• “What is the rate of return to schooling?”

• “Underinvestment or overinvestment in education?”

  Becker (1964)

• The literature broadened this question to consider effects of education on health (longevity and morbidity), crime, voting, and a variety of other socioeconomic outcomes
• What are “the” causal effects of education on a variety of outcomes?

• Actively disputed – Ability bias?
Some Simple Evidence on the Data We Analyze:
Raw and Adjusted Differences in Outcomes
by Educational Status
The Observed Benefits from Education after Controlling for Background and Ability (Outcomes Compared to Dropouts)
The Observed Benefits from Education after Controlling for Background and Ability (Outcomes Compared to Dropouts)

Gains over Dropouts
High School Some College College
Raw Data Background Controls Background and Ability Controls

Health Limits Work

Daily Smoking

Gains over Dropouts
High School Some College College
Raw Data Background Controls Background and Ability Controls
How to Compute Rate of Return? Becker (1964)

Compare earnings at schooling at age \( t(Y_t^s) \) with earnings at schooling \( s' \) at age \( t \) \( (Y_t^{s'}) \):

\[
\sum_{t=0}^{T} \left( \frac{Y_t^{s'} - Y_t^{s}}{(1 + \rho_{s',s})^t} \right) = 0
\]

\[\rho_{s',s} = \text{rate of return}\]

Take schooling \( s' \) if \( \rho_{s',s} \geq r \) (alt. costs of funds)
Mincer Equation for Person $i$

- Developed in Becker and Chiswick (1966)
- Expanded and extensively applied in Mincer (1974)

\[
\ln y_i = \alpha_i + \rho_i s_i + \phi_i(x_i) \tag{1}
\]

$\rho_i$: “rate of return to schooling for person $i$.”
Early Literature Ignored

- Correlation between $\rho_i$ and $s_i$: Correlated Random Coefficient Model (Lewis, 1963; Rosen, 1977)
- Multiple skills and comparative advantage (Willis and Rosen, 1979; Heckman and Sedlacek, 1985)
- Uncertainty: Learning, updating and option values (Altonji, 1993; Arcidiacono, 2004; Cameron and Heckman, 1993; Comay et al., 1973; Levhari and Weiss, 1974; Weisbrod, 1962) focused on ex-post returns
- Measurement of psychic costs
- Role of family in shaping schooling decisions
The Modern Theory of Returns to Education and Educational Choices: Structural Dynamic Discrete Choice Models of Schooling
A Basic Human Capital Model  
(e.g., Keane and Wolpin, 1997)

- Finite decision horizon \((a, \bar{a})\).
- \(M\) alternatives.
- Let \(d_m(a) = 1\) if alternative \(m\) is chosen at age \(a\) and zero otherwise.
- The net reward per period at any age \(a\) (including psychic costs):

\[
R(a) = \sum_{m=1}^{M} R_m(a) \, d_m(a),
\]

Per period reward
• Individual’s state $S(a)$. Discount factor $\delta$:

$$V (S(a), a) = \max_{d_m(a)} E \left[ \sum_{\tau=a}^{A} \delta^{\tau-a} \sum_{m=1}^{M} R_m(\tau)d_m(\tau) \mid S(a) \right].$$

• The value function:

$$V (S(a), a) = \max_{m \in M} \{ V_m (S(a), a) \}.$$
The alternative-specific functions, \( V_m(S(a), a) \), given by

\[
V_m(S(a), a) = R_m(S(a), a) + \delta E \left[ V(S(a + 1), a + 1) \mid S(a), d_m(a) = 1 \right]
\]

Continuation Value

for \( a < \bar{a} \), and \( V_m(S(\bar{a}), \bar{a}) = R_m(S(\bar{a}), \bar{a}) \).
Decision Rule:
Let $m = 1$ denote schooling, and $m = 2$ next best alternative. Continue schooling at age $a$ if

$$V_1(S(a), a) \geq V_2(S(a), a)$$ (2)

But as long as further schooling is undertaken (continuation values sufficiently large), IRR *understates* the rate of return inclusive of continuation values.
Table 1: Structural Approach

<table>
<thead>
<tr>
<th>Benefits</th>
<th>Limitations</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Clearly interpretable economic parameters.</td>
<td>• Imposition of particular assumptions about decision rules and information sets that may not be well justified.</td>
</tr>
<tr>
<td>• Can generate all of the treatment effects reported in the treatment</td>
<td>• As usually implemented invokes arbitrary functional forms with empirical results often dependent on these choices.</td>
</tr>
<tr>
<td>effect literature using a common set of structural parameters.</td>
<td>• Because of complicated functional forms, difficult to replicate or do sensitivity analysis.</td>
</tr>
<tr>
<td>• Estimates distributions of gains and losses among participants.</td>
<td>• Sources of identification are often not clearly specified.</td>
</tr>
<tr>
<td>• Basis for constructing a broad range of economically interpretable</td>
<td>• Many moving parts and separate contributions not clearly identified.</td>
</tr>
<tr>
<td>policy counterfactuals.</td>
<td>• Estimates of psychic costs from risk-neutral models are substantial and drive schooling decisions.</td>
</tr>
<tr>
<td>• Role of observed and unobserved variables clearly delineated.</td>
<td></td>
</tr>
<tr>
<td>• Mechanisms producing treatment effects clearly articulated.</td>
<td></td>
</tr>
<tr>
<td>• Explicitly allows for multiple skills, comparative advantage,</td>
<td></td>
</tr>
<tr>
<td>uncertainty and learning.</td>
<td></td>
</tr>
</tbody>
</table>
Treatment Effect Literature

• Which treatment effect?
• $\bar{\rho}$ (ATE)? Marginal Return? TUT?
• Depends on the policy question.
• Unless instrument is the policy, treatment effects do not answer policy-relevant questions.
• Instrument dependence of estimates.
### Table 2: Treatment Effect Literature

<table>
<thead>
<tr>
<th>Benefits</th>
<th>Limitations</th>
</tr>
</thead>
</table>
| • Simple functional forms used (usually the Mincer equation or a variant). | • “Return” estimated is **instrument dependent** unless $\rho_i = \bar{\rho}$.
| • Clear about the source of identification.                           | • In “RD” analysis only local “effects” identified.                           |
| • Often easily replicated.                                             | • Unclear what economic questions, if any, the parameters estimate.           |
| • Robustness exercises easily conducted.                              | • Role of uncertainty, preferences etc. not clear.                            |
|                                                                        | • Usually **ex-post** analysis.                                               |
|                                                                        | • Does not in general address a variety of economic policy evaluation questions, e.g., *distributions* of outcomes under different policies. |
|                                                                        | • Clear answers to unclear questions.                                         |
• Linearity and absence of essential heterogeneity enable one instrument to identify average return across all schooling levels

• Both essential

• Otherwise, multiple investments required
A Halfway House
• This paper presents a bridge between the structural and the “reduced form” treatment effect literatures.

• It accounts for heterogeneity among persons in the effects of schooling on wages.

• Considers **nonmarket** effects of schooling.

• It allows for comparative advantage, self selection \( \rho_i \) correlated with \( s_i \).

• Consistent with dynamic sequential decision-making.

• Estimates **ex-post** continuation values.

• It addresses who benefits **ex-post** from education.

• It accounts for nonlinearity in effects of education.
• Consistent with rationality but does not impose it.
• Identifies *ex-post* continuation values.
• Estimates the role of cognitive and non-cognitive ability in producing treatment effects.
• Unlike the treatment effect literature, can identify the set of persons affected by the policies and instruments both in terms of observed and unobserved variables.
Figure 1: Sequential Schooling Decisions

- High School Dropout (s=0)
- High School Graduate (s=1)
- Some College (s=3)
- 4-yr College Graduate (s=4)

- GED (s=2)
Notation

- $C_{j,j'}$: choice set available for a person at $j$ choosing between remaining at $j$ or transitioning to $j'$, where $j, j' \in J$
- $D_{j,j'} = 1$ if a person at $j$ chooses $j' \in C_{j,j'}$ at choice node $\{j, j'\}$
- $Q_{j,j'} = 1$ if person gets to decision node $\{j, j'\}$
- $\mathcal{D} = \{D_{j,j'}\}_{j,j' \in J}$ is the set of possible educational decisions taken by an individual over the life cycle
- $S = \{1, \ldots, \bar{s}\}$ denotes set of final schooling states
- $H_s = 1$ if $s$ is the highest schooling level attained
Contributions

- Develop and apply a unified approach to the literature on a variety of causal effects of education for different socioeconomic outcomes.
- Strong evidence of causal effects of education on smoking, depression, health limitations, and wages at all levels of schooling.
- Selection on gains shows up mostly in later educational decisions.
- Linearity in schooling of estimated effects decisively rejected; causal effects vary by level of education.
- Continuation values are important components of causal effects.
- Substantial sorting into schooling both on cognitive and socioemotional skills. These skills have effects beyond their effects on education.
Contributions (Cont’d)

• Ability bias is quantitatively important. Both cognitive and noncognitive skills affect schooling, wages, and health.

• For most outcomes, only high-ability people benefit from graduating from college.

• Dependence among outcomes and choices generating treatment effects arises from unobserved variables that are adequately proxied by our measures of cognitive and noncognitive skills.
A Sequential Model of Educational Attainment
Optimal decision at each decision node

\[
D_{j,j'} = \begin{cases} 
1 & \text{if } I_{j,j'} \geq 0 \\
0 & \text{otherwise}
\end{cases}
\] (3)

where \( I_{j,j'} \) is the perceived value (by the agent) of attaining schooling level \( j' \) for a person in educational state \( j \).

- In principle, agents can make irrational choices or their educational choices could be governed by behavioral anomalies.
Approximate $I_{j,j'}$ using a separable model:

$$I_{j,j'} = \phi_{j,j'}(X_{j,j'}) + \theta' \alpha^j_{j,j'} - \nu_{j,j'}$$

- $\nu_{j,j'} \perp \perp (X_{j,j'}, \theta)$
- Can use general functional forms and be nonparametric as long as separability maintained (see, e.g., Matzkin, 1992, 1993)
- $\theta$ is a vector of cognitive and non-cognitive skills
- Approximates **Decision Rule** (2):
  Continue schooling at age $a$ if

$$V_1(S(a), a) \geq V_2(S(a), a)$$

school next best alternative
Labor Market and Health Outcomes
• $Y^k_s$: the outcome $k \in \{1, \ldots, K\}$ associated with final schooling level $s \in S$:

$$Y^k_s = \tau^k_s(X^k_s) + \theta' \alpha^k_s + \nu^k_s.$$

• $\nu^k_s \perp \perp (X^k_s, \theta)$.

• $H_s = 1$ if $s$ is the highest attained schooling level and 0 otherwise.

$$Y^k = \sum_{s \in S} H_s Y^k_s.$$
Measurement System for Unobserved Cognitive and Socioemotional Endowments
Measurement systems can be fully nonparametric: Cunha et al. (2010a)

\[ T_{s,l}^{C} = \phi_{s,l}^{C}(X_{s,l}^{C}, \theta_{s,l}^{C}, e_{s,l}^{C}) \]  \hspace{1cm} (5)

\[ T_{s,l}^{SE} = \phi_{s,l}^{SE}(X_{s,l}^{SE}, \theta_{s,l}^{SE}, e_{s,l}^{SE}) \]  \hspace{1cm} (6)

\[ T_{s,l}^{C,SE} = \phi_{s,l}^{C,SE}(X_{s,l}^{C,SE}, \theta_{s,l}^{C}, \theta_{s,l}^{SE}, e_{s,l}^{C,SE}) \]  \hspace{1cm} (7)
Identification of Factor Systems for Linear Systems

- For *linear* systems, need a *single dedicated measurement* for only *one* factor
- Factors can be correlated
- All other measurements can load on both factors, and model remains identified (Williams, 2011)
- Need to normalize each factor
- Need at least five measures in total
- Can estimate causal effects of schooling without determining the measurement functions (5), (6) and (7)
Sources of Identification
Conditional Independence and Matching

- Conditional on $\theta, X$, all outcomes and choices independent.
- If $\theta$ is measured with error but all components are proxied, and the identifying restrictions for the factor models are satisfied, we can identify causal effects.
- Model is identified without making any distributional assumptions on the unobservables (see, e.g., Cunha et al., 2010b).
- A generalization of the “random effects” model widely used in the empirical literature.
Matching on Mismeasured Variables
(Carneiro et al., 2003)

\[(Y_1, \ldots Y_s) \perp \perp S | X, \theta\]

- $\theta$ is measured with error but proxies for it are available from estimated measurement system.
- Correct for measurement error using factor models.
• We do not rely exclusively on matching (or conditional independence) on unobservables for identification.
Instrumental Variables

• Have access to transition-specific instruments.

• Variables in $X_{s,s'}$, not in $X_{s'}^k$, assumed to be independent of the model unobservables.

• Under support conditions on the instruments, the model is identified without invoking any distributional assumptions on the unobservables or invoking a factor structure.

• Approximate the distribution of unobservables using mixtures of normal sieve estimators (see Chen, 2007).
Identifying Treatment Effects Using Matching on Unobservables

- Need only to identify the *Span* of $\theta$
- No need to solve the “rotation problem” in factor analysis
- Do not need to identify coefficients of individual factor equations or pick particular normalizations provided that transformations are measure-preserving
Traditional Treatment Effects: Differences Across Final Schooling Levels
- $Y_{s'}$: Outcome at final schooling level $s'$. $Y_s$: Outcome at final schooling level $s$.

- **Average treatment effect** of $s'$ compared to $s$ for the entire population:

  $$\Delta^{ATE}_{s',s} \equiv \int \int E(Y_{s'} - Y_s | X = x, \theta = \bar{\theta}) \, dF_{X,\theta}(x, \bar{\theta}) (8)$$

- For the subset of the population that stops at one of the two schooling levels $s$, $s'$:

  $$\Delta^{ATE*}_{s',s} \equiv \int \int E(Y_{s'} - Y_s | X = x, \theta = \bar{\theta}) \, dF_{X,\theta}(x, \bar{\theta}) | H_s + H_{s'} = 1) (9)$$
Dynamic Treatment Effects
Person-specific treatment effect for an individual making a decision at node \((j, j')\) deciding between going on to \(j'\) or stopping at \(j\):

\[
TE_{j,j'}[Y|X = x, \theta = \bar{\theta}] \equiv E(Y|X = x, \theta = \bar{\theta}, Fix D_{j,j'} = 1, Q_{j,j'} = 1) - E(Y|X = x, \theta = \bar{\theta}, Fix D_{j,j'} = 0, Q_{j,j'} = 1).
\]

Condition on \(X, \theta\) at transition \((j, j')\).
For the population that makes it to that transition:

\[
ATE_{j,j'} \equiv \int \int TE_{j,j'}[Y|X = x, \theta = \bar{\theta}] \, dF_{X,\theta}(x, \bar{\theta} | Q_{j,j'} = 1)
\]  

(10)

- Conditioning on \(Q_{j,j'} = 1\) keeps estimates of treatment effects within sample
Decomposing $TE_{j,j+1}$

The effect of fixing $D_{j,j+1} = 1$ (Fix $D_{j,j+1} = 1$) conditional on $Q_{j,j+1} = 1$. Can be broken up into two parts:

(a) The **Direct Effect** of going from $j$ to $j + 1$: 
   
   $DE_{j,j+1} = Y_{j+1} - Y_j$.

(b) The **Continuation Value** of going beyond $j + 1$: $CON_{j+1}$.

The total effect of fixing $D_{j,j+1}$: $TE_{j,j+1}$

\[
TE_{j,j+1} = DE_{j,j+1} + CON_{j+1} \tag{11}
\]
The Average Marginal Treatment Effect

- Average effect of choosing an additional level of schooling for individuals who are at the margin of indifference between the two levels of schooling, \( j \) and \( j' \):

\[
AMTE_{j,j'} \equiv \int\int TE_{j,j'}[Y|X = x, \theta = \bar{\theta}] \, dF_{x,\theta}(x, \bar{\theta} \mid Q_{j,j'} = 1, |I_{j,j'}| \leq \varepsilon)
\]  

(12)

where \( \varepsilon \) is an arbitrarily small neighborhood around the margin of indifference.

- Many analysts *erroneously* think LATE in general answers this question
• **Policy relevant treatment effect (PRTE):** The average treatment effect for those induced to change educational choices in response to a particular policy intervention.

• Two policies: $p$ and $p'$.

• $Y^p$: the aggregate outcome under policy $p$.

• $S(p)$ is the schooling selected under policy $p$.

• The policy relevant treatment effect from implementing policy $p'$ compared to policy $p$ is:

\[
PRTE_{p', p} \equiv \int \int E(Y^{p'} - Y^p | X = X, \theta = \bar{\theta}) \cdot dF_{X, \theta}(X, \bar{\theta}|\{S(p) \neq S(p')\})
\]
• \{S(p') \neq S(p)\} denotes the set of individuals with their associated \(\theta, X\) values for whom attained schooling levels differ under the two policies (“changers”).

• Changers can come from multiple margins.

• This is a structural version analogous to \textit{LATE}.

• \textit{LATE} answers this question only if the policy is the instrument.

• \textit{PRTE} is more general: no “
\textit{monotonicity assumption}” is required.

• Individuals come from multiple states and go to multiple states.
Data

- The National Longitudinal Survey of Youth (NLSY79).
- Representative sample of men and women born between 1957 and 1964.
- Core sample of males, which after dropping oversample and individuals missing key covariates, contains 2242 individuals.
- We pool across races (effects for whites alone very similar).
<table>
<thead>
<tr>
<th>Outcome</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Economic:</strong></td>
<td></td>
</tr>
<tr>
<td>Log Wages</td>
<td>Log Hourly Wage at Age 30</td>
</tr>
<tr>
<td>Welfare Use</td>
<td>Any Welfare from 1990-2010</td>
</tr>
<tr>
<td>Present Value of Wages</td>
<td>Log Present Value of Wage Income from Age 18 to 40 Discounted at 5%</td>
</tr>
<tr>
<td><strong>Physical Health:</strong></td>
<td></td>
</tr>
<tr>
<td>Smoking</td>
<td>Regular Smoker at Age 30</td>
</tr>
<tr>
<td>Health Constraints</td>
<td>Health has ever limited type or amount of work since age 30</td>
</tr>
<tr>
<td>Physical Health</td>
<td>Self Reported Physical Health by SF-12 at Age 40</td>
</tr>
<tr>
<td><strong>Mental Health:</strong></td>
<td></td>
</tr>
<tr>
<td>Depression</td>
<td>Self Reported by CES-D at Age 40</td>
</tr>
<tr>
<td>Self-Esteem</td>
<td>Rosenberg Scale in 2006 when individuals are in their 40s</td>
</tr>
<tr>
<td><strong>Social:</strong></td>
<td></td>
</tr>
<tr>
<td>Incarceration</td>
<td>The respondent was ever incarcerated at time of interview (1990-2010)</td>
</tr>
<tr>
<td>Trust</td>
<td>Usually or Always Trusting People in 2008</td>
</tr>
<tr>
<td>Voted</td>
<td>Voted in 2006</td>
</tr>
<tr>
<td>Divorce</td>
<td>Ever divorced or separated once married (1990-2010)</td>
</tr>
</tbody>
</table>
Joint Distributions of Endowments

Overall Correlation: 0.24

Distribution of Factors

Cognitive Factor

Social-Emotional Factor
Effects of Cognitive and Noncognitive Factors $\theta$ on Educational Decisions
Figure 2: The Probability of Educational Decisions, by Endowment Levels

Dropping from HS vs. Graduating from HS

- Probability
  - Decile of Cognitive
  - Decile of Socio-Emotional
  - Fraction

Heckman, Humphries, & Veramendi
Market and Nonmarket Benefits
Figure 2: The Probability of Educational Decisions, by Endowment Levels

HS Graduate vs. College Enrollment
Figure 2: The Probability of Educational Decisions, by Endowment Levels

Some College vs. 4-year college degree
Strong Effects of Cognitive and Socioemotional Endowments on ALL Outcomes Even After Fixing Schooling
Causal Effects
Figure 3:  **Causal Versus Observed Differences** by Indicated Schooling Level **Compared to Next Lowest**

Decomposition of Schooling Effects  
Log Wages

<table>
<thead>
<tr>
<th>Margin</th>
<th>Average TE</th>
<th>Observed</th>
<th>Causal Component</th>
</tr>
</thead>
<tbody>
<tr>
<td>High School</td>
<td>0.15</td>
<td>![Observed](p &lt; 0.05)</td>
<td>![Causal Component](p &lt; 0.01)</td>
</tr>
<tr>
<td>Some College</td>
<td>0.10</td>
<td>![Observed](p &lt; 0.05)</td>
<td>![Causal Component](p &lt; 0.01)</td>
</tr>
<tr>
<td>College</td>
<td>0.25</td>
<td>![Observed](p &lt; 0.05)</td>
<td>![Causal Component](p &lt; 0.01)</td>
</tr>
</tbody>
</table>
Figure 4: **Causal Versus Observed Differences** by Indicated Schooling Level Compared to Next Lowest

Decomposition of Schooling Effects

Log PV Wages

Heckman, Humphries, & Veramendi

Market and Nonmarket Benefits
Figure 5: **Causal Versus Observed Differences** by Indicated Schooling Level **Compared to Next Lowest**

Decomposition of Schooling Effects
Health Limits Work

- High School
- Some College
- College

Average TE

Margin

- Observed
- Causal Component

* p < 0.05
* p < 0.01

Heckman, Humphries, & Veramendi

Market and Nonmarket Benefits
Figure 6: **Causal Versus Observed Differences** by Indicated Schooling Level **Compared to Next Lowest**
Treatment Effects by Decision Node
Figure 7: Treatment Effects from Fixing $D_{j,j+1} = 1$ versus Fixing $D_{j,j+1} = 0$ conditional on $Q_{j,j+1} = 1$ where $j + 1$ is the indicated decision node (i.e., comparing those who take transition with those who do not). Transitions reported are at $j + 1$.

Treatment Effects: Log Wages

<table>
<thead>
<tr>
<th>Decision Node</th>
<th>AMTE</th>
<th>ATE (low)</th>
<th>ATE (high)</th>
<th>ATE (high) p &lt; 0.05</th>
<th>ATE (low) p &lt; 0.01</th>
</tr>
</thead>
<tbody>
<tr>
<td>Graduate HS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Enroll in Coll.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Graduate Coll.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure 8: Treatment Effects from $\text{Fixing } D_{j,j+1} = 1$ versus \text{Fixing } $D_{j,j+1} = 0$ conditional on $Q_{j,j+1} = 1$ where $j + 1$ is the indicated decision node (i.e., comparing those who take transition with those who do not).

![Graph showing treatment effects for different decision nodes and outcomes.](image-url)
Figure 9: Treatment Effects from $\text{Fixing } D_{j,j+1} = 1$ versus $\text{Fixing } D_{j,j+1} = 0$ conditional on $Q_{j,j+1} = 1$ where $j + 1$ is the indicated decision node (i.e., comparing those who take transition with those who do not)
Figure 10: Treatment Effects from **Fixing** $D_{j,j+1} = 1$ versus **Fixing** $D_{j,j+1} = 0$ conditional on $Q_{j,j+1} = 1$ where $j + 1$ is the indicated decision node (i.e., comparing those who take transition with those who do not).
Treatment Effects by Quantile of Cognitive and Noncognitive Skills
Figure 10: Transition-Specific Average Treatment Effect of Graduating from a Four-Year College by Outcome

PV Wage Income
Ex-Post Gross Continuation Values

- Indicate that IRR underestimates returns
Figure 11: Treatment Effects from **Fixing** $D_{j,j+1} = 1$ versus **Fixing** $D_{j,j+1} = 0$: Ratio of Gross Value of Continuing to Total Treatment Effect by Decision Node for $Q_{j,j+1} = 1$

![Figure 11](image_url)

Total Effect and Direct Effect:
Log PV Wages

- AMTE
- ATE
- ATE (low)
- ATE (high)

**Average TE**
Graduate HS Enroll in College

- Direct Effect $p < 0.05$
- $p < 0.01$
- $p < 0.05$ for $CV=0$
Distributions of Treatment Effects
Figure 12: Distributions of Expected Treatment Effects

Distribution of Expected TEs
Coll. Grad. on PV–Wage

Notes: Results show the distribution of expected treatment effects conditional on reaching the educational choice. The results do not include the idiosyncratic shocks. TE shows the distribution treatment effects for those with $Q_{j,j'} = 1$, TT shows the distribution of treatment effects for those with $D_{j,j'} = 1$, TUT shows the distribution of treatment effects for those with $D_{j,j'} = 0$. The “AT” line shows the average treatment effect, the “ATT” line shows the average treatment on the treated, and the “ATUT” line shows the average treatment on the untreated.
Linearity in Schooling?

- Is Mincer linearity justified?
Figure 13: The Observed Benefits from Education after Controlling for Background, Ability and Highest Grade Completed
Figure 13: The Observed Benefits from Education after Controlling for Background and Ability
ATE Over Dropouts: Log Wages

Heckman, Humphries, & Veramendi

Market and Nonmarket Benefits
Average TE

ATE (over dropouts) +/−1 Std Error

ATE Over Dropouts: Log PV Wages

Heckman, Humphries, & Veramendi

Market and Nonmarket Benefits
Robustness

• Estimated *treatment effects* are robust to adding a third factor, with no associated measurements.
Comparison of Structural Estimates of ATEs with those Obtained from Matching and IV

- Close approximation with matching
- IV: imprecise estimates
Table 4: Summary of Average Treatment Effects - High School Graduation (Comparing Fix $D_{0,1} = 1$ to Fix $D_{0,1} = 0$)

<table>
<thead>
<tr>
<th></th>
<th>(a) Propensity Score</th>
<th>(b) Nearest Neighbor</th>
<th>Structural Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wages</td>
<td>0.107</td>
<td>0.105</td>
<td>0.10</td>
</tr>
<tr>
<td>SE</td>
<td>0.038</td>
<td>0.035</td>
<td>0.05</td>
</tr>
<tr>
<td>Prison</td>
<td>-0.052</td>
<td>-0.089</td>
<td>-0.07</td>
</tr>
<tr>
<td>SE</td>
<td>0.028</td>
<td>0.027</td>
<td>0.02</td>
</tr>
<tr>
<td>Smoking</td>
<td>-0.197</td>
<td>-0.165</td>
<td>-0.25</td>
</tr>
<tr>
<td>SE</td>
<td>0.045</td>
<td>0.054</td>
<td>0.05</td>
</tr>
<tr>
<td>Health Limits Work</td>
<td>-0.131</td>
<td>-0.076</td>
<td>-0.10</td>
</tr>
<tr>
<td>SE</td>
<td>0.039</td>
<td>0.053</td>
<td>0.04</td>
</tr>
</tbody>
</table>

Model factors come from our estimated model controlling for measurement error. Simple factors come from principal components.
Policy Relevant Treatment Effects
Consider a 2 year tuition subsidy of $7,500.

Policy induces 13% of individuals who graduate from high school but would not enroll in college to enroll in college.
<table>
<thead>
<tr>
<th></th>
<th>PRTE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log Wages</td>
<td>0.13</td>
</tr>
<tr>
<td>PV Log Wages</td>
<td>0.12</td>
</tr>
<tr>
<td>Health Limits Work</td>
<td>-0.04</td>
</tr>
<tr>
<td>Smoking</td>
<td>-0.13</td>
</tr>
</tbody>
</table>
Figure 14: PRTE: Who is Induced to Enroll in College?

Notes: The figure plots the proportion of individuals induced to switch from the policy that are in each decile of $\epsilon_{1,3}$, where $\epsilon_{1,3} = \theta \alpha_{1,3} - \nu_{1,3}$. $\epsilon_{1,3}$ is the unobserved component of the educational choice model. 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.
Simulated Policy Experiment

Evaluating Impacts of Early Childhood Programs Shown to Boost Cognitive and Socioemotional Skills
1. Increase the cognitive skill of those in the lowest decile
2. Increase the non-cognitive skill of those in the lowest decile
   ◦ Such effects are found in early childhood programs
This plot shows the average gains for those in the bottom deciles of cognitive ability (left) and socioemotional ability (right), from an increase in skill.
This plot shows the average gains for those in the bottom deciles of cognitive ability (left) and socioemotional ability (right), from an increase in skill.
Summary
• Develop and apply a unified approach to the literature on a variety of causal effects of education for different socioeconomic outcomes.

• Strong evidence of causal effects of education on smoking, depression, health limitations, and wages at all levels of schooling, as well as trust, etc.

• Causal effects vary by level of education.

• Selection on gains ($\rho$ correlated with $s$) shows up mostly in later educational decisions.

• Linearity in schooling of estimated effects decisively rejected.
• Continuation values are important components of causal effects of schooling on wages, present value of wages, and many other outcomes. Traditional IRR underestimated for those who take schooling.

• Substantial sorting into schooling both on cognitive and socioemotional skills.

• Strong effects of cognitive and noncognitive skills on outcomes even after *fixing* education.
• Ability bias is quantitatively important. Both cognitive and non-cognitive skills affect schooling, wages, and health.

• For most outcomes, only high-ability people benefit from graduating from college.

• Measurement error in skills is empirically important.

• Universal college education is a bad idea. Low ability people do not benefit.