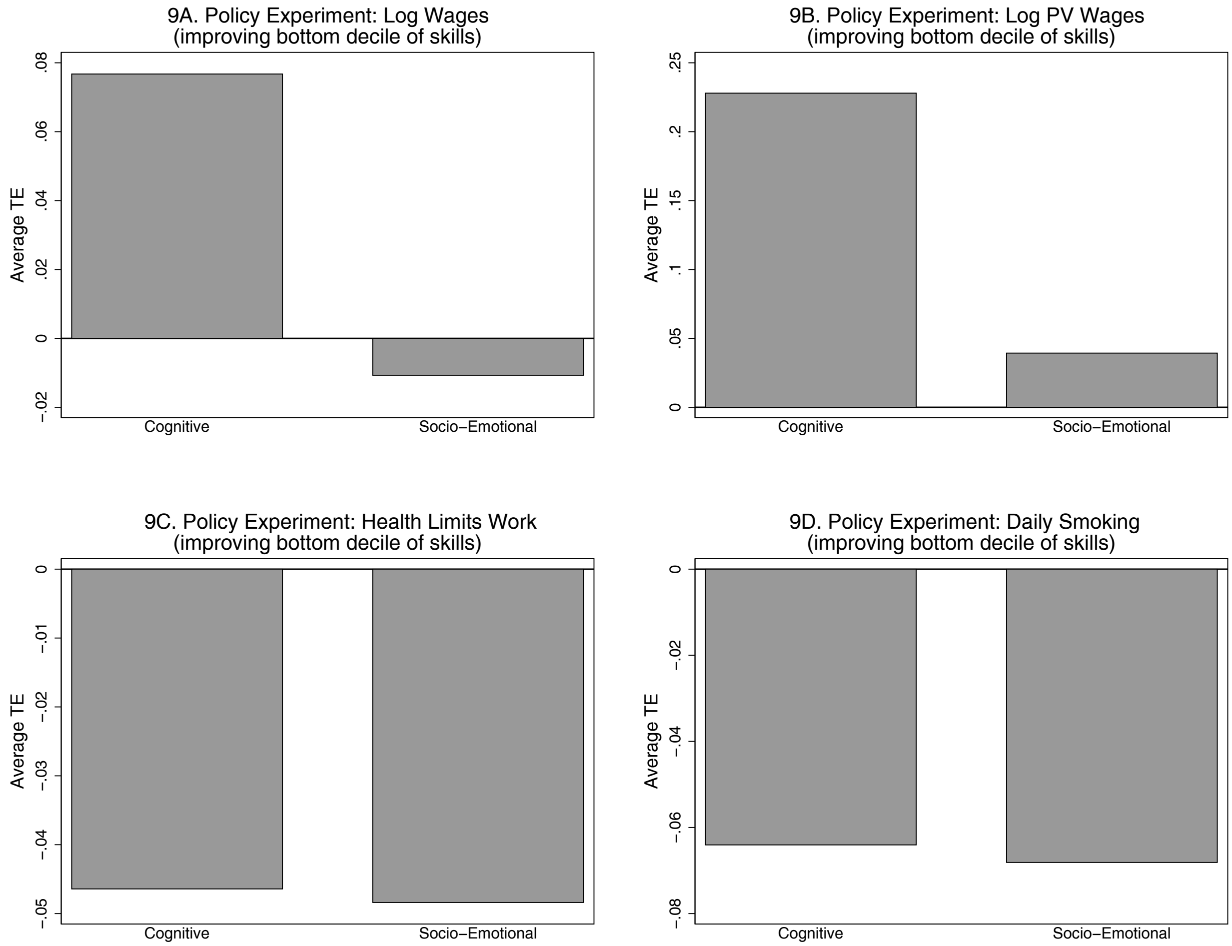
Using simulation, it is possible to conduct counterfactual policy simulations unrelated to any particular set of instruments. For example, some early childhood programs have been shown to have lasting impacts on the cognitive or non-cognitive endowments of low ability children (see Heckman, Pinto, and Savelyev, 2013). We simulate two policy experiments: (1) increasing the cognitive endowment of those in the lowest decile and (2) increasing the non-cognitive endowment of those in the lowest decile.[61]

The panels of Figure 9 show the average gains for increasing the cognitive or non-cognitive endowment of those in the lowest decile of each ability. Increased cognition helps individuals across the board. Increasing socio-emotional endowments has a smaller effect on labor market outcomes but substantial effects on health.

[60] However, a limitation of our model is that we can only estimate the monetary costs and do not estimate psychic costs.

[61] The details of how these simulations were conducted are presented in Web Appendix A.14. Our model does not address general equilibrium effects of such a change in the endowment distribution.

Figure 9: Policy Experiments



Notes: This plot shows the average gains for those in the bottom deciles of cognitive ability (left) and socioemotional ability (right), from an increase in the endowment.

8 Robustness and Comparison of Our Estimates with those Obtained from Other Methods

This section examines the robustness of some of the key assumptions maintained in this paper. It also examines whether simpler methods can be used to obtain average treatment effects. We first test the robustness of our model to relaxing key assumptions. We then consider whether it is possible to obtain reliable estimates using conventional methods in the treatment effect literature.

8.1 Testing the Two Factor Assumption

Throughout this paper, we have assumed that selection of outcomes occurs on the basis of a two component vector $\boldsymbol{\theta}$, where the components can be proxied by our measures of cognitive and noncognitive endowments. An obvious objection is that there may be unproxied endowments that affect both choices and outcomes that we do not measure. For example, one could imagine that the one component of the idiosyncratic error terms in the educational choices represent taste for school. This could generate correlations between the unobservables in the different educational choices and bias our results. We test this assumption in this section.

Cunha and Heckman (2015) estimate a related model using the same data source. They find that a three factor model explains wages and present value of wages. Two of their factors correspond to the factors used in this paper. Their third factor improves the fit of the wage outcome data but does *not* enter agent decision equations or affect selection or sorting bias. Our results are consistent with these findings.

In order to test for the presence of a third factor that influences both choices and outcomes, we test whether the simulated model fits the sample covariances between Y^k and $D_j, j = 1, \ldots, \bar{s}, k = 1, \ldots, 4$. If an important third factor common to both outcome and choice equations has been omitted, the agreement should be poor. In fact, we find close agreement.[62] Like Cunha and Heckman (2015), we find that adding a third factor that appears in outcome equations but not choice equations improves goodness of fit, but has no effect on our estimated treatment effects.[63]

[62]See Web Appendix A.15 and Table A38.

[63]See Web Appendix A.15.

8.2 Comparisons with Alternative Treatment Effect Estimators

Throughout this paper we have exploited the assumption of conditional independence of outcomes and choices given $\boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{\theta}$. This raises the question of how similar our results would be if we had used simple matching methods that also control for $\boldsymbol{\theta}, \boldsymbol{X}, \boldsymbol{Z}$.[64] We estimate Bartlett factor scores based on our measures using standard statistical software.[65] Using these extracted factors, we estimate average treatment effects using (a) Linear regression (with and without factors) and (b) Matching. Table 3 presents our estimates. All models are estimated for individuals who attain each decision node ($Q_j = 1$) and include those who may go on to attain further education in order to make the alternative models comparable to our ATE estimate that includes continuation values.

The first three columns show estimates from linear models. The first two columns introduce schooling by using dummy shifts in intercepts. The first column uses measures of cognitive and noncognitive endowments directly, while the second column includes the extracted factor scores. The third column allows loadings on covariates and factor scores to vary by schooling level.[66]

The fourth and fifth columns show estimates from matching using the Bartlett factor scores previously described as well as an index of covariates. The fourth column shows results from nearest neighbor matching using the 3 nearest neighbors. The fifth column shows results from propensity score matching using the estimated probability of the educational decision as the propensity score.[67]

The estimates differ greatly from the OLS estimates obtained without any adjustment for $\boldsymbol{\theta}$. Controlling for ability has substantial effects on the estimated average treatment effects. Across schooling nodes, all of the other estimates are *roughly* "within the ball park" of the estimates produced from our model, provided that we control for $\boldsymbol{\theta}$. This is good news for

[64] OLS is a version of matching.
[65] See Bartlett (1937, 1938).
[66] See Section A.16.3 for details on these estimators.
[67] Precise specifications of the estimating equations are presented in Web Appendix A.16.

Table 3: Average Treatment Effects - Comparison of Estimates from Our Model to Those from Simpler Methods

	Linear			Matching			Model
HS Grad.	OLS	OLS-P	OLS-F	RA-F	NNM(3)-F	PSM-F	ATE
Wages	0.205	0.073	0.155	0.159	0.106	0.127	0.094
SE	(0.025)	(0.026)	(0.025)	(0.035)	(0.039)	(0.049)	(0.056)
PV-Wage	0.380	0.213	0.318	0.277	0.203	0.227	0.173
SE	(0.030)	(0.031)	(0.030)	(0.041)	(0.053)	(0.055)	(0.059)
Smoking	-0.327	-0.246	-0.281	-0.301	-0.287	-0.286	-0.263
SE	(0.028)	(0.029)	0.028	0.041	0.056	0.068	(0.056)
Health-Limits-Work	-0.178	-0.115	-0.151	-0.150	-0.037	-0.124	-0.108
SE	(0.022)	(0.024)	(0.023)	(0.033)	(0.031)	(0.044)	(0.042)
Coll. Enroll	OLS	OLS-P	OLS-F	RA-F	NNM(3)-F	PSM-F	ATE
Wages	0.223	0.121	0.186	0.190	0.185	0.225	0.134
SE	(0.023)	(0.024)	(0.024)	(0.023)	(0.031)	(0.032)	(0.025)
PV-Wage	0.221	0.109	0.176	0.171	0.198	0.238	0.137
SE	(0.027)	(0.029)	(0.028)	(0.027)	(0.029)	(0.031)	(0.029)
Smoking	-0.177	-0.138	-0.165	-0.170	-0.174	-0.175	-0.139
SE	(0.026)	(0.028)	(0.027)	(0.028)	(0.032)	(0.033)	(0.028)
Health-Limits-Work	-0.085	-0.037	-0.066	-0.057	-0.055	-0.056	-0.037
SE	(0.020)	(0.022)	(0.021)	(0.021)	(0.022)	(0.037)	(0.022)
Coll. Grad	OLS	OLS-P	OLS-F	RA-F	NNM(3)-F	PSM-F	ATE
Wages	0.210	0.146	0.184	0.185	0.176	0.108	0.114
SE	(0.032)	(0.034)	(0.033)	(0.035)	(0.039)	(0.061)	(0.037)
PV-Wage	0.243	0.163	0.208	0.228	0.207	0.251	0.171
SE	(0.037)	(0.040)	(0.038)	(0.037)	(0.040)	(0.042)	(0.040)
Smoking	-0.209	-0.171	-0.195	-0.192	-0.145	-0.157	-0.172
SE	(0.032)	(0.035)	(0.033)	(0.035)	(0.040)	(0.042)	(0.043)
Health-Limits-Work	-0.085	-0.069	-0.078	-0.077	-0.047	-0.059	-0.064
SE	(0.024)	(0.026)	(0.025)	(0.026)	(0.030)	(0.028)	(0.031)

Notes: We estimate the ATE for each outcome and and educational choice using a number of methods. All models are estimated for the population who reaches the choice being considered ($Q_j = 1$), inclusive of those who then may go on to further schooling in order to make them comparable to the ATE from our model that includes continuation value. All models use the full set of controls listed in Table 1. "OLS" estimates a linear model using a schooling dummy, and controls ($Y = D_j + X\beta + \epsilon$). "OLS-P" estimates a linear model using a schooling dummy, controls, the sum of the ASVAB scores used, gpa, and an indicator of risky behavior ($Y = D_j + X\beta + A\gamma + \epsilon$, where A are the proxies for cognitive and socio-emotional endowments). All models ending in "-F" are estimated using Bartlett factor scores (Bartlett (1937, 1938)) estimated using our measurement system, but assuming a bivariate normal distribution and not accounting for schooling at the time of the test. "OLS-F" estimates the model ($Y = D_j + X\beta + \hat{\theta}\gamma + \epsilon$ where $\hat{\theta}$ are the Bartlett factor scores described above. "RA-F" extends OLS-F by letting the loadings on the covariates and factors vary by schooling level as described in Web Appendix A.16.3. "NNM(3)-F" presents the estimated treatment effect of nearest-neighbor matching with 3 neighbors. Neighbors are matched on their Bartlett cognitive factor, Bartlett non-cognitive factor, and an index constructed from their observable characteristics as described in Web Appendix A.16.1. "PSM-F" presents the estimated average treatment effect from propensity score matching, using the Bartlett cognitive factor, Bartlett non-cognitive factor, and control variables as described in Web Appendix A.16.2. "ATE" presents the estimated average treatment effect from the model presented in this paper. See Table A55 for additional comparisons.

applied economists mainly interested in using simple methods to estimate average treatment effects. However, these simple methods are powerless in estimating AMTE and PRTE or answering many of the other questions addressed in this paper.[68]

9 Summary and Conclusion

Gary Becker's seminal research on human capital launched a large and active industry on estimating causal effects and rates of return to schooling. Multiple methodological approaches have been used to secure estimates ranging from reduced form treatment effect methods to fully structural methods. Each methodology has its benefits and limitations.

This paper develops and estimates a robust dynamic causal model of schooling and its consequences for earnings, health, and healthy behaviors. The model recognizes the sequential dynamic nature of educational decisions. We borrow features from both the reduced form treatment effect literature and the structural literature. Our estimated model passes a variety of goodness of fit and model specification tests.

We allow agents to be irrational and myopic in making schooling decisions. Hence we can use our model to test some of the maintained assumptions in the dynamic discrete choice literature on schooling.

We use our dynamic choice model to estimate causal effects from multiple levels of schooling rather than the binary comparisons typically featured in the literature on treatment effects and in many structural papers.[69]

By estimating a sequential model of schooling in a unified framework, we are able analyze the *ex post* returns to education for people at different margins of choice and analyze a variety of interesting policy counterfactuals. We are able to characterize who benefits from education across a variety of market and nonmarket outcomes.

[68] Table A74 of the Web Appendix compares OLS estimates of direct effects and continuation values with our model estimates. The OLS estimates are "within ballpark" for smoking and health limits work, but they are wide of the mark for wages and PV wages.

[69] See, e.g., Willis and Rosen (1979).

The early literature on human capital ignored the dynamics of schooling choices. We decompose these benefits into direct components and indirect components arising from continuation values. We estimate substantial continuation value components especially for high ability individuals. For them, schooling opens up valuable options for future schooling. Standard estimates of the benefits of education based only on direct components underestimate the full benefits of education.

Without imposing rationality, we nonetheless find evidence consistent with it. We find positive sorting into schooling based on gains, especially for higher schooling levels. Schooling has strong causal effects on earnings, health and healthy behaviors even though we also find strong evidence of ability bias at all levels of schooling. Both cognitive and noncognitive endowments affect schooling choices and outcomes for each level of schooling.

We link the structural and matching literatures using conditional independence assumptions. We investigate how simple methods used in the treatment effect literature perform in estimating average treatment effects. They roughly approximate our model estimates provided analysts condition on endowments of cognitive and noncognitive skills, and correct for measurement error in the proxies. However, these simple methods *do not* identify the treatment effects for persons at the margins of different choices, the policy relevant treatment effects, or the continuation values analysed in this paper. Estimates from IV model are very different from the economically interpretable and policy relevant estimates produced from our model.

Our analysis is broadly consistent with the pioneering analysis of Becker (1964) but enriches it. The early research on human capital was casual about agent heterogeneity. It ignored selection bias and comparative advantage in schooling. We quantify the magnitude and sources of selection bias. We find evidence of both ability bias and sorting bias (comparative advantage). Nonetheless, we find strong causal effects of education at most margins for most outcomes.

Our findings thus support the basic insights of Becker (1964). Schooling has strong causal effects on market and nonmarket outcomes. Both cognitive and noncognitive endowments affect schooling choices and outcomes. People tend to sort into schooling based on gains.

References

Abbring, J. H., and J. J. Heckman (2007): "Econometric Evaluation of Social Programs, Part III: Distributional Treatment Effects, Dynamic Treatment Effects, Dynamic Discrete Choice, and General Equilibrium Policy Evaluation," in *Handbook of Econometrics*, ed. by J. Heckman, and E. Leamer, vol. 6B, pp. 5145–5303. Elsevier, Amsterdam.

Adda, J., and R. W. Cooper (2003): *Dynamic Economics: Quantitative Methods and Applications*. The MIT Press, Cambridge, Mass.

Almlund, M., A. 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, pp. 1–181. Elsevier, Amsterdam.

Altonji, J. G. (1993): "The Demand for and Return to Education When Education Outcomes are Uncertain," *Journal of Labor Economics*, 11(1), 48–83.

Angrist, J. D., and G. W. Imbens (1995): "Two-Stage Least Squares Estimation of Average Causal Effects in Models with Variable Treatment Intensity," *Journal of the American Statistical Association*, 90(430), 431–442.

Angrist, J. D., and J.-S. Pischke (2009): *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press, Princeton.

Bartlett, M. S. (1937): "The Statistical Conception of Mental Factors," *British Journal of Psychology*, 28(1), 97–104.

Bartlett, M. S. (1938): "Methods of Estimating Mental Factors," *Nature*, 141, 609–610.

Becker, G. S. (1962): "Irrational Behavior and Economic Theory," *Journal of Political Economy*, 70, 1–13.

Becker, G. S. (1964): *Human Capital: A Theoretical and Empirical Analysis, With Special Reference to Education*. National Bureau of Economic Research, New York.

Becker, G. S., and B. R. Chiswick (1966): "Education and the Distribution of Earnings," *American Economic Review*, 56(1/2), 358–369.

Blevins, J. R. (2014): "Nonparametric Identification of Dynamic Decision Processes with Discrete and Continuous Choices," *Quantitative Economics*, 5, 531–554.

Borghans, L., A. L. Duckworth, J. J. Heckman, and B. ter Weel (2008): "The Economics and Psychology of Personality Traits," *Journal of Human Resources*, 43(4), 972–1059.

Cameron, S. V., and J. J. Heckman (1993): "The Nonequivalence of High School Equivalents," *Journal of Labor Economics*, 11(1, Part 1), 1–47.

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

Cameron, S. V., and C. Taber (2004): "Estimation of Educational Borrowing Constraints Using Returns to Schooling," *Journal of Political Economy*, 112(1), 132–182.

Card, D. (1999): "The Causal Effect of Education on Earnings," in *Handbook of Labor Economics*, ed. by O. Ashenfelter, and D. Card, vol. 5, pp. 1801–1863. North-Holland, New York.

——— (2001): "Estimating the Return to Schooling: Progress on Some Persistent Econometric Problems," *Econometrica*, 69(5), 1127–1160.

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

Carneiro, P., J. J. Heckman, and E. J. Vytlacil (2010): "Evaluating Marginal Policy Changes and the Average Effect of Treatment for Individuals at the Margin," *Econometrica*, 78(1), 377–394.

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

CARNEIRO, P., C. MEGHIR, AND M. PAREY (2013): "Maternal Education, Home Environments, and the Development of Children and Adolescents," *Journal of the European Economic Association*, 11, 123160.

CAWLEY, J., K. CONNEELY, J. J. HECKMAN, AND E. J. VYTLACIL (1997): "Cognitive Ability, Wages and Meritocracy," in *Intelligence, Genes, and Success : Scientists Respond to The Bell Curve*, ed. by B. Devlin, S. Fienberg, D. Resnick, and K. Roeder, pp. 179–192. Springer, New York.

COMAY, Y., A. MELNIK, AND M. A. POLLATSCHEK (1973): "The Option Value of Education and the Optimal Path for Investment in Human Capital," *International Economic Review*, 14(2), 421–435.

CUNHA, F., AND J. J. HECKMAN (2015): "Decomposing Trends in Inequality in Earnings into Forcastable and Uncertain Components* Lazear Festschrift," .

CUNHA, F., J. J. HECKMAN, AND S. NAVARRO (2007): "The Identification and Economic Content of Ordered Choice Models with Stochastic Cutoffs," *International Economic Review*, 48(4), 1273–1309.

CUTLER, D. M., AND A. LLERAS-MUNEY (2010): "Understanding Differences in Health Behaviors by Education," *Journal of Health Economics*, 29(1), 1–28.

DOTHAN, U., AND J. WILLIAMS (1981): "Education as an Option," *The Journal of Business*, 54(1), 117–39.

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

Eisenhauer, P., J. Heckman, and S. Mosso (2015): "Estimation of Dynamic Discrete Choice Models by Maximum Likelihood and the Simulated Method of Moments," *Forthcoming, International Economic Review.*

Eisenhauer, P., J. J. Heckman, and E. J. Vytlacil (2015): "Generalized Roy Model and Cost-Benefit Analysis of Social Programs," Forthcoming, Journal of Political Economy.

Geweke, J., and M. Keane (2001): "Computationally Intensive Methods for Integration in Econometrics," in *Handbook of Econometrics*, ed. by J. J. Heckman, and E. Leamer, vol. 5, pp. 3463–3568. North-Holland, New York.

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

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

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

Hansen, L. P., and T. J. Sargent (1980): "Formulating and Estimating Dynamic Linear Rational Expectations Models," *Journal of Economic Dynamics and Control*, 2(1), 7–46.

Heckman, J., S. Schennach, and B. Williams (2013): "Matching on Proxy Variables," University of Chicago.

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

——— (2010): "Building Bridges Between Structural and Program Evaluation Approaches to Evaluating Policy," *Journal of Economic Literature*, 48(2), 356–398.

Heckman, J. J., P. Carneiro, and E. Vytlacil (2011): "Estimating Marginal Returns to Education," *American Economic Review*, 101(6), 2754–2871.

Heckman, J. J., J. E. Humphries, and T. Kautz (2014a): "The Economic and Social Benefits of GED Certification," in *The Myth of Achievement Tests: The GED and the Role of Character in American Life*, ed. by J. J. Heckman, J. E. Humphries, and T. Kautz, chap. 5, pp. 171–267. University of Chicago Press, Chicago.

Heckman, J. J., J. E. Humphries, and T. Kautz (eds.) (2014b): *The Myth of Achievement Tests: The GED and the Role of Character in American Life*. University of Chicago Press, Chicago.

Heckman, J. J., H. Ichimura, and P. E. Todd (1998): "Matching as an Econometric Evaluation Estimator," *Review of Economic Studies*, 65(2), 261–294.

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

Heckman, J. J., and R. Pinto (2015): "Causal Analysis after Haavelmo," Forthcoming, *Econometric Theory*.

Heckman, J. J., R. Pinto, and P. A. Savelyev (2013): "Understanding the Mechanisms Through Which an Influential Early Childhood Program Boosted Adult Outcomes," *American Economic Review*, 103(6), 1–35.

Heckman, J. J., S. Schennach, and B. Williams (2011): "Matching with Error-Laden Covariates," Unpublished manuscript, University of Chicago, Department of Economics.

Heckman, J. J., J. A. Smith, and N. Clements (1997): "Making the Most Out Of Programme Evaluations and Social Experiments: Accounting for Heterogeneity in Programme Impacts," *Review of Economic Studies*, 64(4), 487–535.

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

Heckman, J. J., and E. J. Vytlacil (1999): "Local Instrumental Variables and Latent Variable Models for Identifying and Bounding Treatment Effects," *Proceedings of the National Academy of Sciences*, 96(8), 4730–4734.

——— (2005): "Structural Equations, Treatment Effects and Econometric Policy Evaluation," *Econometrica*, 73(3), 669–738.

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

——— (2007b): "Econometric Evaluation of Social Programs, Part II: Using the Marginal Treatment Effect to Organize Alternative Economic Estimators to Evaluate Social Programs and to Forecast Their Effects in New Environments," in *Handbook of Econometrics*, ed. by J. Heckman, and E. Leamer, vol. 6B, chap. 71, pp. 4875–5143. Elsevier, Amsterdam.

Kane, T. J., and C. E. Rouse (1993): "Labor Market Returns to Two- and Four-Year Colleges: Is a Credit a Credit and Do Degrees Matter?," 1993 4268, National Bureau of Economic Research.

Keane, M. P., and K. I. Wolpin (1997): "The Career Decisions of Young Men," *Journal of Political Economy*, 105(3), 473–522.

Kling, J. R. (2001): "Interpreting Instrumental Variables Estimates of the Returns to Schooling," *Journal of Business and Economic Statistics*, 19(3), 358–364.

Lochner, L. (2011): "Non-Production Benefits of Education: Crime, Health, and Good Citizenship," Working Paper 16722, National Bureau of Economic Research.

McMahon, W. W. (2000): "Externalities, non-market effects, and trends in returns to educational investments," in *The Appraisal of Investments in Educational Facilities*, pp. 51–83. European Investment Bank/OECD, Paris.

——— (2009): *Higher Learning, Greater Good*. Johns Hopkins University Press, Baltimore.

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

Oreopoulos, P., and U. Petronijevic (2013): "Making College Worth It: A Review of Research on the Returns to Higher Education," Working Paper 19053, National Bureau of Economic Research, Cambridge.

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

Quandt, R. E. (1958): "The Estimation of the Parameters of a Linear Regression System Obeying Two Separate Regimes," *Journal of the American Statistical Association*, 53(284), 873–880.

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

Rust, J. (1994): "Structural Estimation of Markov Decision Processes," in *Handbook of Econometrics, Volume*, ed. by R. Engle, and D. McFadden, pp. 3081–3143. North-Holland, New York.

Schennach, S., H. White, and K. Chalak (2012): "Local indirect least squares and average marginal effects in nonseparable structural systems," *Journal of Econometrics*, 166(2), 282–302.

Vytlacil, E. J. (2002): "Independence, Monotonicity, and Latent Index Models: An Equivalence Result," *Econometrica*, 70(1), 331–341.

Weisbrod, B. A. (1962): "Education and Investment in Human Capital," *Journal of Political Economy*, 70(5, Part 2: Investment in Human Beings), 106–123.

White, H., and K. Chalak (2009): "Settable Systems: An Extension of Pearl's Causal Model with Optimization, Equilibrium, and Learning," *Journal of Machine Learning Research*, 10, 1759–1799.

Willis, R. J., and S. Rosen (1979): "Education and Self-Selection," *Journal of Political Economy*, 87(5, Part 2), S7–S36.