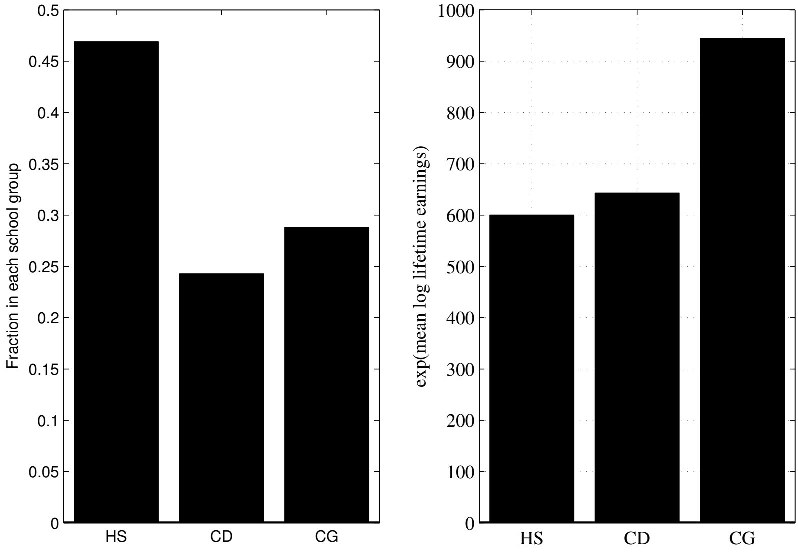
The Return to College: Selection Bias and Dropout Risk

Lutz Hendricks
Oksana Leukhina

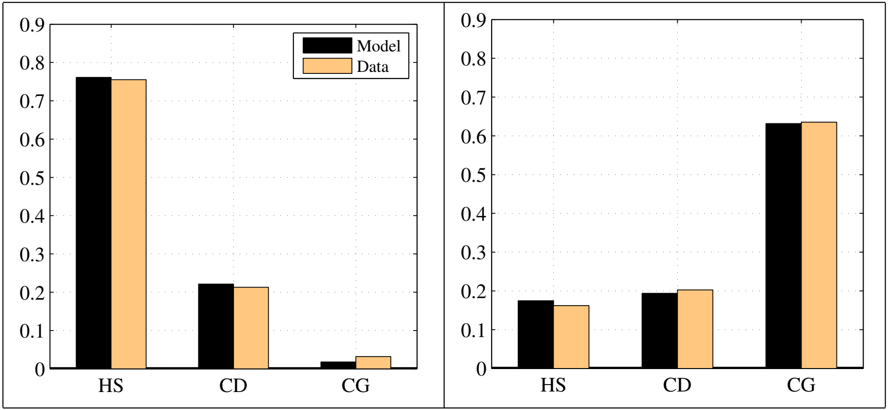
February 20, 2014

Motivation



Data: NLSY79, men.

Motivation



AFQT quartile 1 | AFQT quartile 4

Questions

1. How much of the college lifetime earnings premium is due to selection?
2. Why do so many students drop out of college?
3. Why do students with low graduation probabilities try college?

Focus: College Completion Risk

The idea:

- Graduating from college requires sufficient college credits.
- More able students are more likely to graduate.
- Graduation is uncertain.

Selection occurs at 2 levels:

1. At entry: low ability students are deterred by poor graduation prospects.
2. In college: low ability students fail to progress and drop out.

Implication: Large ability gaps between college graduates and high school graduates.

Main result: Roughly **half** of the college earnings premium is **selection**.

Is This New?

A birds-eye view of existing models of college choice:

1. Roy models:
   Anyone can graduate from college in 4 years.
   "Psychic costs" keep some students out of college.
2. Risky college completion models where ability does not affect wages:
   Altonji (1993), Caucutt and Kumar (2003), ...
3. Keane-Wolpin (1997) type models:
   Anyone can graduate from college in 4 years.
   Some students are hit by shocks (such as wage offers) and choose to drop out.

Why Is This Important?

A puzzling empirical finding:

- In response to a $1,000 tuition subsidy, 3-4% more students enter college.
- A $1,000 subsidy amounts to roughly 1% of the college earnings premium.
- Why is the response so large?

Our answer:

- Low ability students enter college *knowing* that they won't graduate.
- For them, the financial benefits (wage gain) and the costs (tuition etc) are small.
- These students respond to small changes in college costs or borrowing opportunities.
- But they don't graduate.

Outline

1. Model
2. Parametrization
3. Results
    1. Selection and the college premium
    2. Understanding dropouts
4. Counterfactual experiments
    1. Tuition subsidy
    2. Increased borrowing limits
    3. Dual enrollment programs

The Model

Model Outline

- Partial equilibrium.
- 1 cohort.
- Students enter the model at high school graduation $(t = 1)$.
- They draw endowments and choose whether to try college or work as high school graduates.
- In college:
  - students attempt credits with random success
  - they update their beliefs about how long it would take to graduate
  - they decide to drop out or continue studying
- At work:
  - individuals consume their lifetime earnings

Endowments

1. $n_1 = 0$ completed college credits
2. learning ability $a \in \{\hat{a}_1, ..., \hat{a}_{N_a}\}$ with $\hat{a}_1 = 0$ and $\hat{a}_{i+1} > \hat{a}_i$ not observed until the start of work
3. type $j \in \{1, ..., J\}$ which determines
    1. initial assets $k_1 = \hat{k}_j$
    2. ability signal $m = \hat{m}_j$
    3. net price of attending college $q = \hat{q}_j$

Work

State vector $(k_\tau, n_\tau, a, s, \tau)$

- $s \in \{HS, CD, CG\}$
- τ: age

Worker's problem:

$$V(k_\tau, n_\tau, a, s, \tau) = \max_{\{c_t\}} \sum_{t=\tau}^{T} \beta^{t-\tau} u(c_t) + U_s$$

subject to the budget constraint

$$\underbrace{\exp\left(\phi_s a + \mu n_\tau + y_s\right)}_{\text{lifetime earnings}} + Rk_\tau = \sum_{t=\tau}^{T} c_t R^{\tau-t}.$$

U_s: non-monetary utility from working with skill type s

Assumptions: $y_{CD} = y_{HS}$ and $\phi_{CD} = \phi_{HS}$

Work

Before ability is revealed:

$$V_W(k_\tau, n_\tau, j, s, \tau) = \sum_{i=1}^{N_a} \Pr(\hat{a}_i | n_\tau, j, \tau) V(k_\tau, n_\tau, \hat{a}_i, s, \tau).$$

We show: (n_τ, j, τ) is a sufficient statistic for the worker's beliefs about his ability

College

State vector (n,i,j,t)

- j determines assets and college costs
- t is age
- i fixes consumption at $c_{i,j}$ (see below)

Attempt n_c credits

Earn each credit with logistic probability $\Pr_c(a) = \gamma_{min} + \frac{\gamma_{max} - \gamma_{min}}{1+\gamma_1 e^{-\gamma_2 a}}$

Update beliefs about a.

Sufficient statistic for beliefs: n_{t+1}, j, t

Consume $c_{i,j}$ and pay tuition $\hat{q}_j$.

Borrowing constraint: $k_{t+1} = Rk_t - c_{i,j} - \hat{q}_j \geq -k_{min}$.

College: Value Function

Value of studying in period t:

$$V_C(n,i,j,t) = u(c_{i,j}) + \beta \sum_{n'} \Pr(n'|n,j,t)\, V_{EC}(n',i,j,t+1),$$

with continuation value

$$V_{EC}(n,i,j,t) = \max\left\{\underbrace{V_C(n,i,j,t) + \pi p_c}_{\text{study}}, \underbrace{V_W(k_{i,j,t},n,j,s(n),t) + \pi p_w}_{\text{work}}\right\} - \pi\bar{\gamma},$$

where

- p_c and p_w are independent draws from standard type I extreme value distributions
- $\bar{\gamma}$ is the Euler–Mascheroni constant
- $s(n)$ is the schooling level associated with n college credits (CG if $n \geq n_{grad}$ and CD otherwise).

Continuation Value

Cases:

1. If $n \geq n_{grad}$: work as a CG.
2. If $t = T_c$ and $n < n_{grad}$: work as CD.
3. Otherwise: choose between working as CD or studying next period.

Choices at High School Graduation

1. Choose fixed consumption while in college, $c_{i,j}$
   Admissible values exhaust borrowing limits after $1, ..., T_c$ periods
2. Choose whether to work as HS or try college

Choices are subject to type I extreme value shocks (for continuity).

Details

Model Summary

1. Graduate from high school
2. Draw financial resources and ability signal
3. If ability signal is low or money is tight: work as high school graduate
4. Otherwise enter college
    1. pre-commit to consumption
5. In each period:
    1. earn credits
    2. update beliefs
    3. if beliefs indicate low ability or money runs out: drop out
6. Work as a permanent income consumer

Setting Model Parameters

Data Sources

NLSY79

- representative sample of men born between 1957 and 1964
- annual interviews until 1994; then biannual
- wages, schooling, AFQT scores

High School & Beyond (HS&B)

- high school sophomores in 1980
- high school GPAs, college transcripts

Endowment Distributions

Approximate a joint Normal distribution for

- $\left(\ln\left(\hat{k}_j\right), \quad \hat{q}_j, \quad \hat{m}_j\right)$
- $a|m$

$J = 120$ types

Details

Measurement

College attendance:

- a student attempts at least 9 non-vocational credits in a year.

College credits.

- n_t/n_c = [earned college credits] / [full course load]
- full course load = number of credits attempted by students who eventually graduate from college

Test score quartiles.

- NLSY79: AFQT
- HS&B: High school GPA
- In the model: IQ is a noisy measure of m

Measurement

College costs q:

- all college related payments that are *conditional* on attending college
- tuition and fees net of scholarships, grants, and labor earnings
- "other" college expenditures (books, supplies, and transportation)
- key fact: mean q is close to 0

Assets k_1:

- financial resources the student receives *regardless* of college attendance
- financial assets, parental transfers within 6 years of HS graduation

Borrowing limit: Stafford loans ($19,750)

Setting Parameters

20 calibrated parameters:

1. Endowment distributions
2. Lifetime earnings: ϕ_s, y_s, μ
3. Preferences: U_s, π
4. Probability of passing a course $\Pr(a)$

Simulate 100,000 person histories.

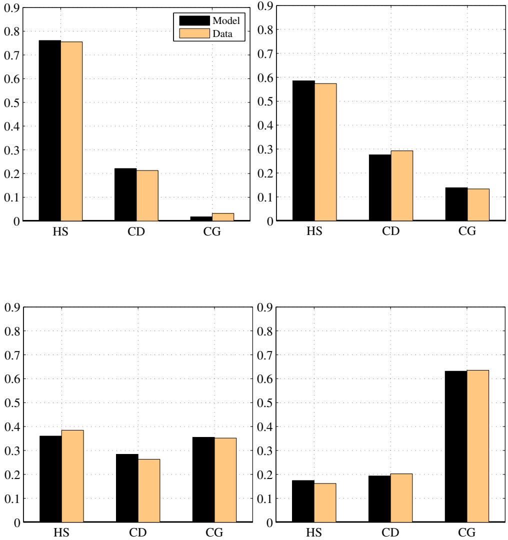
Minimize the weighted sum of squared deviations between model and data moments.

Details

Calibration Targets

1. Fraction in population, by (test score quartile, schooling)
2. Lifetime earnings, by (test score quartile, schooling)
3. Dropout rate, by (test score quartile, year in college)
4. Fraction of credits passed, by graduation status and year
5. Mean and standard deviation of k_1 (HS and college)
6. Mean and standard deviation of q (college)
7. Fraction of students in debt, by year in college
8. Mean student debt, by year in college
9. Average time to BA degree (years)

Fit: Schooling and Test Scores



Fit: Debt

	Mean debt		Fraction with debt	
Year	Model	Data	Model	Data
1	5,827	3,549	15.3	27.7
2	6,740	6,060	27.6	36.0
3	7,907	8,045	49.8	42.5
4	11,000	9,740	72.1	48.0

Mean debt: conditional on having debt.

More

Results

Selection and Earnings

Mean log lifetime earnings of school group s:

$$\mathbb{E}[\phi_s a + \mu n_\tau + y_s + \ln(R^{-\tau})|s], \tag{1}$$

Decomposing the gap relative to high school graduates:

1. prices: $y_s - y_{HS} + (\phi_s - \phi_{HS})\mathbb{E}(a|s)$;
2. credits: $\mathbb{E}(\mu n_\tau|s)$;
3. delayed labor market entry: $\mathbb{E}\{\ln R^\tau|s\} - \ln R^{-1} = \mathbb{E}\{\ln R^{1-\tau}|s\}$;
4. ability selection: $\phi_{HS}[\mathbb{E}(a|s) - \mathbb{E}(a|HS)]$.

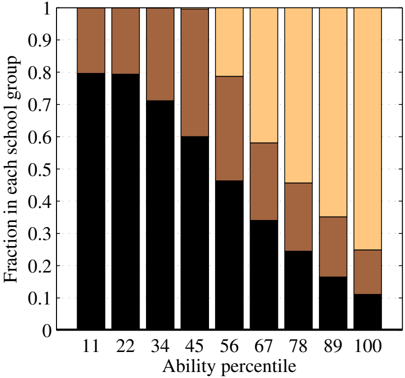
Selection and Earnings

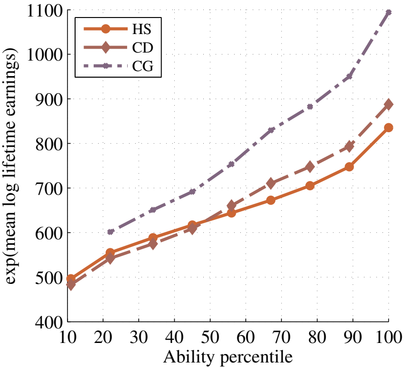
Gap relative to HS (in log points)	College dropouts Gap	Fraction	College graduates Gap	Fraction
Total gap	8	–	45	–
Delayed labor market entry	-9	-124	-18	-39
Prices: y_s and ϕ_s	0	0	11	24
Credits	11	143	30	67
Ability selection	6	81	22	48

Selection at entry accounts for 2/3 of the CG/HS ability gap

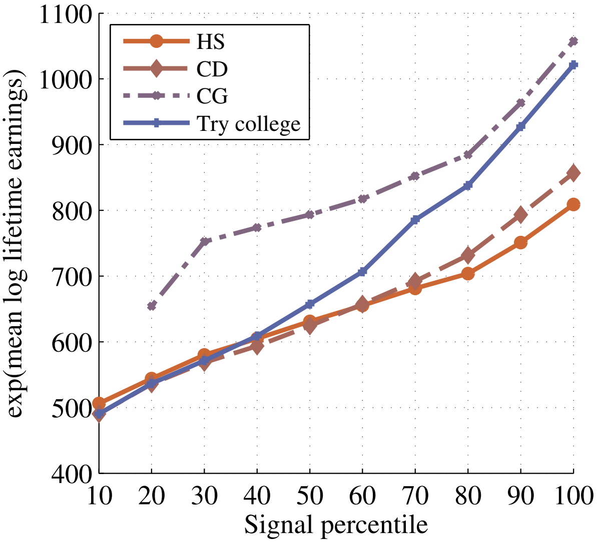
Understanding College Entry

Ability and schooling	Lifetime earnings





Signals and Lifetime Earnings



Understanding Dropouts

Why do nearly half of all college entrants fail to earn a degree?

Our model offers 3 reasons:

1. Money:
   Low ability students know that they will not graduate.
   They enter college because it is cheap.
   Key feature of the data: mean q is close to 0. ▸ Details
2. Luck:
   Medium ability students drop out if they receive poorer than expected "grades." ▸ Details
3. Preference shocks:
   Shutting down preference shocks reduces dropout rates from 46% to 40%.

Counterfactual Experiments

Tuition Subsidy

Experiment:

- reduce mean q by \$1,000
- for comparison: raise y_{CG} by 4 log points
- both produce similar changes in college enrollment

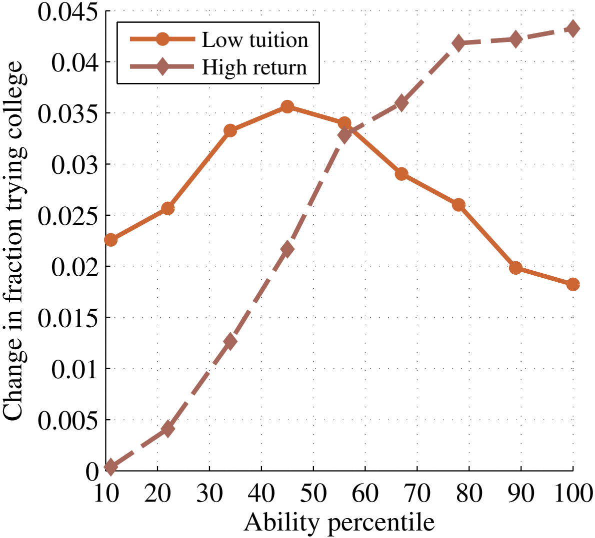
Result:

- college enrollment rises by 2.7 percentage points
- Dynarski (2003): in the data, a \$1,000 tuition subsidy raises enrollment by 3-4 percentage points

Puzzle:

- Why is the response so large, given that the subsidy is so small (1% of the college earnings premium)?
- Why is the response the same for the much larger change in y_{CG}?

Tuition Subsidy



Relax Borrowing Limits

Experiment:

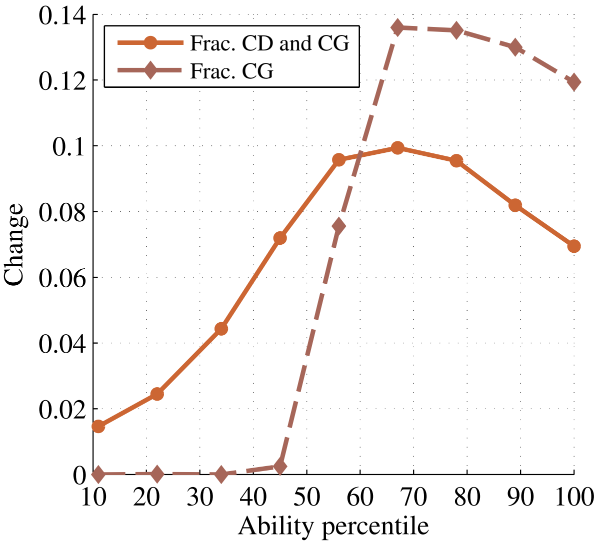
- double k_{min}

Result:

- college enrollment rises by 7 percentage points
- even though (in model and data) very few students are close to their borrowing limits

Two effects:

- low to median ability students with low q and low k_1 enter college (most drop out)
- high ability students with high q or low k_1 can now try to graduate (they would have dropped out otherwise)



Dual Enrollment Programs

Conclusion

- The question: What fraction of the college earnings premium is selection?
- The approach:
  - Focus on risky college completion.
  - Selection occurs at 2 levels: in college and at entry.
- Main result: about **half** of the college earnings premium is selection.
- Asymmetric incentives:
  - **high ability** students attend college as investment
    they respond to the college premium
  - **low ability** students attend college for consumption
    they respond to the direct cost of college

Choices at High School Graduation
Consumption choice

Students commit to fixed consumption while in college

- For each type j there are T_c consumption levels
- Level i exhausts the borrowing limit after i periods in college

Consumption choice problem:

$$i = \arg\max_{\hat{i}} \left\{ V_C\left(0, \hat{i}, j, 1\right) + \pi_c (p_{\hat{i}} - \bar{\gamma}) \right\}. \tag{2}$$

where $p_{\hat{i}}$ is drawn from a standard type I extreme value distribution.

Choices at High School Graduation

College entry decision

The college/work decision is made after consumption has been chosen.

The agent solves

$$\max\left\{V_C(0,i,j,1)+\pi p_c,\, V_W(\hat{k}_j,0,j,HS,1)+\pi p_w\right\}-\pi\bar{\gamma}, \tag{3}$$

Endowment Distributions

The goal: heterogeneity in several variables, but computationally efficient.

Types:

- $J = 120$
- drawn from a joint Normal distribution

Abilities:

- $N_a = 9$ abilities with equal mass
- $\Pr(\hat{a}_i|j)$ approximates a joint Normal distribution for (a, m)

Endowments

Draw 3 independent standard Normal random vectors of length J: ε_k, ε_q, and ε_m.

$\ln \hat{k}_j = \mu_k + \sigma_k \varepsilon_{k,j}$, where $\varepsilon_{k,j}$ is the j^{th} element of ε_k.

$\hat{q}_j = \mu_q + \sigma_q \frac{\alpha_{q,k}\varepsilon_{k,j}+\varepsilon_{q,j}}{\left(\alpha_{q,k}^2+1\right)^{1/2}}$

$\hat{m}_j = \frac{\alpha_{m,k}\varepsilon_k+\alpha_{m,q}\varepsilon_{q,j}+\varepsilon_{m,j}}{\left(\alpha_{m,k}^2+\alpha_{m,q}^2+1\right)^{1/2}}$

Ability grid:

A discrete approximation of the joint normal distribution

$$a = \bar{a} + \frac{\alpha_{a,m} m + \varepsilon_a}{\left(\alpha_{a,m}^2 + 1\right)^{1/2}}, \tag{4}$$

Set $\hat{a}_i = \mathbb{E}\{a|a \in \Omega_i\}$ where $\Omega_i = \left\{a : \frac{i-1}{N_a} \leq \Phi(a-\bar{a}) < \frac{i}{N_a}\right\}$

Set $\Pr(\hat{a}_i|j) = \Pr(a \in \Omega_i | m = \hat{m}_j)$.

Fixed Model Parameters

Parameter	Description	Value
Preferences		
β	Discount factor	0.98
π_c	Scale of preference shocks at consumption choice	0.20
College		
T_c	Maximum duration of college	6
n_{grad}	Number of credits required to graduate	20
n_c	Number of credits attempted each year	5
k_{min}	Borrowing limit	-$19,750
Other		
R	Gross interest rate	1.04

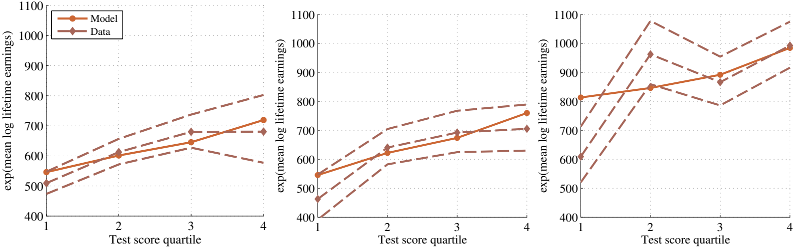
Calibrated Parameters

Parameter	Description	Value
Endowments		
μ_k, σ_k	Marginal distribution of $\ln(k_1)$	$0.41, 1.17$
μ_q, σ_q	Marginal distribution of q	$3.01, 5.81$
$\alpha_{m,k}, \alpha_{m,q}, \alpha_{q,k}, \alpha_{a,m}, \alpha_{IQ,m}$	Endowment correlations	$0.23, -0.11, -0.44, 2.97, 1.78$
Lifetime earnings		
ϕ_{HS}, ϕ_{CG}	Effect of ability on lifetime earnings	$0.153, 0.194$
y_{HS}, y_{CG}	Lifetime earnings factors	$3.90, 3.91$
μ	Earnings gain for each college credit	0.014
Other parameters		
π	Scale of preference shocks	0.767
U_{CD}, U_{CG}	Preference for job type s	$-1.11, -2.98$
$\gamma_1, \gamma_2, \gamma_{min}$	Probability of passing a course	$0.68, 7.89, 0.42$

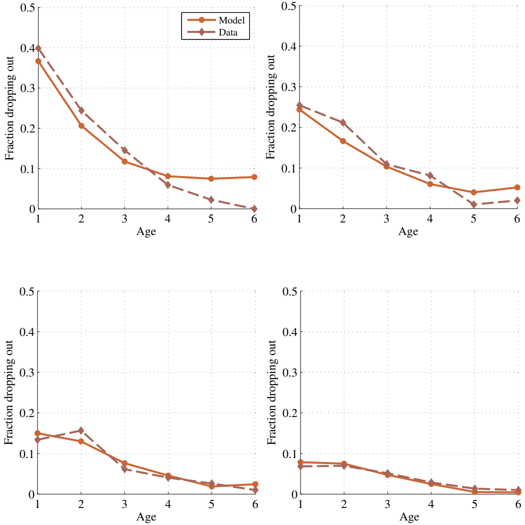
Fit: Schooling and Lifetime Earnings

	School group		
	HS	CD	CG
Fraction			
Data	46.9	24.3	28.8
Model	47.1	24.4	28.5
Gap (pct)	0.4	0.4	-1.0
Lifetime earnings			
Data	600	643	944
Model	596	643	934
Gap (pct)	-0.7	-0.0	-1.0

Fit: Lifetime Earnings



Fit: Dropout Rates (IQ quartiles)



Fit: Credit Passing Rates

	College dropouts		College graduates	
Year	Model	Data	Model	Data
1	66.7	67.7	95.9	98.7
2	67.9	71.8	95.8	96.3
3	64.7	66.9	95.8	95.7
4	57.7	63.8	95.8	94.9

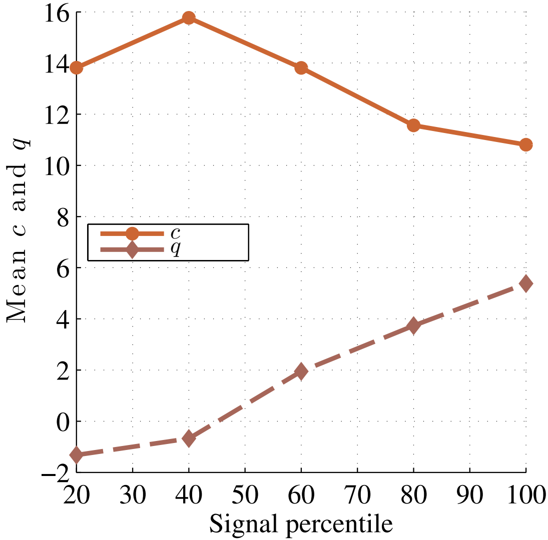
Fit: Financial Moments

	Model	Data
Distribution of k_1, HS		
mean	16,770	16,630
standard deviation	22,867	23,266
Distribution of k_1, college		
mean	38,011	37,390
standard deviation	37,329	38,475
Distribution of q, college		
mean	-740	-584
standard deviation	4,928	5,787

Fit: Financial Moments

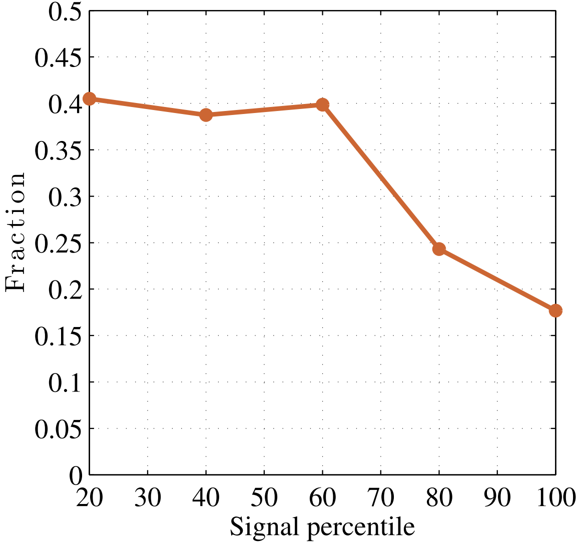
	Mean q		Standard deviation		
Test score quartile	Model	Data	Model	Data	N
1	-2,934	-2,266 (678)	5,075	5,253	60
2	-1,362	-1,741 (454)	4,896	6,121	182
3	-560	-509 (308)	4,924	5,692	341
4	-173	-20 (253)	4,764	5,704	510

Dropouts: Money



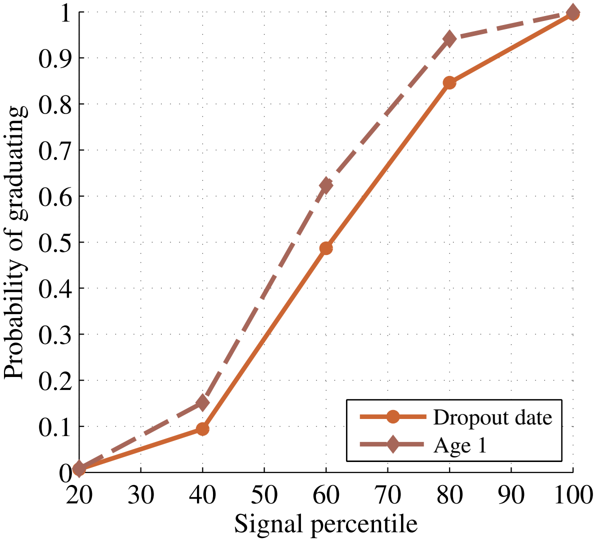
Consumption and college costs among dropouts

Dropouts: Money



Fraction of college students who choose c so high that graduation is not

Dropouts: Luck



Beliefs among dropouts

Dual Enrollment Programs

Experiment:

- allow high school students to take 40% of an annual course load

Result:

- almost no change in college attendance or completion

Reasons:

1. For many students, college entry decisions are not sensitive to more precise information about ability
   Very high (low) ability students almost always (never) try college
2. The rate of learning is slow
   The experiment does not change beliefs much

Dual Enrollment Programs

College enrollment	Beliefs