Returns to Education: The Causal Effects of Education on Earnings, Health and Smoking

JPE, September 2018

James J. Heckman and John Eric Humphries
The University of Chicago

Greg Veramendi
Arizona State University

THE UNIVERSITY OF CHICAGO

Econ 350, Winter 2021

Introduction

Rate of Return

- Key parameter for evaluating the effectiveness of human capital investments.

Literature Review on Education and Health

Link to Appendix

Our Approach

- A *middle ground* between the reduced form treatment approach and the fully structural dynamic discrete choice approach.

- Like the structural literature, estimate causal effects at clearly identified margins of choice.
- 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, agnostic about the precise rules used by agents to make decisions.
- Unlike that literature, recognize the possibility that people somehow make decisions and account for the consequences of their choices.
- Approximate agent decision rules.
- Do not impose the cross-equation restrictions that are the hallmark of the structural approach.

- Model approximates a dynamic discrete choice model without taking a stance on exactly what agents are maximizing or their information sets.
- Model can be identified through multiple sources of variation (both IV and Matching).
- 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.
- Correct the matching variables for measurement error and the bias introduced into the measurements by family background.
- Also can use exclusion restrictions to identify our model as in the IV and control function literatures.

Decomposing *ex post* treatment effects:

- (i) Direct benefits of going from one level of schooling to the next

and

- (ii) Continuation values arising from access to additional education beyond the next step.

Empirical Findings

1. Substantial returns/causal effects of education on wages, the present value of wages, health, and smoking.
2. Continuation values arising from sequential choices are empirically important components of returns to education.
    - Low-ability individuals gain mostly from graduating high school and stopping there (little continuation value).
    - High-ability individuals have substantial post-high school continuation values.
3. Estimated returns (causal effects) differ by schooling level and depend on observed and unobserved characteristics of individuals.
    - Graduating high school benefits all—and especially low-ability persons.
    - Only high-ability individuals receive substantial benefits from college graduation.
    - Positive sorting on gains only at higher educational levels.

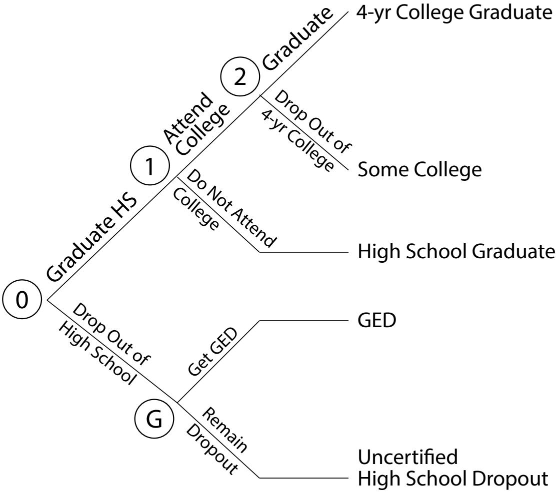
4. People sort on *ex post* gains, especially more able people at higher schooling levels, confirming a core tenet of human capital theory.
    - Yet, at the same time, people do not know or act on publicly available information when making decisions about high school graduation.
5. Paper contributes to an emerging literature on the importance of both cognitive and non-cognitive skills in shaping life outcomes.
    - Consistent with the recent literature, we find that both categories of ability are important predictors of educational attainment.
    - Within schooling levels, cognitive and non-cognitive abilities have impacts on most outcomes.
6. Selection bias arising from both observed and unobserved variables accounts for a substantial portion (typically over one half) of the raw differences in wage outcomes classified by education.
    - This finding runs counter to a common interpretation in the literature based on comparing IV and OLS.

Conduct Two Policy Experiments

1. Examines the impact of a tuition subsidy on college enrollment.
    - We identify who is affected by the policy, how their decisions change, and how much they benefit.
2. Analyze a policy that improves the ability 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.

Model

Figure 1: A Multistage Dynamic Decision Model



- $\mathcal{J}$: 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.
- A person may remain a dropout or get the GED.
- For $D_0 = 1$, we define the attainable set as $\{0, G\}$.

- $\mathcal{S} = \{G, 0, \ldots, \overline{s}\}$ denotes the set of stopping states with $S = s$ if the agent stops at $s \in \mathcal{S}$.
- Define $\overline{s}$ as the highest attainable element in $\mathcal{S}$ in the ordered subset $\{0, \ldots, \bar{s}\} = \mathcal{S} \backslash \{G\}$.
- $Q_j = 1$: agent *gets to* decision node j and acquires at least the education associated with j.
- $Q_j = 0$ if the person never gets there.

- The history of nodes visited by an agent described by the collection of the Q_j such that $Q_j = 1$.
- Observe that $D_s = 1$ is equivalent to $S = s$ for $s \in \{1, \ldots, \bar{s}\}$ and $D_{\bar{s}} = 1$ if $D_j = 0, \forall j \in \mathcal{S} \backslash \{\bar{s}\}$.
- For notational convenience, we assign $D_j = 0$ for all $j > s$.
- $D_0 = 1$ and $D_G = 0$ is equivalent to $S = G$.

A Sequential Decision Model

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

THE UNIVERSITY OF CHICAGO

- I_j: agent's *perceived* value at node j of going on to the next node.
- $Q_j = 1$ ensures that agents are able to make the transition at j by conditioning on the population eligible to make the transition.
- Associated with each final state $s \in \mathcal{S}$ is a set of K_s potential outcomes for each agent with indices $k \in \mathcal{K}_s$.

- We define the $\tilde{Y}_s^k$ as latent variables that map into potential outcomes Y_s^k:

$$Y_s^k = \left\{ \begin{array}{ll} \tilde{Y}_s^k & \text{if } Y_s^k \text{ is continuous} \\ \mathbf{1}(\tilde{Y}_s^k \geq 0) & \text{if } Y_s^k \text{ is a binary outcome} \end{array} \right\} \text{ for } k \in \mathcal{K}_s, \quad s \in \mathcal{S}. \tag{2}$$

- Switching regression framework of Quandt (1958, 1972).
- Observed outcome Y^k for a k common across all decision nodes:

$$Y^k = \left(\sum_{\mathcal{S} \setminus \{0,G\}} D_s Y_s^k \right) (1 - D_0) + \left(Y_0^k D_G + Y_G^k (1 - D_G) \right) D_0. \quad (3)$$

Parameterizations of the Decision Rules and Potential Outcomes for Final States

$$I_j = \phi_j \underbrace{(Z)}_{\substack{\text{Observed} \\ \text{by analyst}}} - \underbrace{\eta_j}_{\substack{\text{Unobserved} \\ \text{by analyst}}}, \quad j \in \mathcal{J} \tag{4}$$

- ***Z***: a vector of variables observed by the analyst, components of which determine transition decisions.

$$\tilde{Y}_s^k = \tau_s^k \underbrace{(\boldsymbol{X})}_{\substack{\text{Observed} \\ \text{by analyst}}} + \underbrace{U_s^k}_{\substack{\text{Unobserved} \\ \text{by analyst}}}, \quad k \in \mathcal{K}_s, \quad s \in \mathcal{S}, \tag{5}$$

- $\boldsymbol{X}$ is a vector of observed determinants of outcomes and U_s^k is unobserved by the analyst.
- Separability not strictly required in the structural or discrete choice literatures.

Assumptions about the Unobservables

- A finite dimensional vector $\boldsymbol{\theta}$.

$$\eta_j = -(\boldsymbol{\theta}' \boldsymbol{\lambda}_j - \nu_j), \;\; j \in \mathcal{J} \tag{6}$$

and

$$U_s^k = \boldsymbol{\theta}' \boldsymbol{\alpha}_s^k + \omega_s^k, \;\; k \in \mathcal{K}_s, s \in \mathcal{S} \tag{7}$$

- ν_j: an idiosyncratic error term for transition j.
- ω_s^k represents: idiosyncratic error term for outcome k in state s.

- $\boldsymbol{\nu} = (\nu_G, \nu_0, \nu_1, \ldots, \nu_{\bar{s}-1})$.
- $\boldsymbol{\eta} = (\eta_G, \eta_0, \ldots, \eta_{\bar{s}-1})$.
- $\boldsymbol{\omega}_s = (\omega_s^1, \ldots, \omega_s^{Ks})$.
- $\boldsymbol{U}_s = (U_s^1, \ldots, U_s^{Ks})$.
- $\boldsymbol{U}_s$ into $\boldsymbol{U} = (\boldsymbol{U}_G, \boldsymbol{U}_0, \ldots, \boldsymbol{U}_{\bar{s}})$.

- "$\perp\!\!\!\perp$" denotes statistical independence.
- Condition everywhere on $\boldsymbol{X}$.

$$\nu_j \perp\!\!\!\perp \nu_l, \quad \forall l \neq j \quad l, j \in \mathcal{J} \tag{A-1a}$$

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

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

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

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

Measurement System for Unobserved Factors θ

- Allow for the possibility that θ cannot be measured precisely.
- $\boldsymbol{M}$: vector of N_M measurements on θ.

$$\boldsymbol{M} = \Phi(\boldsymbol{X}, \theta, \boldsymbol{e}) \tag{8}$$

$\boldsymbol{X}$: observed variables
θ: factors

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

$$\boldsymbol{e} = (e_1, \ldots, e_{N_M})$$

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

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

- For the purpose of identifying treatment effects, we do not need to identify each equation of system (8).
- Just need to identify the *span* of θ that preserves the information on θ in (8).

Defining Returns/Causal Effects of Education

- No single “causal effect” of education.
- A variety of causal effects.
- In the spirit of credible econometrics, we define such treatment effects conditional on $Q_j = 1$.
- This approach blends structural and treatment effect approaches.

- Treatment effect T_j^k for outcome k for an individual selected from the population $Q_j = 1$ with characteristics $\boldsymbol{X} = \boldsymbol{x}, \boldsymbol{Z} = \boldsymbol{z}, \theta = \overline{\theta}$, 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:

$$T_j^k[Y^k|\boldsymbol{X} = \boldsymbol{x}, \boldsymbol{Z} = \boldsymbol{z}, \theta = \overline{\theta}] := (Y^k|\boldsymbol{X} = \boldsymbol{x}, \boldsymbol{Z} = \boldsymbol{z}, \theta = \overline{\theta}, Q_j = 1, Fix\ D_j = 0) \\ -(Y^k|\boldsymbol{X} = \boldsymbol{x}, \boldsymbol{Z} = \boldsymbol{z}, \theta = \overline{\theta}, Q_j = 1, Fix\ D_j = 1). \quad (9)$$

Direct Effects and Continuation Values

- *Direct effect* of going from j to $j+1$: $DE_j^k = Y_{j+1}^k - Y_j^k$.
- *Continuation value* of going beyond $j+1$ for persons with $D_0 = 0$:

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

- *Total effect* of fixing $D_j = 0$ on Y^k.

$$T_j^k = DE_j^k + C_{j+1}^k. \tag{10}$$

$$ATE_j^k := \int \cdots \int E(T_j^k[Y^k|\boldsymbol{X} = \boldsymbol{x}, \boldsymbol{Z} = \boldsymbol{z}, \theta = \overline{\theta}])\, dF_{\boldsymbol{X},\boldsymbol{Z},\theta}(\boldsymbol{x}, \boldsymbol{z}, \overline{\theta} \mid Q_j = 1). \tag{11}$$

Continuation Value:

$$E_{\boldsymbol{X},\boldsymbol{Z},\theta}(C^k_{j+1}) =$$

$$E_{\boldsymbol{X},\boldsymbol{Z},\theta}\Bigg[\sum_{l=j+1}^{\bar{s}-1}\Big\{E(Y^k_{l+1} - Y^k_l|\boldsymbol{X}=\boldsymbol{x},\boldsymbol{Z}=\boldsymbol{z},\theta=\overline{\theta},Q_{l+1}=1,\ Fix\ Q_{j+1}=1)$$

$$\cdot Pr(Q_{l+1}=1|\boldsymbol{X}=\boldsymbol{x},\boldsymbol{Z}=\boldsymbol{z},\theta=\overline{\theta},Q_j=1,Fix\ Q_{j+1}=1)\Big\}|Q_j=1\Bigg] \quad (12)$$

- $Q_{\bar{s}} = 1$ if $S = \bar{s}$.

- Population distribution of total effects.

$$Pr(T_j^k < t_j^k | \boldsymbol{X} = \boldsymbol{x}, \boldsymbol{Z} = \boldsymbol{z}, \theta = \overline{\theta}, Q_j = 1) \qquad (13)$$

Average Marginal Treatment Effects

$$AMTE_j^k :=$$
$$\iiint E\left[T_j^k\left(Y^k|\boldsymbol{X}=\boldsymbol{x},\boldsymbol{Z}=\boldsymbol{z},\theta=\overline{\theta}\right)\right]\, dF_{\boldsymbol{X},\boldsymbol{Z},\theta}(\boldsymbol{x},\boldsymbol{z},\bar{\theta}\mid Q_j=1,|I_j|\leq\varepsilon) \tag{14}$$

Policy-Relevant Treatment Effects

- Let $Y^k(p)$: aggregate outcome under policy p for outcome k.
- $S(p)$ be the final state selected by an agent under policy p.

$$PRTE^k_{p,p'} := \iiint E(Y^k(p') - Y^k(p)|\boldsymbol{X} = \boldsymbol{x}, \boldsymbol{Z} = \boldsymbol{z}, \theta = \overline{\theta})dF_{\boldsymbol{X},\boldsymbol{Z},\theta}(\boldsymbol{x}, \boldsymbol{z}, \overline{\theta}|S(p) \neq S(p')) \tag{15}$$

- $S(p) \neq S(p')$ denotes the set of the characteristics of people for whom attained states differ under the two policies.
- PRTE is often confused with LATE.
- In general, they are different unless the proposed policy change coincides with the instrument used to define LATE.

Differences Across Final Schooling Levels

- The mean gain for the subset of the population that completes one of the two adjacent schooling level $S \in \{s, s'\}$:

$$ATE^k_{s,s'} := \iiint E(Y^k_{s'} - Y^k_s | \boldsymbol{X} = \boldsymbol{x}, \boldsymbol{Z} = \boldsymbol{z}, \theta = \overline{\theta}) \, dF_{\boldsymbol{X},\boldsymbol{Z},\theta}(\boldsymbol{x}, \boldsymbol{z}, \overline{\theta} \,|\, S \in \{s, s'\}). \quad (16)$$

- Ignores continuation values.

Decomposing Raw Differences in Outcomes into Selection Bias, Sorting Gains, and Average Treatment Effects

- Focus on the upper branch of Figure 1 ($D_0 = 0$).
- Two basic models used in the empirical literature estimating returns to schooling.

1. Comparisons of pairwise final schooling levels (s_0, s) attained by agents $(D_{s_0} + D_s = 1)$, $s_0 \neq s$.
2. Gains and ability bias in terms of benefits associated with attaining (and possibly exceeding) given schooling levels $(Q_j = 1)$.

- The effect of additional schooling starting at s_0 and stopping at s:

$$Y^k = Y^k_{s_0} + \sum_{s \in \mathcal{S}} \rho^k_{s_0,s} D_s. \tag{17}$$

$$\rho^k_{s_0,s} = Y^k_s - Y^k_{s_0} = \tau^k_s(\boldsymbol{X}) - \tau^k_{s_0}(\boldsymbol{X}) + \boldsymbol{\theta}'(\boldsymbol{\alpha}^k_s - \boldsymbol{\alpha}^k_{s_0}) + \omega^k_s - \omega^k_{s_0}.$$

- $E(\rho^k_{s_0,s})$: one version of the returns to schooling compared to benchmark s_0 defined for the entire population.

Griliches (1977): "Ability Bias"

- Ignores sorting on gains and only considers ability bias.
- Conditions on $\boldsymbol{X}$ (in levels and in interactions with D_s), sorting gains arise only if $\boldsymbol{\alpha}_s^k - \boldsymbol{\alpha}_{s_0}^k \neq 0$ and $\boldsymbol{\lambda}_s \neq 0$.
- Even if $\boldsymbol{\alpha}_s^k - \boldsymbol{\alpha}_{s_0}^k = 0$, as long as $\boldsymbol{\alpha}_{s_0}^k \neq 0$, ability bias will arise in estimating the mean of the gains $\rho_{s_0,s}^k$ in (17), provided $\boldsymbol{\lambda}_s \neq 0$.

Decomposing Observed Differences

$$\underbrace{E[Y^k_{j+1}|S=j+1]-E[Y^k_j|S=j]}_{\textbf{Observed difference}}$$
$$=\underbrace{E[Y^k_{j+1}-Y^k_j|S=j+1]}_{\textbf{Treatment on the Treated}\,:TT_{j,j+1}}$$
$$+\underbrace{E[Y^k_j|S=j+1]-E[Y^k_j|S=j]}_{\substack{\textbf{Selection bias}\,:SB_{j,j+1}\\ \text{from base state } j}} \tag{18}$$

Decomposing Treatment on the Treated

$$\underbrace{E[Y^k_{j+1} - Y^k_j | S \in \{j, j+1\}]}_{\substack{\text{Pairwise average treatment effect } ATE_{j,j+1} \\ \text{for people in conditioning set } \{j,j+1\}}}$$

$$+ \underbrace{E[Y^k_{j+1} - Y^k_j | S = j+1] - E[Y^k_{j+1} - Y^k_j | S \in \{j, j+1\}]}_{\text{Sorting gains } SG_{j,j+1}} \tag{18}$$

Another Decomposition Based on Attainment Levels:

- Components associated with stopping at j and possibly continuing beyond j.
- Assume ($D_0 = 0$), i.e., stay on upper branch of Figure 1.

$$Y^k = Y_0^k + \sum_{j \geq 1}^{\bar{s}} \rho_{j-1,j}^k Q_j \tag{19}$$

$$\rho_{j-1,j}^k = Y_j^k - Y_{j-1}^k .$$

- The expected future gain for a person at j (≥ 1) is

$$E_j\left(\sum_{l>j}^{\bar{s}} \rho^k_{l-1,l} Q_l \middle| Q_j = 1\right)$$
$$= \sum_{l>j}^{\bar{s}} \left[E_j(\rho^k_{l-1,l} | Q_l = 1) P(Q_l = 1 | Q_j = 1)\right] \quad j \geq 1$$

$$
\begin{aligned}
&\underbrace{E[Y^k|D_j=0,Q_j=1]-E[Y^k|D_j=1,Q_j=1]}_{\text{Observed difference}}\\
=&\underbrace{E[Y^k|D_j=0,Q_j=1]-E[Y^k|D_j=0,Q_j=1,\mathit{Fix}\ D_j=1]}_{\textbf{Dynamic Treatment on the Treated for those at } j}\\
+&\underbrace{E[Y^k|D_j=0,Q_j=1,\mathit{Fix}\ D_j=1]-E[Y^k|D_j=1,Q_j=1]}_{\textbf{Selection Bias for those at } j}
\end{aligned}
\tag{20}
$$

Treatment on the Treated

$$\underbrace{E[Y^k|Q_j = 1, Fix\ D_j = 0] - E[Y^k|Q_j = 1, Fix\ D_j = 1]}_{\textbf{ATE for those at } j}$$

$$+ \underbrace{\left\{ \begin{array}{l} \left(E[Y^k|D_j = 0, Q_j = 1] - E[Y^k|D_j = 0, Q_j = 1, Fix\ D_j = 1]\right) \\ - \left(E[Y^k|Q_j = 1, Fix\ D_j = 0] - E[Y^k|Q_j = 1, Fix\ D_j = 1]\right) \end{array} \right\}}_{\textbf{TT - ATE: Sorting gain at } j \text{ for those who transit to } j+1} \quad (20)$$

Identification and Model Likelihood

See Heckman, Humphries, and Veramendi (2016) Dynamic Treatment Effects

ARTICLE IN PRESS

Journal of Econometrics ▮ (▮▮▮▮) ▮▮▮–▮▮▮

ELSEVIER

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Dynamic treatment effects[☆]

James J. Heckman[a,b,∗], John Eric Humphries[a,1], Gregory Veramendi[c,2]

[a] Department of Economics, University of Chicago, 1126 East 59th Street, Chicago, IL 60637, United States

[b] The American Bar Foundation, 750 North Lake Shore Drive, Chicago, IL, 60611-4403, United States

[c] Arizona State University, 501 East Orange Street, CPCOM 412A, Tempe, AZ, 85287-9801, United States

Link to Web Appendix A.4

THE UNIVERSITY OF CHICAGO

Goodness of Fit

Link to Web Appendix A.5

Data and Exclusion Restrictions

Control Variables and Exclusion Restrictions

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	

Benchmark OLS Study: Conditioning on X and θ

Figure 2: Observed and Adjusted Benefits from Education

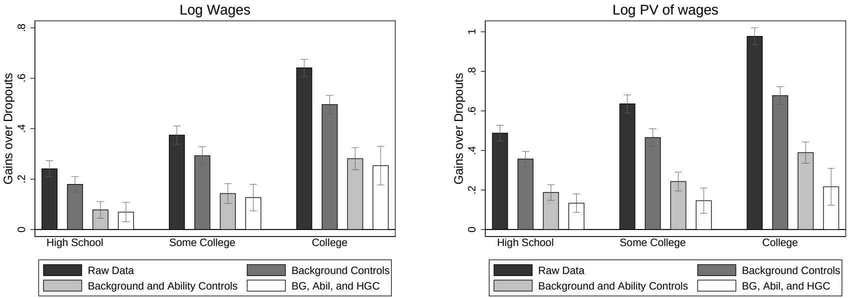
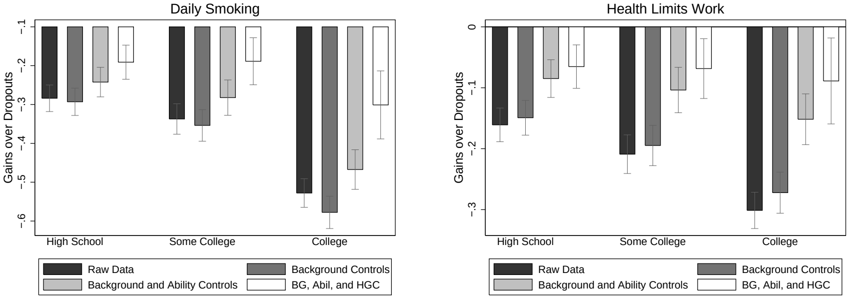


Figure 2: Observed and Adjusted Benefits from Education, Cont'd



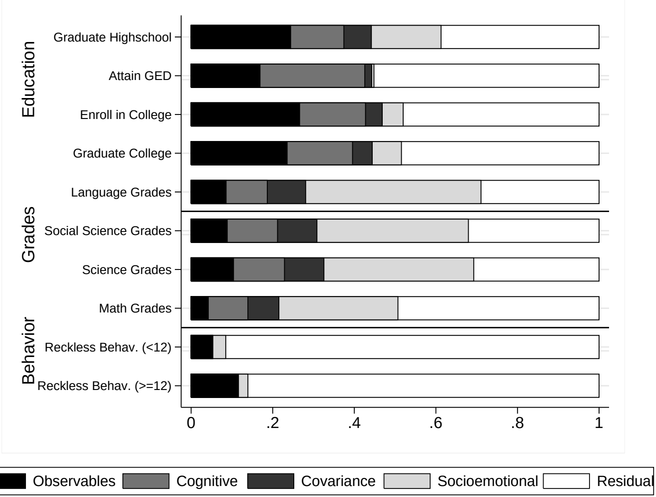
Notes: The bars represent the coefficients from a regression of the designated outcome on dummy variables for educational attainment, where the omitted category is high school dropout. Regressions are run adding successive controls for background and proxies for ability. Background controls include race,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

Measurement System

- Latent skills are measured from behaviors that can also be affected by incentives and other traits. Even after controlling for these incentives, some normalizations are necessary to operationalize the measures of skills, and distinguish one skill from another.

- Following Hansen et al. (2004), control for the effect of schooling at the time the measurements are taken on the measurements to control for feedback from schooling to measured traits.
- ASVAB are used for measures of cognition.
- Specifically, consider scores from Arithmetic Reasoning, Coding Speed, Paragraph Comprehension, Word Knowledge, Math Knowledge, and Numerical Operations.
- Academic success (measured by GPA) depends on cognitive ability

Figure 3: Decomposing Variances in the Measurement System



Joint Distributions of Endowments

Overall Correlation: 0.24

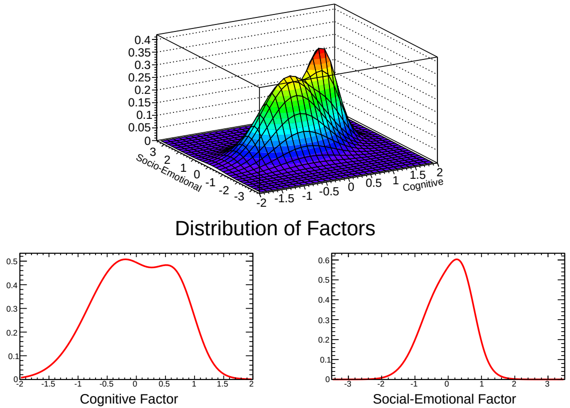
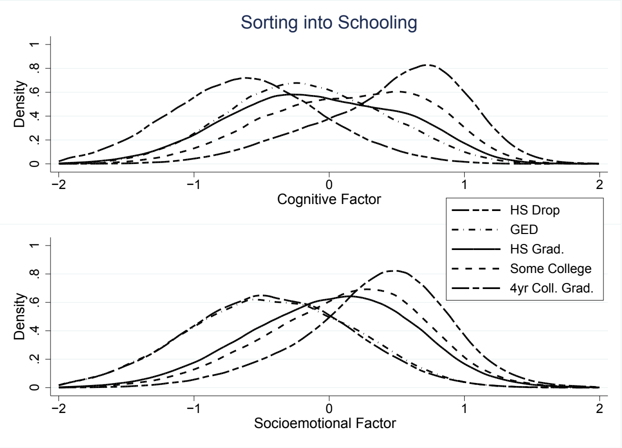


Figure 4: Distribution of factors by schooling level

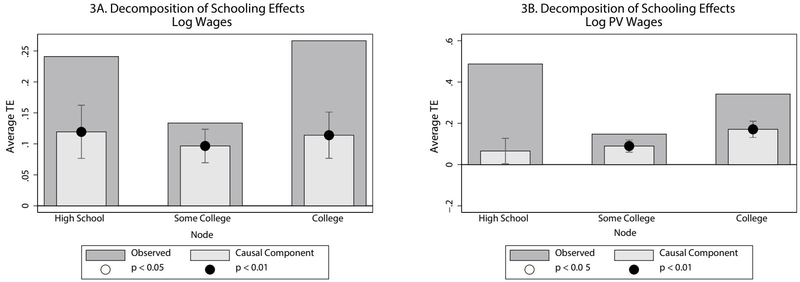


Note: The factors are simulated from the estimates of the model. The simulated data contain 1 million observations.

Estimated Causal Effects

The Estimated Average Causal Effect of Educational Choices by Pairwise Final Schooling Levels ($ATE_{s,s'}$)

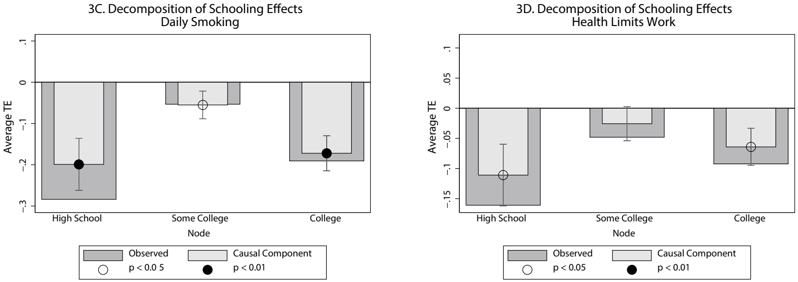
Figure 5: Causal Versus Observed Differences by final schooling level (compared to next lowest level)



Notes: These figures report pairwise treatment effect (16) for the indicated schooling nodes. Each bar compares the mean outcomes from a particular schooling level j and the next lowest level $j-1$ defined for the set of persons who complete schooling at $j-1$ or j. The "Observed" bar displays the observed differences in the data. The "Causal Component" bar displays the estimated average treatment effect to those who get treated (ATE) for the indicated group. The difference between the observed and causal treatment effect is attributed to the effect of selection and ability. Selection includes sorting on gains. The error bars and significance levels for the estimated ATE are calculated using 200 bootstrap samples. Error bars show one standard deviation and correspond to the 15.87th and 84.13th percentiles of the bootstrapped estimates, allowing for asymmetry. Significance at the 5% and 1% levels is shown by open and filled circles on the plots, respectively.

THE UNIVERSITY OF CHICAGO

Figure 5: Causal Versus Observed Differences by final schooling level (compared to next lowest level)



Notes: These figures report pairwise treatment effect (16) for the indicated schooling nodes. Each bar compares the mean outcomes from a particular schooling level j and the next lowest level $j-1$ defined for the set of persons who complete schooling at $j-1$ or j. The "Observed" bar displays the observed differences in the data. The "Causal Component" bar displays the estimated average treatment effect to those who get treated (ATE) for the indicated group. The difference between the observed and causal treatment effect is attributed to the effect of selection and ability. Selection includes sorting on gains. The error bars and significance levels for the estimated ATE are calculated using 200 bootstrap samples. Error bars show one standard deviation and correspond to the 15.87th and 84.13th percentiles of the bootstrapped estimates, allowing for asymmetry. Significance at the 5% and 1% levels is shown by open and filled circles on the plots, respectively.

- Average treatment effects $ATE_{s-1,s}$ (16).
- Web Appendix A.14.1 reports traditional treatment effects (treatment on the treated, treatment on the untreated, as well as the ATEs displayed in Figure 5).
- Web Appendix A.15.1 and A.15.2 present estimates of decomposition (18) for all four outcomes.

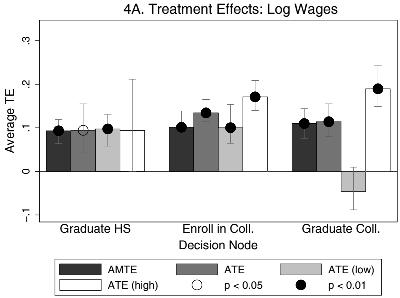
Traditional Treatment Effects

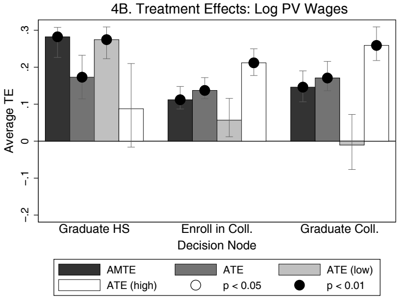
Link to Web Appendix A.14.1

Decompositions:
Into ATE, Sorting Gains and Selection Bias
Link to Web Appendix A.15

Dynamic Treatment Effects

Figure 6: Treatment Effects of Outcomes by Decision Node $E(Y^k|\text{Fix } D_j = 0, Q_j = 1) - E(Y^k|\text{Fix } D_j = 1, Q_j = 1)$

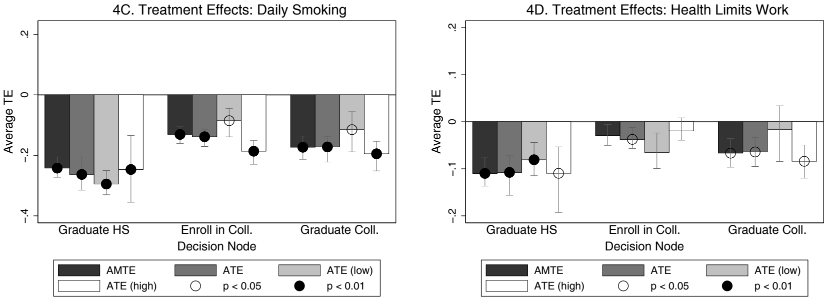




Sorting on Ability

	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

Figure 7: Treatment Effects of Outcomes by Decision Node $E(Y^k|\text{Fix } D_j = 0, Q_j = 1) - E(Y^k|\text{Fix } D_j = 1, Q_j = 1)$



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

- In Appendix A.14.2, we present a variety of treatment effects, including treatment on the treated (TT).

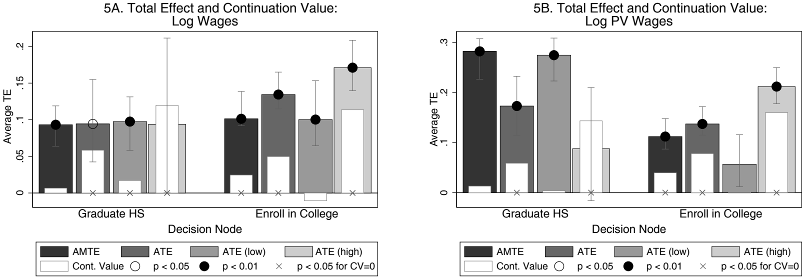
Link to Web Appendix A.14.2

- In Appendix A.15.3, we go further and decompose observed differences in the raw data into average treatment effects, sorting gains, and selection bias (Equation (20)).

Link to Web Appendix A.15.3

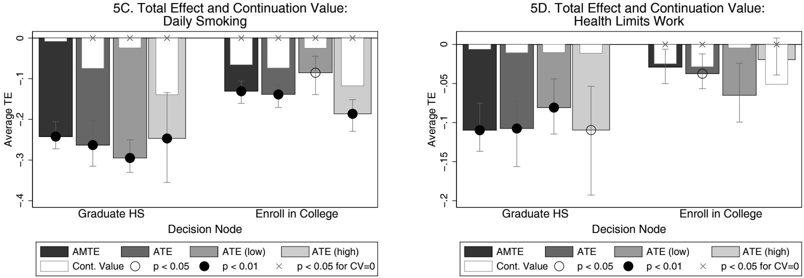
Continuation Values

Figure 8: 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 socio-emotional endowments. Low-ability individuals are those in the bottom 50% of the distributions of both cognitive and socio-emotional 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 30 provides details on how the continuation values and treatment effects are defined.

Figure 8: Dynamic Treatment Effects: Continuation Values and Total Treatment Effects by Node, Cont'd



Notes: High-ability individuals are those in the top 50% of the distributions of both cognitive and socio-emotional endowments. Low-ability individuals are those in the bottom 50% of the distributions of both cognitive and socio-emotional 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 30 provides details on how the continuation values and treatment effects are defined.

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

Figure 9: Average Treatment Effect of Graduating from a Four-Year College by Outcome

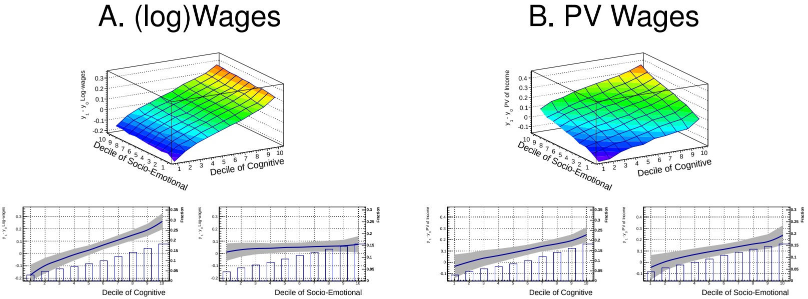
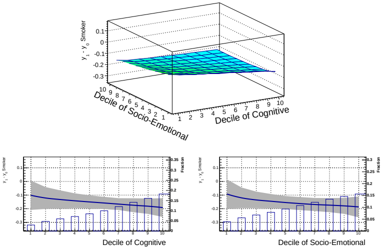
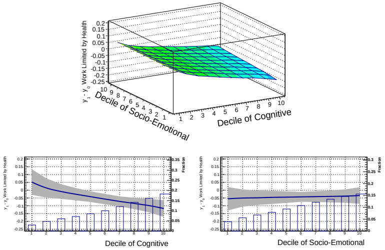


Figure 9: Average Treatment Effect of Graduating from a Four-Year College by Outcome, Cont'd

C. Smoking



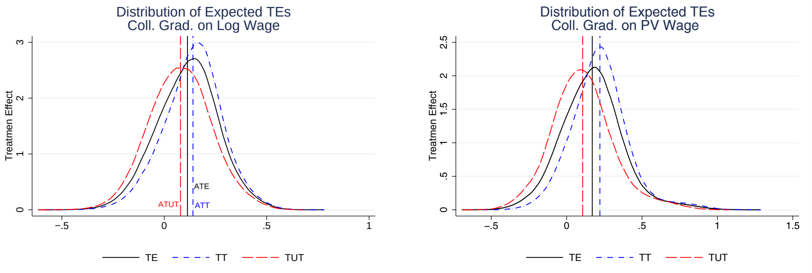
D. Health Limits Work



Link to Appendix A.7

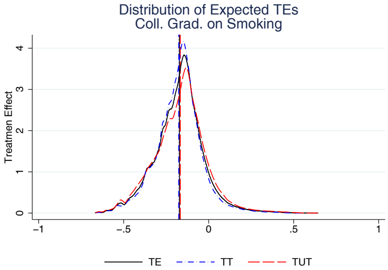
Distributions of Treatment Effects

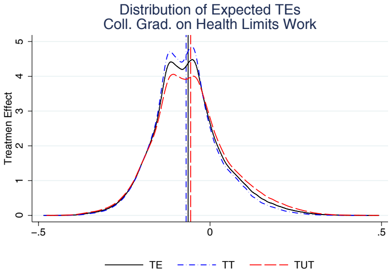
Figure 10: Distributions of Expected Treatment Effects: College Graduation



Notes: Distributions of treatment effects including continuation values for those who reach the educational choice. The vertical lines represent the average treatment effects (ATE, ATT, and ATUT) for each of the distributions.

Figure 10: Distributions of Expected Treatment Effects: College Graduation, Cont'd





Notes: Distributions of treatment effects including continuation values for those who reach the educational choice. The vertical lines represent the average treatment effects (ATE, ATT, and ATUT) for each of the distributions.

Link to Appendix A.9

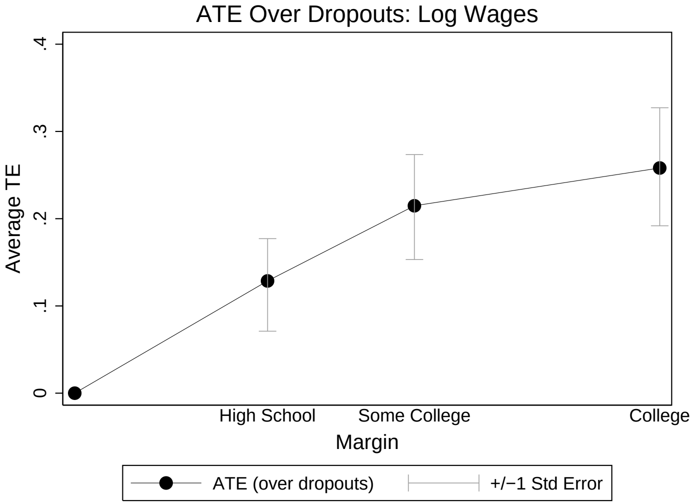
- Strong evidence against rank invariance invoked in quantile treatment effect literature.

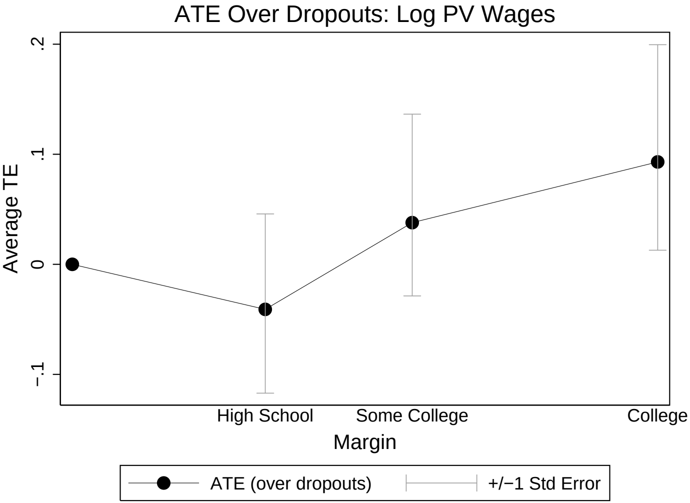
Taking Stock of the Becker-Chiswick-Mincer Model

- OLS-regression adjusted versions are not linear in years of schooling.
- The correlation between ρ_i and S_i is a centerpiece of the modern IV literature.
- The correlation varies across transitions.
- Estimated correlation patterns are consistent with our evidence on sorting gains presented in Web Appendix A.16.

Linearity

Link to Appendix A.8


ATE Over Dropouts: Log Wages
Average TE
0
.1
.2
.3
.4
High School
Some College
College
Margin
ATE (over dropouts)
+/−1 Std Error


ATE Over Dropouts: Log PV Wages
Average TE
.2
.1
0
-.1
High School
Some College
College
Margin
ATE (over dropouts)
+/−1 Std Error

Patterns of Covariance

Link to Appendix A.13

Sorting Gains

Link to Appendix A.16

Summarizing Our Analysis of Causal Effects of Education

1. *Substantial causal benefits for all outcomes analyzed from education, except for GED certification.*
2. *Continuation values are an important component of causal effects for most outcomes except health limits work.*
3. *Substantial benefits from graduating high school that are especially strong for the less able, many of whom currently do not graduate. This suggests strong gains from programs promoting high school graduation.*
4. *For the wage outcomes we study, there is evidence on sorting on gains from graduating college for high-ability persons.*
5. *There are no causal effects of college graduation for low-ability people. College graduation is not for all.*

THE UNIVERSITY OF CHICAGO

6. *There are strong marginal benefits of education for those at the margin of indifference at all nodes. These are largely direct effects with little contribution from continuation values.*
7. *We estimate strong causal effects for the non-monetary outcomes studied. They are particularly strong for high school graduation. There is little evidence of sorting on gains in either non-monetary outcome examined. Continuation values are largely absent for our measure of health. For smoking, continuation values are most pronounced among higher-ability persons. Selection bias is less empirically important for smoking, but is substantial for health limits work.*

Policy Simulations

Policy-Relevant Treatment Effects of Tuition Subsidy

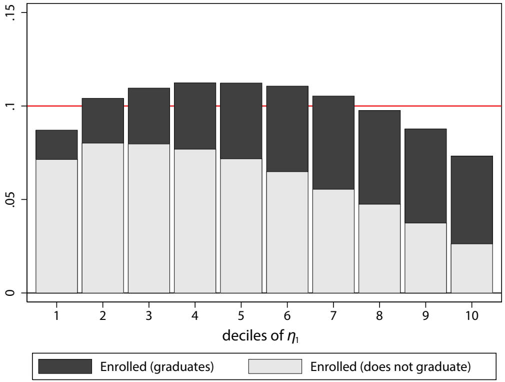
Table 2: PRTE: Standard Deviation Decrease in Tuition

	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 per annum). The PRTE is the average treatment effect of those induced to change educational choices as a result of the policy:
$PRTE^k_{p,p'} := \int\int\int E(Y^k(p') - Y^k(p)|\boldsymbol{X} = \boldsymbol{x}, \boldsymbol{Z} = \boldsymbol{z}, \boldsymbol{\theta} = \overline{\boldsymbol{\theta}})dF_{\boldsymbol{X},\boldsymbol{Z},\boldsymbol{\theta}}(\boldsymbol{x}, \boldsymbol{z}, \overline{\boldsymbol{\theta}}|S(p) \neq S(p'))$. 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.

THE UNIVERSITY OF CHICAGO

Figure 11: PRTE: Who is induced to switch?



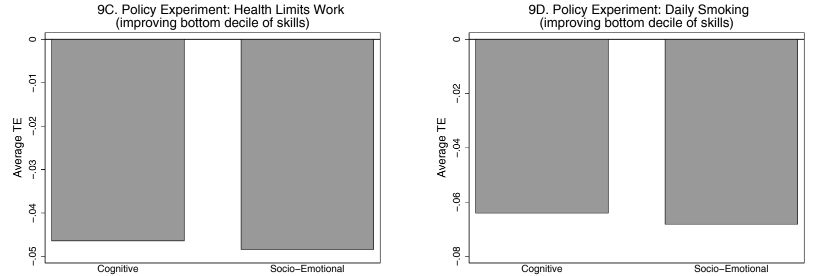
Notes: The figure plots the proportion of individuals induced to switch from the policy that lay in each decile of η_2, where $\eta_1 = -(\boldsymbol{\theta}'\boldsymbol{\lambda}_1 - \nu_2)$. η_2 is the unobserved component of the educational choice model. The deciles are conditional on $Q_1 = 1$, so 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.

Boosting Cognitive and Non-Cognitive Endowments

Figure 12: Policy Experiments

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

Figure 12: Policy Experiments, Cont'd



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

Testing the Two Factor Assumption

Comparisons with Alternative Treatment Effect Estimators

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

	Linear Regression				Matching		Model
HS Grad.	OLS	OLS-P	OLS-F	OLS-FI	NNM(3)-F	PSM-F	ATE_j^k *
Wages	0.205	0.073	0.155	0.159	0.098	0.132	0.094
SE	(0.025)	(0.026)	(0.025)	(0.035)	(0.037)	(0.051)	(0.056)
PV-Wage	0.380	0.213	0.318	0.277	0.196	0.226	0.173
SE	(0.030)	(0.031)	(0.030)	(0.041)	(0.053)	(0.058)	(0.059)
Smoking	-0.327	-0.246	-0.281	-0.301	-0.260	-0.271	-0.263
SE	(0.028)	(0.029)	0.028	0.041	(0.058)	(0.060)	(0.056)
Health-Limits-Work	-0.178	-0.115	-0.151	-0.150	-0.048	-0.095	-0.108
SE	(0.022)	(0.024)	(0.023)	(0.033)	(0.029)	(0.036)	(0.042)

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

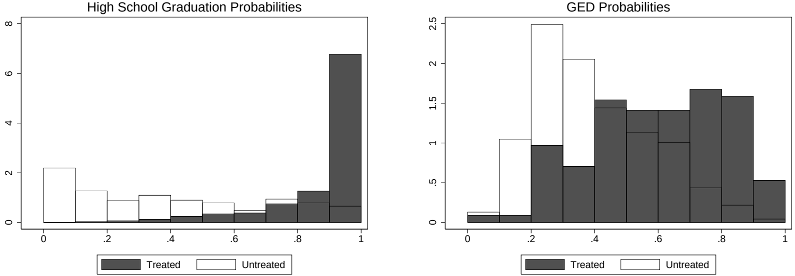
	Linear Regression				Matching		Model
Coll. Enroll	OLS	OLS-P	OLS-F	OLS-FI	NNM(3)-F	PSM-F	ATE_j^k
Wages	0.223	0.121	0.186	0.190	0.177	0.207	0.134
SE	(0.023)	(0.024)	(0.024)	(0.023)	(0.029)	(0.031)	(0.025)
PV-Wage	0.221	0.109	0.176	0.171	0.188	0.226	0.137
SE	(0.027)	(0.029)	(0.028)	(0.027)	(0.030)	(0.032)	(0.029)
Smoking	-0.177	-0.138	-0.165	-0.170	-0.129	-0.144	-0.139
SE	(0.026)	(0.028)	(0.027)	(0.028)	(0.029)	(0.058)	(0.028)
Health-Limits-Work	-0.085	-0.037	-0.066	-0.057	-0.029	-0.042	-0.037
SE	(0.020)	(0.022)	(0.021)	(0.021)	(0.022)	(0.029)	(0.022)

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

	Linear Regression				Matching		Model
Coll. Grad	OLS	OLS-P	OLS-F	OLS-FI	NNM(3)-F	PSM-F	ATE_j^k
Wages	0.210	0.146	0.184	0.185	0.173	0.143	0.114
SE	(0.032)	(0.034)	(0.033)	(0.035)	(0.041)	(0.051)	(0.037)
PV-Wage	0.243	0.163	0.208	0.228	0.191	0.269	0.171
SE	(0.037)	(0.040)	(0.038)	(0.037)	(0.039)	(0.042)	(0.040)
Smoking	-0.209	-0.171	-0.195	-0.192	-0.132	-0.161	-0.172
SE	(0.032)	(0.035)	(0.033)	(0.035)	(0.039)	(0.039)	(0.043)
Health-Limits-Work	-0.085	-0.069	-0.078	-0.077	-0.048	-0.051	-0.064
SE	(0.024)	(0.026)	(0.025)	(0.026)	(0.026)	(0.027)	(0.031)

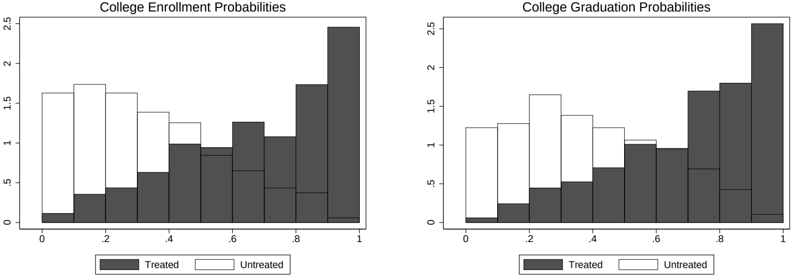
Notes: We estimate the ATE inclusive of continuation values for each outcome and and educational choice using a variety of methods. All models are estimated for populations that reach the node being analyzed ($Q_j = 1$), inclusive of those who go on to further schooling in order to make them comparable to the ATE from our model that includes continuation values (Equation (19)). All OLS models use the full set of controls listed in Table 1. "OLS" estimates a linear model using a schooling dummy (Q_{j+1}), and controls ($Y = Q_{j+1} b_j + \boldsymbol{X}'\boldsymbol{\beta} + \epsilon$). "OLS-P" estimates a linear model using a schooling dummy, a vector of controls, and three measures of abilities arrayed in a vector $\boldsymbol{A}$: summed ASVAB scores, GPA, and an indicator of risky behavior ($Y = Q_{j+1} b_j + \boldsymbol{X}'\boldsymbol{\beta} + \boldsymbol{A}'\boldsymbol{\alpha} + \epsilon$). All models ending in "-F" are estimated using Bartlett factor scores (Bartlett, 1937, 1938) estimated using our measurement system, but using the built-in routine for estimating factor models in STATA via maximum likelihood, not accounting for schooling at the time of the test. "OLS-F" estimates the model $Y = Q_{j+1} b_j + \boldsymbol{X}'\boldsymbol{\beta} + \hat{\boldsymbol{\theta}}'\boldsymbol{\alpha} + \epsilon$ where $\hat{\boldsymbol{\theta}}$ are the Bartlett factor scores described above. "OLS-FI" is similar to "OLS-F" except that Q_{j+1} is interacted with the $\boldsymbol{X}$ and $\hat{\boldsymbol{\theta}}$ allowing the coefficients on the controls and abilities to vary by education level. "NNM(3)-F" is the estimated treatment effect from nearest-neighbor matching with 3 neighbors. Neighbors are matched on their Bartlett cognitive factor, Bartlett non-cognitive factor, and an index constructed from their observed characteristics ($\boldsymbol{Z}$) generating choices as described in the Web Appendix. "PSM-F" presents the estimated average treatment effect from propensity score matching where propensity scores are estimated using Bartlett cognitive factors, Bartlett non-cognitive factors, the full set of control variables, and the full set of node-specific instruments. "ATE_j^k" presents the estimated average treatment effect from the model presented in this paper (inclusive of continuation value), corresponding to Equation (11).

Figure 13: Supports At Each Decision Node



Notes: Each plot is for the population who reaches that decision node in the data. “Treated” are those who choose to complete the reported level of schooling, while “Untreated” are those who choose to not complete the reported level of schooling (but reach the decision node). Probabilities are estimated by a probit model that controls for the set of control variables and decision specific instruments used and reported in the paper.

Figure 13: Supports At Each Decision Node, Cont'd



Notes: Each plot is for the population who reaches that decision node in the data. "Treated" are those who choose to complete the reported level of schooling, while "Untreated" are those who choose to not complete the reported level of schooling (but reach the decision node). Probabilities are estimated by a probit model that controls for the set of control variables and decision specific instruments used and reported in the paper.

Summary and Conclusion

- Develop and estimate a robust dynamic model of schooling and its causal consequences for earnings, health, and healthy behaviors.
- Borrow features from both the reduced form treatment effect literature and the structural literature.
- Estimated model passes a variety of goodness of fit and model specification tests.

- Agents can be irrational and myopic in making schooling decisions.
- Use model to test some of the maintained rationality and information processing assumptions in the dynamic discrete choice literature on education.
- Use dynamic choice model to estimate causal effects arising from multiple levels of schooling rather than just the binary comparisons typically featured in the literature on treatment effects and in many structural papers.
- Analyze the *ex post* returns to education for people at different margins of choice and analyze a variety of economically interesting policy counterfactuals.
- We decompose the benefits of schooling at different levels into direct components and indirect components arising from continuation values.

- Standard estimates of the benefits of education based only on direct components of treatment effects underestimate the full benefits of education.
- Without imposing rationality, we nonetheless find evidence consistent with it.
- 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 of average treatment effects, provided we condition on endowments of cognitive and non-cognitive skills and correct for measurement error in the proxies.
- We use our estimated model to conduct two policy experiments.

- We determine the groups that benefit from a tuition reduction policy and what those benefits are.
- The early research on human capital was casual about agent heterogeneity.
- It ignored selection bias and sorting gains from schooling.
- Later work by Griliches (1977) focused on selection bias (“ability bias”), but ignored sorting gains.
- In this paper, we quantify both components of outcome equations.
- Schooling has strong causal effects on market and non-market outcomes.
- Both cognitive and non-cognitive endowments affect schooling choices and outcomes.
- People sort into schooling based on realized incremental gains.

Web Appendix A.4
Precise Parameterization of the Model and Our Likelihood

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

Parameterization of the Model

- Following a well-established tradition in the literature, in this paper we approximate I_j using a linear-in-the-parameters model:

$$I_j = \boldsymbol{Z}_j' \beta_j + \boldsymbol{\theta}' \boldsymbol{\alpha}_j - \nu_j, \; j \in \{0, \dots, \overline{s} - 1\} \tag{21}$$

where $\boldsymbol{Z}_j$ is a vector of variables (and functions of these variables) observed by the economist that determine the schooling transition decision of the agent with schooling level j and $\boldsymbol{\theta}$ is a vector of unobserved (by the economist) endowments.

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

- Outcomes are also approximated by a linear-in-the-parameters model.

$$\tilde{Y}_s^k = \boldsymbol{X}_s^k \boldsymbol{\beta}_s^k + \boldsymbol{\theta}' \boldsymbol{\alpha}_s^k + \omega_s^k, \tag{22}$$

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

THE UNIVERSITY OF CHICAGO

Measurement System for Unobserved Endowments θ

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

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

- Let θ^C and θ^{SE} denote the levels of cognitive and socio-emotional endowments and suppose $\boldsymbol{\theta} = (\theta^C, \theta^{SE})$.
- We allow θ^C and θ^{SE} to be correlated.
- Let $t^C_{m,s}$ be the m^{th} cognitive test score and $t^{C,SE}_{m,s}$ the m^{th} measure influenced by both cognitive and socio-emotional endowments, all measured at schooling level s.
- Parallel to the treatment of the index and outcome equations, we assume linear measurement systems for these variables:

$$t^C_{m,s} = \boldsymbol{X}^C_{m,s}\boldsymbol{\beta}^C_{m,s} + \theta^C \alpha^C_{m,s} + e^C_{m,s}, \tag{23}$$

$$t^{C,SE}_{m,s} = \boldsymbol{X}^{C,SE}_{m,s}\boldsymbol{\beta}^{C,SE}_{m,s} + \theta^C \tilde{\alpha}^C_{m,s} + \theta^{SE}\tilde{\alpha}^{SE}_{m,s} + e^{C,SE}_{m,s}. \tag{24}$$

- The structure assumed in Equations (23) and (24) is identified even when allowing for correlated factors, if we have one measure that is a determinant of cognitive endowments ($t^C_{m,s}$) and at least four measures that load on both cognitive ability and socio-emotional ability, and conventional normalizations are assumed.
- In the main text we report results from models that use measurements to proxy θ. Let $H^m_{i,s}$ be an indicator for if an individual i took test t at schooling level s.

THE UNIVERSITY OF CHICAGO

Specification of the measurement system

- When estimating the factor model, we must make normalizations and exclusion restrictions.
- There is no precise method for determining these restrictions.
- As laid out below, we use a collection of empirical evidence and theory for determining our measurement system.

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

- Following Heckman et al. (2006), the model imposes that the ASVAB measures do not load on socio-emotional factors.
- If any particular ASVAB score is excluded, it does not substantively change the analysis.
- Course grades are assumed to load on both the cognitive and socio-emotional factors.
- As discussed in the main paper, this assumption by Duckworth and Seligman (2005) and Borghans et al. (2011), who both find that grades are largely determined by endowments other than cognitive ability.

- As discussed above, the identification strategy used in the paper requires one measure that loads exclusively on cognitive ability.
- We assume ASVAB tests only measure cognition.
- Subject-specific 9th grade GPA, educational choices, and early risky behavior are assumed to depend on both factors.
- We include violent behavior, smoking regularly by age 15, drinking regularly by age 15, ever smoking marijuana by age 15, and sexual intercourse by age 15 as early "outcomes" in our model.
- These do not inform the cognitive or socio-emotional factor but provide a robustness check of our interpretation of our factors and aiding in interpretation.

Likelihood

- We estimate our model in two stages using maximum likelihood.
- The measurement system, the distribution of latent endowments, and the model of schooling decisions are estimated in the first stage.
- The outcome equations are estimated in the second stage using estimates from the first stage.
- We follow Hansen et al. (2004), and correct estimated factor distributions for the causal effect of choices on measurements by jointly estimating the choice and measurement equations in the first stage.

- The distribution of the latent factors is estimated only using data on educational choices and measurements.
- This allows us to interpret the factors as cognitive and socio-emotional endowments.
- It links our estimates to an emerging literature on the economics of personality and psychological traits but the link is not strictly required if we only seek to control for selection in schooling choices and do not seek to identify the system of measurement equations presented in the text.
- We do not use the final outcome system to estimate the distribution of factors, thus avoiding tautologically strong predictions of outcomes from the system of estimated factors.

- Let $\mathcal{J}$ denote the set of possible terminal states.
- Let $D_j \in \mathcal{D}$ be the set of possible transition decisions that can be taken by the individual over the decision horizon.
- Let $\mathcal{S}$ denote the finite and bounded set of stopping states with $S = s$ if the agent stops at $s \in \mathcal{S}$.
- Define $\bar{s}$ as the highest attainable element in $\mathcal{S}$. $Q_j = 1$ indicates that an agent *gets to* decision node j.
- $Q_j = 0$ if the person never gets there.
- The history of nodes visited by an agent can be described by the collection of the Q_j such that $Q_j = 1$.

- To ensure consistent notation, we define $Q_0 := 1$. $\boldsymbol{Y}_i$, $\boldsymbol{D}_i$, and $\boldsymbol{M}_i$ are vectors of individual i's outcomes, educational decisions and measurements of endowments, respectively.
- $\boldsymbol{Z}$ is a vector of observed determinants of decisions, $\boldsymbol{X}$ is a vector of observed determinants of outcomes, and $\boldsymbol{\theta}$ is the vector of unobserved endowments.
- The $\boldsymbol{Z}$ can include all variables in $\boldsymbol{X}$.
- When instrumental variable methods are used to identify components of the model, it is assumed that there are some variables in $\boldsymbol{Z}$ not in $\boldsymbol{X}$.

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

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

- where $f(\cdot)$ denotes a probability density function. The last step is justified from the assumptions (A-1a) – (A-1g).

- For the first stage, the sample likelihood is

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

- where we integrate over the distributions of the latent factors.

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

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

- The log-likelihood can be rewritten as

$$\log \mathcal{L}^1 = \sum_i \log \int_{\overline{\theta} \in \Theta} \left[\prod_{j \in \mathcal{S} \setminus \overline{s}} f(\boldsymbol{D}_{i,j} | \boldsymbol{Z}_{i,j}, \boldsymbol{\theta} = \overline{\boldsymbol{\theta}}; \gamma_j)^{Q_{i,j}} \right] \times \left[\prod_{m=1}^{N_M} \prod_{s \in \mathcal{S}^M} f(\boldsymbol{M}_{i,m,s} | \boldsymbol{X}_{i,m,s}, \boldsymbol{\theta} = \overline{\boldsymbol{\theta}}; \gamma_{m,s})^{H^m_{i,s}} \right] \times \left[\sum_\ell \rho_\ell f^\ell_\theta(\overline{\boldsymbol{\theta}}; \gamma^\ell_\theta) \right] d\overline{\boldsymbol{\theta}}$$

$$= \sum_i \log \left\{ \sum_\ell \rho_\ell \int_{\overline{\theta} \in \Theta} \left[\prod_{j \in \mathcal{S} \setminus \overline{s}} f(\boldsymbol{D}_{i,j} | \boldsymbol{Z}_{i,j}, \boldsymbol{\theta} = \overline{\boldsymbol{\theta}}; \gamma_j)^{Q_{i,j}} \right] \left[\prod_{m=1}^{N_M} \prod_{s \in \mathcal{S}^M} f(\boldsymbol{M}_{i,m,s} | \boldsymbol{X}_{i,m,s}, \boldsymbol{\theta} = \overline{\boldsymbol{\theta}}; \gamma_{m,s})^{H^m_{i,s}} \right] f^\ell_\theta(\overline{\boldsymbol{\theta}}; \gamma^\ell_\theta) d\overline{\boldsymbol{\theta}} \right\}.$$

THE UNIVERSITY OF CHICAGO

- We use Gauss-Hermite quadrature to numerically evaluate the integral.
- Although there are a number of ways to numerically evaluate an integral, one advantage of Gaussian quadrature is that it gives analytical expressions for the integral.
- Analytical expressions for the gradient and hessian can then be calculated which allows for the use of efficient second-order optimization routines.
- Since the models are very smooth, a second-order optimization strategy leads to faster convergence.

- Given that we are using a mixture of normals, $f_\theta^\ell(\boldsymbol{\theta}; \gamma_\theta^\ell) = \phi(\boldsymbol{\theta}; \boldsymbol{\mu}_\theta^\ell, \boldsymbol{\sigma}_\theta^\ell)$ is a multivariate normal, where we assume for now that the components are independent.
- This assumption can easily be relaxed, but keeping it simplifies notation.
- The Gauss-Hermite quadrature rule is $\int f(v)e^{-v^2}dv = \sum_n \lambda_n f(v_n)$, where the weights, λ_n, and nodes, v_n, are defined by the quadrature rule depending on the number of points used (Judd, 1998).

- Applying the Gauss-Hermite rule and making a change of variables ($\bar{\theta} = \sqrt{2}\boldsymbol{\sigma}_\theta^\ell \circ \boldsymbol{v}_n + \boldsymbol{\mu}_\theta^\ell$), we can rewrite the likelihood as

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

- where $\boldsymbol{v_n} = \begin{pmatrix} v_{n1}, & v_{n2} \end{pmatrix}$ represents the vector of nodes.

- Multivariate normal variables with correlated components can be rewritten as the sum of independent standard normal variables and then one can use the same procedure.
- The goal of the first stage is then to maximize $\log \mathcal{L}^1$ and obtain estimates $\hat{\gamma}_j$, $\hat{\gamma}_{m,s}$, $\hat{\boldsymbol{\sigma}}_\theta^\ell$, $\hat{\boldsymbol{\mu}}_\theta^\ell$, and $\hat{\rho}_\ell$ for $j \in J^{MS}$.
- If a density $f(\cdot)$ cannot be calculated either because of missing data or because that model does not apply to individual i, then $f(\cdot) = 1$.

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

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

- Secondly, we can correct for measurement error in the outcome equations by integrating over the PDF of the latent factor.
- The likelihood for the outcome equations is

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

- where $H_{i,s}$ is an indicator for the highest level of schooling attained by individual i.

THE UNIVERSITY OF CHICAGO

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

Return to main text

Web Appendix A.5
Goodness of Fit

- This section tests the goodness of fit of our various measurement, education, and outcome equations.
- For continuous models we compare means and standard deviations while for discrete outcomes we compare proportions.
- Table 9 jointly test for equality of means in the outcomes for 16 unique sub-populations.

Table 4: Goodness of Fit - Schooling Choice

Schooling Level	Data	Model	p-value
High School Dropout	0.131	0.122	0.980
High School Graduate	0.370	0.377	0.989
Some College	0.168	0.176	0.982
College Graduate	0.230	0.222	0.986

Notes: The simulated data (Model) contains one million observations generated from the model estimates. The actual data (Actual) contains 2242 observations from the NLSY79 sample of Males.
(a) Goodness of fit is tested using a χ^2 test that the two proportions are equal, where the Null Hypothesis is *Model=Data*.

Table 5: Goodness of Fit - Early Risky and Reckless Behavior

Outcome	Actual	Model	*p*-value[a]
Early marijuana[c]	0.338	0.338	0.999
Early daily smoking[c]	0.187	0.186	0.999
Early drinking[c]	0.188	0.188	0.999
Early intercourse[c]	0.163	0.161	0.994
Early Reckless (9th-11th)[b]	0.607	0.599	0.987
Early Reckless (12th)[b]	0.533	0.541	0.988

Notes: The simulated data (Model) contains one million observations generated from the model estimates. The actual data (Actual) contains 2242 observations from the NLSY79 sample of Males.
(a) Goodness of fit is tested using a χ^2 test that the two proportions are equal, where the Null Hypothesis is that the model fits the data. (b) The reckless and violent variables are taken from the NSLY 1980 Illegal Activities Supplement. (c) Early is defined as engaging in risky behavior before 15 years old.

Table 6: Goodness of Fit - ASVAB and Grade models

	Mean		Std Dev		
	Data	Model	Data	Model	*p*-value
ASVAB Tests					
Arithmetic Reasoning ($<$ 12)	-0.291	-0.354	0.932	0.898	0.035
Word Knowledge ($<$ 12)	-0.448	-0.530	1.082	1.057	0.017
Paragraph Comprehension ($<$ 12)	-0.513	-0.588	1.180	1.166	0.049
Numerical Operations ($<$ 12)	-0.519	-0.574	0.963	0.927	0.074
Math Knowledge ($<$ 12)	-0.320	-0.389	0.887	0.835	0.015
Coding Speed ($<$ 12)	-0.599	-0.643	0.782	0.767	0.075
Arithmetic Reasoning ($=$ 12)	0.196	0.186	0.862	0.823	0.720
Word Knowledge ($=$ 12)	0.132	0.126	0.778	0.735	0.837
Paragraph Comprehension ($=$ 12)	0.039	0.029	0.796	0.751	0.699
Numerical Operations ($=$ 12)	-0.012	-0.020	0.890	0.848	0.787
Math Knowledge ($=$ 12)	0.001	-0.022	0.812	0.745	0.417
Coding Speed ($=$ 12)	-0.163	-0.167	0.773	0.749	0.866

Notes: The simulated data (Model) contains one million observations generated from the Model's estimates. The actual data (Actual) contains 2242 observations from the NLSY79 sample of Males.The numbers inside the parentheses describe the years of schooling at the time of the test. The ASVAB models are estimated separately for those with less than twelve years ($<$ 12), those who are high school graduates (=12), and those who have attended college ($>$ 12) at the time they took the ASVAB tests. (a) The *p*-values reported are from a T-test for the equivalence of the means where the null hypothesis is that *Actual* $=$ *Model*.

Table 6: Goodness of Fit - ASVAB and Grade models, Cont'd

	Mean		Std Dev		
	Data	Model	Data	Model	*p*-value
ASVAB Tests					
Arithmetic Reasoning ($>$ 12)	0.942	0.905	0.665	0.643	0.258
Word Knowledge ($>$ 12)	0.744	0.754	0.328	0.307	0.552
Paragraph Comprehension ($>$ 12)	0.636	0.622	0.324	0.314	0.377
Numerical Operations ($>$ 12)	0.580	0.560	0.589	0.571	0.475
Math Knowledge ($>$ 12)	0.975	0.947	0.736	0.720	0.438
Coding Speed ($>$ 12)	0.474	0.460	0.654	0.639	0.665
9th Grade GPA					
GPA Language	-0.117	-0.175	0.969	0.973	0.012
GPA Social Sciences	-0.012	-0.074	0.985	0.993	0.018
GPA Science	0.026	-0.017	0.955	0.939	0.085
GPA Math	-0.011	-0.050	0.977	0.975	0.083

Notes: The simulated data (Model) contains one million observations generated from the Model's estimates. The actual data (Actual) contains 2242 observations from the NLSY79 sample of Males.The numbers inside the parentheses describe the years of schooling at the time of the test. The ASVAB models are estimated separately for those with less than twelve years ($<$ 12), those who are high school graduates (=12), and those who have attended college ($>$ 12) at the time they took the ASVAB tests. (a) The *p*-values reported are from a T-test for the equivalence of the means where the null hypothesis is that $Actual = Model$.

Table 7: Goodness of Fit - Discrete Outcomes

Outcome	Actual	Model	*p*-value[a]
Smoking Age 30	0.385	0.387	0.997
High school dropouts	0.674	0.650	0.959
High school graduates	0.390	0.383	0.989
Some college	0.337	0.339	0.995
Four-year college graduate	0.146	0.166	0.955
Health Limits Work	0.227	0.226	0.997
High school dropouts	0.392	0.412	0.968
High school graduates	0.232	0.229	0.994
Some college	0.184	0.179	0.992
Four-year college graduate	0.091	0.099	0.980

Notes: The simulated data (Model) contains one million observations generated from the Model's estimates. The actual data (Actual) contains 2242 observations from the NLSY79 sample of Males.
(a) Goodness of fit is tested using a χ^2 test that the two proportions are equal, where the Null Hypothesis is that the model predictions fits the data.

Table 8: Goodness of Fit - Continuous Outcomes

	Mean		Std Dev		
	Actual	Model	Actual	Model	*p*-value
Log Wages (30)	2.612	2.604	0.229	0.223	0.132
High school dropouts	2.291	2.247	0.135	0.130	0.000
High school graduates	2.531	2.528	0.184	0.182	0.637
Some college	2.665	2.677	0.207	0.200	0.283
Four-year college graduate	2.932	2.949	0.188	0.186	0.039
PVLog Wages (30)	12.315	12.317	0.397	0.395	0.876
High school dropouts	11.787	11.681	0.366	0.391	0.000
High school graduates	12.275	12.275	0.273	0.262	0.983
Some college	12.422	12.432	0.257	0.255	0.499
Four-year college graduate	12.764	12.817	0.266	0.272	0.000

Notes: The simulated data (Model) contains one million observations generated from the Model's estimates. The actual data (Actual) contains 2242 observations from the NLSY79 sample of Males.
(a) The *p*-values reported are from a T-test for the equivalence of the means where the null hypothesis is that the model predictions fits the data.

Table 9: χ^2 Test for equality of the means across sub-populations

	p-value
Log Wages (30)	0.44
Log PV Wage Income (30)	0.07
Health Limits Work	0.40
Smoking (30)	0.06

Notes: This table jointly tests if the observed and simulated outcome means are equal for 16 unique sub-populations. Subpopulations are the unique groups defined by the binary variables: white, southern residence at age 14, family income greater than $20,500 in 1979, and mother's highest grade completed is less than 12. The reported *p*-value is for the χ^2-test against the null hypothesis that the means are equal for the observed and simulated data.

Return to main text

Web Appendix A.14.1
Treatment Effects Across Final Schooling Levels

- We first present the traditional treatment effects across adjacent final levels corresponding Figure 5 in the text.
- Tables 10-13 report traditional treatment effects by final schooling level: $\text{ATE}_{s,s'}$ (shown in Figure 5), treatment on the treated ($\text{TT}_{s,s'}$), and treatment on the untreated ($\text{TUT}_{s,s'}$).
- We also display the raw difference ("observed") also shown in Figure 5 and ATEs derived from our model but computed for the entire population ($\text{ATE}^{\dagger}$).

- These tables show the gains from switching from one final schooling level to another.
- All education levels are compared to dropouts as well as the level of education directly below it for both branches of Figure 1.
- $\text{ATE}_{s,s'}$ is the ATE computed from our model over the entire population.
- The other treatment parameters are defined for populations at the two final schooling levels.
- The difference between $\text{ATE}^{\dagger}$ and ATE is a measure of how different the characteristics are for those in the general population from those at the indicated pair of final schooling states.

- The differences between TT and ATE are measures of sorting gains.
- The differences between TUT and ATE are measures of sorting losses.
- Thus, in Table 10 the characteristics of people at the node deciding between the GED and dropping out are substantially less favorable than those in the general population, but there are little sorting gains or losses for this pair of alternatives.
- At the same time, there are substantial sorting gains (and losses) for those choosing between graduating college and not completing college.
- Moreover, the characteristics of people at this margin of choice are far more favorable.

Table 10: The Effects of Education on Log Wages, by *Final Schooling Level* using High School Dropouts and Adjacent Schooling Levels as Baselines

	Observed	$ATE^{\dagger}_{s,s'}$	$ATE_{s,s'}$	$TT_{s,s'}$	$TUT_{s,s'}$	OLS
GED vs. HS Dropout	0.14	0.12	0.06	0.06	0.06	0.05
		(0.08)	(0.05)	(0.05)	(0.05)	(0.04)
HS Graduate vs. HS Dropout	0.24	0.13*	0.12**	0.12**	0.11**	0.08*
		(0.05)	(0.04)	(0.05)	(0.04)	(0.03)
Some College vs. HS Dropout	0.37	0.21**	0.21**	0.21**	0.22**	0.14**
		(0.06)	(0.05)	(0.07)	(0.06)	(0.04)
Four Year College Degree vs. HS Dropout	0.64	0.26**	0.27**	0.34**	0.15*	0.28**
		(0.07)	(0.07)	(0.10)	(0.07)	(0.04)
Some College vs. HS Graduate	0.13	0.09**	0.10**	0.07**	0.11**	0.06*
		(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Four Year College Degree vs. Some College	0.26	0.04	0.11**	0.14**	0.08**	0.13**
		(0.04)	(0.04)	(0.04)	(0.04)	(0.03)

Table 11: The Effects of Education on Log PV of wages, by *Final Schooling Level* using High School Dropouts and Adjacent Schooling Levels as Baselines

	Observed	$ATE^{\dagger}_{s,s'}$	$ATE_{s,s'}$	$TT_{s,s'}$	$TUT_{s,s'}$	OLS
GED vs. HS Dropout	0.17	-0.20*	-0.11	-0.14*	-0.08	-0.01
		(0.10)	(0.06)	(0.06)	(0.08)	(0.05)
HS Graduate vs. HS Dropout	0.49	-0.04	0.07	-0.01	0.30**	0.19**
		(0.08)	(0.06)	(0.08)	(0.05)	(0.04)
Some College vs. HS Dropout	0.64	0.04	0.15*	-0.06	0.45**	0.24**
		(0.08)	(0.07)	(0.10)	(0.08)	(0.04)
Four Year College Degree vs. HS Dropout	0.98	0.09	0.06	-0.09	0.33**	0.39**
		(0.09)	(0.10)	(0.13)	(0.10)	(0.05)
Some College vs. HS Graduate	0.15	0.08*	0.09**	0.06*	0.11**	0.06
		(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Four Year College Degree vs. Some College	0.34	0.06	0.17**	0.22**	0.11*	0.15**
		(0.06)	(0.04)	(0.05)	(0.05)	(0.04)

Table 12: The Effects of Education on Daily Smoking, by *Final Schooling Level* using High School Dropouts and Adjacent Schooling Levels as Baselines

	Observed	$ATE^{\dagger}_{s,s'}$	$ATE_{s,s'}$	$TT_{s,s'}$	$TUT_{s,s'}$	OLS
GED vs. HS Dropout	-0.05	0.04	0.02	0.01	0.02	-0.03
		(0.09)	(0.05)	(0.06)	(0.05)	(0.05)
HS Graduate vs. HS Dropout	-0.28	-0.16*	-0.20**	-0.18*	-0.26**	-0.24**
		(0.08)	(0.06)	(0.08)	(0.05)	(0.04)
Some College vs. HS Dropout	-0.34	-0.22**	-0.23**	-0.20*	-0.27**	-0.28**
		(0.09)	(0.07)	(0.10)	(0.08)	(0.05)
Four Year College Degree vs. HS Dropout	-0.53	-0.38**	-0.38**	-0.36**	-0.41**	-0.47**
		(0.09)	(0.09)	(0.14)	(0.09)	(0.05)
Some College vs. HS Graduate	-0.05	-0.05*	-0.06*	-0.07*	-0.05	-0.04
		(0.03)	(0.03)	(0.03)	(0.04)	(0.03)
Four Year College Degree vs. Some College	-0.19	-0.16**	-0.17**	-0.18**	-0.17**	-0,19**
		(0.04)	(0.04)	(0.05)	(0.04)	(0.04)

Table 13: The Effects of Education on Health Limits Work, by *Final Schooling Level* using High School Dropouts and Adjacent Schooling Levels as Baselines

	Observed	$ATE^{\dagger}_{s,s'}$	$ATE_{s,s'}$	$TT_{s,s'}$	$TUT_{s,s'}$	OLS
GED vs. HS Dropout	-0.01	-0.03	0.06	0.05	0.06	0.04
		(0.08)	(0.05)	(0.06)	(0.06)	(0.04)
HS Graduate vs. HS Dropout	-0.16	-0.13*	-0.11*	-0.13*	-0.06	-0.08**
		(0.07)	(0.05)	(0.06)	(0.04)	(0.03)
Some College vs. HS Dropout	-0.21	-0.15**	-0.16**	-0.16*	-0.16*	-0.10**
		(0.07)	(0.06)	(0.09)	(0.07)	(0.04)
Four Year College Degree vs. HS Dropout	-0.30	-0.20**	-0.20**	-0.21**	-0.18*	-0.15**
		(0.08)	(0.09)	(0.13)	(0.09)	(0.04)
Some College vs. HS Graduate	-0.05	-0.02	-0.03	-0.00	-0.04	-0.02
		(0.03)	(0.03)	(0.02)	(0.03)	(0.03)
Four Year College Degree vs. Some College	-0.09	-0.05	-0.06*	-0.07*	-0.06*	-0.05
		(0.05)	(0.03)	(0.03)	(0.04)	(0.03)

Return to main text

Web Appendix A.14.2
Treatment Effects Across Nodes
(Including Continuation Values)

- We next present the traditional treatment effects including continuation values.
- We show two tables for each outcome analyzed for populations conditional on $Q_j = 1$.
- The first in the format similar to that of Tables 10-13 and shows the population-wide average treatment effect, the average treatment effect for those who reach the node, treatment on the treated, treatment on the untreated (conditional on making it to the decision), and the average marginal treatment effect.
- These results are shown for all four branches of Figure 1.

- Each treatment effect is further broken into low-ability and high-ability samples where low ability individuals are in the bottom half of both cognitive and non-cognitive ability, while high-ability individuals are in the top half of both cognitive and non-cognitive individuals.
- The second table for each outcome shows the various treatment effects (population-wide average treatment effect, average treatment effect for those who reach the node, treatment on the treated, treatment on the untreated (conditional on making it to the decision), and the average marginal treatment effect) and decomposes them into their total effect and their direct effect (excluding option value).
- This is shown for each educational node.

Table 14: The Effects of Education on Log Wages, by *Decision Node*

	%	$ATE_j^{\dagger}$	ATE_j	TT_j	TUT_j	$AMTE_j$
A. Graduating from HS vs. Dropping from HS						
All		0.09*	0.09*	0.09	0.10**	0.09**
		(0.06)	(0.06)	(0.07)	(0.03)	(0.03)
Low Ability	0.31	0.10**	0.10**	0.09*	0.10**	
		(0.04)	(0.04)	(0.05)	(0.04)	
High Ability	0.31	0.09	0.09	0.10	0.07	
		(0.11)	(0.11)	(0.11)	(0.07)	
B. **Getting a GED** vs. **HS Dropout**						
All		0.12	0.06	0.06	0.06	0.06
		(0.08)	(0.05)	(0.05)	(0.05)	(0.05)
Low Ability	0.61	0.02	0.01	0.01	0.02	
		(0.06)	(0.05)	(0.06)	(0.06)	
High Ability	0.06	0.23	0.24*	0.23*	0.25*	
		(0.15)	(0.11)	(0.11)	(0.11)	

Table 14: The Effects of Education on Log Wages, by *Decision Node*, Cont'd

	%	ATE_j †	ATE_j	TT_j	TUT_j	$AMTE_j$
C. College Enrollment vs. **HS Graduate**						
All		0.13**	0.13**	0.14**	0.13**	0.10**
		(0.03)	(0.03)	(0.03)	(0.03)	(0.02)
Low Ability	0.22	0.10*	0.10**	0.08**	0.11**	
		(0.05)	(0.04)	(0.04)	(0.04)	
High Ability	0.38	0.17**	0.17**	0.18**	0.15**	
		(0.03)	(0.03)	(0.04)	(0.03)	
D. **4-year college degree** vs. **Some College**						
All		0.04	0.11**	0.14**	0.08**	0.11**
		(0.04)	(0.04)	(0.04)	(0.04)	(0.03)
Low Ability	0.14	-0.08	-0.05	-0.04	-0.05	
		(0.07)	(0.05)	(0.06)	(0.05)	
High Ability	0.51	0.18**	0.19**	0.19**	0.18**	
		(0.04)	(0.05)	(0.05)	(0.04)	

Table 15: The Effects of Education on Log Wages, by *Decision Node* (Total and Direct Effects)

	ATE_j †	(Dir)	ATE_j	(Dir)	TT_j	(Dir)
A. Graduating from HS vs. Dropping from HS	0.094*	0.036	0.094*	0.036	0.093	0.021
	(0.056)	(0.056)	(0.056)	(0.056)	(0.072)	(0.068)
C. College Enrollment vs. **HS Graduate**	0.126**	0.086**	0.134**	0.085**	0.140**	0.062
	(0.027)	(0.027)	(0.025)	(0.029)	(0.031)	(0.040)
D. **4-year college degree** vs. **Some College**	0.044		0.114**		0.141**	
	(0.044)		(0.037)		(0.042)	

Table 15: The Effects of Education on Log Wages, by *Decision Node* (Total and Direct Effects), Cont'd

	TUT_j	(Dir)	$AMTE_j$	(Dir)
A. Graduating from HS vs. Dropping from HS	0.100**	0.089**	0.093**	0.087**
	(0.029)	(0.031)	(0.028)	(0.032)
C. College Enrollment vs. **HS Graduate**	0.128**	0.109**	0.101**	0.077**
	(0.026)	(0.030)	(0.023)	(0.028)
D. **4-year college degree** vs. **Some College**	0.079**		0.110**	
	(0.036)		(0.034)	

Table 16: The Effects of Education on Log PV of wages, by *Decision Node*

	%	$ATE^{\dagger}_j$	ATE_j	TT_j	TUT_j	$AMTE_j$
A. Graduating from HS vs. Dropping from HS						
All		0.17**	0.17**	0.14**	0.29**	0.28**
		(0.06)	(0.06)	(0.07)	(0.04)	(0.04)
Low Ability	0.31	0.27**	0.27**	0.22**	0.35**	
		(0.04)	(0.04)	(0.05)	(0.05)	
High Ability	0.31	0.09	0.09	0.09	0.12	
		(0.11)	(0.11)	(0.11)	(0.09)	
B. **Getting a GED** vs. **HS Dropout**						
All		-0.20*	-0.11	-0.14*	-0.08	-0.14*
		(0.10)	(0.06)	(0.06)	(0.08)	(0.06)
Low Ability	0.61	-0.19*	-0.13*	-0.17*	-0.10	
		(0.07)	(0.08)	(0.08)	(0.08)	
High Ability	0.06	-0.21	-0.04	-0.06	0.01	
		(0.19)	(0.16)	(0.15)	(0.17)	

Table 16: The Effects of Education on Log PV of wages, by *Decision Node*, Cont'd

	%	$ATE_j^{\dagger}$	ATE_j	TT_j	TUT_j	$AMTE_j$
C. College Enrollment vs. **HS Graduate**						
All		0.14**	0.14**	0.14**	0.14**	0.11**
		(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Low Ability	0.22	0.09*	0.06	0.01	0.08*	
		(0.06)	(0.05)	(0.05)	(0.05)	
High Ability	0.38	0.21**	0.21**	0.21**	0.21**	
		(0.04)	(0.04)	(0.04)	(0.04)	
D. **4-year college degree** vs. **Some College**						
All		0.06	0.17**	0.22**	0.11*	0.15**
		(0.06)	(0.04)	(0.05)	(0.05)	(0.04)
Low Ability	0.14	-0.11	-0.01	0.04	-0.04	
		(0.10)	(0.07)	(0.07)	(0.08)	
High Ability	0.51	0.23**	0.26**	0.28**	0.21**	
		(0.05)	(0.05)	(0.05)	(0.05)	

THE UNIVERSITY OF CHICAGO

Table 17: The Effects of Education on Log PV of wages, by *Decision Node* (Total and Direct Effects)

	$ATE_j^{\dagger}$	(Dir)	ATE_j	(Dir)	TT_j	(Dir)
A. Graduating from HS vs. Dropping from HS	0.173**	0.114*	0.173**	0.114*	0.138**	0.067
	(0.059)	(0.057)	(0.059)	(0.057)	(0.071)	(0.072)
C. College Enrollment vs. **HS Graduate**	0.138**	0.077*	0.137**	0.059*	0.139**	0.015
	(0.033)	(0.029)	(0.029)	(0.031)	(0.031)	(0.039)
D. **4-year college degree** vs. **Some College**	0.056		0.171**		0.222**	
	(0.057)		(0.040)		(0.048)	

Table 17: The Effects of Education on Log PV of wages, by *Decision Node* (Total and Direct Effects), Cont'd

	TUT_j	(Dir)	AMTE_j	(Dir)
A. Graduating from HS vs. Dropping from HS	0.295**	0.279**	0.282**	0.269**
	(0.039)	(0.042)	(0.041)	(0.042)
C. College Enrollment vs. **HS Graduate**	0.136**	0.106**	0.112**	0.072**
	(0.031)	(0.033)	(0.031)	(0.031)
D. **4-year college degree** vs. **Some College**	0.106*		0.146**	
	(0.047)		(0.042)	

Table 18: The Effects of Education on Daily Smoking, by *Decision Node*

	%	$ATE_j^{\dagger}$	ATE_j	TT_j	TUT_j	$AMTE_j$
A. Dropping from HS vs. Graduating from HS						
All		-0.26**	-0.26**	-0.27**	-0.26**	-0.24**
		(0.06)	(0.06)	(0.07)	(0.04)	(0.03)
Low Ability	0.31	-0.29**	-0.29**	-0.30**	-0.29**	
		(0.04)	(0.04)	(0.05)	(0.04)	
High Ability	0.31	-0.25**	-0.25**	-0.25**	-0.15*	
		(0.11)	(0.11)	(0.11)	(0.07)	
B. **Getting a GED** vs. **HS Dropout**						
All		0.04	0.02	0.01	0.02	0.01
		(0.09)	(0.05)	(0.06)	(0.05)	(0.05)
Low Ability	0.61	-0.00	0.00	-0.00	0.01	
		(0.05)	(0.05)	(0.05)	(0.06)	
High Ability	0.06	0.08	0.07	0.07	0.07	
		(0.17)	(0.14)	(0.14)	(0.13)	

Table 18: The Effects of Education on Daily Smoking, by *Decision Node*, Cont'd

	%	$ATE_j^{\dagger}$	ATE_j	TT_j	TUT_j	$AMTE_j$
C. College Enrollment vs. **HS Graduate**						
All		-0.12**	-0.14**	-0.18**	-0.10**	-0.13**
		(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Low Ability	0.22	-0.06	-0.09*	-0.12**	-0.07	
		(0.06)	(0.05)	(0.05)	(0.05)	
High Ability	0.38	-0.18**	-0.19**	-0.21**	-0.13**	
		(0.04)	(0.04)	(0.04)	(0.03)	
D. **4-year college degree** vs. **Some College**						
All		-0.16**	-0.17**	-0.18**	-0.17**	-0.17**
		(0.04)	(0.04)	(0.05)	(0.04)	(0.04)
Low Ability	0.14	-0.12	-0.12*	-0.12	-0.11*	
		(0.08)	(0.07)	(0.08)	(0.07)	
High Ability	0.51	-0.20**	-0.19**	-0.19**	-0.20**	
		(0.05)	(0.05)	(0.05)	(0.04)	

Table 19: The Effects of Education on Daily Smoking, by *Decision Node* (Total and Direct Effects)

	$ATE_j^{\dagger}$	(Dir)	ATE_j	(Dir)	TT_j	(Dir)
A. Graduating from HS vs. Dropping from HS	-0.263**	-0.189**	-0.263**	-0.189**	-0.265**	-0.174**
	(0.056)	(0.058)	(0.056)	(0.058)	(0.071)	(0.069)
C. College Enrollment vs. **HS Graduate**	-0.115**	-0.054*	-0.139**	-0.065*	-0.178**	-0.080*
	(0.031)	(0.031)	(0.028)	(0.033)	(0.033)	(0.045)
D. **4-year college degree** vs. **Some College**	-0.164**		-0.172**		-0.176**	
	(0.044)		(0.043)		(0.051)	

Table 19: The Effects of Education on Daily Smoking, by *Decision Node* (Total and Direct Effects), Cont'd

	TUT_j	(Dir)	$AMTE_j$	(Dir)
A. Graduating from HS vs. Dropping from HS	-0.255**	-0.240**	-0.242**	-0.234**
	(0.036)	(0.038)	(0.033)	(0.038)
C. College Enrollment vs. **HS Graduate**	-0.096**	-0.049	-0.131**	-0.065**
	(0.032)	(0.037)	(0.027)	(0.033)
D. **4-year college degree** vs. **Some College**	-0.167**		-0.173**	
	(0.038)		(0.038)	

Table 20: The Effects of Education on Health Limits Work, by *Decision Node*

	%	$ATE_j^{\dagger}$	ATE_j	TT_j	TUT_j	$AMTE_j$
A. Graduating from HS vs. Dropping from HS						
All		-0.11**	-0.11**	-0.11**	-0.09**	-0.11**
		(0.04)	(0.04)	(0.05)	(0.03)	(0.03)
Low Ability	0.31	-0.08**	-0.08**	-0.09**	-0.07	
		(0.04)	(0.04)	(0.04)	(0.04)	
High Ability	0.31	-0.11*	-0.11*	-0.11*	-0.12**	
		(0.07)	(0.07)	(0.07)	(0.06)	
B. **Getting a GED** vs. **HS Dropout**						
All		-0.03	0.06	0.05	0.06	0.06
		(0.08)	(0.05)	(0.06)	(0.06)	(0.05)
Low Ability	0.61	0.06	0.09*	0.09	0.10	
		(0.06)	(0.05)	(0.05)	(0.06)	
High Ability	0.06	-0.11	-0.05	-0.06	-0.04	
		(0.15)	(0.12)	(0.12)	(0.13)	

Table 20: The Effects of Education on Health Limits Work, by *Decision Node*, Cont'd

	%	$ATE_j^{\dagger}$	ATE_j	TT_j	TUT_j	$AMTE_j$
C. College Enrollment vs. **HS Graduate**						
All		-0.05	-0.04*	-0.02	-0.05*	-0.03
		(0.03)	(0.02)	(0.02)	(0.03)	(0.02)
Low Ability	0.22	-0.08	-0.07	-0.04	-0.08	
		(0.05)	(0.04)	(0.03)	(0.04)	
High Ability	0.38	-0.02	-0.02	-0.02	-0.02	
		(0.02)	(0.02)	(0.02)	(0.03)	
D. **4-year college degree** vs. **Some College**						
All		-0.05	-0.06*	-0.07*	-0.06*	-0.07*
		(0.05)	(0.03)	(0.03)	(0.04)	(0.03)
Low Ability	0.14	-0.01	-0.02	-0.02	-0.01	
		(0.09)	(0.06)	(0.06)	(0.06)	
High Ability	0.51	-0.09**	-0.08*	-0.08*	-0.09**	
		(0.04)	(0.04)	(0.04)	(0.04)	

Table 21: The Effects of Education on Health Limits Work, by *Decision Node* (Total and Direct Effects)

	$ATE_j^{\dagger}$	(Dir)	ATE_j	(Dir)	TT_j	(Dir)
A. Graduating from HS vs. Dropping from HS	-0.108**	-0.097**	-0.108**	-0.097**	-0.113**	-0.101**
	(0.042)	(0.045)	(0.042)	(0.045)	(0.050)	(0.054)
C. College Enrollment vs. **HS Graduate**	-0.046	-0.023	-0.037*	-0.009	-0.023	0.016
	(0.025)	(0.028)	(0.022)	(0.024)	(0.019)	(0.029)
D. **4-year college degree** vs. **Some College**	-0.051		-0.064*		-0.070*	
	(0.048)		(0.031)		(0.034)	

Table 21: The Effects of Education on Health Limits Work, by *Decision Node* (Total and Direct Effects), Cont'd

	TUT_j	(Dir)	$AMTE_j$	(Dir)
A. Graduating from HS vs. Dropping from HS	-0.090**	-0.084**	-0.110**	-0.104**
	(0.034)	(0.036)	(0.031)	(0.033)
C. College Enrollment vs. **HS Graduate**	-0.053*	-0.035	-0.029	-0.005
	(0.027)	(0.032)	(0.022)	(0.025)
D. **4-year college degree** vs. **Some College**	-0.057*		-0.067*	
	(0.036)		(0.030)	

Return to main text

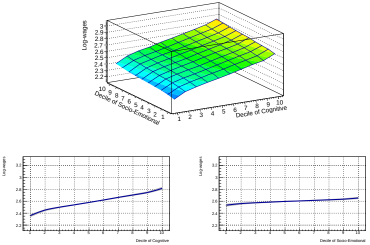
Web Appendix A.7
The Measurement of Endowments and Their Effects on Outcomes

The Role of Endowments on Later Life Outcomes

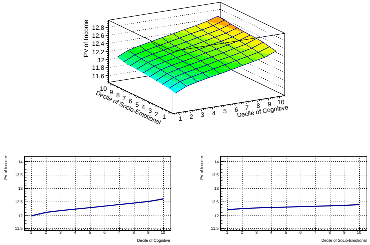
- The latent endowments have statistically significant effects on labor market and health outcomes.
- Figure 14 plots the effects of the latent endowments on (log) wages, log present value of log wage income, daily, work limitations, and daily smoking The cognitive endowment affects all four outcomes, while the effect of the socioemotional endowment is statistically significant only in the equations for wages and smoking.
- Moving someone from the lowest decile to the highest decile in both cognitive and socioemotional ability, increases their wages by 0.6 log points, lowers the probability of being a smoker by 60%, increases their self-esteem by one standard deviation and increases their health by half a standard deviation.

Figure 14: The Effect of Cognitive and Socioemotional Endowments

A. (log)Wages



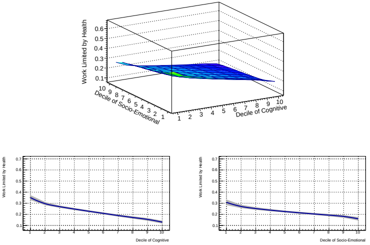
B. PV Wages



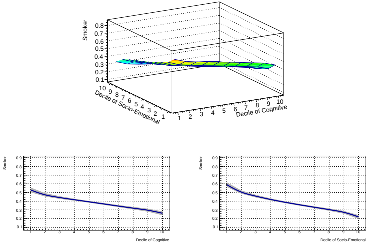
Notes: For each of the four outcomes, we present three figures that study the impact of cognitive and socioemotional endowments. The top figure in each panel displays the levels of the outcome as a function of cognitive and socioemotional endowments. In particular, we present the average level of outcomes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

Figure 14: The Effect of Cognitive and Socioemotional Endowments, Cont'd

C. Health Limits Work



D. Smoking



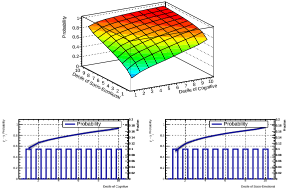
Notes: For each of the four outcomes, we present three figures that study the impact of cognitive and socioemotional endowments. The top figure in each panel displays the levels of the outcome as a function of cognitive and socioemotional endowments. In particular, we present the average level of outcomes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

The Role of Endowments on Later Life Outcomes

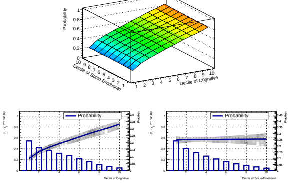
- Figure 15 presents the probabilities of making the indicated educational choice at various levels of agent latent endowments.
- Figure 17 shows the distribution of the factors by final schooling level.
- Individuals sort on both cognitive and socioemotional endowments into increasing schooling levels.
- The only exception are the GEDs, who have cognitive ability distributions similar to terminal high school graduates but socioemotional distributions similar to dropouts.

Figure 15: The Probability of Educational Decisions, by Endowment Levels (Final Schooling Levels are Highlighted Using Bold Letters)

A. Dropping from HS vs. Graduating from HS



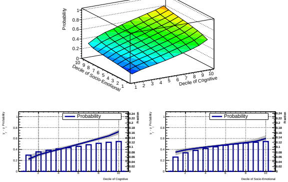
B. **HS Dropout** vs. **Getting a GED**



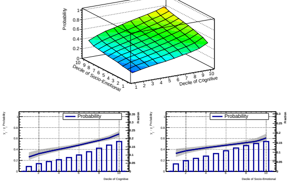
Notes: For each of the four educational choices, we present three figures that study the probability of that specific educational choice. Final schooling levels do not allow for further options. For each pair of schooling levels j and $j+1$, the first subfigure (top) presents $Prob(D_j = 0|d^C, d^{SE})$ where d^C and d^{SE} denote the cognitive and socioemotional deciles computed from the marginal distributions of cognitive and socioemotional endowments. $Prob(D_j = 0|d^C, d^{SE})$ is computed for those who reach the decision node involving a decision between levels j and $j+1$. The bottom left subfigures present $Prob(D_j = 0|d^C)$ where the socioemotional factor is integrated out. The bars in these figures display, for a given decile of cognitive endowment, the fraction of individuals visiting the node leading to the educational decision involving levels j and $j+1$. The bottom right subfigures present $Prob(D_j = 0|d^{SE})$ for a given decile of socioemotional endowment, as well as the fraction of individuals visiting the node leading to the educational decision involving levels j and $j+1$.

Figure 15: The Probability of Educational Decisions, by Endowment Levels (Final Schooling Levels are Highlighted Using Bold Letters), Cont'd

C. **HS Graduate** vs. College Enrollment

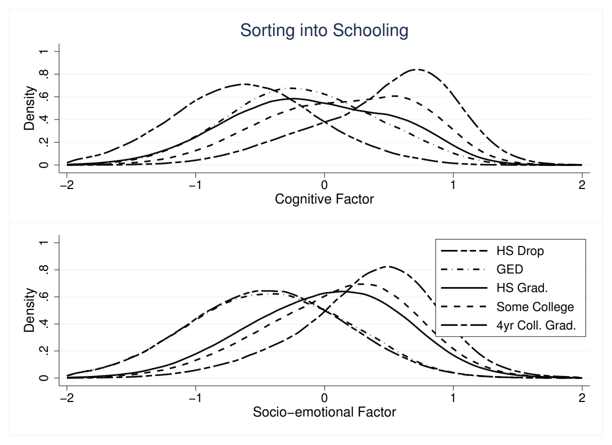


D. **Some College** vs. **4-year college degre**



Notes: For each of the four educational choices, we present three figures that study the probability of that specific educational choice. Final schooling levels do not allow for further options. For each pair of schooling levels j and $j+1$, the first subfigure (top) presents $Prob(D_j = 0 | d^C, d^{SE})$ where d^C and d^{SE} denote the cognitive and socioemotional deciles computed from the marginal distributions of cognitive and socioemotional endowments. $Prob(D_j = 0 | d^C, d^{SE})$ is computed for those who reach the decision node involving a decision between levels j and $j+1$. The bottom left subfigures present $Prob(D_j = 0 | d^C)$ where the socioemotional factor is integrated out. The bars in these figures display, for a given decile of cognitive endowment, the fraction of individuals visiting the node leading to the educational decision involving levels j and $j+1$. The bottom right subfigures present $Prob(D_j = 0 | d^{SE})$ for a given decile of socioemotional endowment, as well as the fraction of individuals visiting the node leading to the educational decision involving levels j and $j+1$.

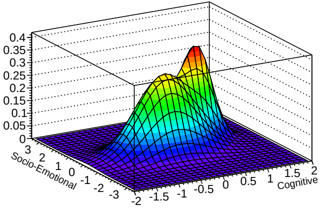
Figure 16: Distribution of Factors by Schooling Level



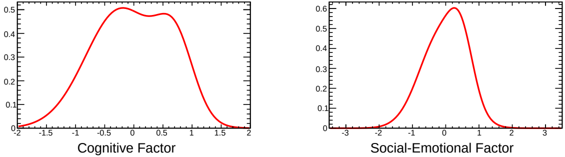
Note: The factors are simulated from the estimates of the model. The simulated data contain 1 million observations.

- The estimates reveal clear evidence of sorting into education by both cognitive and socioemotional endowments.
- At the same time, these endowments have significant impacts on adult outcomes.
- Together these results imply strong selection biases in observed differences in outcomes by education level.
- This highlights the importance of accounting for observed and latent traits when estimating the causal impact of education.

Figure 17: Joint Distribution of Cognitive and Socioemotional Ability



Distribution of Factors



$$\blacksquare_1 = \begin{pmatrix} 0.0971459 & 0 \\ 0 & 0.128441 \end{pmatrix}, \quad \blacksquare_2 = \begin{pmatrix} 0.367792 & 0 \\ 0 & 0.421223 \end{pmatrix},$$

Table 22: Means and Weights for Mixtures

	Mixture for Cognitive θ	Mixture for Non-cognitive θ
μ_1	0.721206	-0.218251
μ_2	0.487487	-0.147523
Weight on First Component	0.232316	0.767684

Sorting on Observables

- Table 23 the means and standard deviations of our controls by education level.
- There is sorting in nearly every background characteristic, except for age in 1980.
- The sorting is very strong in the cognitive factor, the socioemotional factor, parental education, parental income, number of siblings and growing up in a broken home.

THE UNIVERSITY OF CHICAGO

Table 23: Educational Sorting on Observables

	Dropout	GED	High School	Some Coll.	Coll. Grad.
Cog	-0.651	-0.170	-0.119	0.162	0.522
	(0.535)	(0.651)	(0.667)	(0.641)	(0.533)
Socioemotional	-0.677	-0.840	0.074	0.157	0.440
	(0.742)	(0.894)	(0.756)	(0.737)	(0.664)
Black	0.181	0.179	0.119	0.104	0.060
	(0.386)	(0.384)	(0.324)	(0.305)	(0.238)
Hisp	0.113	0.074	0.064	0.080	0.031
	(0.317)	(0.263)	(0.245)	(0.271)	(0.174)
Broken Home	0.423	0.358	0.211	0.229	0.142
	(0.495)	(0.480)	(0.408)	(0.421)	(0.349)
Num. Siblings	4.181	3.777	3.379	2.923	2.538
	(2.640)	(2.585)	(2.177)	(2.122)	(1.807)
Mom's HGC	9.966	10.670	11.250	11.928	13.281
	(2.518)	(2.382)	(2.252)	(2.419)	(2.496)
Dad's HGC	9.768	10.577	11.115	12.312	14.290
	(3.099)	(2.997)	(2.872)	(3.251)	(3.397)
Fam. Inc 1979	13.675	16.378	19.680	20.922	27.075
	(8.033)	(9.696)	(10.514)	(12.097)	(16.152)
Urban age 14	0.771	0.790	0.715	0.755	0.804
	(0.421)	(0.408)	(0.452)	(0.430)	(0.397)
South age 14	0.420	0.406	0.277	0.298	0.243
	(0.494)	(0.492)	(0.448)	(0.458)	(0.429)
Age in 1980	19.096	18.808	19.230	19.226	19.206
	(2.103)	(2.129)	(2.173)	(2.232)	(2.227)

mean coefficients; sd in parentheses
* $p < 0.10$, ** $p < 0.05$, ** $p < 0.01$

THE UNIVERSITY OF CHICAGO

- There may be some concern that the socioemotional factor is describing something like academic ability, since it is partly based on grades.
- Table 24 provides additional support for our interpretation of the socioemotional factor.
- The table estimates the impact of observables and unobservables on early risky behavior, but is not used in estimating our factors.
- As we see, our non-cognitive factor plays an important role in each of the early risky outcomes.
- If the socioemotional factor were measuring purely academic behavior, we would not expect it to be so predictive in explaining early risky behaviors.

Table 24: Early Outcomes: Estimates for "Early Risky Behaviors"

Variable	Tried Marijuana[a]		Daily Smoking[a]		Regular Drinking[a]		Intercourse[a]	
	β	Std Err.	β	Std Err.	β	Std Err.	β	Std Err.
Black	-0.321	0.101	-0.341	0.112	-0.237	0.108	0.605	0.099
Hispanic	-0.160	0.125	-0.496	0.150	-0.010	0.130	-0.034	0.140
Broken Home	0.421	0.073	0.417	0.081	0.236	0.077	0.366	0.081
Number of Siblings	0.030	0.014	0.033	0.015	0.028	0.015	0.011	0.016
Mother's Education	0.011	0.015	-0.021	0.017	0.001	0.016	-0.022	0.017
Father's Education	-0.011	0.011	-0.036	0.013	-0.004	0.012	-0.027	0.013
Family Income	0.001	0.003	-0.003	0.003	-0.001	0.003	-0.003	0.003
Intercept	-0.165	2.473	-4.411	2.829	1.067	2.620	3.064	2.873
Age	0.022	0.257	0.384	0.293	-0.203	0.273	-0.406	0.298
Age^2	-0.003	0.007	-0.009	0.008	0.005	0.007	0.010	0.008
Urban	0.271	0.072	0.113	0.081	0.096	0.077	0.211	0.087
South	-0.110	0.067	-0.025	0.075	0.066	0.071	0.103	0.076
Cognitive	-0.102	0.048	-0.209	0.054	-0.137	0.052	-0.277	0.057
Socio-emotional	-0.616	0.060	-0.527	0.064	-0.288	0.061	-0.403	0.066
N	2239		2176		2231		2218	

The numbers in this table represent the estimated coefficients and standard errors associated with binary choice models of early risky behaviors on the set of controls presented in rows. Information about living in the West and Northeast is only available in 1979. [a] The dependent variable takes a value of one if the individual has reported the behavior before age 15, and zero otherwise.

Return to main text

Web Appendix A.8
Evidence in the Text: Linearity of The Returns to Schooling

THE UNIVERSITY OF CHICAGO

- In this section, we use OLS to test the assumption of linearity in schooling for our four outcomes.
- We find significant sheepskin effects in all outcomes and specifications rejecting the linear returns to schooling assumption.
- Specifically, we run Mincer regressions, then add our dummies for schooling levels and conduct an F-test of the null hypothesis that the coefficients are jointly equal to zero.
- In figure 18, we display the ATE over being a dropout for each schooling level spaced by the difference in the years of schooling.

Table 25: Years of Schooling Regression: Log Wage

	Log Wage					
Highest Grade Comp.	0.083 ***	0.037 ***	0.073 ***	0.035 ***	0.040 ***	0.012
	(0.004)	(0.011)	(0.005)	(0.011)	(0.006)	(0.011)
HS Grad		0.149 ***		0.145 ***		0.081 *
		(0.041)		(0.043)		(0.042)
GED		0.044		0.011		0.038
		(0.049)		(0.051)		(0.048)
Enroll Coll		0.071 **		0.054		0.059 *
		(0.034)		(0.034)		(0.033)
Grad. Coll		0.156 ***		0.142 ***		0.108 **
		(0.044)		(0.045)		(0.043)
Includes Factors			X	X	X	X
Includes Controls					X	X
JointTest	.	0.000	.	0.000	.	0.027

JointTest provides the p-value from an F-test for if education dummies are jointly equal to zero
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

THE UNIVERSITY OF CHICAGO

Table 26: Years of Schooling Regression: PV-Wage

	PV-Wage					
Highest Grade Comp.	0.122 ***	0.077 ***	0.103 ***	0.076 ***	0.042 ***	0.036 ***
	(0.005)	(0.014)	(0.007)	(0.014)	(0.007)	(0.014)
HS Grad		0.303 ***		0.260 ***		0.133 **
		(0.052)		(0.055)		(0.052)
GED		-0.019		-0.109 *		-0.070
		(0.061)		(0.063)		(0.058)
Enroll Coll		0.017		-0.002		-0.001
		(0.044)		(0.044)		(0.041)
Grad. Coll		0.104 *		0.090		0.024
		(0.058)		(0.058)		(0.054)
Includes Factors			X	X	X	X
Includes Controls					X	X
JointTest	.	0.000	.	0.000	.	0.000

JointTest provides the p-value from an F-test for if education dummies are jointly equal to zero
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 27: Years of Schooling Regression: Health Limits Work

	Health Limits Work					
Highest Grade Comp.	-0.038 ***	-0.021 **	-0.025 ***	-0.016	-0.014 **	-0.010
	(0.004)	(0.010)	(0.005)	(0.010)	(0.006)	(0.011)
HS Grad		-0.109 ***		-0.068 *		-0.052
		(0.036)		(0.039)		(0.040)
GED		0.032		0.069		0.071
		(0.043)		(0.045)		(0.045)
Enroll Coll		-0.016		-0.012		-0.008
		(0.031)		(0.031)		(0.032)
Grad. Coll		-0.027		-0.026		-0.018
		(0.041)		(0.041)		(0.042)
Includes Factors			X	X	X	X
Includes Controls					X	X
JointTest	.	0.000	.	0.001	.	0.007

JointTest provides the p-value from an F-test for if education dummies are jointly equal to zero
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

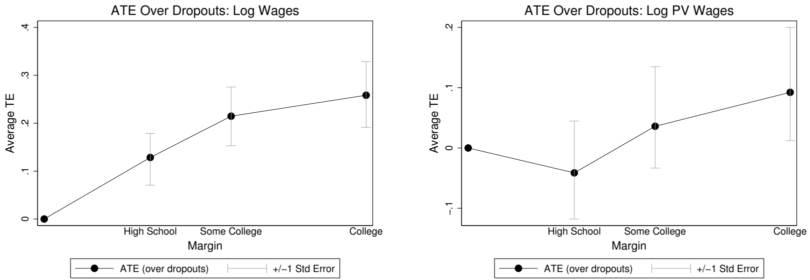
Table 28: Years of Schooling Regression: Smoking

	Smoking					
Highest Grade Comp.	-0.065 ***	-0.039 ***	-0.047 ***	-0.032 **	-0.049 ***	-0.037 ***
	(0.004)	(0.012)	(0.006)	(0.013)	(0.007)	(0.013)
HS Grad		-0.194 ***		-0.151 ***		-0.150 ***
		(0.045)		(0.048)		(0.049)
GED		0.040		0.042		0.045
		(0.053)		(0.055)		(0.055)
Enroll Coll		0.019		0.022		0.020
		(0.038)		(0.038)		(0.038)
Grad. Coll		-0.076		-0.071		-0.081
		(0.049)		(0.051)		(0.051)
Includes Factors			X	X	X	X
Includes Controls					X	X
JointTest	.	0.000	.	0.000	.	0.000

JointTest provides the p-value from an F-test for if education dummies are jointly equal to zero
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

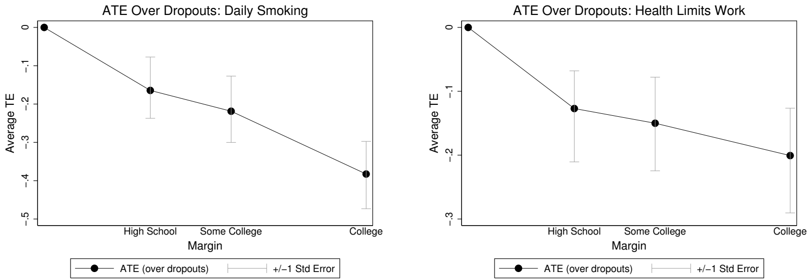
THE UNIVERSITY OF CHICAGO

Figure 18: Plotting Average Treatment Effects by average years of schooling



Notes: Figure displays the estimated pair-wise ATE for the full population ($Y_j^k - Y_{j-1}^k$). The ATEs are spaced out according to the average difference in highest grade completed between each educational group.

Figure 18: Plotting Average Treatment Effects by average years of schooling, Cont'd



Notes: Figure displays the estimated pair-wise ATE for the full population ($Y_j^k - Y_{j-1}^k$). The ATEs are spaced out according to the average difference in highest grade completed between each educational group.

Testing the Linearity of the Average Treatment Effect

- This section tests if the annualized average treatment effect are linear across schooling decisions.
- Results are provided in Table 29.
- Results are reported for both the full-population ATE and the ATE restricted to those who reach the decision node.
- We can reject linearity for wages for the conditional ATE at the 0.05 level.
- We can also reject linearity for smoking in the conditional population and linearity for wages in the full population at the 0.10 level, but not the 0.05 level.

Table 29: Testing Linearity of the ATE

	ATE ($Q_j = 1$)	ATE (full pop)
Wages	0.011	0.090
PV Wages	0.306	0.367
Smoking	0.055	0.653
Health Limits Work	0.353	0.113

Notes: This table reports *p*-values from a Wald Test for the null hypothesis that the average returns to a year of schooling are linear across schooling decisions. The first column reports the test for the ATE conditional on being at the decision node, while the second column reports the test for Specifically we test if $\frac{ATE_1}{\bar{q}_1 - \bar{q}_0} - \frac{ATE_2}{\bar{q}_2 - \bar{q}_1} = 0$ and $\frac{ATE_2}{\bar{q}_2 - \bar{q}_1} - \frac{ATE_3}{\bar{q}_3 - \bar{q}_2} = 0$ where the covariance between $\frac{ATE_1}{\bar{q}_1 - \bar{q}_0} - \frac{ATE_2}{\bar{q}_2 - \bar{q}_1}$ and $\frac{ATE_2}{\bar{q}_2 - \bar{q}_1} - \frac{ATE_3}{\bar{q}_3 - \bar{q}_2}$ is estimated using 200 bootstrapped samples. $\bar{q}_j$ is the average years of completed schooling for those at schooling level j.

Return to main text

Web Appendix A.13
Decomposing the Correlation Between ρ and S: Are Those Who Go to School the Ones Who Benefit from It?

- The correlation between ρ_i and S_i is a measure of the quality of sorting of people into schooling by their gain from it—a topic Becker investigated in depth in his Woytinsky lecture (1967,1991).
- We have already established that the distributions of returns differ across schooling levels and the returns across schooling levels are far from perfectly correlated.
- It is thus of interest to push our analysis a bit further and investigate the correlation of annualized returns with attained schooling levels.
- We consider this question for direct returns and for total returns inclusive of continuation values.
- Table 30 shows the correlations between educational choices and the node-specific annualized (direct terminal) gains $\frac{(Y_j - Y_{j-1})}{(q_j - q_{j-1})}$ as well as the overall correlation.

- The correlations between the ρ_j and S are shown in column 1.
- Columns 2 through 4 show the correlations between the individual treatment effects ρ_j and choices at node D_j.
- For columns 2 through 4, each correlation is estimated conditional on the population that makes it to the specific decision ($Q_j = 1$).

Table 30: Correlation Between Annualized Returns and Educational Choices

	$Corr(\rho, S)$	$Corr(\rho_1, (1-D_1))$ Grad. HS	$Corr(\rho_2, 1-D_2))$ Enroll in College	$Corr(\rho_3, (1-D_3))$ Grad. College
Wage	0.069	0.011	-0.041	0.053
	(0.011)	(0.061)	(0.002)	(0.018)
PVwage	-0.080	-0.193	-0.068	0.084
	(0.037)	(0.034)	(0.023)	(0.021)
Smoking	-0.082	0.202	-0.110	-0.034
	(0.069)	(0.171)	(0.030)	(0.063)
Health Limits Work	-0.102	-0.225	0.227	-0.065
	(0.064)	(0.029)	(0.022)	(0.056)

Notes: Let q_j be the years of schooling associated with node j. The annualized terminal node j return is $\rho_j := \frac{Y_j - Y_{j-1}}{q_j - q_{j-1}}$ and we define $\rho = \frac{Y_j - Y_0}{q_j - q_0}$. Total years of schooling is $S = \sum_{j=1}^{\bar{s}} q_j D_j$. Note $D_1 = 1$ if individuals stop their education as a high school graduate. $D_2 = 1$ and $D_3 = 1$ denote stopping at some college and college respectively. Standard errors are estimated using 200 bootstrap samples and show the standard deviation of the estimate across the samples.

- The overall correlation and the correlation by node differ substantially.
- The general pattern is that for wages, people sort on terminal gains although the effect is only strong for graduating college, and for most outcomes it is perverse for some college.
- The sorting is *negative* for PV wages, except for college graduation.
- For smoking, the overall effect is negative, but is positive for high school graduation.
- For health limits work, the correlations differ but are negative except for the anomalous correlation for some college.

- Table 31 decomposes the correlation between S and ρ in a fashion similar to what is reported in Table 30, except we work with dynamic treatment effects (T_j) inclusive of continuation values.
- This better represents the gains that agents use to make choices rather than the benefit associated with the comparison between terminal outcomes at j and $j-1$.
- The patterns are roughly similar across the two tables.
- The correlations are consistently negative for smoking across all transitions.
- The strongest negative correlation for health limits work is for high school graduation.
- The correlation with Some College is anomalous.
- Using either terminal level treatment effects or dynamic treatment effects sorting is generally positive, broadly consistent with a meritocratic society.

Table 31: Correlation Between Returns and Educational Choice Including Continuation Values

	$Corr(\rho, S)$	$Corr(T_1, (1-D_1))$ Grad. HS	$Corr(T_2, 1-D_2))$ Enroll in College	$Corr(T_3, (1-D_3))$ Grad. College
Wage	0.069	-0.007	0.011	0.053
	(0.030)	(0.062)	(0.030)	(0.033)
PVwage	-0.080	-0.102	0.002	0.084
	(0.030)	(0.052)	(0.027)	(0.032)
Smoking	-0.082	-0.030	-0.298	-0.034
	(0.089)	(0.130)	(0.084)	(0.101)
Health Limits Work	-0.102	-0.089	0.162	-0.065
	(0.079)	(0.132)	(0.111)	(0.110)

Notes: Let q_j be the years of schooling associated with node j. The annualized terminal node j return is $\rho_j := \frac{Y_j - Y_{j-1}}{q_j - q_{j-1}}$ and we define $\rho = \sum_{j=1}^{\bar{s}} \rho_j(1 - D_j)$. Total years of schooling is $S = \sum_{j-1}^{\bar{s}} q_j(1 - D_j)$. $1 - D_1 = 1$ stopping at high school, with $1 - D_2$ and $1 - D_3$ denoting stopping at some college and college, respectively.

THE UNIVERSITY OF CHICAGO

Return to main text

Web Appendix A.15
Decomposing Observed Differences Into Average Treatment Effects, Sorting Gains, and Selection Bias

Web Appendix A.15.1
Decompositions by Final Schooling Level

- We first decompose raw differences by final schooling level.
- We then decompose effects defined on $Q_j = 1$ that include continuation values.
- Equation (18) can be written explicitly in terms of $\boldsymbol{X}$ and θ as follows:

$$\underbrace{E[\tau^k_{j+1}(\boldsymbol{X}) - \tau^k_j(\boldsymbol{X})|S \in \{j, j+1\}] + E[\boldsymbol{\theta}'(\boldsymbol{\alpha}^k_{j+1} - \boldsymbol{\alpha}^k_j)|S \in \{j, j+1\}]}_{\text{ATE}}$$
$$+ \underbrace{\Big(E[\tau^k_{j+1}(\boldsymbol{X}) - \tau^k_j(\boldsymbol{X})|S = j+1] + E[\boldsymbol{\theta}'(\boldsymbol{\alpha}^k_{j+1} - \boldsymbol{\alpha}^k_j)|S = j+1] - E[\tau^k_{j+1}(\boldsymbol{X}) - \tau^k_j(\boldsymbol{X})|S \in \{j, j+1\}] - E[\boldsymbol{\theta}'(\boldsymbol{\alpha}^k_{j+1} - \boldsymbol{\alpha}^k_j)|S \in \{j, j+1\}]\Big)}_{\text{Sorting Gains}}$$
$$+ \underbrace{\Big(E[\tau^k_j(\boldsymbol{X})|S = j+1] + E[\boldsymbol{\theta}'\boldsymbol{\alpha}^k_j|S = j+1] - [E[\tau^k_j(\boldsymbol{X})|S = j] + E[\boldsymbol{\theta}'\boldsymbol{\alpha}^k_j|S = j]]\Big)}_{\text{Selection Bias}}. \quad (25)$$

Appendix A.15.2
Decompositions: The Pairwise Observed Differences by Final Schooling Level

Table 32: Decomposition the Observed Difference in Log Wages (pairwise comparison)

	Observed	Average Treatment Effects	Sorting on Gains	Selection Bias
HS-DO	0.25	0.12	0.00	0.13
	(0.03)	(0.04)	(0.01)	(0.05)
SC-HS	0.14	0.10	-0.03	0.07
	(0.03)	(0.03)	(0.02)	(0.02)
Coll-SC	0.23	0.11	0.03	0.09
	(0.03)	(0.04)	(0.02)	(0.03)
GED-DO	0.15	0.06	0.00	0.09
	(0.04)	(0.05)	(0.03)	(0.04)

Notes: All numbers are from simulations of our model. Average Treatment Effects, Sorting on Gains and Selection Bias sum to the "Observed" column for each row. Each decomposition decomposes the observed difference in outcomes between people with final schooling levels j or $j+1$ into the various components above.

Table 33: Decomposition the Observed Difference in PV Log Wages (pairwise comparison)

	Observed	Average Treatment Effects	Sorting on Gains	Selection Bias
HS-DO	0.50	0.07	-0.08	0.51
	(0.04)	(0.06)	(0.02)	(0.08)
SC-HS	0.15	0.09	-0.03	0.09
	(0.03)	(0.03)	(0.02)	(0.02)
Coll-SC	0.31	0.17	0.05	0.09
	(0.04)	(0.04)	(0.02)	(0.04)
GED-DO	0.20	-0.11	-0.04	0.34
	(0.06)	(0.06)	(0.04)	(0.06)

Notes: All numbers are from simulations of our model. Average Treatment Effects, Sorting on Gains and Selection Bias sum to the “Observed” column for each row. Each decomposition decomposes the observed difference in outcomes between people with final schooling levels j or $j + 1$ into the various components above.

Table 34: Decomposition the Observed Difference in Smoking (pairwise comparison)

	Observed	Average Treatment Effects	Sorting on Gains	Selection Bias
HS-DO	-0.27	-0.20	0.02	-0.09
	(0.04)	(0.07)	(0.02)	(0.08)
SC-HS	-0.06	-0.06	-0.01	0.01
	(0.03)	(0.03)	(0.02)	(0.02)
Coll-SC	-0.19	-0.17	-0.00	-0.02
	(0.03)	(0.04)	(0.02)	(0.04)
GED-DO	-0.03	0.02	-0.01	-0.04
	(0.05)	(0.05)	(0.03)	(0.05)

Notes: All numbers are from simulations of our model. Average Treatment Effects, Sorting on Gains and Selection Bias sum to the “Observed” column for each row. Each decomposition decomposes the observed difference in outcomes between people with final schooling levels j or $j+1$ into the various components above.

Table 35: Decomposition the Observed Difference in Health Limits Work
(pairwise comparison)

	Observed	Average Treatment Effects	Sorting on Gains	Selection Bias
HS-DO	-0.17	-0.11	-0.02	-0.04
	(0.03)	(0.05)	(0.02)	(0.07)
SC-HS	-0.04	-0.03	0.02	-0.04
	(0.03)	(0.03)	(0.02)	(0.01)
Coll-SC	-0.09	-0.06	-0.01	-0.02
	(0.03)	(0.03)	(0.01)	(0.02)
GED-DO	-0.02	0.06	-0.01	-0.08
	(0.04)	(0.05)	(0.03)	(0.04)

Notes: All numbers are from simulations of our model. Average Treatment Effects, Sorting on Gains and Selection Bias sum to the “Observed” column for each row. Each decomposition decomposes the observed difference in outcomes between people with final schooling levels j or $j+1$ into the various components above.

THE UNIVERSITY OF CHICAGO

Decomposing the components into observed characteristics and latent ability (wage outcomes only)

Table 36: Decomposition the Observed Difference in Log Wages (decomposed pairwise comparison)

		Average Treatment Effects			Sorting on Gains			Selection Bias		
	Observed	Total	Obs ($\boldsymbol{X}$)	Abil ($\boldsymbol{\theta}$)	Total	Obs ($\boldsymbol{X}$)	Abil ($\boldsymbol{\theta}$)	Total	Obs ($\boldsymbol{X}$)	Abil ($\boldsymbol{\theta}$)
HS-DO	0.25	0.12	0.13	-0.01	0.00	-0.01	0.01	0.13	0.10	0.02
	(0.03)	(0.04)	(0.05)	(0.01)	(0.01)	(0.01)	(0.01)	(0.05)	(0.03)	(0.04)
SC-HS	0.14	0.10	0.09	0.01	-0.03	-0.01	-0.01	0.07	0.04	0.03
	(0.04)	(0.05)	(0.07)	(0.06)	(0.03)	(0.02)	(0.02)	(0.04)	(0.02)	(0.02)
Coll-SC	0.23	0.11	0.04	0.07	0.03	-0.00	0.03	0.09	0.07	0.02
	(0.03)	(0.03)	(0.03)	(0.01)	(0.02)	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)
GED-DO	0.15	0.06	0.14	-0.08	0.00	-0.01	0.01	0.09	0.05	0.04
	(0.03)	(0.04)	(0.04)	(0.02)	(0.02)	(0.01)	(0.01)	(0.03)	(0.02)	(0.02)

Notes: All numbers are from simulations of our model. The Total column of Average Treatment Effects, Sorting on Gains and Selection Bias sum to the "Observed" column for each row. The "Obs" ($\boldsymbol{X}$) and "Abil" ($\boldsymbol{\theta}$) columns decompose their respective totals into the part coming from observable characteristics and the part coming from the unobserved abilities. Each decomposition decomposes the observed difference in outcomes between people with final schooling levels j or $j+1$ into the various components above.

Table 37: Decomposition the Observed Difference in PV Log Wages (decomposed pairwise comparison)

		Average Treatment Effects			Sorting on Gains			Selection Bias		
	Observed	Total	Obs ($\boldsymbol{X}$)	Abil ($\boldsymbol{\theta}$)	Total	Obs ($\boldsymbol{X}$)	Abil ($\boldsymbol{\theta}$)	Total	Obs ($\boldsymbol{X}$)	Abil ($\boldsymbol{\theta}$)
HS-DO	0.50	0.07	-0.00	0.07	-0.08	-0.03	-0.04	0.51	0.29	0.22
	(0.04)	(0.06)	(0.08)	(0.02)	(0.02)	(0.01)	(0.02)	(0.08)	(0.05)	(0.06)
SC-HS	0.15	0.09	0.08	0.01	-0.03	-0.02	-0.01	0.09	0.06	0.03
	(0.06)	(0.06)	(0.09)	(0.07)	(0.04)	(0.02)	(0.03)	(0.06)	(0.05)	(0.04)
Coll-SC	0.31	0.17	0.09	0.09	0.05	0.01	0.04	0.09	0.07	0.02
	(0.03)	(0.03)	(0.03)	(0.01)	(0.02)	(0.01)	(0.01)	(0.02)	(0.02)	(0.01)
GED-DO	0.20	-0.11	-0.12	0.01	-0.04	-0.03	-0.00	0.34	0.15	0.19
	(0.04)	(0.04)	(0.05)	(0.03)	(0.02)	(0.01)	(0.01)	(0.04)	(0.03)	(0.02)

Notes: All numbers are from simulations of our model. The Total column of Average Treatment Effects, Sorting on Gains and Selection Bias sum to the "Observed" column for each row. The "Obs" ($\boldsymbol{X}$) and "Abil" ($\boldsymbol{\theta}$) columns decompose their respective totals into the part coming from observable characteristics and the part coming from the unobserved abilities. Each decomposition decomposes the observed difference in outcomes between people with final schooling levels j or $j+1$ into the various components above.

Table 38: Decomposition the Observed Difference in Smoking (decomposed pairwise comparison)

		Average Treatment Effects			Sorting on Gains			Selection Bias		
	Observed	Total	Obs ($\boldsymbol{X}$)	Abil ($\boldsymbol{\theta}$)	Total	Obs ($\boldsymbol{X}$)	Abil ($\boldsymbol{\theta}$)	Total	Obs ($\boldsymbol{X}$)	Abil ($\boldsymbol{\theta}$)
HS-DO	-0.27	-0.20	-0.16	-0.03	0.02	-0.01	0.02	-0.09	0.03	-0.14
	(0.04)	(0.07)	(0.09)	(0.03)	(0.02)	(0.01)	(0.01)	(0.08)	(0.04)	(0.08)
SC-HS	-0.06	-0.06	-0.06	0.00	-0.01	-0.01	-0.00	0.01	0.02	-0.01
	(0.05)	(0.05)	(0.07)	(0.07)	(0.03)	(0.02)	(0.02)	(0.05)	(0.03)	(0.05)
Coll-SC	-0.19	-0.17	-0.16	-0.03	-0.00	0.01	-0.01	-0.02	0.00	-0.02
	(0.03)	(0.03)	(0.03)	(0.01)	(0.02)	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)
GED-DO	-0.03	0.02	0.04	-0.03	-0.01	-0.00	-0.00	-0.04	0.01	-0.06
	(0.03)	(0.04)	(0.04)	(0.03)	(0.02)	(0.01)	(0.01)	(0.04)	(0.03)	(0.02)

Notes: All numbers are from simulations of our model. The Total column of Average Treatment Effects, Sorting on Gains and Selection Bias sum to the "Observed" column for each row. The "Obs" ($\boldsymbol{X}$) and "Abil" ($\boldsymbol{\theta}$) columns decompose their respective totals into the part coming from observable characteristics and the part coming from the unobserved abilities. Each decomposition decomposes the observed difference in outcomes between people with final schooling levels j or $j+1$ into the various components above.

Table 39: Decomposition the Observed Difference in Health Limits Work (decomposed pairwise comparison)

		Average Treatment Effects			Sorting on Gains			Selection Bias		
	Observed	Total	Obs	Abil	Total	Obs	Abil	Total	Obs	Abil
			($\boldsymbol{X}$)	($\boldsymbol{\theta}$)		($\boldsymbol{X}$)	($\boldsymbol{\theta}$)		($\boldsymbol{X}$)	($\boldsymbol{\theta}$)
HS-DO	-0.17	-0.11	-0.12	0.02	-0.02	0.00	-0.02	-0.04	-0.04	-0.00
	(0.03)	(0.05)	(0.07)	(0.02)	(0.02)	(0.01)	(0.01)	(0.07)	(0.04)	(0.06)
SC-HS	-0.04	-0.03	-0.02	-0.00	0.02	0.01	0.02	-0.04	-0.01	-0.03
	(0.04)	(0.05)	(0.08)	(0.07)	(0.03)	(0.02)	(0.02)	(0.04)	(0.03)	(0.04)
Coll-SC	-0.09	-0.06	-0.05	-0.02	-0.01	0.00	-0.01	-0.02	-0.01	-0.01
	(0.03)	(0.03)	(0.03)	(0.01)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
GED-DO	-0.02	0.06	-0.01	0.07	-0.01	-0.00	-0.00	-0.08	-0.03	-0.05
	(0.03)	(0.03)	(0.03)	(0.02)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.01)

Notes: All numbers are from simulations of our model. The Total column of Average Treatment Effects, Sorting on Gains and Selection Bias sum to the "Observed" column for each row. The "Obs" ($\boldsymbol{X}$) and "Abil" ($\boldsymbol{\theta}$) columns decompose their respective totals into the part coming from observable characteristics and the part coming from the unobserved abilities. Each decomposition decomposes the observed difference in outcomes between people with final schooling levels j or $j+1$ into the various components above.

- We further decompose the wage and PV effects into components of ability (cognitive and non-cognitive). See Tables 41 and 42.

Table 40: Decomposition of the Observed Difference in Wages (fully decomposed pairwise comparison)

		Average Treatment Effects				Sorting on Gains			
	Observed	Total	Obs ($\boldsymbol{X}$)	Cog	Non-Cog	Total	Obs ($\boldsymbol{X}$)	Cog	Non-Cog
HS-DO	0.25	0.12	0.13	-0.01	-0.00	0.00	-0.01	0.01	0.00
	(0.03)	(0.04)	(0.05)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
SC-HS	0.14	0.10	0.09	0.00	0.00	-0.03	-0.01	-0.02	0.01
	(0.04)	(0.05)	(0.07)	(0.03)	(0.04)	(0.03)	(0.02)	(0.02)	(0.01)
Coll-SC	0.23	0.11	0.04	0.06	0.01	0.03	-0.00	0.03	0.00
	(0.03)	(0.03)	(0.03)	(0.00)	(0.00)	(0.02)	(0.01)	(0.01)	(0.01)
GED-DO	0.15	0.06	0.14	-0.02	-0.06	0.00	-0.01	0.01	0.00
	(0.03)	(0.04)	(0.04)	(0.02)	(0.01)	(0.02)	(0.01)	(0.01)	(0.01)

Notes: All numbers are from simulations of our model. The Total column of Average Treatment Effects, Sorting on Gains and Selection Bias sum to the "Observed" column for each row. The "Obs" and "Abil" columns decompose their respective totals into the part coming from observable characteristics and the part coming from the unobserved abilities. Each decomposition decomposes the observed difference in outcomes between people with final schooling levels j or $j+1$ into the various components above.

Table 41: Decomposition of the Observed Difference in Wages (fully decomposed pairwise comparison), Cont'd

		Selection Bias			
	Observed	Total	Obs (***X***)	Cog	Non-Cog
HS-DO	0.25	0.13	0.10	0.05	-0.03
	(0.03)	(0.05)	(0.03)	(0.03)	(0.02)
SC-HS	0.14	0.07	0.04	0.04	-0.01
	(0.04)	(0.04)	(0.02)	(0.02)	(0.01)
Coll-SC	0.23	0.09	0.07	0.01	0.01
	(0.03)	(0.02)	(0.01)	(0.01)	(0.00)
GED-DO	0.15	0.09	0.05	0.04	-0.00
	(0.03)	(0.03)	(0.02)	(0.01)	(0.01)

Notes: All numbers are from simulations of our model. The Total column of Average Treatment Effects, Sorting on Gains and Selection Bias sum to the “Observed” column for each row. The “Obs” and “Abil” columns decompose their respective totals into the part coming from observable characteristics and the part coming from the unobserved abilities. Each decomposition decomposes the observed difference in outcomes between people with final schooling levels j or $j+1$ into the various components above.

Table 41: Decomposition of the Observed Difference in PV Wages (fully decomposed pairwise comparison)

		Average Treatment Effects				Sorting on Gains			
	Observed	Total	Obs ($\boldsymbol{X}$)	Cog	Non-Cog	Total	Obs ($\boldsymbol{X}$)	Cog	Non-Cog
HS-DO	0.50	0.07	-0.00	0.06	0.01	-0.08	-0.03	-0.03	-0.01
	(0.04)	(0.06)	(0.08)	(0.02)	(0.01)	(0.02)	(0.01)	(0.01)	(0.01)
SC-HS	0.15	0.09	0.08	0.00	0.00	-0.03	-0.02	-0.02	0.01
	(0.03)	(0.03)	(0.03)	(0.00)	(0.00)	(0.02)	(0.01)	(0.01)	(0.01)
Coll-SC	0.31	0.17	0.09	0.05	0.04	0.05	0.01	0.02	0.02
	(0.04)	(0.04)	(0.05)	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)
GED-DO	0.20	-0.11	-0.12	0.01	0.00	-0.04	-0.03	-0.00	-0.00
	(0.06)	(0.06)	(0.09)	(0.04)	(0.06)	(0.04)	(0.02)	(0.03)	(0.01)

Notes: All numbers are from simulations of our model. The Total column of Average Treatment Effects, Sorting on Gains and Selection Bias sum to the "Observed" column for each row. The "Obs" and "Abil" columns decompose their respective totals into the part coming from observable characteristics and the part coming from the unobserved abilities. Each decomposition decomposes the observed difference in outcomes between people with final schooling levels j or $j+1$ into the various components above.

Table 42: Decomposition of the Observed Difference in PV Wages (fully decomposed pairwise comparison), Cont'd

		Selection Bias			
	Observed	Total	Obs ($\boldsymbol{X}$)	Cog	Non-Cog
HS-DO	0.50	0.51	0.29	0.22	0.01
	(0.04)	(0.08)	(0.05)	(0.05)	(0.03)
SC-HS	0.15	0.09	0.06	0.04	-0.01
	(0.03)	(0.02)	(0.02)	(0.01)	(0.01)
Coll-SC	0.31	0.09	0.07	0.03	-0.00
	(0.04)	(0.04)	(0.03)	(0.02)	(0.01)
GED-DO	0.20	0.34	0.15	0.19	0.00
	(0.06)	(0.06)	(0.05)	(0.04)	(0.01)

Notes: All numbers are from simulations of our model. The Total column of Average Treatment Effects, Sorting on Gains and Selection Bias sum to the "Observed" column for each row. The "Obs" and "Abil" columns decompose their respective totals into the part coming from observable characteristics and the part coming from the unobserved abilities. Each decomposition decomposes the observed difference in outcomes between people with final schooling levels j or $j+1$ into the various components above.

Return to main text

Web Appendix A.15.3
Decompositions in Observed Differences of Arriving at $j(Q_j = 1)$ Including Continuation Values

- Parallel to the decomposition (18) in the text, we decompose the values of being at j into components associated with stopping at j and continuing beyond j where, for the upper branch of Figure 1 ($D_0 = 0$),

$$Y^k = Y_0^k + \sum_{j \geq 1}^{\bar{s}} \rho_{j-1,j}^k Q_j, \tag{26}$$

- where $\rho_{j-1,j}^k = Y_j^k - Y_{j-1}^k$.

- The expected future gain for a person at j (≥ 1) is

$$E_j\left(\sum_{l>j}^{\bar{s}} \rho^k_{l-1,l} Q_l | Q_j = 1\right)$$
$$= \sum_{l>j} \left[E_j(\rho^k_{l-1,l} | Q_l = 1) P(Q_l = 1 | Q_j = 1)\right], \quad j \geq 1,$$

- where the conditioning $D_0 = 0$ is kept implicit.

- Introducing D_0 and note that at the initial node, $Q_0 := 1$ and

$$Y^k = Y_0 + \left(\sum_{j \in \mathcal{S} \setminus \{0,G\}} \rho^k_{j-1,j} Q_j \right) (1 - D_0) + \rho^k_{0,G} Q_G (D_0)$$

- where $\rho^k_{0,G} = (Y^k_G - Y^k_0)$.

- Thus, the expected future gain for a person at $j = 0$ is

$$E_0\left[\left(\sum_{l\geq 1}\rho^k_{l-1,l}Q_l(1-D_0)|Q_l=1,D_0=0\right)P(Q_l=1|D_0=0)+\rho^k_{0,G}Q_G(1-D_0)\right] \tag{27}$$

$$=\sum_{l\geq 1}E_0\left(\rho^k_{l-1,l}|Q_l=1,D_0=0\right)P(Q_l=1|D_0=0)+E_0\left(\rho^k_{0,G}|Q_G=1,D_0=1\right)P(Q_G=1|D_0=1).$$

- Specifically for the kth outcome at node j:

$$
\begin{aligned}
&\underbrace{E[Y^k|D_j=0,Q_j=1]-E[Y^k|D_j=1,Q_j=1]}_{\text{Raw difference}}\\
&=\underbrace{E[Y^k|D_j=0,Q_j=1]-E[Y^k|D_j=0,Q_j=1,\mathit{Fix}\ D_j=1]}_{\text{Dynamic treatment on the treated for those at } j}\\
&+\underbrace{E[Y^k|D_j=0,Q_j=1,\mathit{Fix}\ D_j=1]-E[Y^k|D_j=1,Q_j=1]}_{\text{Selection bias for those at } j}\\
&=\underbrace{E[Y^k|Q_j=1,\mathit{Fix}\ D_j=0]-E[Y^k|Q_j=1,\mathit{Fix}\ D_j=1]}_{\text{ATE for those at } j}\\
&+\underbrace{\left\{\begin{array}{l}\left(E[Y^k|D_j=0,Q_j=1]-E[Y^k|D_j=0,Q_j=1,\mathit{Fix}\ D_j=1]\right)\\-\left(E[Y^k|Q_j=1,\mathit{Fix}\ D_j=0]-E[Y^k|Q_j=1,\mathit{Fix}\ D_j=1]\right)\end{array}\right\}}_{\text{TT - ATE: Sorting gain at } j \text{ for those who transit to } j+1}\\
&+\underbrace{E[Y^k|D_j=0,Q_j=1,\mathit{Fix}\ D_j=1]-E[Y^k|D_j=1,Q_j=1]}_{\text{Selection bias}}. \qquad (28)
\end{aligned}
$$

- The node-specific ATE_j is defined for the population at $Q_j = 1$ and considers moving the entire group from j to $j + 1$
(i.e, *Fix* $D_j = 1$ and *Fix* $D_j = 0$, respectively).
- The sorting gain is the net gain beyond ATE_j to those who actually take the transition ($D_j = 0$).

Table 42: Decomposition of the Observed Difference in Wage (including continuation values)

	Observed	Average Treatment Effects	Sorting on Gains	Selection Bias
Graduate HS	0.32	0.09	-0.00	0.22
	(0.02)	(0.06)	(0.02)	(0.07)
Enroll in Coll	0.27	0.13	0.01	0.13
	(0.02)	(0.03)	(0.01)	(0.02)
Graduate Coll	0.23	0.11	0.03	0.09
	(0.03)	(0.04)	(0.02)	(0.03)
Get GED	0.15	0.06	0.00	0.09
	(0.04)	(0.05)	(0.03)	(0.04)

Notes: All numbers are from simulations of our model. Average Treatment Effects, Sorting on Gains and Selection Bias sum to the “Observed” column for each row. Each decomposition decomposes the observed difference in outcomes between people who make and do not make a particular decision (conditional on reaching the decision).

Table 43: Decomposition the Observed Difference in PV Wage (including continuation values)

	Observed	Average Treatment Effects	Sorting on Gains	Selection Bias
Graduate HS	0.58	0.17	-0.04	0.44
	(0.03)	(0.06)	(0.02)	(0.07)
Enroll in Coll	0.32	0.14	0.00	0.18
	(0.03)	(0.03)	(0.02)	(0.03)
Graduate Coll	0.31	0.17	0.05	0.09
	(0.04)	(0.04)	(0.02)	(0.04)
Get GED	0.20	-0.11	-0.04	0.34
	(0.06)	(0.06)	(0.04)	(0.06)

Notes: All numbers are from simulations of our model. Average Treatment Effects, Sorting on Gains and Selection Bias sum to the "Observed" column for each row. Each decomposition decomposes the observed difference in outcomes between people who make and do not make a particular decision (conditional on reaching the decision).

Table 44: Decomposition the Observed Difference in Smoking (including continuation values)

	Observed	Average Treatment Effects	Sorting on Gains	Selection Bias
Graduate HS	-0.34	-0.26	-0.00	-0.08
	(0.03)	(0.06)	(0.01)	(0.06)
Enroll in Coll	-0.16	-0.14	-0.04	0.01
	(0.02)	(0.03)	(0.02)	(0.02)
Graduate Coll	-0.19	-0.17	-0.00	-0.02
	(0.03)	(0.04)	(0.02)	(0.04)
Get GED	-0.03	0.02	-0.01	-0.04
	(0.05)	(0.05)	(0.03)	(0.05)

Notes: All numbers are from simulations of our model. Average Treatment Effects, Sorting on Gains and Selection Bias sum to the “Observed” column for each row. Each decomposition decomposes the observed difference in outcomes between people who make and do not make a particular decision (conditional on reaching the decision).

Table 45: Decomposition the Observed Difference in Health Limits Work (including continuation values)

	Observed	Average Treatment Effects	Sorting on Gains	Selection Bias
Graduate HS	-0.21	-0.11	-0.00	-0.09
	(0.02)	(0.04)	(0.01)	(0.04)
Enroll in Coll	-0.09	-0.04	0.01	-0.07
	(0.02)	(0.02)	(0.01)	(0.02)
Graduate Coll	-0.09	-0.06	-0.01	-0.02
	(0.02)	(0.03)	(0.01)	(0.02)
Get GED	-0.02	0.06	-0.01	-0.08
	(0.04)	(0.05)	(0.03)	(0.04)

Notes: All numbers are from simulations of our model. Average Treatment Effects, Sorting on Gains and Selection Bias sum to the "Observed" column for each row. Each decomposition decomposes the observed difference in outcomes between people who make and do not make a particular decision (conditional on reaching the decision).

Return to main text

Appendix A.16: Sorting Gains

Table 46: Decomposition the Observed Difference in Log Wages (decomposed pairwise comparison)

		Average Treatment Effects			Sorting on Gains			Selection Bias		
	Observed	Total	Obs ($\boldsymbol{X}$)	Abil ($\boldsymbol{\theta}$)	Total	Obs ($\boldsymbol{X}$)	Abil ($\boldsymbol{\theta}$)	Total	Obs ($\boldsymbol{X}$)	Abil ($\boldsymbol{\theta}$)
HS-DO	0.25	0.12	0.13	-0.01	0.00	-0.01	0.01	0.13	0.10	0.02
	(0.03)	(0.04)	(0.05)	(0.01)	(0.01)	(0.01)	(0.01)	(0.05)	(0.03)	(0.04)
SC-HS	0.14	0.10	0.09	0.01	-0.03	-0.01	-0.01	0.07	0.04	0.03
	(0.04)	(0.05)	(0.07)	(0.06)	(0.03)	(0.02)	(0.02)	(0.04)	(0.02)	(0.02)
Coll-SC	0.23	0.11	0.04	0.07	0.03	-0.00	0.03	0.09	0.07	0.02
	(0.03)	(0.03)	(0.03)	(0.01)	(0.02)	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)
GED-DO	0.15	0.06	0.14	-0.08	0.00	-0.01	0.01	0.09	0.05	0.04
	(0.03)	(0.04)	(0.04)	(0.02)	(0.02)	(0.01)	(0.01)	(0.03)	(0.02)	(0.02)

Notes: All numbers are from simulations of our model. The Total column of Average Treatment Effects, Sorting on Gains and Selection Bias sum to the “Observed” column for each row. The “Obs” ($\boldsymbol{X}$) and “Abil” ($\boldsymbol{\theta}$) columns decompose their respective totals into the part coming from observable characteristics and the part coming from the unobserved abilities. Each decomposition decomposes the observed difference in outcomes between people with final schooling levels j or $j+1$ into the various components above.

Table 47: Decomposition the Observed Difference in PV Log Wages (decomposed pairwise comparison)

		Average Treatment Effects			Sorting on Gains			Selection Bias		
	Observed	Total	Obs ($\boldsymbol{X}$)	Abil ($\boldsymbol{\theta}$)	Total	Obs ($\boldsymbol{X}$)	Abil ($\boldsymbol{\theta}$)	Total	Obs ($\boldsymbol{X}$)	Abil ($\boldsymbol{\theta}$)
HS-DO	0.50	0.07	-0.00	0.07	-0.08	-0.03	-0.04	0.51	0.29	0.22
	(0.04)	(0.06)	(0.08)	(0.02)	(0.02)	(0.01)	(0.02)	(0.08)	(0.05)	(0.06)
SC-HS	0.15	0.09	0.08	0.01	-0.03	-0.02	-0.01	0.09	0.06	0.03
	(0.06)	(0.06)	(0.09)	(0.07)	(0.04)	(0.02)	(0.03)	(0.06)	(0.05)	(0.04)
Coll-SC	0.31	0.17	0.09	0.09	0.05	0.01	0.04	0.09	0.07	0.02
	(0.03)	(0.03)	(0.03)	(0.01)	(0.02)	(0.01)	(0.01)	(0.02)	(0.02)	(0.01)
GED-DO	0.20	-0.11	-0.12	0.01	-0.04	-0.03	-0.00	0.34	0.15	0.19
	(0.04)	(0.04)	(0.05)	(0.03)	(0.02)	(0.01)	(0.01)	(0.04)	(0.03)	(0.02)

Notes: All numbers are from simulations of our model. The Total column of Average Treatment Effects, Sorting on Gains and Selection Bias sum to the "Observed" column for each row. The "Obs" ($\boldsymbol{X}$) and "Abil" ($\boldsymbol{\theta}$) columns decompose their respective totals into the part coming from observable characteristics and the part coming from the unobserved abilities. Each decomposition decomposes the observed difference in outcomes between people with final schooling levels j or $j+1$ into the various components above.

Table 48: Decomposition the Observed Difference in Smoking (decomposed pairwise comparison)

		Average Treatment Effects			Sorting on Gains			Selection Bias		
	Observed	Total	Obs ($\boldsymbol{X}$)	Abil ($\boldsymbol{\theta}$)	Total	Obs ($\boldsymbol{X}$)	Abil ($\boldsymbol{\theta}$)	Total	Obs ($\boldsymbol{X}$)	Abil ($\boldsymbol{\theta}$)
HS-DO	-0.27	-0.20	-0.16	-0.03	0.02	-0.01	0.02	-0.09	0.03	-0.14
	(0.04)	(0.07)	(0.09)	(0.03)	(0.02)	(0.01)	(0.01)	(0.08)	(0.04)	(0.08)
SC-HS	-0.06	-0.06	-0.06	0.00	-0.01	-0.01	-0.00	0.01	0.02	-0.01
	(0.05)	(0.05)	(0.07)	(0.07)	(0.03)	(0.02)	(0.02)	(0.05)	(0.03)	(0.05)
Coll-SC	-0.19	-0.17	-0.16	-0.03	-0.00	0.01	-0.01	-0.02	0.00	-0.02
	(0.03)	(0.03)	(0.03)	(0.01)	(0.02)	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)
GED-DO	-0.03	0.02	0.04	-0.03	-0.01	-0.00	-0.00	-0.04	0.01	-0.06
	(0.03)	(0.04)	(0.04)	(0.03)	(0.02)	(0.01)	(0.01)	(0.04)	(0.03)	(0.02)

Notes: All numbers are from simulations of our model. The Total column of Average Treatment Effects, Sorting on Gains and Selection Bias sum to the "Observed" column for each row. The "Obs" ($\boldsymbol{X}$) and "Abil" ($\boldsymbol{\theta}$) columns decompose their respective totals into the part coming from observable characteristics and the part coming from the unobserved abilities. Each decomposition decomposes the observed difference in outcomes between people with final schooling levels j or $j+1$ into the various components above.

Table 49: Decomposition the Observed Difference in Health Limits Work (decomposed pairwise comparison)

		Average Treatment Effects			Sorting on Gains			Selection Bias		
	Observed	Total	Obs	Abil	Total	Obs	Abil	Total	Obs	Abil
			($\boldsymbol{X}$)	($\boldsymbol{\theta}$)		($\boldsymbol{X}$)	($\boldsymbol{\theta}$)		($\boldsymbol{X}$)	($\boldsymbol{\theta}$)
HS-DO	-0.17	-0.11	-0.12	0.02	-0.02	0.00	-0.02	-0.04	-0.04	-0.00
	(0.03)	(0.05)	(0.07)	(0.02)	(0.02)	(0.01)	(0.01)	(0.07)	(0.04)	(0.06)
SC-HS	-0.04	-0.03	-0.02	-0.00	0.02	0.01	0.02	-0.04	-0.01	-0.03
	(0.04)	(0.05)	(0.08)	(0.07)	(0.03)	(0.02)	(0.02)	(0.04)	(0.03)	(0.04)
Coll-SC	-0.09	-0.06	-0.05	-0.02	-0.01	0.00	-0.01	-0.02	-0.01	-0.01
	(0.03)	(0.03)	(0.03)	(0.01)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
GED-DO	-0.02	0.06	-0.01	0.07	-0.01	-0.00	-0.00	-0.08	-0.03	-0.05
	(0.03)	(0.03)	(0.03)	(0.02)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.01)

Notes: All numbers are from simulations of our model. The Total column of Average Treatment Effects, Sorting on Gains and Selection Bias sum to the “Observed” column for each row. The “Obs” ($\boldsymbol{X}$) and “Abil” ($\boldsymbol{\theta}$) columns decompose their respective totals into the part coming from observable characteristics and the part coming from the unobserved abilities. Each decomposition decomposes the observed difference in outcomes between people with final schooling levels j or $j+1$ into the various components above.

Return to main text

Appendix:
Literature Review on Education and Health

Literature Review

Physical Health Conditions

Measure of Education	Health Outcome	Effect Estimated	Exclusions	Inclusions
		Adams (2002), *Education Economics*		
Years of education	Can do with ease:			
	climb stairs	0.038 (0.005) [M]; 0.040 (0.006) [F]	quarter of birth, parental education, number of siblings, % female siblings, dummy youngest child, dummy middle child	year of birth, race, marital status, region of birth, quadratic in age, parents alive, height, missings dummy
	walk a block	0.017 (0.003) [M]; 0.023 (0.003) [F]		
	take a bath	0.008 (0.002) [M]; 0.007 (0.002) [F]		
	pick up a dime	0.008 (0.002) [M]; 0.006 (0.002) [F]		
	stoop, kneel, crouch	0.026 (0.005) [M]; -0.014 (0.006) [F]		
		Arkes (2003), *RAND Discussion Paper*		
Years of education	work-limit condition	-0.126 (0.041) [-0.026] [M&F]	average state unemployment rate at ages 15-17	state per-capita income, teacher salary/per-capita income, age, state of birth, student-teacher ratio
	mobility limitation	-0.020 (0.028) [-0.0012] [M&F]		
	require personal care	-0.097 (0.046) [-0.0067] [M&F]		

Numbers in parentheses are standard errors. Numbers in brackets are marginal effects. M: males; F: females.

Literature Review

Physical Health Conditions (ctd.)				
Measure of Education	**Health Outcome**	**Effect Estimated**	**Exclusions**	**Inclusions**
Auld and Sidhu (2005), *Health Economics*				
Years of education	health limitation health limitation health limitation	-0.001 (0.005) [M&F] [AFQT in the *X* set] -0.023 (0.003) [M&F] [AFQT in the *IV* set] -0.015 (0.005) [M&F] [AFQT excluded]	parents' schooling, father's occupation, unemployment rate in 1979, AFQT (in some specifications)	cohort, race, no. of sibs, region at 14, previous health, marital status, family size, family income
Berger and Leigh (1989), *Journal of Human Resources*				
Years of education	work-limit condition functional limitation	-0.062 (0.018) [M&F] (probit coefficients) -0.046 (0.015) [M&F] (probit coefficients)	IQ, Knowledge of Work Test, parents' schooling	gender, quadratic in age, race, marital status, hh size, SMSA residence, industry illness&injury rate, past health
Kaestner and Callison (2011), *JHC*				
High School	Short-Form Physical	1.21 (0.88) [M]; 2.34 (1.28) [F]	n/a n/a	AFQT, Rosenberg and Rotter scales, church attendance, stealing history, use of tobacco, alcohol and marijuana by 14, family background
Some College		0.71 (1.04) [M]; 4.20 (1.43) [F]	n/a n/a	
BA or more		2.08 (1.17) [M]; 5.53 (1.62) [F]	n/a n/a	

Numbers in parentheses are standard errors. M: males; F: females.

Literature Review

Physical Health Conditions (ctd.)				
Measure of Education	**Health Outcome**	**Effect Estimated**	**Exclusions**	**Inclusions**
Mazumder (2008), *Economic Perspectives*				
Years of education	trouble seeing trouble hearing trouble speaking health limitation	-0.056 (0.025) -0.050 (0.025) -0.019 (0.008) -0.074 (0.035)	compulsory schooling laws	gender, sob, rob, rob × cohort, sob characteristics, state × cohort trends age cubic × year
Oreopoulos (2006), *American Economic Review*				
Years of education	health disability mobility limitation	-0.025 (0.006) -0.043 (0.007)	compulsory schooling laws	birth year, region, survey year, sex, quartic in age, race, %urban, % in labor force, % in manufacturing
Clark and Royer (2013), *American Economic Review*				
Years of education	reduced activity longstanding illness	-0.005 (0.008) 0.002 (0.011)	1947, 1972 UK school reform cubic in quarter of birth	sex, month-of-birth dummies month- and year-of interview dummies and cubic in age (in months)
Silles (2009), *Economics of Education Review*				
Years of education	no long-term illness no activity-limiting illness no work-preventing illness	0.055 (0.009) 0.046 (0.008) 0.009 (0.005)	1947 and 1973 UK school reforms	quadratic in age, survey year dummy, gender

Numbers in parentheses are standard errors. M: males; F: females. sob: state of birth; rob: region of birth.

Literature Review

Self-Reported Health				
Measure of Education	**Health Outcome**	**Effect Estimated**	**Exclusions**	**Inclusions**
Adams (2002), *Education Economics*				
Years of education	SRH: good health SRH: very good health SRH: excellent health	0.044 (0.005) [M]; 0.048 (0.005) [F] 0.022 (0.018) [M]; 0.063 (0.007) [F] 0.032 (0.006) [M]; 0.042 (0.006) [F]	quarter of birth, parents' ed, no. siblings, % female sibs, dummy youngest +middle child	year of birth, race, marital status, rob, height, quadratic in age, parents alive, missings dummy

Numbers in parentheses are standard errors. M: males; F: females. sob: state of birth; rob: region of birth.

Literature Review

Self-Reported Health (ctd.)				
Measure of Education	**Health Outcome**	**Effect Estimated**	**Exclusions**	**Inclusions**
Mazumder (2008), *Economic Perspectives*				
Yrs of education	fair/poor health	-0.082 (0.034)	compulsory schooling laws sob characteristics, state $\times$ cohort trends age cubic $\times$ year	gender, sob, rob, rob $\times$ cohort,
Clark and Royer (2013), *American Economic Review*				
Yrs of education	fair/bad health	-0.0018 (0.0022)	1947 UK reform	sex and month-of -birth dummies
Oreopoulos (2006), *American Economic Review*				
Years of education	poor health good health	-0.032 (0.011) 0.060 (0.015)	1947 UK reform	birth year, region, survey year, sex, quartic in age, race, %urban, % in labor force, % in manufacturing
Silles (2009), *Economics of Education Review*				
Years of education	good health	0.045 (0.009)	1947&1973 UK reforms	quadratic in age, year dummy, gender

Numbers in parentheses are standard errors. M: males; F: females. sob: state of birth; rob: region of birth.

Literature Review

Mortality				
Measure of Education	**Health Outcome**	**Effect Estimated**	**Exclusions**	**Inclusions**
Lleras-Muney (2005), *Review of Economic Studies*				
Years of education	10-year mortality	-0.037 (0.006)	compulsory schooling laws	gender, sob, rob, rob $\times$ cohort, sob characteristics,
Mazumder (2008), *Economic Perspectives*				
Years of education	10-year mortality	-0.016 (0.024)	compulsory schooling laws	gender, sob, rob, rob $\times$ cohort, sob characteristics, state $\times$ cohort trends age cubic $\times$ year
Clark and Royer (2013), *American Economic Review*				
Years of education	mortality 1970-2003	6.98 (3.30) [M]; 0.05 (0.045) [F]	1947 UK reform	sex and month-of-birth dummies

Numbers in parentheses are standard errors. M: males; F: females. sob: state of birth; rob: region of birth.

Literature Review

Blood Pressure				
Measure of Education	**Health Outcome**	**Effect Estimated**	**Exclusions**	**Inclusions**
Berger and Leigh (1989), *Journal of Human Resources*				
Years of education	systolic BP diastolic BP	-0.567 (0.250) [M&F] -0.192 (0.150) [M&F]	per-capita income and expenditure in education in the childhood state	gender, quadratic in age, race marital status, hh size, SMSA residence, industry illness&injury rate, past health
Mazumder (2008), *Economic Perspectives*				
Years of education	hypertension	0.038 (0.012)	schooling laws	gender, sob, rob, rob $\times$ cohort, sob characteristics, state $\times$ cohort trends age cubic $\times$ year
Clark and Royer (2013), *American Economic Review*				
Years of education	hypertension	-0.029 (0.018)	1947, 1972 UK school reform, cubic in quarter of birth	sex, month-of-birth dummies month- and year-of interview dummies in age (in months)

Numbers in parentheses are standard errors. M: males; F: females. sob: state of birth; rob: region of birth.

Literature Review

Measure of Education	Health Outcome	Effect Estimated	Exclusions	Inclusions
Smoking				
Sander (1995), *Review of Economics and Statistics*				
Years of education	quit smoking	0.058 [M]; (0.027) 0.088 [F] (0.036)	parents' schooling, rural residence at 16, region at 16, no. of siblings	age, black, dummies for survey year, current region, current rural residence
Kenkel et al. (2006), *Journal of Labor Economics*				
High school	current smoker	-0.229 [M] (0.088); -0.102 [F] (0.124)	no. of courses required for for high school graduation	age, age squared, race, work-limiting health condition, state fixed effects, cigarette taxes,
GED	current smoker	0.068 [M] (0.110); 0.116 [F] (0.108);	minimum graduation requirements at school district, fraction of	anti-smoking sentiments at state level, parents' schooling, presence of a
High school	former smoker	0.063 [M] (0.117); 0.189 [F] (0.130)	same-age youth who took the GED index of GED policies	magazine at home at 14, AFQT, Rotter scale
GED	former smoker	0.030 [M] (0.107); 0.099 [F] (0.094)	12-year average of per-capita education spending.	
Clark and Royer (2013), *American Economic Review*				
Years of education	current smoker	-0.004 (0.012)	1947, 1972 UK school reform, cubic in quarter month	sex, month-of-birth dummies month- and year-of interview dummies and cubic in age (in months)

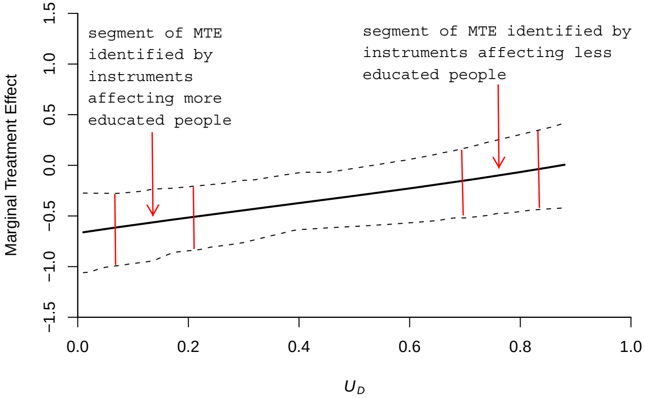
Numbers in parentheses are standard errors. M: males; F: females.

Literature Review

Obesity				
Measure of Education	**Health Outcome**	**Effect Estimated**	**Exclusions**	**Inclusions**
Kenkel et al. (2006), *Journal of Labor Economics*				
High school	overweight	0.110 [M] (0.083); 0.098 [F] (0.120)	no. of courses required for for high school graduation	age, age squared, race, work-limiting health condition, state fixed effects, cigarette taxes,
GED	overweight	0.006 [M] (0.101); 0.080 [F] (0.095);	minimum graduation requirements at school district, fraction of	anti-smoking sentiments at state level, parents' schooling, presence of a
High school	obesity	-0.008 [M] (0.082); -0.021 [F] (0.139)	same-age youth who took the GED index of GED policies	magazine at home at 14, AFQT, Rotter scale
GED	obesity	0.033 [M] (0.091); 0.022 [F] (0.095)	12-year average of per-capita education spending.	
Clark and Royer (2013), *American Economic Review*				
Years of education	BMI obesity	0.205 (0.164) 0.016 (0.015)	1947, 1972 UK school reform, cubic in quarter of birth	sex, month-of-birth dummies month- and year-of interview dummies and cubic in age (in months)

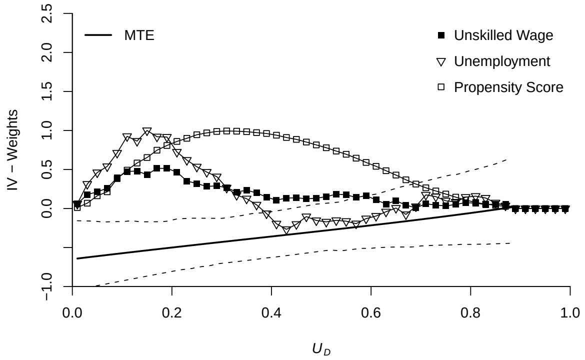
Numbers in parentheses are standard errors. M: males; F: females.

MTE Smoking, Males



Solid line depicts the MTE for the sample, while the dashed lines indicate, 95% confidence bands. The confidence intervals are computed based on 200 bootstrap runs. We estimate the MTE within the common support on a discrete grid with step size 0.01.

MTE Smoking and Weights for Instrumental Variables, Males

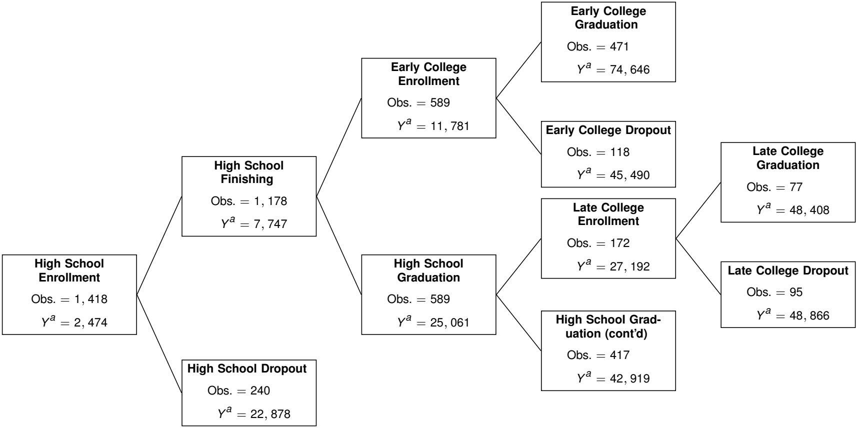


The scale of the y-axis is the scale of the MTE, not the scale of the weights, which are scaled to fit the picture.

Return to main text

Appendix:
Structural Dynamic Discrete Choice Model of Schooling

Figure 19: Decision Tree

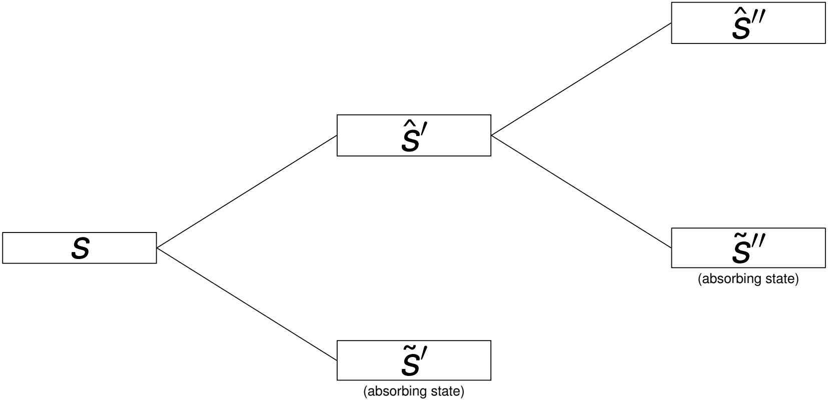


Notes: Y^a refers to average annual earnings in the state in 2005 dollars. Obs. refers to the number of observations in the state.

Setup

- Current state $s \in \mathcal{S} = \{s_1, \ldots, s_N\}$.
- $\mathcal{S}^v(s) \subseteq \mathcal{S}$: set of visited states.
- $\mathcal{S}^f(s) \subseteq \mathcal{S}$ the set of feasible states that can be reached from s.
- Choice set of the agent in state s: $\Omega(s) = \{s' \mid s' \in \mathcal{S}^f(s)\}$.
- Consider binary choices only, so $\Omega(s)$ has at most two elements.
- *Ex post*, the agent receives per period rewards $R(s') = Y(s') - C(s', s)$.
- Costs $C(s', s)$ associated with moving from state s to state s'.

Figure 20: Generic Decision Problem



Payoffs and Costs

$$Y(s) = \mu_s(X(s)) + \theta' \alpha_s + \epsilon(s)$$

$$C(s', s) \quad = K_{s',s}(Q(s', s)) + \theta' \varphi_{s',s} + \eta(s', s)$$

System of Measurement Equations for θ

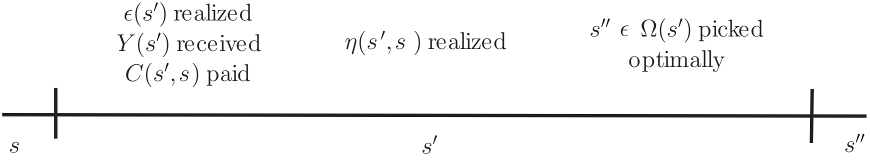
$$M(j) \quad = \quad X(j)'\kappa_j + \theta'\gamma_j + \nu(j) \qquad \forall \quad j \in \mathcal{M}$$

θ is unobserved ability vector (cognitive and noncognitive)

Information

Timing

In period s':



Information Set

$$\left.\begin{array}{ll} \text{for all } s \in \text{ contained in } \mathcal{S}^v(s) & \eta(\hat{s}', s); \epsilon(s) \\ \text{for } s' \in \mathcal{S}^f(s) & \\ \text{and for all } s, s' & X(s); Q(s', s); \theta \end{array}\right\} \in \mathcal{I}(s).$$

Value Function

$$V(s \,|\, \mathcal{I}(s)) = Y(s) + \max_{s' \in \Omega(s)} \left\{ \frac{1}{1+r} \Big(-\underbrace{C(s', s) + \mathbb{E}[V(s' \,|\, \mathcal{I}(s')) \,\big|\, \mathcal{I}(s)\,]}_{\text{Continuation value}} \Big) \right\}$$

Decision Rule: Pick $s' \in \Omega(s)$ if

$$\mathbb{E}\left[V(s') \,\middle|\, \mathcal{I}(s)\right] - C(s', s) \geq \mathbb{E}\left[V(s) | \mathcal{I}(s)\right]$$

Ex Ante Net Return

$$NR(s', s) = \frac{\mathbb{E}\left[V(s') - V(s) \,\middle|\, \mathcal{I}(s)\right] - C(s', s)}{\mathbb{E}\left[V(s') \,\middle|\, \mathcal{I}(s)\right]}$$

Ex Ante Gross Returns

$$GR(s', s) = \frac{\mathbb{E}\left[V(s') - \tilde{V}(s) \,\middle|\, \mathcal{I}(s)\right]}{\mathbb{E}\left[\tilde{V}(\tilde{s}') \,\middle|\, \mathcal{I}(s)\right]}$$

$\tilde{V} :=$ Gross of Cost Value Function

Option Value (Weisbrod, 1962)

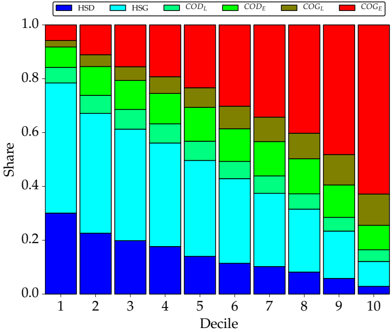
$$OV(s', s) =$$
$$\frac{1}{1+r}\,\mathbb{E}\left[\underbrace{\max_{s''\in\Omega(s')}\left\{E_{s'}(V(s'') - C(s'', s'))\right\}}_{\text{value of options arising from } s'} - \underbrace{\Big(V(s')\Big)}_{\text{fallback value}} \;\Bigg|\; \mathcal{I}(s)\right]$$

$E_{s'}$ is value with respect to information at s'.

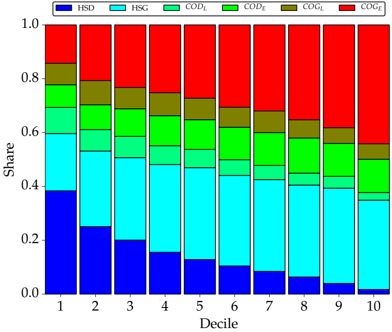
Return to main text

Appendix:
Empirical Results from Structural Models

Figure 21: Ability Distributions by Final Education



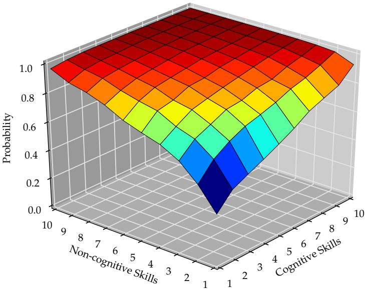
(a) Cognitive



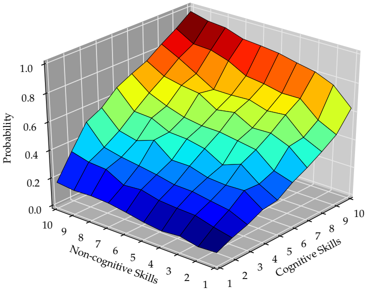
(b) Non-cognitive

Notes: We simulate a sample of 50,000 agents based on the estimates of the model.

Figure 22: Transition Probabilities by Abilities

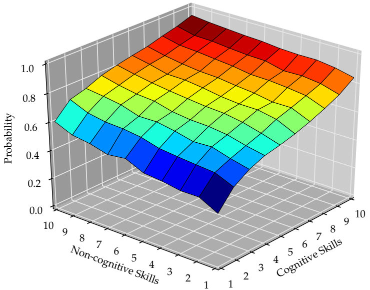


(a) High School Finishing



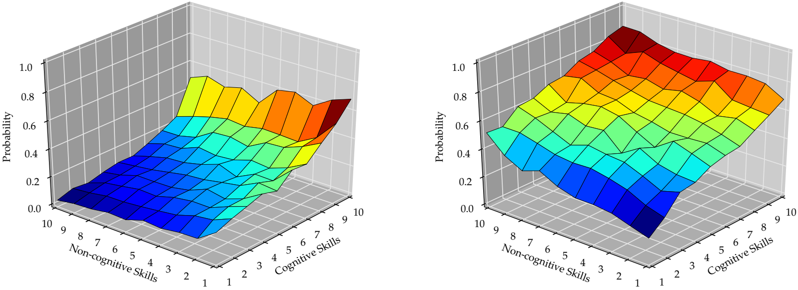
(b) Early College Enrollment

Figure 22: Transition Probabilities by Abilities



(c) Early College Graduation

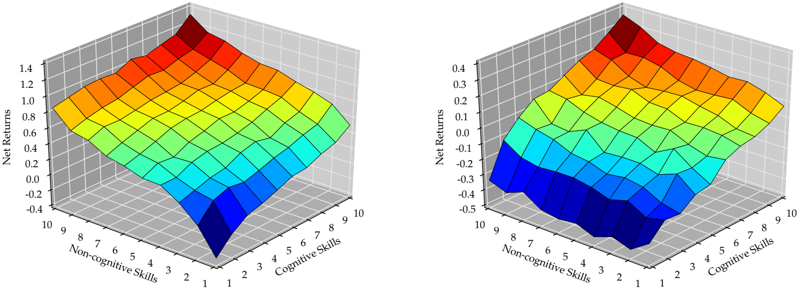
Figure 22: Transition Probabilities by Abilities



(d) Late College Enrollment (e) Late College Graduation

Notes: We simulate a sample of 50,000 agents based on the estimates of the model. In each subfigure, we condition on the agents that actually visit the relevant decision state.

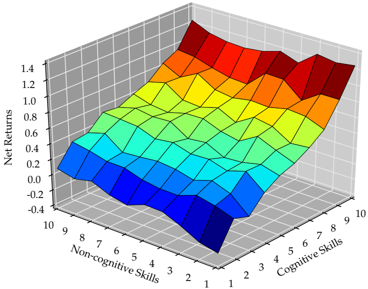
Figure 23: *Ex Ante* Net Returns by Abilities



(a) High School Finishing
$NR^a = 0.64$
$GR^a = 0.27$

(b) Early College Enrl.
$NR^a = -0.03$
$GR^a = 0.14$

Figure 23: *Ex Ante* Net Returns by Abilities

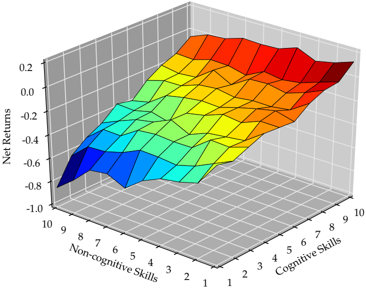


(c) Early College Grad.

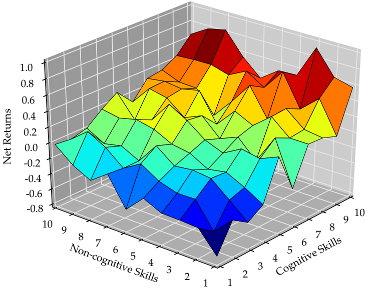
$NR^a = 0.50$

$GR^a = 0.75$

Figure 23: *Ex Ante* Net Returns by Abilities



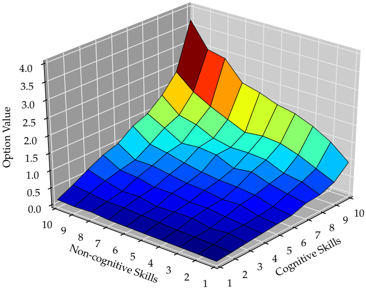
(d) Late College Enrl.
NR^a = -0.21
GR^a = 0.29



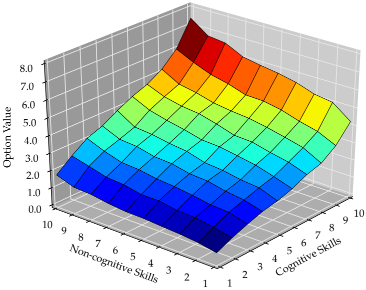
(e) Late College Grad.
NR^a = 0.10
GR^a = 0.24

Notes: We simulate a sample of 50,000 agents based on the estimates of the model. In each subfigure, we condition on the agents that actually visit the relevant decision state. Enrl. = Enrollment, Grad. = Graduation.

Figure 24: Option Values by Abilities

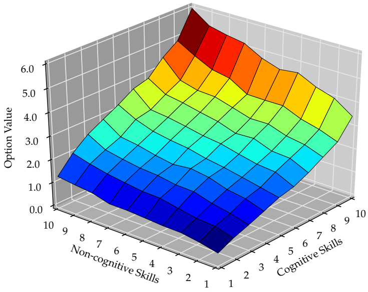


(a) High School Finishing
$OV = 0.52$
$OVC = 0.07$



(b) Early College Enrollment
$OV = 3.06$
$OVC = 0.30$

Figure 24: Option Values by Abilities



(c) Late College Enrollment

$$OV \quad = 1.87$$
$$OVC = 0.17$$

Notes: We simulate a sample of 50,000 agents based on the estimates of the model. In each subfigure, we condition on the agents that actually visit the relevant decision state. In units of $100,000.

Table 50: Cross Section Model Fit

State	Average Earnings		State Frequencies	
	Observed	ML	Observed	ML
High School Graduates	4.29	3.83	0.29	0.32
High School Dropouts	2.29	2.57	0.17	0.14
Early College Graduates	7.47	6.77	0.33	0.29
Early College Dropouts	4.55	3.84	0.08	0.11
Late College Graduates	4.84	6.16	0.05	0.08
Late College Dropouts	4.89	4.95	0.07	0.06

Table 51: Conditional Model Fit

State	Number of Children	Baby in Household	Parental Education	Broken Home
High School Dropout	0.77	0.26	0.37	0.03
High School Finishing	0.88	0.73	0.55	0.35
High School Graduation	0.91	0.94	0.65	0.91
High School Graduation (cont'd)	0.95	0.33	0.40	0.85
Early College Enrollment	0.46	0.54	0.01	0.15
Early College Graduation	0.06	0.86	0.00	0.14
Early College Dropout	0.33	0.27	0.54	0.75
Late College Enrollment	0.80	0.23	0.90	0.60
Late College Graduation	0.90	0.39	0.90	0.60
Late College Dropout	0.89	0.42	0.91	0.76

Table 52: Internal Rates of Return

		All	
High School Graduation	vs.	High School Dropout	215%
Early College Graduation	vs.	Early College Dropout	24%
Early College Graduation	vs.	High School Graduation (cont'd)	19%
Late College Dropout	vs.	High School Graduation (cont'd)	10%
Late College Graduation	vs.	High School Graduation (cont'd)	17%
Late College Dropout	vs.	High School Graduation (cont'd)	16%

Notes: The calculation is based on 1,407 individuals in the observed data. The Mincer rate of return is 11.6%.

Table 53: Net Returns

State	All	Treated	Untreated
High School Finishing	64%	75%	-27%
Early College Enrollment	-3%	24%	-28%
Early College Graduation	50%	82%	-44%
Late College Enrollment	-21%	22%	-38%
Late College Graduation	10%	62%	-51%

Notes: We simulate a sample of 50,000 agents based on the estimates of the model.

Table 54: Gross Returns

State	All	Treated	Untreated
High School Finishing	27%	29%	16%
Early College Enrollment	14%	20%	8%
Early College Graduation	75%	84%	49%
Late College Enrollment	29%	28%	29%
Late College Graduation	24%	36%	9%

Notes: We simulate a sample of 50,000 agents based on the estimates of the model.

Table 55: Regret

State	All	Treated	Untreated
High School Finishing	7%	4%	24%
Early College Enrollment	15%	28%	2%
Early College Graduation	29%	33%	19%
Late College Enrollment	21%	27%	19%
Late College Graduation	27%	34%	18%

Notes: We simulate a sample of 50,000 agents based on the estimates of the model.

Table 56: Option Value Contribution

State	All	Treated	Untreated
High School Finishing	7%	8%	2%
Early College Enrollment	30%	37%	23%
Late College Enrollment	17%	24%	15%

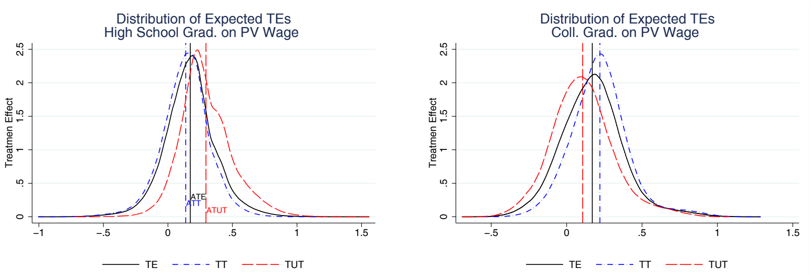
Notes: We simulate a sample of 50,000 agents based on the estimates of the model.

Return to main text

Web Appendix A.9
Distributions of Treatment Effects

- Using the full model it is possible to estimate the distribution of various treatment effects.
- Figure 25 shows the distribution of expected treatment effects at the choice to graduate from high school and the choice to graduate from college for the log present value of wages.
- Expectations are computed over the idiosyncratic error terms (ω_s^k).
- The individual's expected treatment effect is $E_\omega(Y_{s'}^k - Y_s^k) = (\tau_{s'}^k(\boldsymbol{X}) + \boldsymbol{\theta}'\boldsymbol{\alpha}_{s'}^k) - (\tau_s^k(\boldsymbol{X}) + \boldsymbol{\theta}'\boldsymbol{\alpha}_s^k)$, where the variation in the expected treatment effect is coming from the observables ($\boldsymbol{X}$) and the unobserved endowments ($\boldsymbol{\theta}$).
- The figure shows the distribution of expected treatment effects for everyone at the decision node, the distribution of those that choose to go on ($D_j = 0$), and those that choose not to ($D_j = 1$).

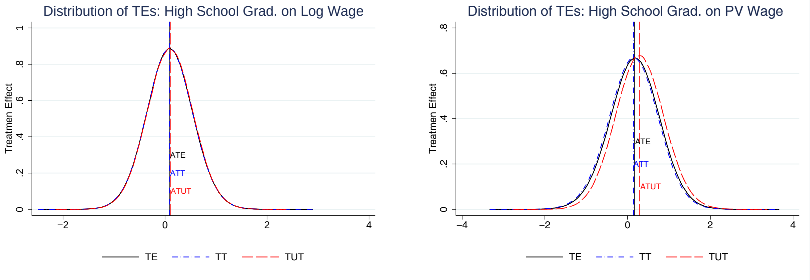
Figure 25: Distributions of Expected Treatment Effects: Log PV Wages



Notes: Distributions of expected treatment effects are for those who reach the educational choice. The expectation is computed over the idiosyncratic error terms (ω_s^k). The individual's expected treatment effect is $E_\omega(Y_{s'}^k - Y_s^k) = \tau_{s'}^k(\boldsymbol{X}) + \boldsymbol{\theta}'\boldsymbol{\alpha}_{s'}^k - (\tau_s^k(\boldsymbol{X}) + \boldsymbol{\theta}'\boldsymbol{\alpha}_s^k)$, where the variation in the expected treatment effect is coming from the observables ($\boldsymbol{X}$) and the unobserved endowments ($\boldsymbol{\theta}$). "TT" stands for average treatment on the treated and "TUT" stands for average treatment on the untreated. The vertical lines show the average treatment effect for each distribution. Note that the plots show expected benefits and do not include the idiosyncratic shocks realized *ex-post*.

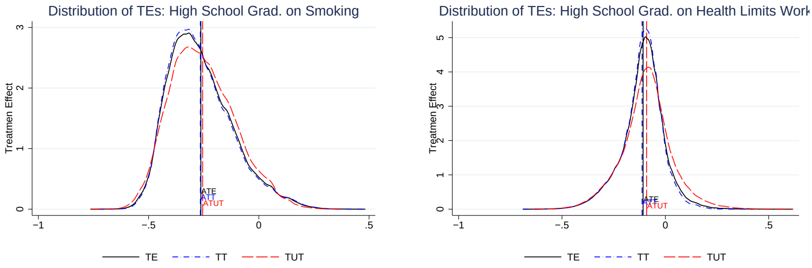
- Given the distributions, it is also possible to estimate the percent of individuals who benefit (or are expected to benefit) from a given transition.
- The model does not impose that individuals make educational choices based on expected gains, making it a testable hypothesis.
- Examining Figure 25, a portion of each distribution is to the left of 0.
- This represents the portion of the population that is expected to have lower present value of wages from making the transition.
- Many individuals do not make the transitions in spite of expected *ex-post* gains, while others make the transitions in spite of expected *ex-post* losses in the present value of wages.
- We find that the proportion of individuals who benefit is higher for those that choose to graduate from college,

Figure 26: Distributions of Treatment Effects: High School Graduation



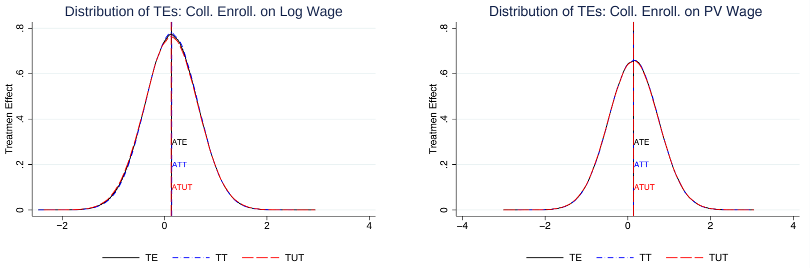
Notes: Distributions of treatment effects are for those who reach the educational choice. "TT" stands for average treatment on the treated and "TUT" stands for average treatment on the untreated. The vertical lines show the average treatment effect for each distribution. Note that the plots show expected benefits and do not include the idiosyncratic shocks realized *ex-post*.

Figure 26: Distributions of Treatment Effects: High School Graduation, Cont'd



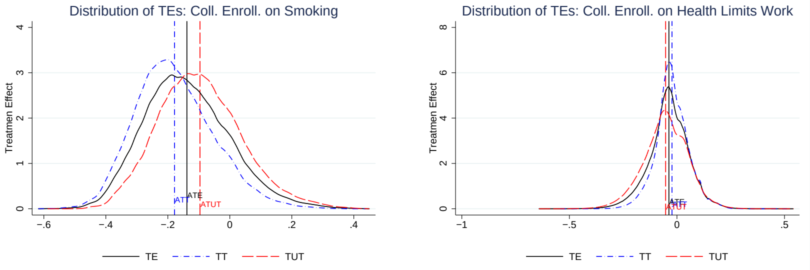
Notes: Distributions of treatment effects are for those who reach the educational choice. “TT” stands for average treatment on the treated and “TUT” stands for average treatment on the untreated. The vertical lines show the average treatment effect for each distribution. Note that the plots show expected benefits and do not include the idiosyncratic shocks realized *ex-post*.

Figure 27: Distributions of Treatment Effects: College Enrollment



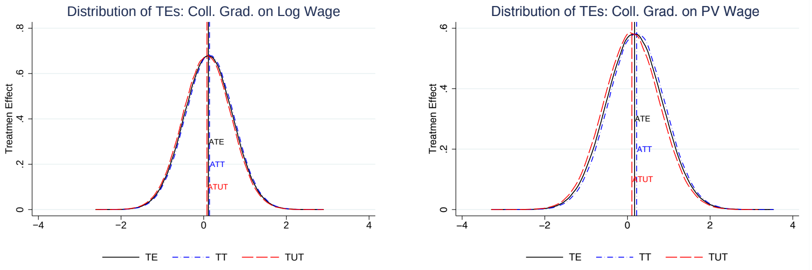
Notes: Distributions of treatment effects are for those who reach the educational choice. “TT” stands for average treatment on the treated and “TUT” stands for average treatment on the untreated. The vertical lines show the average treatment effect for each distribution. Note that the plots show expected benefits and do not include the idiosyncratic shocks realized *ex-post*.

Figure 27: Distributions of Treatment Effects: College Enrollment, Cont'd



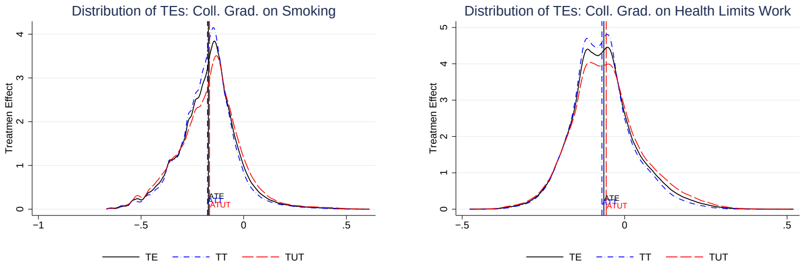
Notes: Distributions of treatment effects are for those who reach the educational choice. "TT" stands for average treatment on the treated and "TUT" stands for average treatment on the untreated. The vertical lines show the average treatment effect for each distribution. Note that the plots show expected benefits and do not include the idiosyncratic shocks realized *ex-post*.

Figure 28: Distributions of Treatment Effects: College Graduation



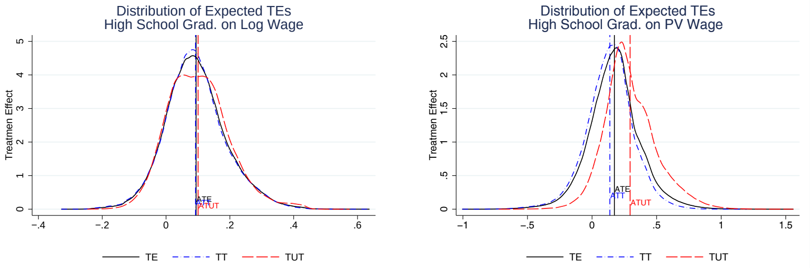
Notes: Distributions of treatment effects are for those who reach the educational choice. “TT” stands for average treatment on the treated and “TUT” stands for average treatment on the untreated. The vertical lines show the average treatment effect for each distribution. Note that the plots show expected benefits and do not include the idiosyncratic shocks realized *ex-post*.

Figure 28: Distributions of Treatment Effects: College Graduation, Cont'd



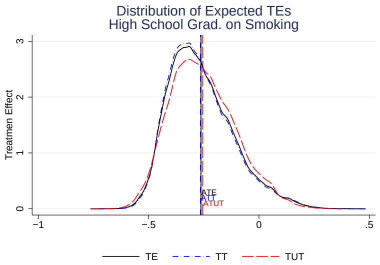
Notes: Distributions of treatment effects are for those who reach the educational choice. "TT" stands for average treatment on the treated and "TUT" stands for average treatment on the untreated. The vertical lines show the average treatment effect for each distribution. Note that the plots show expected benefits and do not include the idiosyncratic shocks realized *ex-post*.

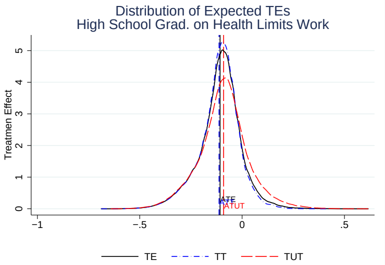
Figure 29: Distributions of Expected Treatment Effects: High School Graduation



Notes: Distributions of treatment effects are for those who reach the educational choice. “TT” stands for average treatment on the treated and “TUT” stands for average treatment on the untreated. The vertical lines show the average treatment effect for each distribution. Note that the plots show expected benefits and do not include the idiosyncratic shocks realized *ex-post*.

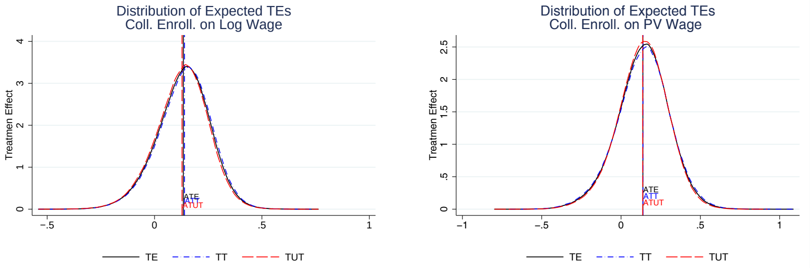
Figure 29: Distributions of Expected Treatment Effects: High School Graduation, Cont'd





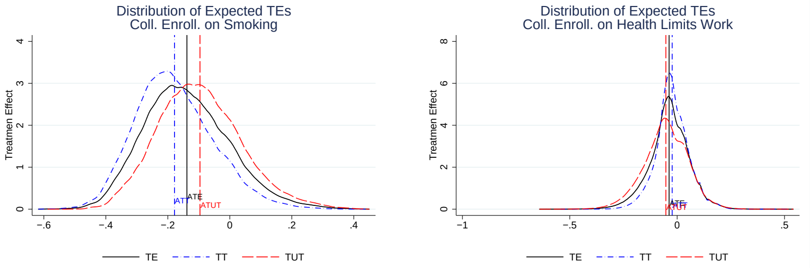
Notes: Distributions of treatment effects are for those who reach the educational choice. "TT" stands for average treatment on the treated and "TUT" stands for average treatment on the untreated. The vertical lines show the average treatment effect for each distribution. Note that the plots show expected benefits and do not include the idiosyncratic shocks realized *ex-post*.

Figure 30: Distributions of Expected Treatment Effects: College Enrollment



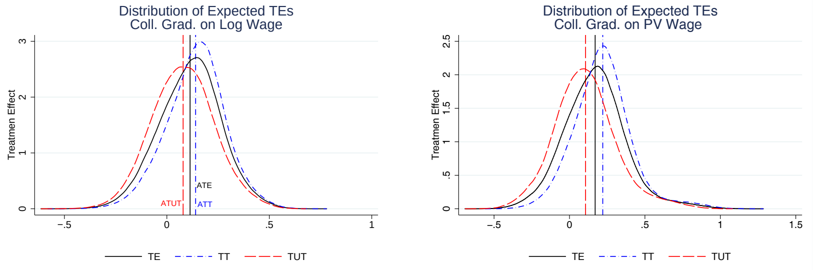
Notes: Distributions of treatment effects are for those who reach the educational choice. "TT" stands for average treatment on the treated and "TUT" stands for average treatment on the untreated. The vertical lines show the average treatment effect for each distribution. Note that the plots show expected benefits and do not include the idiosyncratic shocks realized *ex-post*.

Figure 30: Distributions of Expected Treatment Effects: College Enrollment, Cont'd



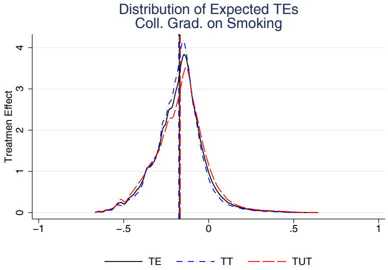
Notes: Distributions of treatment effects are for those who reach the educational choice. "TT" stands for average treatment on the treated and "TUT" stands for average treatment on the untreated. The vertical lines show the average treatment effect for each distribution. Note that the plots show expected benefits and do not include the idiosyncratic shocks realized *ex-post*.

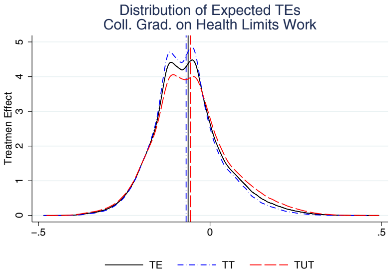
Figure 31: Distributions of Expected Treatment Effects: College Graduation



Notes: Distributions of treatment effects are for those who reach the educational choice. “TT” stands for average treatment on the treated and “TUT” stands for average treatment on the untreated. The vertical lines show the average treatment effect for each distribution. Note that the plots show expected benefits and do not include the idiosyncratic shocks realized *ex-post*.

Figure 31: Distributions of Expected Treatment Effects: College Graduation, Cont'd





Notes: Distributions of treatment effects are for those who reach the educational choice. "TT" stands for average treatment on the treated and "TUT" stands for average treatment on the untreated. The vertical lines show the average treatment effect for each distribution. Note that the plots show expected benefits and do not include the idiosyncratic shocks realized *ex-post*.

Table 57: Estimated Percent Who Benefit

Full Population: $Pr(Y^k_{j+1} - Y^k_j > 0)$			
	HS Graduation	Enroll in Coll	Grad. College
Log Wages	0.58	0.59	0.53
PV Log Wages	0.62	0.59	0.53
Health Limits Work	0.89	0.68	0.72
Daily Smoking	0.94	0.79	0.88
Conditional on Being at the Decision Node: $Pr(Y^k_{j+1} - Y^k_j > 0 \mid Q_j = 1)$			
	HS Graduation	Enroll in Coll	Grad. College
Log Wages	0.58	0.60	0.58
PV Log Wages	0.62	0.59	0.60
Health Limits Work	0.89	0.67	0.78
Daily Smoking	0.94	0.84	0.91
Conditional on Taking the Transition			
	HS Graduation	Enroll in Coll	Grad. College
Log Wages	0.58	0.61	0.60
PV Log Wages	0.59	0.59	0.63
Health Limits Work	0.91	0.65	0.80
Daily Smoking	0.94	0.91	0.93
Transition Probabilities: $Pr(D_j = 0 \mid Q_j = 1)$			
	HS Graduation	Enroll in Coll	Grad. College
Prob. of Taking Transition	.775	.514	.558

Notes: Results show the estimated percent who benefit. “Benefit” is defined as reduced probability of smoking, reduced probability of health limiting work, increased wages, or increased PV wages.

THE UNIVERSITY OF CHICAGO

Table 58: Spearman Correlations for Counterfactual States Using Simulated Expected Log Wages (age 30)

	Dropout	GED	Hs. Grad.	Some Coll.	Coll. Grad
Dropout	1.0000				
GED	0.7195	1.0000			
HS Grad.	0.7994	0.8558	1.0000		
Some Coll.	0.8535	0.7338	0.7407	1.0000	
Coll. Grad	0.7077	0.7824	0.7767	0.6986	1.0000

Notes: Table shows the Spearman correlation between the expected outcome for each level of schooling from a simulation of our model.

Table 59: Spearman Correlations for Counterfactual States Using Simulated Expected Log PV Wages

	Dropout	GED	Hs. Grad.	Some Coll.	Coll. Grad
Dropout	1.0000				
GED	0.8985	1.0000			
HS Grad.	0.9015	0.8327	1.0000		
Some Coll.	0.8321	0.6571	0.7647	1.0000	
Coll. Grad	0.8409	0.8194	0.7275	0.6069	1.0000

Notes: Table shows the Spearman correlation between the expected outcome for each level of schooling from a simulation of our model.

Table 60: Spearman Correlations for Counterfactual States Using Simulated Expected Smoking Age 30

	Dropout	GED	Hs. Grad.	Some Coll.	Coll. Grad
Dropout	1.0000				
GED	0.6072	1.0000			
HS Grad.	0.4166	0.1254	1.0000		
Some Coll.	0.5503	0.2852	0.1235	1.0000	
Coll. Grad	0.4985	0.4516	0.3892	0.1921	1.0000

Notes: Table shows the Spearman correlation between the expected outcome for each level of schooling from a simulation of our model.

Table 61: Spearman Correlations for Counterfactual States Using Simulated Expected Health Limits Work

	Dropout	GED	Hs. Grad.	Some Coll.	Coll. Grad
Dropout	1.0000				
GED	0.4076	1.0000			
HS Grad.	0.3104	0.7074	1.0000		
Some Coll.	-0.0823	0.2178	0.1219	1.0000	
Coll. Grad	0.6080	0.6147	0.5162	0.1916	1.0000

Notes: Table shows the Spearman correlation between the expected outcome for each level of schooling from a simulation of our model.

Return to main text