Unobserved Heterogeneity, Self-selection, and Program Evaluation

Advanced Microeconometrics

Instructor: Cristián Sánchez

Spring 2022

Outline

1. Introduction and Motivation
2. Heckman, Stixrud, and Urzua (2006). *The Effect of Cognitive and Noncognitive Skills on Labor Market Outcomes and Human Behavior*
3. Rau, Sánchez, and Urzua (2020). *The Schooling and Labor Market Effects of Vouchers*

Introduction

- Empirical economic models are commonly constructed under the assumption that the econometrician has information on all relevant variables determining outcome(s).

- Example
$$\ln Y_j = \alpha + \beta S_j + \gamma X_j + U_j \text{ for } j = 1, \ldots, N$$
where Y_j ($=$ wage), S_j ($=$ years of schooling), X_j ($=$ additional controls) represent the observables, and U_j is the residual or unobservable.

- The assumption $(S, X) \perp\!\!\!\perp U$ means that there are no additional determinants of Y left out.

Introduction

- But, what if U in fact *contains* information known by the individuals but unavailable to the econometrician and this information was used by the agent when deciding his schooling level?
- In this case, U and S would be correlated (even after controlling for X) and the regression model would suffer endogeneity problems.
- Natural candidates for unobserved components: Cognitive and Noncognitive abilities.

Roy Model Framework

- We can use the Roy Model framework as a departure point.
- The Roy Model provides a simple economic framework to analyze the effects of unobserved abilities on outcomes.
- The basic idea:
  1. Individuals select between two alternatives based on the associated benefits (e.g. receiving or not a vaccination, continuing or not to the next schooling level, and in general, receiving or not treatment versus control).
  2. Individuals make decisions using information which is not necessarily available for the econometrician (unobserved abilities).

Roy Model Framework

- Two potential outcomes

$$\begin{aligned} Y_1 &= \beta_1 X + U_1 \qquad \text{(Treatment)} \\ Y_0 &= \beta_0 X + U_0 \qquad \text{(Control)} \end{aligned}$$

- Decision rule

$$D^* = \mathbf{1}(D^* \geq 0)$$

- Specifically, let

$$D^* = Z_\gamma - V$$

where Z is a vector of variables (instruments).

Endogeneity

- Notice that
$$(U_1, U_0) \text{ are not necessarily independent of } V$$

- Consider for example the Generalized Roy Model:
$$D = 1 \Leftrightarrow Y_1 - Y_0 - C(Q) \geq 0 \quad (\Leftrightarrow Z_\gamma - V \geq 0)$$
where $C(Q)$ represents costs.

- Let $C(Q) = \delta Q + \varepsilon$, thus
$$\begin{aligned} Z_\gamma &= \beta_1 X - \beta_0 X - \delta Q \\ V &= U_1 - U_0 - \varepsilon \end{aligned}$$
and (U_1, U_0) and V are correlated by construction.

Unobserved Abilities

- More generally, suppose that U_1, U_0, V contain information on unobserved abilities. This would naturally produce endogeneity.
- Let θ represent unobserved abilities (cognitive and/or noncognotive)
- θ affect potential outcomes and decision rules:

$$Y_1 = \beta_1 X + U_1 = \beta_1 X + \underbrace{\alpha_1 \theta + \varepsilon_1}_{=U_1}$$

$$Y_0 = \beta_0 X + U_0 = \beta_0 X + \underbrace{\alpha_0 \theta + \varepsilon_0}_{=U_0}$$

$$D^* = Z_\gamma - V = Z_\gamma - \underbrace{(\alpha_V \theta + \varepsilon_v)}_{=V}$$

where $\varepsilon_V \perp\!\!\!\perp \varepsilon_1 \perp\!\!\!\perp \varepsilon_0$, and $\theta \perp\!\!\!\perp (\varepsilon_V, \varepsilon_1, \varepsilon_0)$

Linking Unobserved Abilities to Test Scores

Let T denote a test score (e.g. math score, IQ test)

$$T = \beta_T X_T + \underbrace{\alpha_T \theta + \varepsilon_T}_{=U_T}$$

where $(X_T, \varepsilon_T) \perp\!\!\!\perp \theta$.

- Thus, we can link unobserved abilities to ability measures.
- The set up recognizes that T is not a direct measure of θ.
- Furthermore, we can include the effect of other observables (X_T) (mother's and father's education for example).

Likelihood Function

- We observe Y_j, T_j, D_j for $j = 1, ..., N$, with

$$Y_j = D_j Y_{1,j} + (1 - D_j) Y_{0,j}$$

- Key Insight: Conditional on unobserved abilities, U_1, U_0, V and U_T are mutually independent. Thus,

$$\prod_{j=1}^{N} f(Y_j, T_j, D_j \mid X, X_T, Z) = \prod_{j=1}^{N} \int f(Y_j, T_j, D_j \mid X, X_T, Z, \theta)\, d\theta$$

  where we can write

$$f(Y_j, T_j, D_j \mid X, X_T, Z, \theta) = f(Y_j, D_j \mid X, Z, \theta) f(T_j \mid X_T, \theta)$$

- Notice that:

$$f(Y_j, D_j \mid X, Z, \theta) = \left[f(Y_{1,j} \mid X, \theta)\Pr(D_j = 1 \mid Z, \theta)\right]^{D_j}$$
$$\left[f(Y_{0,j} \mid X, \theta)\Pr(D_j = 0 \mid Z, \theta)\right]^{1-D_j}$$

where

$$f(Y_{1,j} \mid X, \theta) = f_{\varepsilon_1}(Y_{1,j} - \beta_1 X_j - \alpha_1\theta)$$
$$f(Y_{0,j} \mid X, \theta) = f_{\varepsilon_0}(Y_{0,j} - \beta_0 X_j - \alpha_0\theta)$$

and

$$\Pr(D_j \mid Z, \theta) = \Pr(Z_j\gamma - \alpha_V\theta - \varepsilon_V \geq 0 \mid Z, \theta)$$
$$= F_{\varepsilon_V}(Z_j\gamma - \alpha_V\theta)$$

- Finally,

$$f(T_j \mid X_T, \theta) = f_{\varepsilon_T}(T_j - \beta_T X_T - \alpha_T\theta)$$

Normal Case

- Assume

$$\varepsilon_1 \sim N(0,\sigma_1^2), \quad \varepsilon_0 \sim N(0,\sigma_0^2), \quad \varepsilon_V \sim N(0,\sigma_V^2), \quad \varepsilon_T \sim N(0,\sigma_T^2).$$

- Thus, the likelihood function is given by:

$$\Gamma(Y,T,D \mid X, X_T, Z) =$$

$$\prod_{j=1}^{N} \int \left[\frac{1}{\sqrt{2\pi}\sigma_T} \exp\left\{ -\frac{1}{2\sigma_T^2} (T_j - \beta_T X_{j,T} - \alpha_T \theta)^2 \right\} \right]$$

$$\left[\frac{1}{\sqrt{2\pi}\sigma_1} \exp\left\{ -\frac{1}{2\sigma_1^2} (Y_{1,j} - \beta_1 X_j - \alpha_1 \theta)^2 \right\} \Phi\left(\frac{Z_j\gamma - \alpha_V \theta}{\sigma_V^2} \right) \right]^{D_j}$$

$$\left[\frac{1}{\sqrt{2\pi}\sigma_0} \exp\left\{ -\frac{1}{2\sigma_0^2} (Y_{0,j} - \beta_0 X_j - \alpha_0 \theta)^2 \right\} \left(1 - \Phi\left(\frac{Z_j\gamma - \alpha_V \theta}{\sigma_V^2} \right) \right) \right]^{1-D_j} dF(\theta)$$

Estimation of the Model

- Maximum Likelihood:
  - Advantage: Widely used and any software can be used (including STATA-MATA)
  - Disadvantage: We need to deal with numerical integration
- Markov Chain Monte Carlo
  - Advantage: We avoid the computation of the integral
  - Disadvantage: Bayesian methods have just recently become empirically appealing.

Identification of the Model

- The identification of the model relies on the information available (data).
- Let assume that at least three test scores (T_1, T_2, T_3) are available in the data.
- For simplicity, suppose that θ is a scalar and omit X from the discussion.
- Thus,

$$\begin{aligned} T_1 &= \alpha_{T_1}\theta + \varepsilon_{T_1} \\ T_2 &= \alpha_{T_2}\theta + \varepsilon_{T_2} \\ T_3 &= \alpha_{T_3}\theta + \varepsilon_{T_3} \end{aligned}$$

- We can compute

$$
\begin{aligned}
Cov(T_1, T_2) &= \alpha_{T_1}\alpha_{T_2}\sigma_\theta^2 \\
Cov(T_1, T_3) &= \alpha_{T_1}\alpha_{T_3}\sigma_\theta^2 \\
Cov(T_2, T_3) &= \alpha_{T_2}\alpha_{T_3}\sigma_\theta^2
\end{aligned}
$$

- Since we observe the left hand side, we can form

$$
\begin{aligned}
\frac{Cov(T_1, T_2)}{Cov(T_2, T_3)} &= \frac{\alpha_{T_1}}{\alpha_{T_3}} \\
\frac{Cov(T_1, T_2)}{Cov(T_1, T_3)} &= \frac{\alpha_{T_2}}{\alpha_{T_3}}
\end{aligned}
$$

- By normalizing $\alpha_{T_3} = 1$, we get α_{T_1} and α_{T_2}

- Finally, we can rewrite the system as:

$$\frac{T_1}{\alpha_{T_1}} = \theta + \frac{\varepsilon_{T_1}}{\alpha_{T_1}} = \theta + \varepsilon'_{T_1}$$

$$\frac{T_2}{\alpha_{T_2}} = \theta + \frac{\varepsilon_{T_2}}{\alpha_{T_2}} = \theta + \varepsilon'_{T_2}$$

  and we can apply Kotlarski's Theorem (Kotlarski, 1967) to identify

$$f_{\varepsilon_{T_1}}(\cdot), f_{\varepsilon_{T_2}}(\cdot), f_\theta(\cdot)$$

- We can identify the whole model applying this logic.

The Effects of Cognitive and Noncognitive Abilities on Labor Outcomes and Social Behavior

by Heckman, Stixrud and Urzua, 2006

Introduction

- Although the importance of cognitive skills for success in a variety of dimensions of social and economic life is well established, the importance of noncognitive skills has largely been overlooked.

- Because cognitive and non-cognitive abilities are shaped early in the lifecycle, differences in these abilities are persistent, and both are crucial to the social and economic success of an individual, gaps among income and racial groups begin early and persist.

- Jencks (1979), Osborne (2000), Bowles, Gintis, and Osborne (2001), Heckman and Rubinstein (2001), among others.

- **Early interventions**, such as enriched childcare centers coupled with home visitations, have been successful in alleviating some of the initial disadvantages of children born into adverse conditions. The success of these interventions has primarily been due not to their success in improving the cognitive skills (IQ) of these children, but rather to their success in boosting non-cognitive skills and increasing child motivation (Cunha, Heckman, Lochner and Masterov, 2005).

Our Framerwork

- We posit the existence of two underlying latent abilities:

f^C	Cognitive
f^N	Noncognitive

- We take f^C and f^N as initial conditions.

- We assume that levels of both abilities are known by each individual but not by the researcher.

- We focus the analysis on the impact of these abilities on the following outcomes:
  - Schooling attainment
  - Wages given schooling and overall
  - Work experience
  - Occupational choice
  - Social behaviors and risky correlated behaviors:
    - Crime and incarceration
    - Teenage pregnancy
    - Drug use
    - Smoking

Data

- The National Longitudinal Survey of Youth (NLSY79) is a representative sample of young Americans between the ages of 14 and 21 at the time of the first interview in 1979. We use the random sample of 6111 noninstitutionalized civilian youths.

- The NLSY collects information on parental background, schooling decisions, labor market experiences, cognitive and noncognitive test scores and other behavioral measures of these individuals on an annual basis.

- Cognitive measures: arithmetic reasoning, word knowledge, paragraph comprehension, numerical operations, math knowledge and coding speed. Noncognitive measures: Rotter Locus of control and Rosemberg self-esteem scales.

Table 1. Rotter Internal-External Locus of Control Scale

Question 1 (Rotter 1)

(a) What happens to me is my own doing.

(b) Sometimes I feel that I don't have enough control over the direction my life is taking.

Question 2 (Rotter 2)

When I make plans,

(a) I am almost certain that I can make them work.

(b) It is not always wise to plan too far ahead, because many things turn out to be a matter of good or bad fortune anyhow.

Question 3 (Rotter 3)

(a) Getting what I want has little or nothing to do with luck.

(b) Many times we might just as well decide what to do by flipping a coin

Question 4 (Rotter 4)

(a) Many times I feel that I have little influence over the things that happen to me.

(b) It is impossible for me to believe that chance or luck plays an important role in my life.

Table 2. Rosenberg Self-Esteem Scale

Question 1

I feel that I'm a person of worth, at least on an equal basis with others.

Question 2

I feel that I have a number of good qualities.

Question 3

All in all, I am inclined to feel that I am a failure.

Question 4

I am able to do things as well as most other people.

Question 5

I feel I do not have much to be proud of.

Question 6

I take a positive attitude toward myself.

Question 7

On the whole, I am satisfied with myself.

Question 8

I wish I could have more respect for myself.

Question 9

I certainly feel useless at times.

Question 10

At times I think I am no good at all.

Traditional Empirical Approach

Table 3. Estimated Coefficients from Log Wage Regressions

NLSY79 - Males and Females at Age 30[a]

Variables[b]	Males (A)	Males (B)	Females (A)	Females (B)
GED	0.017		-0.002	
	(0.048)		(0.056)	
High School Graduate	0.087		0.059	
	(0.035)		(0.044)	
Some College	0.146		0.117	
	(0.044)		(0.052)	
2yr College Graduate	0.215		0.233	
	(0.058)		(0.058)	
4yr College Graduate	0.292		0.354	
	(0.046)		(0.054)	
Cognitive	0.121	0.190	0.169	0.251
	(0.016)	(0.013)	(0.017)	(0.014)
Non-Cognitive	0.042	0.052	0.028	0.041
	(0.011)	(0.012)	(0.013)	(0.013)
Constant	2.558	2.690	2.178	2.288
	(0.057)	(0.050)	(0.063)	(0.052)

Problems with This Approach and Our Solution

- Naive regressions of earnings on test scores (cognitive and noncognitive) are problematic.
- Problem is reverse causality: Schooling may cause both earnings and test scores.
- The recent literature notes that schooling and age may influence cognitive measures and corrects for the effect of schooling on ability (Hansen, Heckman and Mullen, 2004).

The Empirical Model

- We posit the existence of two factors: f^C and f^N.
- The levels of an individual's factors may result from some combination of inherited ability, the quality of the environment provided by his parents, early effort on his part, and the effects of any early interventions.
- Our sample starts at age 14 so we cannot investigate the effects of early environments in this study. We take f^C and f^N as initial conditions.

- We assume that levels of both abilities are known by each individual but not by the researcher.
- For convenience we assume that latent abilities are mutually independent, and both determine the individual's wage, schooling decision and behavioral outcomes.
- This does not mean that the manifest abilities are independent.

The Model of Schooling and Wages

- Hedonic Model of (log) wages:

$$\ln Y_s = \beta_{Y,s} X_Y + \alpha^C_{Y,s} f^C + \alpha^N_{Y,s} f^N + e_{Y,s}$$

for $s = 1, ..., \bar{S}$, where $e_{Y,s} \perp\!\!\!\perp (f^N, f^C, X_Y)$.

- Schooling choice:

$$s^* = \underset{s=\{1,\ldots,\bar{S}\}}{\text{argmax}} \{I_1, ..., I_{\bar{S}}\}$$

where

$$I_s = \beta_s X_s + \alpha^C_s f_C + \alpha^N_s f^N + e_s \quad \text{for} \quad s = 1, ..., \bar{S}$$

is a reduced form net utility, where $e_s \perp\!\!\!\perp (f^C, f^N, X_s)$.

Additional Labor and Behavioral Outcomes

- We use binary models for occupational choice, crime and incarceration, drug use and smoking:
$$D_{B_k} = 1[I_{B_k} > 0]$$
where $I_{B_k} = \beta_{B_k} X_{B_k} + \alpha^C_{B_k} f^C + \alpha^N_{B_k} f^N + e_{B_k}$ and $e_{B_k} \perp\!\!\!\perp (f^C, f^N, X_{B_k})$ for $k = 1, ..., \bar{K}$.

- For work experience we use linear models by schooling level:
$$E_s = \beta_{E,s} X_E + \alpha^C_{E,s} f^C + \alpha^N_{E,s} f^N + e_{E,s}$$
for $s = 1, ..., \bar{S}$, where $e_{E,s} \perp\!\!\!\perp (f^C, f^N, X_E)$.

The Measurement System

- The identification strategy assumes the existence of a set of cognitive and noncognitive measures (test scores).
- We address the potential problem of reverse causality between schooling and test scores and schooling and attitude scales.

The basic idea:

- We observe test scores, schooling at time of the test and final schooling level for the same individuals.
- Then, we can compare the results of people with the same final schooling level but with different schooling levels at the time of the tests.

- Our measurement system:

$$C_i(s_T) = \beta_{C_i}(s_T)X_C + \alpha_{C_i}(s_T)f^C + e_{C_i}(s_T)$$

with $i = 1, ..., n_C$, $s_T = 1, ..., \bar{S}_T$.

$$N_i(s_T) = \beta_{N_i}(s_T)X_N + \alpha_{N_i}(s_T)f^N + e_{N_i}(s_T)$$

with $i = 1, ..., n_N$, $s_T = 1, ..., \bar{S}_T$.

- There are no natural units for latent ability. Therefore, for some C_i and N_i we set $\alpha_{C_i} = \alpha_{N_i} = 1$.

- This extends traditional factor analysis by having endogenous loadings $(\alpha_{C_i}(s_T), \alpha_{N_i}(s_T))$

Empirical Implementation and Specification

- We use Bayesian MCMC methods to compute the sample likelihood.
- The use of Bayesian methods is only a computational convenience.
- We use robust mixture of normal approximations to the underlying distributions.
- The evidence argues strongly against normality.

Table 4A. Variables in the empirical implementation of the model Outcome Equations

Variables	Log of Hourly Wage, Employment and Occupational Choice Models	Educational Choice Model (Multinomial Probit): HS Dropouts	GED	HS Graduates	Some College, No Degree	2-yr. degree	4-yr. degree	Behavioral Outcomes (Multinomial Probit) and Fertililty Choice Model
Black (Dummy)	Yes	Yes	Yes	Yes	Yes	Yes	-	Yes
Hispanic (Dummy)	Yes	Yes	Yes	Yes	Yes	Yes	-	Yes
Region of Residence (Dummy Variables)	Yes	-		-	-	-	-	-
Urban Residence (Dummy)	Yes	-		-	-	-	-	-
Local Unemployment Rate at age 30	Yes	-		-	-	-	-	-
Living in a Urban area at age 14 (Dummy)	-	Yes	Yes	Yes	Yes	Yes	-	Yes
Living in the South at age 14 (Dummy)	-	Yes	Yes	Yes	Yes	Yes	-	Yes
Family income in 1979	-	Yes	Yes	Yes	Yes	Yes	-	Yes
Broken home at Age 14 (Dummy)	-	Yes	Yes	Yes	Yes	Yes	-	Yes
Number of Siblings at Age 17 (Dummy)	-	Yes	Yes	Yes	Yes	Yes	-	Yes
Mother Highest Grade Completed at Age 17	-	Yes	Yes	Yes	Yes	Yes	-	Yes
Father Highest Grade Completed at Age 17	-	Yes	Yes	Yes	Yes	Yes	-	Yes
Local Wage of High School Dropouts at Age 17	-	Yes	-	-	-	-	-	-
Local Unemployment Rate of High School Dropouts at Age 17	-	Yes	-	-	-	-	-	-
Local Wage of High School Graduates at Age 17	-	-	-	Yes	-	-	-	-
Local Unemployment Rate of High School Graduates at Age 17	-	-	-	Yes	-	-	-	-
Local Wage of Attendees of Some College at Age 17	-	-	-	-	Yes	-	-	-
Local Unemployment Rate of Attendees of Some College at Age 17	-	-	-	-	Yes	-	-	-
Local Wage for College Graduates at Age 17	-	-	-	-	-	-	Yes	
Local Unemployment for College Graduates at Age 17			-		-	-	Yes	
Tuition at Two Year College at Age 17	-	-	-	-	-	Yes	-	-
Tuition at Four Year College at Age 17		-	-	-	-	-	Yes	-
GED Costs	-	-	Yes	-	-	-	-	-
Cohort Dummies	Yes	Yes	Yes	Yes	Yes	Yes	-	Yes
Factors								
Cognitive Factor	Yes	Yes		Yes	Yes	Yes	-	Yes
Non-cognitive Factor	Yes	Yes		Yes	Yes	Yes	-	Yes

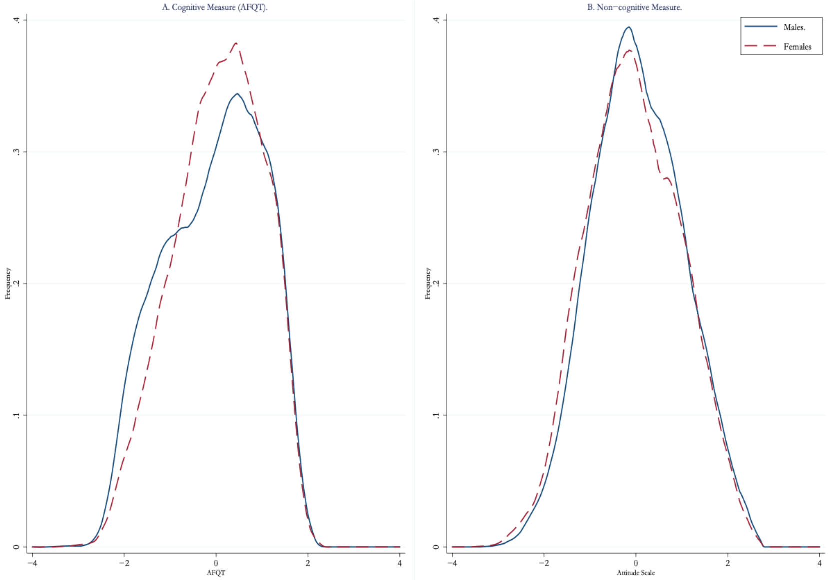
Table 4B. Variables in the empirical implementation of the model
Auxiliary Measures

	Test Scores (Cognitive Measures[(a)])	Attitude Scales (Noncognitive Measures[b])
Black (Dummy)	Yes	Yes
Hispanic (Dummy)	Yes	Yes
Living in a Urban area at age 14 (Dummy)	Yes	Yes
Living in the South at age 14 (Dummy)	Yes	Yes
Mother Highest Grade Completed at Age 17	Yes	Yes
Father Highest Grade Completed at Age 17	Yes	Yes
Number of Siblings at Age 17 (Dummy)	Yes	Yes
Family income in 1979	Yes	Yes
Broken home (Dummy)	Yes	Yes
Cohort Dummies	Yes	Yes
Factors		
Cognitive Factor	Yes	-
Non-cognitive Factor	-	Yes

Empirical Evidence

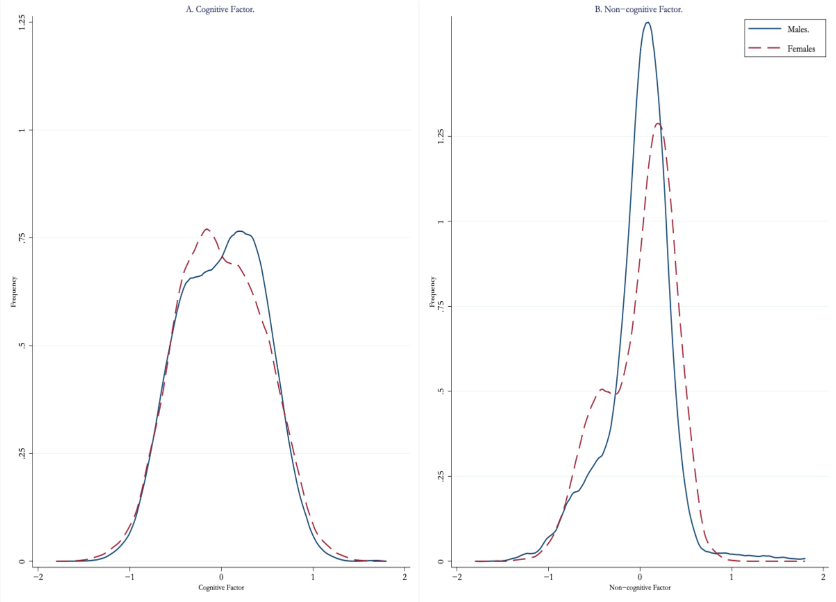
- Distributions of the test scores versus factors.
- The effect of reverse causality on test scores.
- The impact of cognitive and noncognitive abilities on outcomes.

Figure 1A. Distribution of Test Scores by Gender



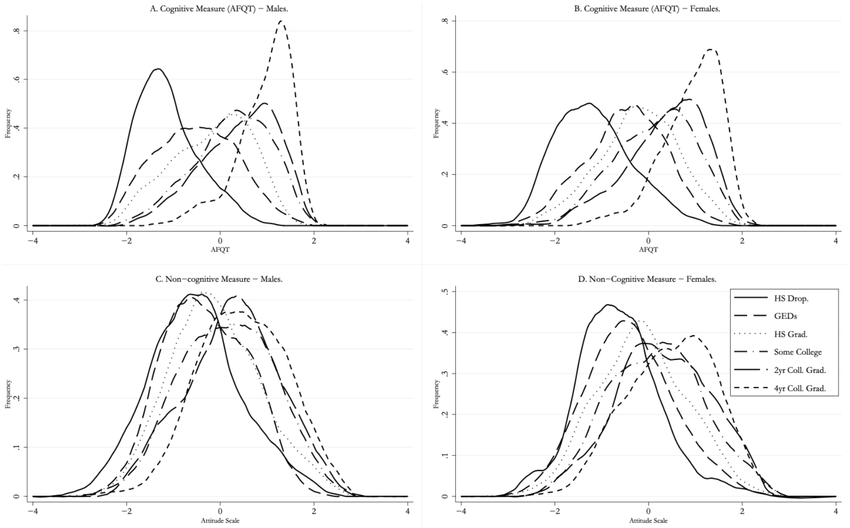
Notes: The AFQT is the mean raw score computed using ASVAB tests.The Attitude Scale is the average raw score between the Rosenberg scale of Self−Steem and the Rotter scale of internal−external locus of control.

Figure 1B. Distribution of Factors by Gender



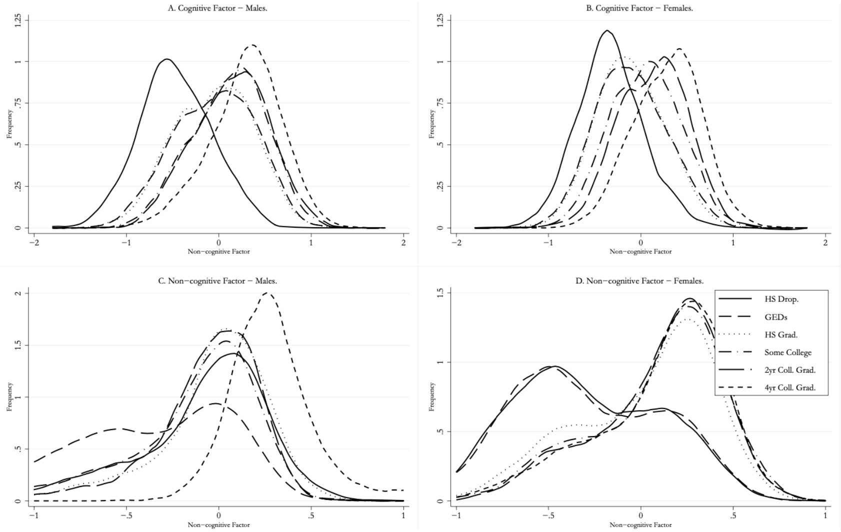
Notes: The factors are simulated from the estimates of the model. The simulated data contain 19,600 observations.

Figure 2A. Distribution of Test Scores by Gender and Schooling Level



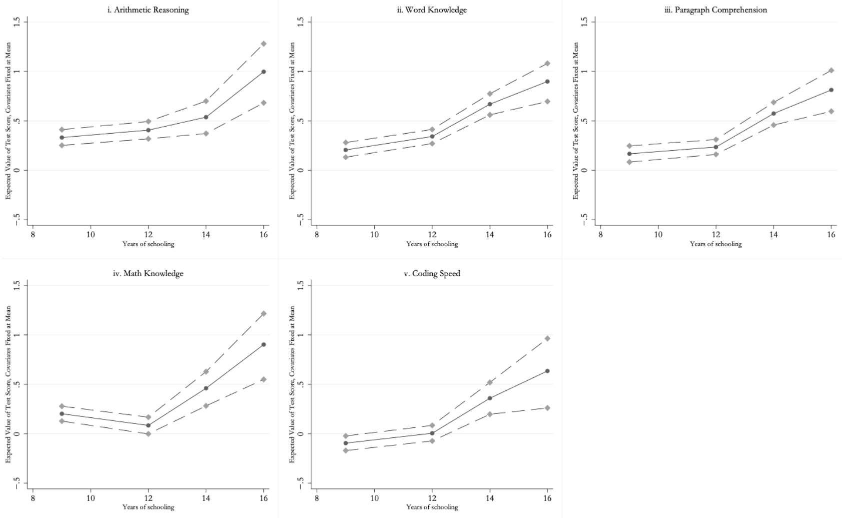
Notes: The cognitive measure represents the standardized average over the ASVAB scores (arithmetic reasoning, word knowledge, paragraph comprehension, numerical operations and coding speed). The Noncognitive measure is computed as a (standardized) average of the Rosenberg self-esteem scale and Rotter internal-external locus of control. The schooling levels represent the observed schooling level by age 30 in the NLSY79 sample (See Appendix A for details).

Figure 2B. Distribution of Factors by Gender and Schooling Level



Notes: The factors are simulated from the estimates of the model. The schooling levels represent the predicted schooling level by age 30. These schooling levels are obtained. from the structure and estimates of the model and our sample of the NLSY79 (See Appendix A for details). The simulated data contain 19,600 observations.

Figure 3A. Effect of schooling on ASVAB Components for person with average ability with 95% confidence bands - Males



Notes: We standardize the test scores to have within-sample mean 0, variance 1. The model is estimated using the Age 30 NLSY79 Sample (See Appendix A for details).

Figure 4. Mean Log Wages by Age 30 - Males

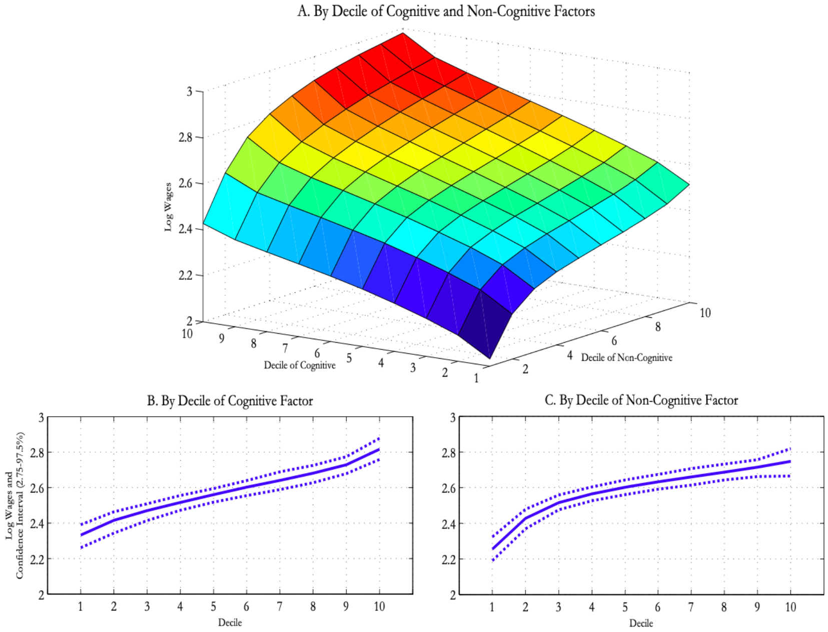


Figure 5. Mean Log Wages of High School Dropouts by Age 30 - Males

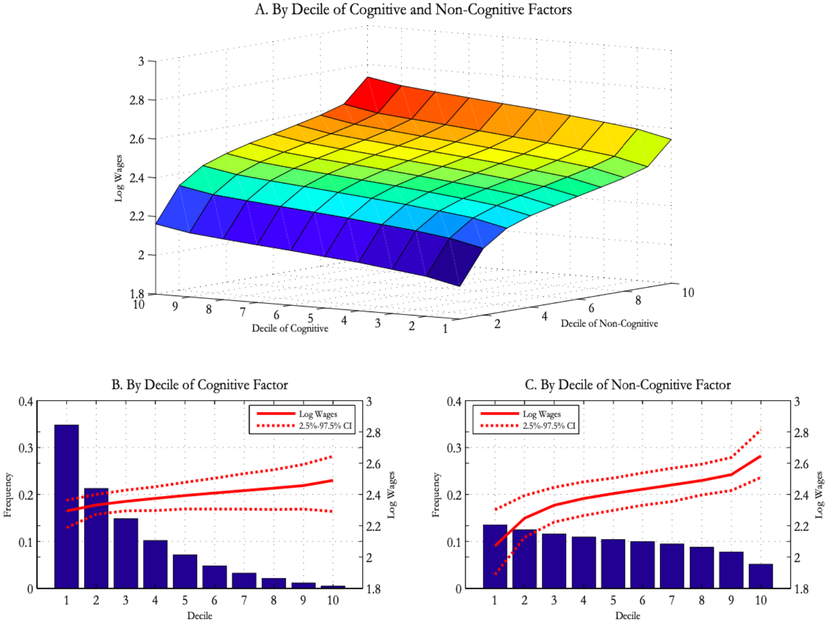


Figure 6. Probability of Being a High School Dropout by Age 30 - Males

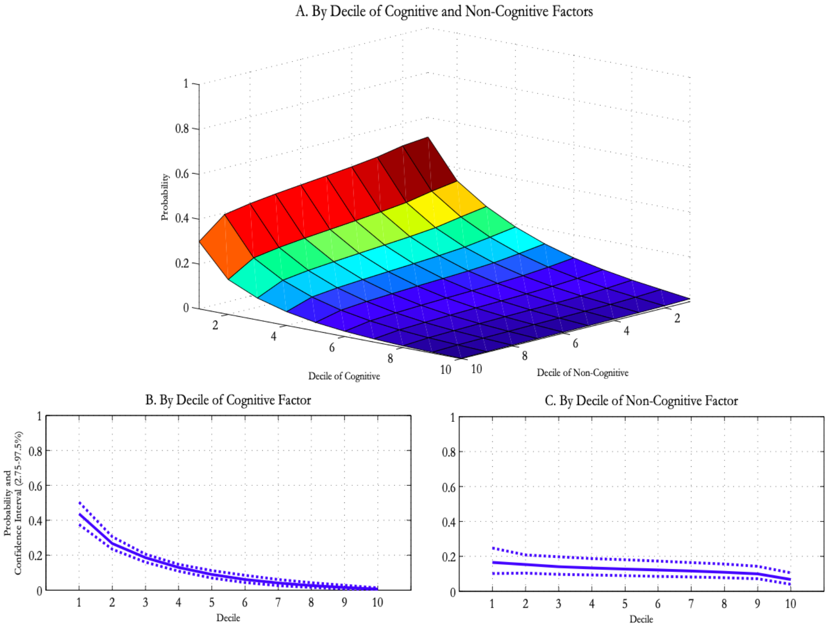


Figure 7. Probability of Being a High School Graduate by Age 30 - Males

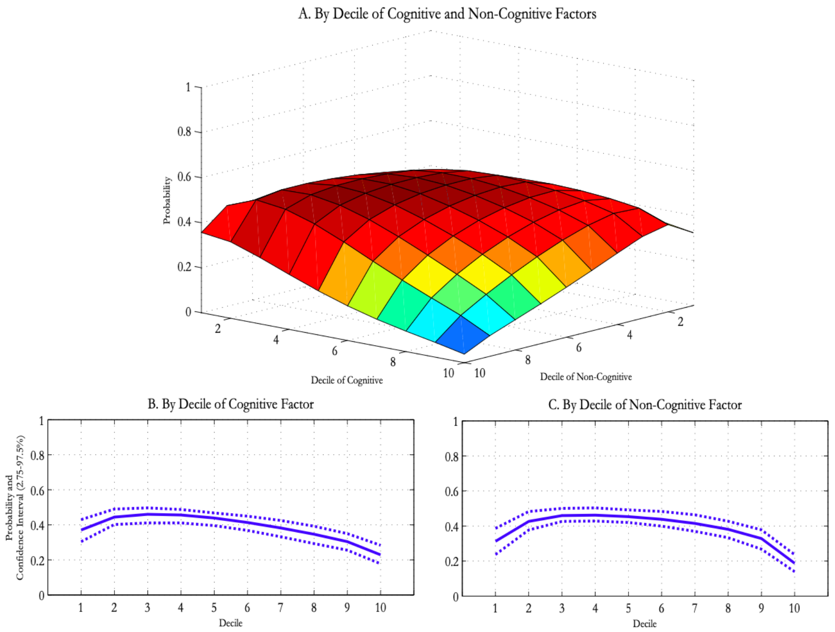


Figure 8. Probability of Being a 4-yr College Graduate by Age 30 - Males

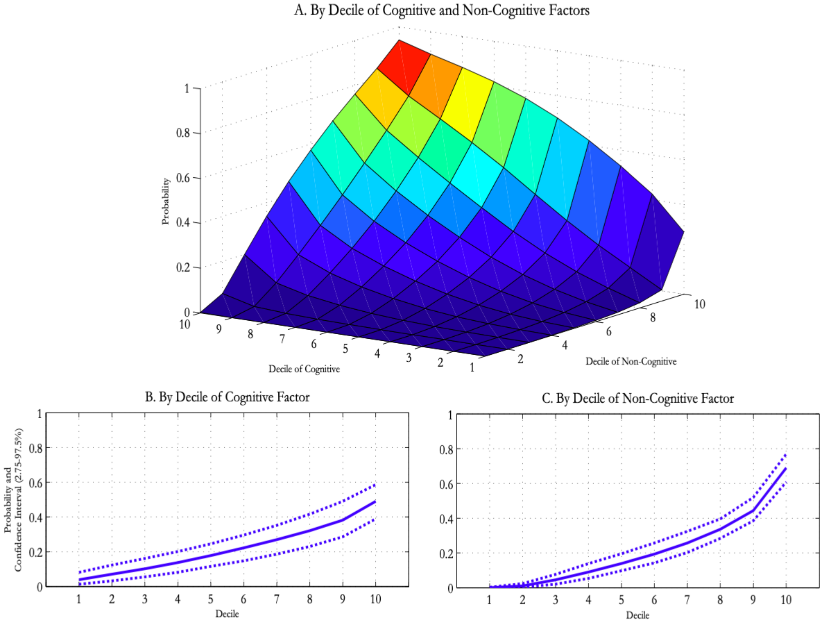


Figure 9. Probability of Employment by Age 30 - Males

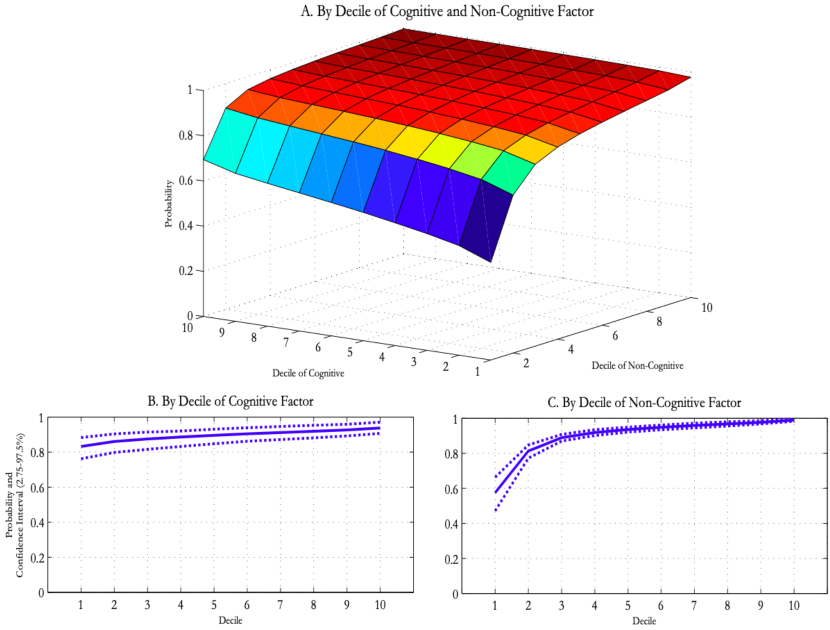


Figure 10. Probability of Participating in Illegal Activities during the Year 1979 - Males

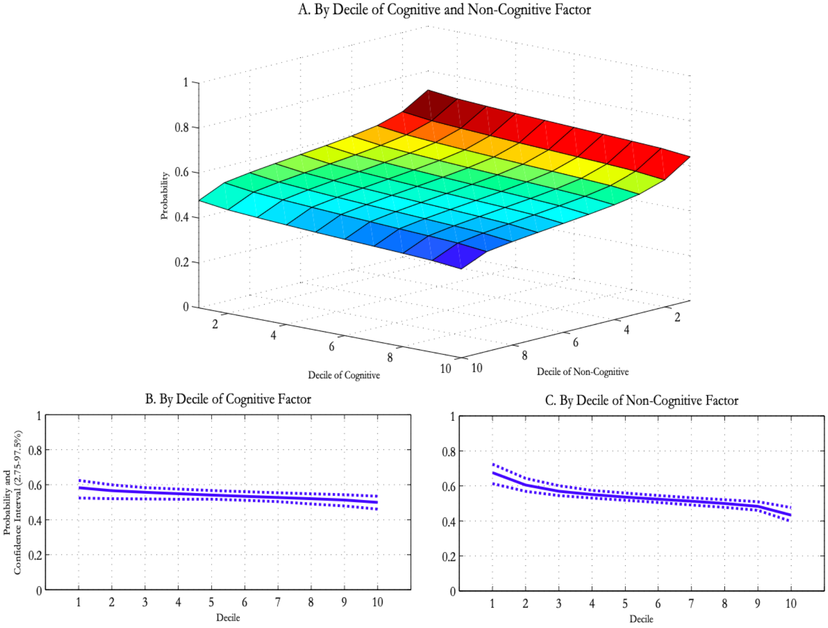
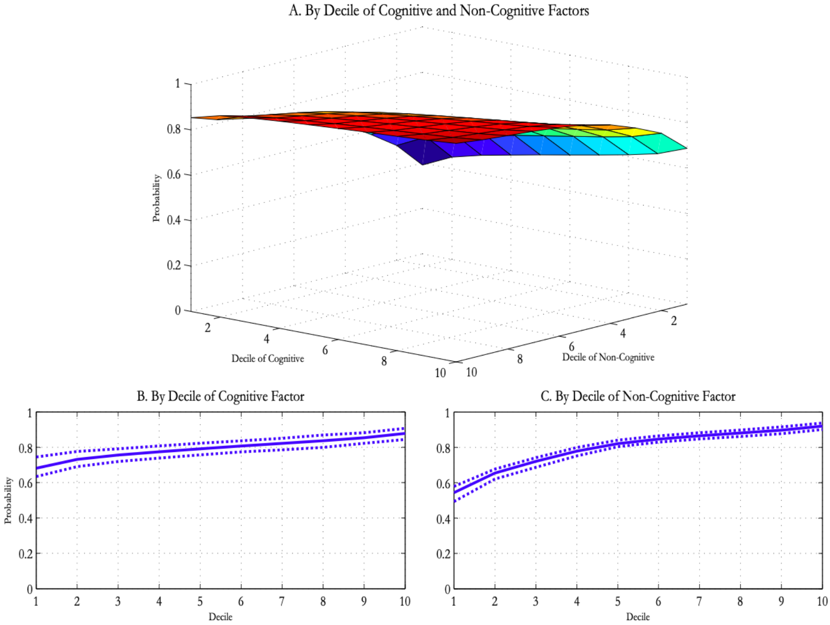


Figure 11. Probability Of Being Single With No Child - Females



Conclusion

- Low dimensional model for two latent abilities explains a diverse array of behaviors controlling for reverse causality and selection
- We move beyond looking only at effects of cognitive and noncognitive skills on wages.
- For many dimensions of behavior, noncognitive ability is more important than, or as important as (in the sense of effects of movements from the top to the bottom of the distribution) cognitive ability.

- Noncognitive ability affects acquisition of skills and a variety of behaviors as well as market productivity as measured by wages.

- Cognitive ability affects market productivity, skill acquisition and a variety of behaviors.

- We find by allowing for reverse causality that schooling has important effects on the cognitive and non-cognitive test scores.

The Schooling and Labor Market Effects of Vouchers

by Rau, Sánchez, and Urzua, 2020

Introduction

- Fundamental question in the vouchers literature: What are the effects of vouchers on the students who use them?
- Two decades of important progress in answering this question (Epple et al, 2017)
  - From Angrist et al. (2002) to Abdulkadiroglu et al. (2018)
- However, focus on short-term results (i.e. test scores)
- Longer-term outcomes (e.g. college enrollment, employment) are arguably more important
- This paper:
  - Long-term effects of vouchers in Chile
  - Focus on heterogeneity and schooling transitions

This Paper

- Sequential model of schooling decisions and outcomes for Chilean students
  - Constructed from reduced-form equation
  - Well-defined treatment effect parameters
- Rich administrative data for Chile
  - Track students from age 14 to 27
  - Schooling data merged to labor market data
    - Schooling data (Min Edu): primary, HS, college; test scores, family background
    - Labor market data (UI): monthly wages, employment

Literature Review

- Comprehensive lit. review: Epple et al. (2017)
  - Mixed effects on test scores (from Angrist et al., 2002 to Abdulkadiroglu et al., 2018)
- Very few papers on long-term effects of vouchers
  - US: high school graduation, college enrollment (Wolf et al., 2010, Chingos and Peterson, 2015, Chingos, 2018)
  - Colombia: college admission exams, high school graduation (Angrist et al., 2006); employment and wages (Bettinger et al., 2019)

Chile's Universal Voucher System

- Since 1981, the educational system operates under a nationwide voucher agenda
- Per-student subsidy paid directly to the school of student's choice
- Funding comes from the government
- Universal eligibility
- Public and private schools are equally funded by the per-student voucher
- Top-up fees:
  - Public schools, not allowed
  - Private-voucher schools, allowed, though only $\sim 50\%$ charge some positive tuition
- Today, more than 50% of students attend private-voucher schools
- Third group of schools: elite private-non-voucher. Only 7% of enrollment

Empirical Approach

- Suppose we want to estimate the effect of a certain schooling choice, $D = 1$, on an outcome Y
- We posit a causal relationship between Y and D, and other variables affecting Y, X:

$$Y = \beta_0 + \beta_1 D + f(X; \beta_2) + U$$

  - Ex. 1: The wage return of schooling

$$\ln(w) = \beta_0 + \beta_1 D + f(X; \beta_2) + U$$

  - Ex. 2: The effect of schooling on test scores

$$T = \beta_0 + \beta_1 D + f(X; \beta_2) + U$$

- With random assignment of D, β_1 is consistently estimated by OLS
- Without random assignment, D is very likely endogenous

Empirical Approach

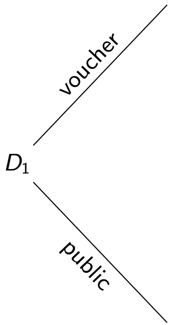
- With endogenous D, need to understand what determines $D = 1$
- Roy model:
  - Two potential outcomes,
$$\begin{aligned} Y^1 &= X\beta^1 + U^1 \\ Y^0 &= X\beta^0 + U^0 \end{aligned}$$
  - A decision rule,
$$\begin{aligned} D^* &= Z\theta + U^D \\ D(Z) &= 1\left[D^*(Z) \geq 0\right] \end{aligned}$$
  - $Y = DY^1 + (1 - D)Y^0$

Empirical Approach

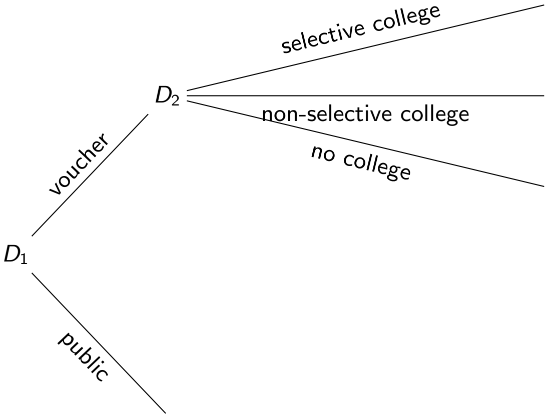
- Roy model allows estimation of effect of **vouchers** (D) on **wages** (Y) (Heckman and Honoré, 1990)

- But, what about other schooling choices along the way (e.g. college/no college, graduate/dropout)?

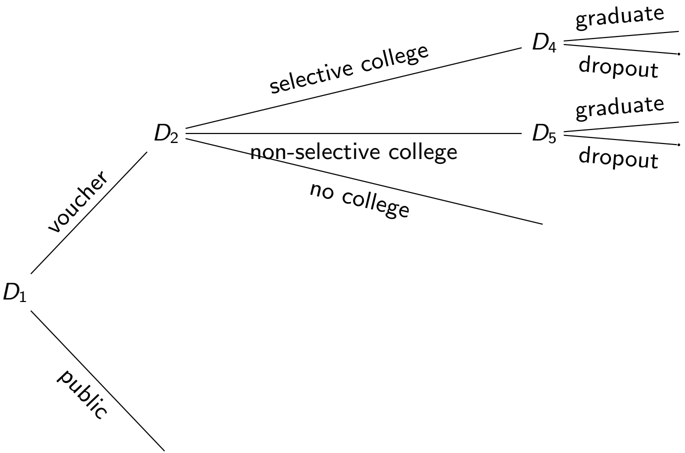
Schooling Decisions



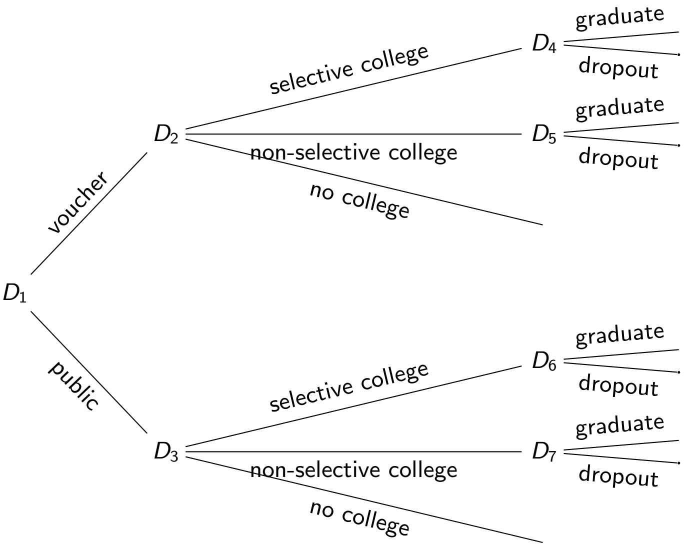
Schooling Decisions



Schooling Decisions



Schooling Decisions



Empirical Approach

- Model of sequential schooling decisions and outcomes
- Generalization of the Roy framework
- We model each schooling decision using a flexible discrete choice model
- Key role of latent ability in determining choices and outcomes
- Identification comes from accounting for unobserved ability and the use of exclusion restrictions
- Model allows the definition of various treatment effect parameters

Model: Schooling Decisions and Outcomes

- Decision of attending a schooling level j' given that the student attained schooling level j:

$$D_{j'|j} = \begin{cases} 1 & \text{if } I_{j'|j} \geq 0 \\ 0 & \text{otherwise.} \end{cases}$$

- Performance in college admission exams:

$$Y = D_1 Y^1 + (1 - D_1) Y^0$$

- Log wages:

$$\ln w = \sum_{j'} D_{j'|j} \ln w_{j'}$$

- Decision of taking the exams and employment is also modeled

Model: Parameterization

- Value of transitioning from schooling level j to schooling level j',

$$I_{j'|j} = Z_{j'|j}\gamma_{j'|j} + \theta\lambda_{j'|j} + \nu_{j'|j},$$

where θ is student's *unobserved heterogeneity*

- Performance in college admission exams,

$$Y^k = X^Y\beta_k^Y + \theta\alpha_k + \eta^k,$$

- Wages,

$$\ln w_{j'} = X^w\beta_{j'}^w + \theta\alpha_{j'}^w + \eta_{j'}^w,$$

Model: Measurement System for the Unobserved Heterogeneity

- Measurement system of equations to help identify the distribution of θ, as well as to facilitate its interpretation:

$$M_l = X^M \delta_l + \theta \psi_l + \epsilon_l,$$

- Measurements taken before the student makes the decisions relevant to our model

Model: Identification

- Identification strategy follows Carneiro et al. (2003), Hansen et al., 2004, Heckman et al. (2017)
- It uses a version of matching on (Z, X, θ)
  - Assume θ is independent from all covariates and error terms
  - Conditional on θ, all error terms are independent from each other
- It also benefits from the use of instruments (i.e. schooling decision shifters)

Estimation

- Estimate the model by maximum likelihood:

$$\begin{aligned}\mathcal{L} &= \prod_{i=1}^{N} \int f(Y_i, D_i, M_i \mid Z_i, X_i, \theta) f(\theta) d(\theta) \\ &= \prod_{i=1}^{N} \int f(Y_i, D_i \mid Z_i, X_i, \theta) f(M_i \mid X_i, \theta) f(\theta) d(\theta),\end{aligned}$$

- Assume normal distributions for the error terms, and approximate the distribution of the factor using a mixture of two normals:

$$\theta \sim pN(\mu_1, \sigma_1^2) + (1-p)N(\mu_2, \sigma_2^2),$$

Definition of Treatment Effects

- Average treatment effect:

$$ATE = \int \int E(Y_1 - Y_0 \mid X = x, \theta = \bar{\theta}) dF_{X,\theta}(x, \bar{\theta}).$$

- Average treatment effect on the treated:

$$TT = \int \int E(Y_1 - Y_0 \mid X = x, \theta = \bar{\theta}, D = 1) dF_{X,\theta|D=1}(x, \bar{\theta}),$$

- Average treatment effect on the untreated:

$$TUT = \int \int E(Y_1 - Y_0 \mid X = x, \theta = \bar{\theta}, D = 0) dF_{X,\theta|D=0}(x, \bar{\theta}),$$

Data

- Administrative data for 95,672 students
  - 8th grade in 2000
  - Finishing primary school in 2000
  - Primary school does not offer secondary grades
- Measurement system: standardized test scores taken at the end of primary school (verbal, math, soc. sc, nat. sc.)
- Disregard students in private-non-voucher schools

Summary Statistics: Endogenous Variables

	mean	std. dev.	min	max
schooling decisions:				
voucher high school	0.36	0.48	0.00	1.00
take college admission exams	0.47	0.50	0.00	1.00
college enrollment	0.57	0.49	0.00	1.00
college degree[a]	0.32	0.46	0.00	1.00
college degree[b]	0.55	0.50	0.00	1.00
employed	0.77	0.42	0.00	1.00
log annual earnings	8.12	1.15	-2.39	10.91
college admission exams:				
verbal	-0.42	0.88	-3.13	2.90
math	-0.43	0.86	-3.18	3.17
measurement system:				
verbal	-0.11	0.91	-2.81	2.90
math	-0.11	0.91	-2.68	2.63
social sciences	-0.11	0.92	-2.79	3.06
natural sciences	-0.11	0.92	-2.76	2.81

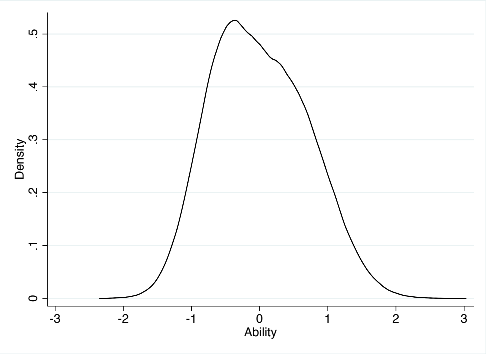
Summary Statistics: Exogenous Variables

	mean	std. dev.	min	max
covariates:				
male	0.50	0.50	0.00	1.00
mother's education	9.0	3.3	0.0	22.0
father's education	9.4	3.4	0.0	22.0
broken home	0.26	0.44	0.00	1.00
log income per capita (age 14)	3.73	0.84	1.06	7.41
log income per capita (age 18)	4.13	0.58	1.92	7.47
region: north (age 14)	0.18	0.38	0.00	1.00
region: center (age 14)	0.48	0.50	0.00	1.00
region: north (age 18)	0.18	0.38	0.00	1.00
region: center (age 18)	0.48	0.50	0.00	1.00
exclusion restrictions:				
% voucher schools in municipality	0.54	0.24	0.00	1.00
avg. tuition in voucher schools	0.12	0.08	0.00	0.35
Δ local unemployment (high–low skill, age 17)	-0.05	0.06	-0.27	0.26
Δ local log wage (high–low skill, age 17)	0.74	0.31	-0.39	2.55
avg. tuition in local selective college (age 17)	2.57	0.69	0.36	6.31
avg. tuition in local non-selective college (age 17)	1.86	0.65	0.16	5.54
Δ local unemployment (high–low skill, age 20)	-0.03	0.06	-0.30	0.46
Δ local log wage (high–low skill, age 20)	0.74	0.28	-0.21	2.13
Δ local unemployment (high–low skill, age 23)	-0.04	0.06	-0.30	0.33
Δ local log wage (high–low skill, age 23)	0.69	0.27	-0.13	2.15
college in municipality (age 21)	0.61	0.49	0.00	1.00

Results

- Use our estimates to simulate 500,000 observations
- We use our simulations to:
  - Recover the distribution of the unobserved ability
  - Construct counterfactuals
  - Compute treatment effects

Results: Distribution of Ability



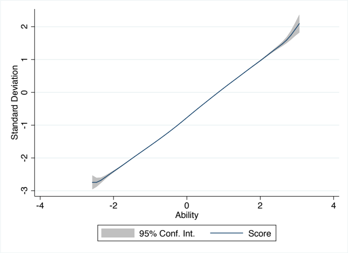
$$f \sim pN(\mu_1, \sigma_1^2) + (1-p)N(\mu_2, \sigma_2^2)$$

where

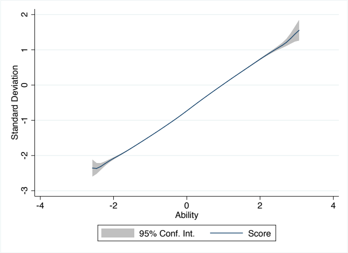
$$(\mu_1, \mu_2) = (0.44, -0.08)$$
$$(\sigma_1, \sigma_2) = (0.58, 0.44)$$
$$p = 0.19$$

Results: Effects of Ability on Outcomes

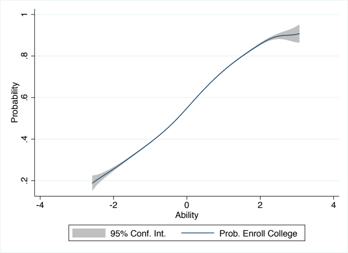
A. Verbal Score



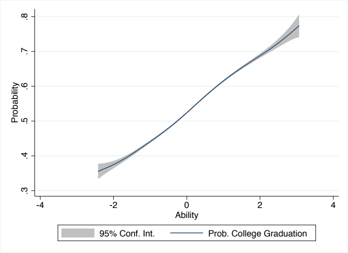
B. Math Score



C. College Enrollment

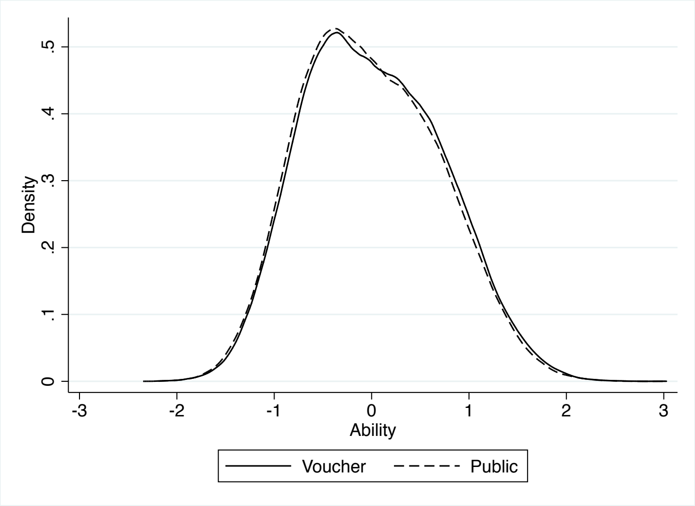


D. College Graduation



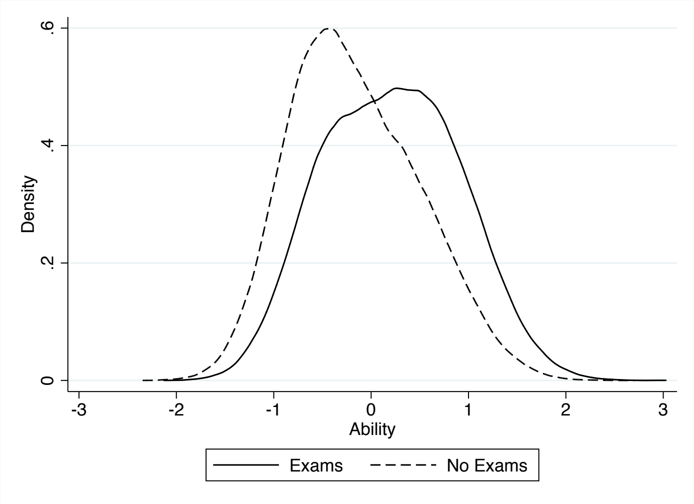
Results: Sorting on Ability

Figure: Distribution of Ability by Voucher/Public High School



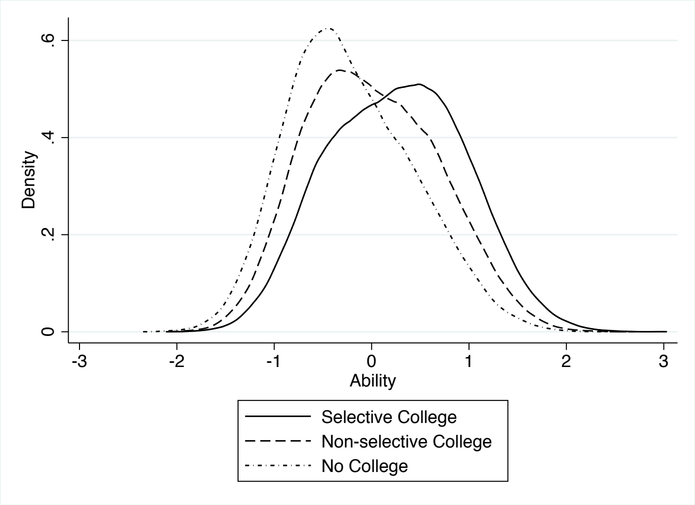
Results: Sorting on Ability

Figure: Distribution of Ability by College Admission Exams Takeup



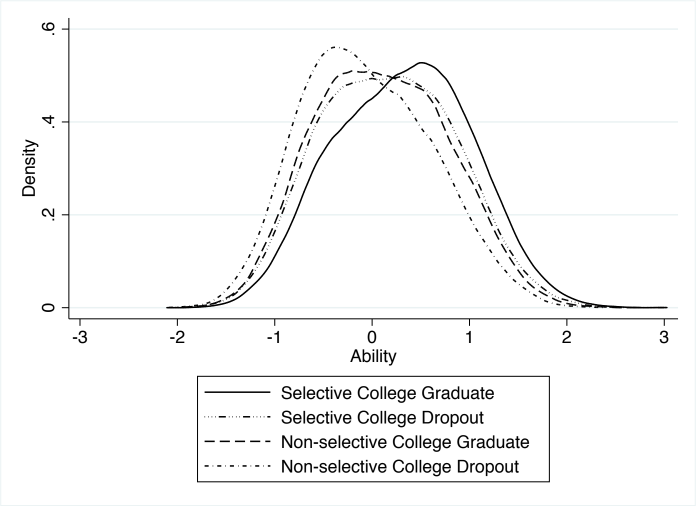
Results: Sorting on Ability

Figure: Distribution of Ability by College Enrollment



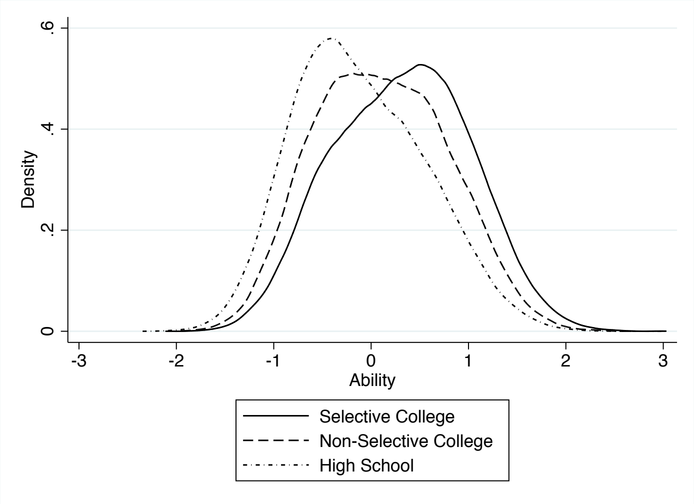
Results: Sorting on Ability

Figure: Distribution of Ability by College Degree



Results: Sorting on Ability

Figure: Distribution of Ability by Final Schooling Level

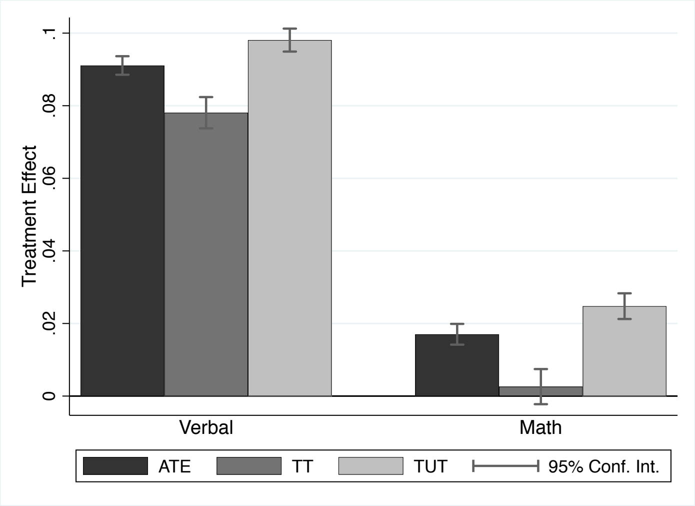


Results: The Effects of Vouchers

- From the simulations, we observe:
  - Actual school choices, and their corresponding consequences/outcomes
  - Counterfactual choices and outcomes
- For each student, we compute:
  - TE: $Y_1 - Y_0$
  - TT: $Y_1 - Y_0 | D = 1$
  - TUT: $Y_1 - Y_0 | D = 0$
- Aggregating over the relevant distributions, we obtain the average versions of the treatment effects

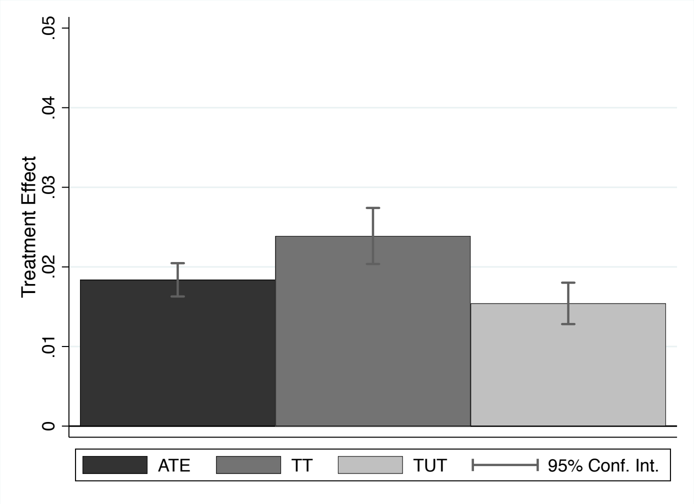
Results: The Effects of Vouchers

Figure: Treatment Effects: Test Scores



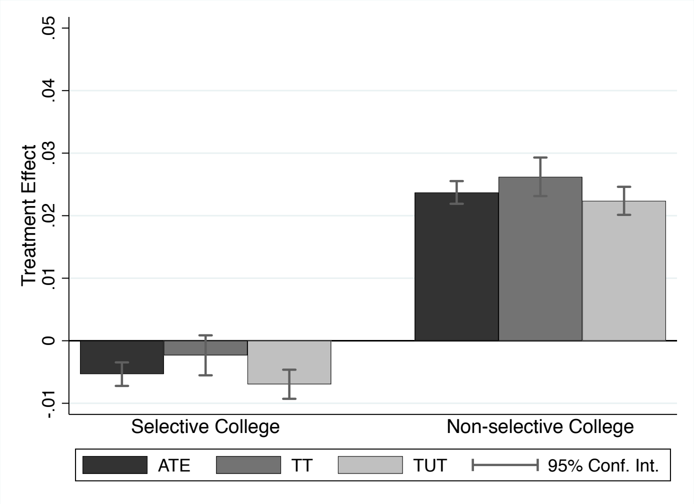
Results: The Effects of Vouchers

Figure: Treatment Effects: College Enrollment



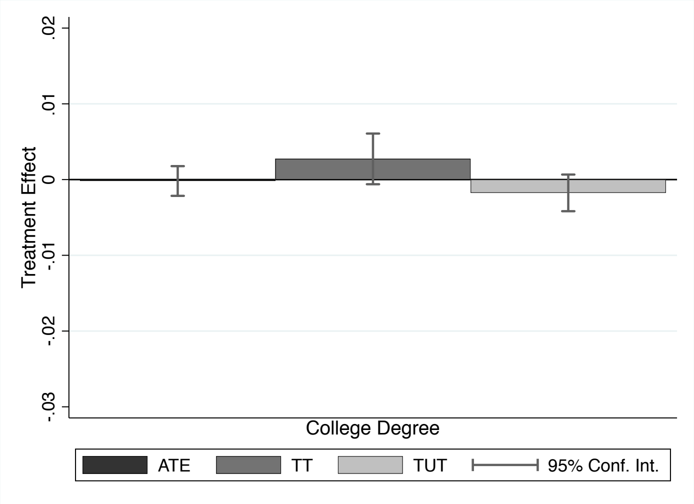
Results: The Effects of Vouchers

Figure: Treatment Effects: Selective and Non-selective College Enrollment



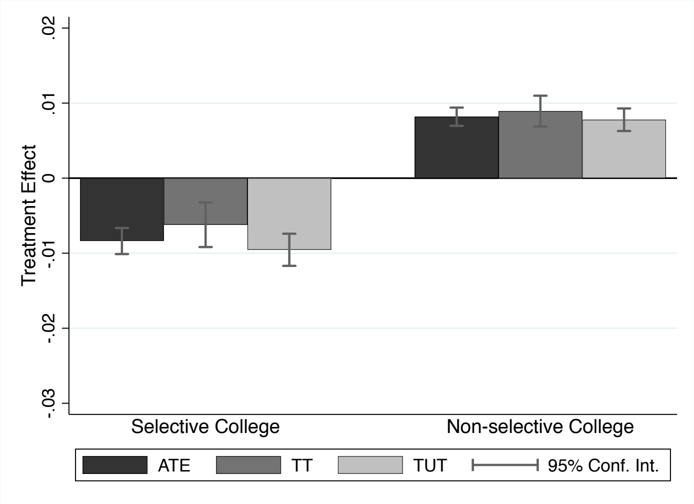
Results: The Effects of Vouchers

Figure: Treatment Effects: College Degree



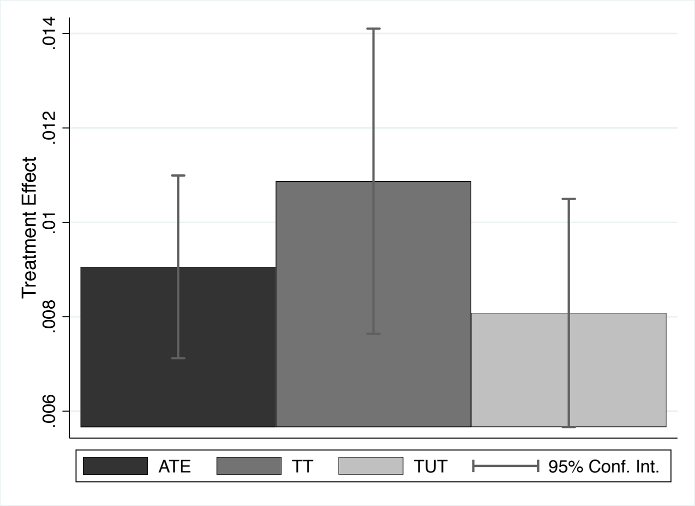
Results: The Effects of Vouchers

Figure: Treatment Effects: Selective and Non-selective College Degree



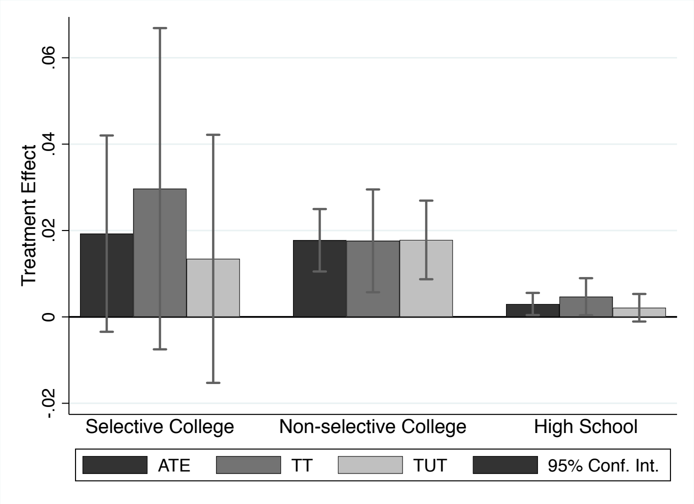
Results: The Effects of Vouchers

Figure: Treatment Effects: Employment



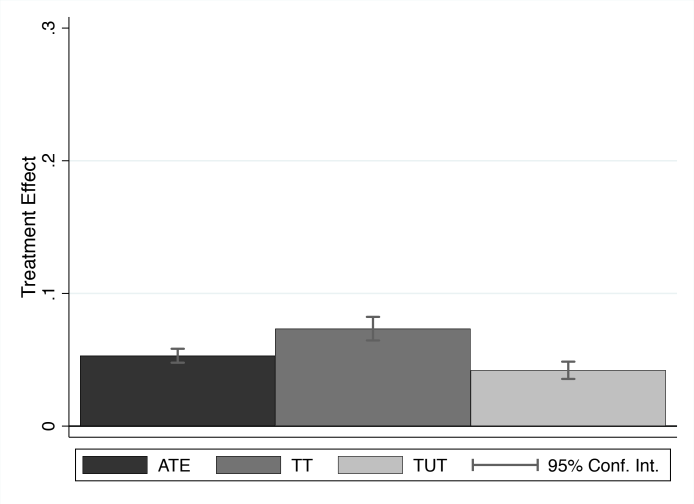
Results: The Effects of Vouchers

Figure: Treatment Effects: Employment Conditional on Final Schooling Level



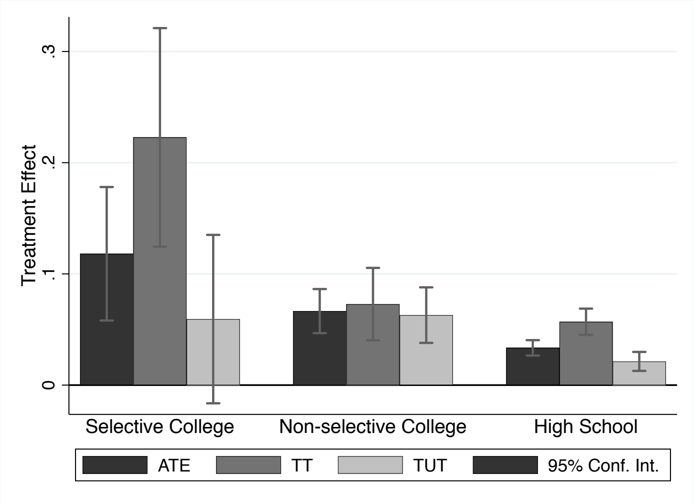
Results: The Effects of Vouchers

Figure: Treatment Effects: Log Annual Earnings



Results: The Effects of Vouchers

Figure: Treatment Effects: Log Annual Earnings Conditional on Final Schooling Level



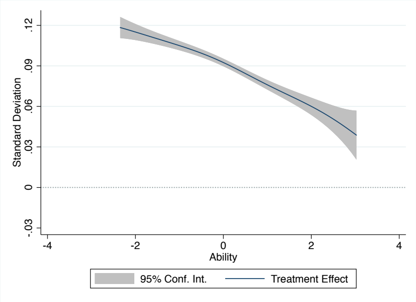
Results: The Effect of Ability on Treatment Effects

- We are interested in heterogeneity in the treatment effects
- We investigate whether more/less able individuals benefit more/less from attending a voucher high school

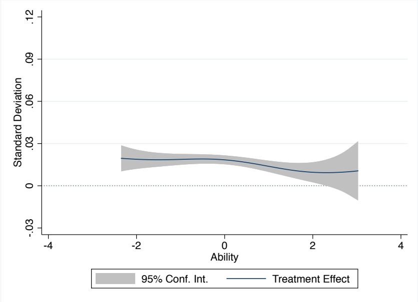
Results: The Effect of Ability on Treatment Effects

Figure: Test Scores

A. Verbal Score

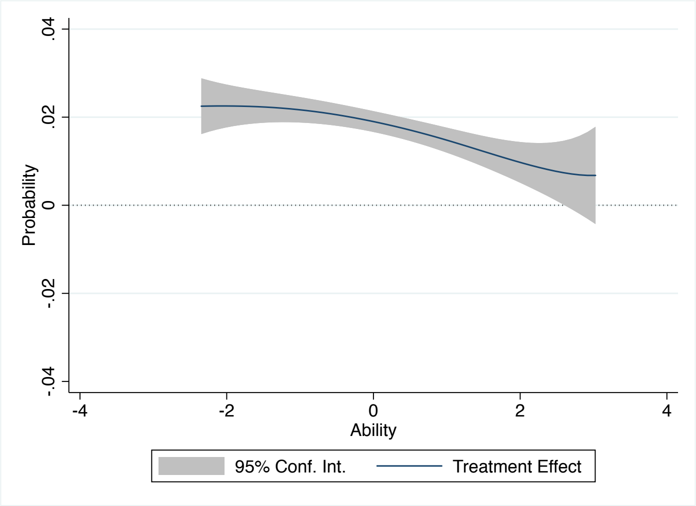


B. Math Score



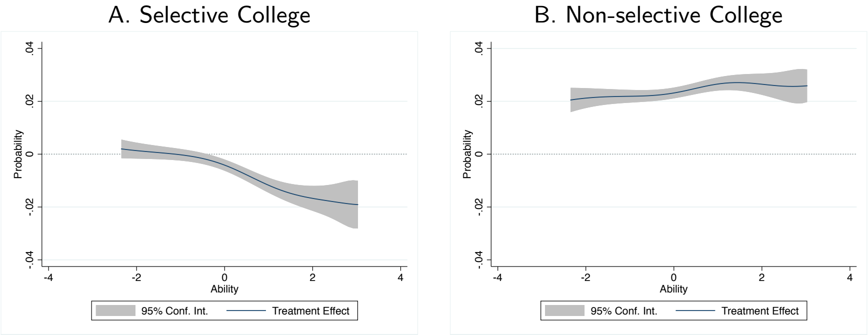
Results: The Effect of Ability on Treatment Effects

Figure: College Enrollment



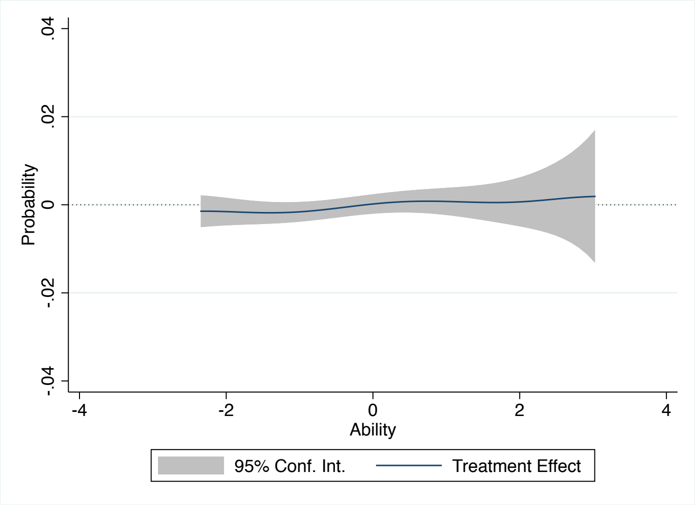
Results: The Effect of Ability on Treatment Effects

Figure: Selective/Non-selective College Enrollment



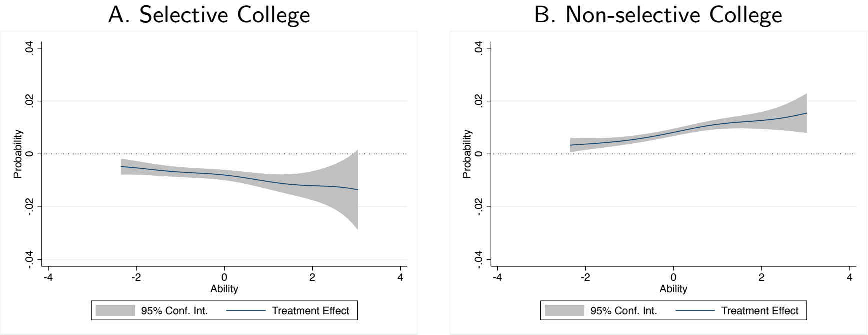
Results: The Effect of Ability on Treatment Effects

Figure: College Degree



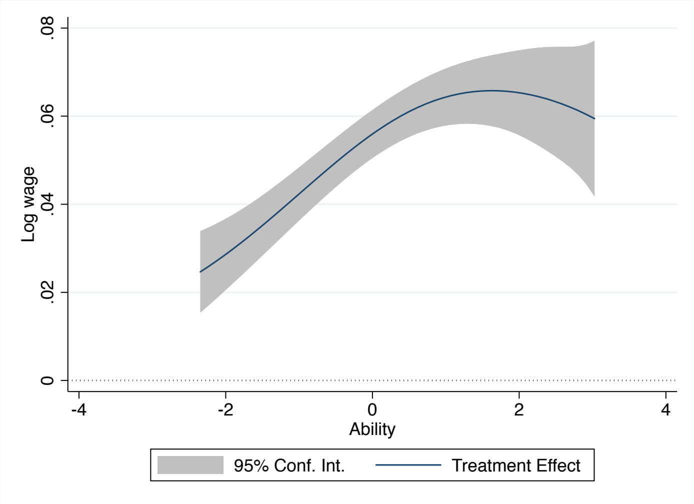
Results: The Effect of Ability on Treatment Effects

Figure: Selective/Non-selective College Degree



Results: The Effect of Ability on Treatment Effects

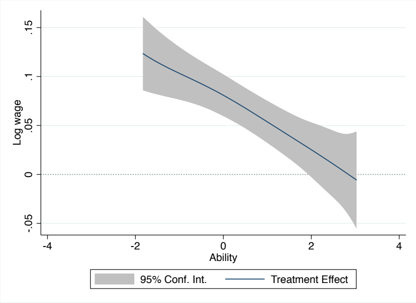
Figure: Log Annual Earnings



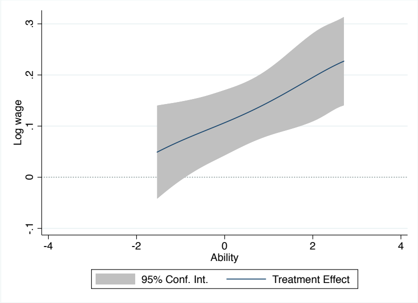
Results: The Effect of Ability on Treatment Effects

Figure: Log Annual Earnings by Final Schooling Level

A. Selective College



B. Non-selective College



Conclusions

- Examine the short- and long-term effects of attending a voucher school in Chile
- Positive effect:
  - Test scores
  - College enrollment (- selective, + non-selective)
  - Employment and wages
- Null effect: college degree (- selective, + non-selective)
- Latent ability plays a key role in determining choices and outcomes
- Important treatment effect heterogeneity