Information and Preferences

in Household Demand for School Value Added

Robert Ainsworth, Rajeev Dehejia, Kiki Pop-Eleches, Miguel Urquiola

July 2020

Introduction

- How do households choose schools?
  - Do they favor those with the highest value added?

- Analogous questions arise in recent work; do households choose:
  - Neighborhoods that boost their children's adult outcomes?

    Chetty, Friedman, Hendren, Jones, Porter 2020
  - Effective hospitals?

    Chandra, Finkelstein, Sacarny, Syverson 2016
  - Schools with high value added?

    Abdulkadiroglu, Pathak, Schellenberg, Walters 2020; Chetty, Friedman, Saez, Turner, Yagan 2020

- This work suggests households leave good options "on the table"
  - "Bargain" neighborhoods combine good outcomes/low housing prices
  - Some "mid tier" colleges produce good outcomes

Introduction

- Why do hhlds. not always favor options that economists deem productive?
  - They might *prefer* other dimensions of service quality
  - They might lack *information*, i.e., they do not observe productivity

- Distinguishing between **preferences** and **information** is important
  - If information is the issue, policy could improve household choices
  - It would also incentivize providers to compete on value added

- To explore this, we study high school admissions in Romania
  - Use admin. data, a survey, and an experiment

  → Info. has a positive, modest effect on the VA of schools hhlds. use

  → Preferences constrain the demand for value added

Four questions

1. Do households choose high value added schools?
   - Leverage admin. data to get new type of descriptive results

   → Students could access schools with 1 s.d. worth of additional VA

2. Do households have accurate beliefs regarding value added?
   - Leverage admin. & survey data to get new type of descriptive results

   → Beliefs are imperfect; there is substantial room for improvement

3. Does providing information on value added change their choices?
   - Leverage experiment to get causal results

   → Information has a modest effect on the VA of schools hhlds. use

   → Driven by hhlds. unable to access their top choices

4. Do household preferences enhance/limit the impact of such policy?
   - Leverage survey to exploit new type of data

   → Households prioritize traits in addition to VA

   → Correcting all beliefs would increase demand for VA, but only partially

Relation to literature

- Value added and the effects of school choice
  - Friedman 1955; Hoxby 2000, 2002; Rothstein 2006; Abdulkadiroglu, Pathak, Schellenberg, Walters 2020; Figlio, Hart, Karbownik 2020
- Information and school markets
  - Hastings, Weinstein 2008; Imberman, Lovenheim 2016; Andrabi, Das, Khwaja 2017; Corcoran, Jennings, Cohodes, Sattin-Bajaj 2018; Allende, Gallego, Neilson 2020; Bergman, Chetty, Deluca, Hendren, Katz, Palmer 2020; Bergman, Chan, Kapor 2020
- Information and educational choices
  - Rockoff, Staiger, Kane, Taylor 2011; Hastings, Neilson, Zimmerman 2018; Ajayi, Friedman, Lucas 2019
- Measurement and validation of value added
  - Bacher-Hicks, Kane, Staiger 2014; Chetty, Friedman, Rockoff 2014; Angrist, Hull, Pathak, Walters 2017
- Multidimensionality of school attributes
  - Hoxby 2009; Beuerman, Jackson, Navarro-Sola, Pardo 2019; Kraft 2019; Gilraine, Petronijevic, Singleton 2019; Riehl, Saavedra, Urquiola 2019

Outline

- Preliminaries

1. Do households choose high value added schools?
2. Do households have accurate beliefs regarding value added?
3. Does providing information on value added change their choices?
4. Do household preferences enhance/limit the impact of such policy?

Outline

- **Preliminaries**

1. Do households choose high-value added schools?
2. Do households have accurate beliefs regarding value added?
3. Does providing information on value added change their choices?
4. Do household preferences enhance/limit the impact of such policy?

High school admissions

- Romanian students have choice over high school *tracks*
  - Self-contained units within high schools
  - Have a set curricular focus (e.g., lang., natural sciences, technical)

- Students gain admission based on achievement in middle school
  - *Transition score*: grade 5-8 GPA + score on national entrance exam

- Students are assigned to tracks via a serial dictatorship
  - A household submits a *preference ranking* over tracks in its town
  - Government allocates students in order of their transition scores
  - Student gets assigned to her most-preferred *feasible* track

  $\rightarrow$ Dominant for households to truthfully reveal preferences

Administrative data

- Student-level data on high school admissions and achievement
- Includes:
  - Track assignment
  - Covariates – demographics, transition score
  - Performance on national *exit* exam ("baccalaureate exam")

- Use to:
  1. Calculate value added for each track
  2. Examine whether households attend high-VA tracks

- Data on the 2004-2019 admissions cohorts
  - Average cohort: 142,000 students choosing among 3,800 tracks

Value added, V_{jt}

- Calculate a track's effect on *baccalaureate exam* outcomes
  - Exam is optional for 12th graders
  - High-stakes: required for diploma, college
- Selection into exam-taking $\Rightarrow$ focus on VA on *passing* the exam
  - Main results unchanged if use VA on exam score
- **Estimation**: conventional "selection-on-observables" approach

$$p_i = \gamma_t' \cdot f(X_i) + \sum_j V_{jt} \cdot d_{ij} + e_i, \quad \text{where:}$$

$p_i = \mathbb{1}\{i \text{ passes bacc.}\}$, $d_{ij} = \mathbb{1}\{i \text{ attends track } j\}$, X_i are covariates

  - Results are robust to using a logit
- **Validation**: use admissions-cutoff RDs from serial dictatorship
  - Adaptation of Angrist, Hull, Pathak, Walters 2017 link
- **Forecast** VA for recent cohorts using a local linear forest
  - Athey, Friedberg, Tibshirani, Wager 2019 link

$\rightarrow$ 1 s.d. $\uparrow$ in VA $\approx$ 10 %-point $\uparrow$ in probability of passing the bacc.

Baseline survey

- Visited middle schools in May 2019; one month before allocation
  - Interviewed parents of 8th-graders during "application nights"

- Sample selection:
  - Moderately sized towns (7-28 tracks) with predictable VA (R-sq.)

  $\rightarrow$ 3,898 students in 228 middle schools, 49 towns

- Collected households'
  1. Intended track preference rankings
  2. Beliefs about VA and track characteristics; asked for 1-5 scores, e.g.:

Quality dimension	**This track . . .**
Value added	will help my child pass the baccalaureate exam
Peer quality	attracts academically gifted students
Location	is conveniently located for me
Curriculum	has a curriculum my child will enjoy

Experiment

- Randomly assigned middle schools to treatment, control groups
  - Well balanced
- Finished baseline survey by giving parents flyers:
  - Treatment and control: link to government admissions website
    - Lists prior-year minimum transition scores, MTS_{jt-1}
  - Treatment: an additional page with:
    1. Explanation of VA.[1]
    2. A ranking of tracks in the town by predicted VA, V^P_{jt}
- After the allocation, we:
  - Obtained track assignments from the Ministry of Education
  - Conducted an endline phone survey asking parents:
    - For submitted track preference rankings
    - To again score tracks on quality dimensions (1-5 scale)

[1] A high-VA track "effectively improves a student's chance of passing the baccalaureate exam relative to his or her 9^{th}-grade starting point"

Outline

- Preliminaries

1. **Do households choose high-value added tracks?**
2. Do households have accurate beliefs regarding value added?
3. Does providing information on value added change their choices?
4. Do household preferences enhance/limit the impact of such policy?

1. Do households choose tracks with high value added?

- Use administrative data to describe *outcomes* of choice behavior
  - Analysis doesn't reveal whether hhlds. make choices *based on* VA
- Aim to see:
  1. Do households gain academic benefits from their choices?
  2. By how much could they increase their benefits by switching choices?

1. Households' choices within their feasible sets

- How do households choices compare with the available options?

	Value
Percent of students attending the best feasible track	
Value added, V_{jt}	17.3
Selectivity, MTS_{jt}	32.5
Mean percentile rank of student's track among feasible tracks	
Value added, V_{jt}	70.0
Selectivity, MTS_{jt}	81.5
Mean potential increase (in std. dev.) among feasible tracks	
Value added, V_{jt}	1.05
Selectivity, MTS_{jt}	0.34
Number of students	2,158,020

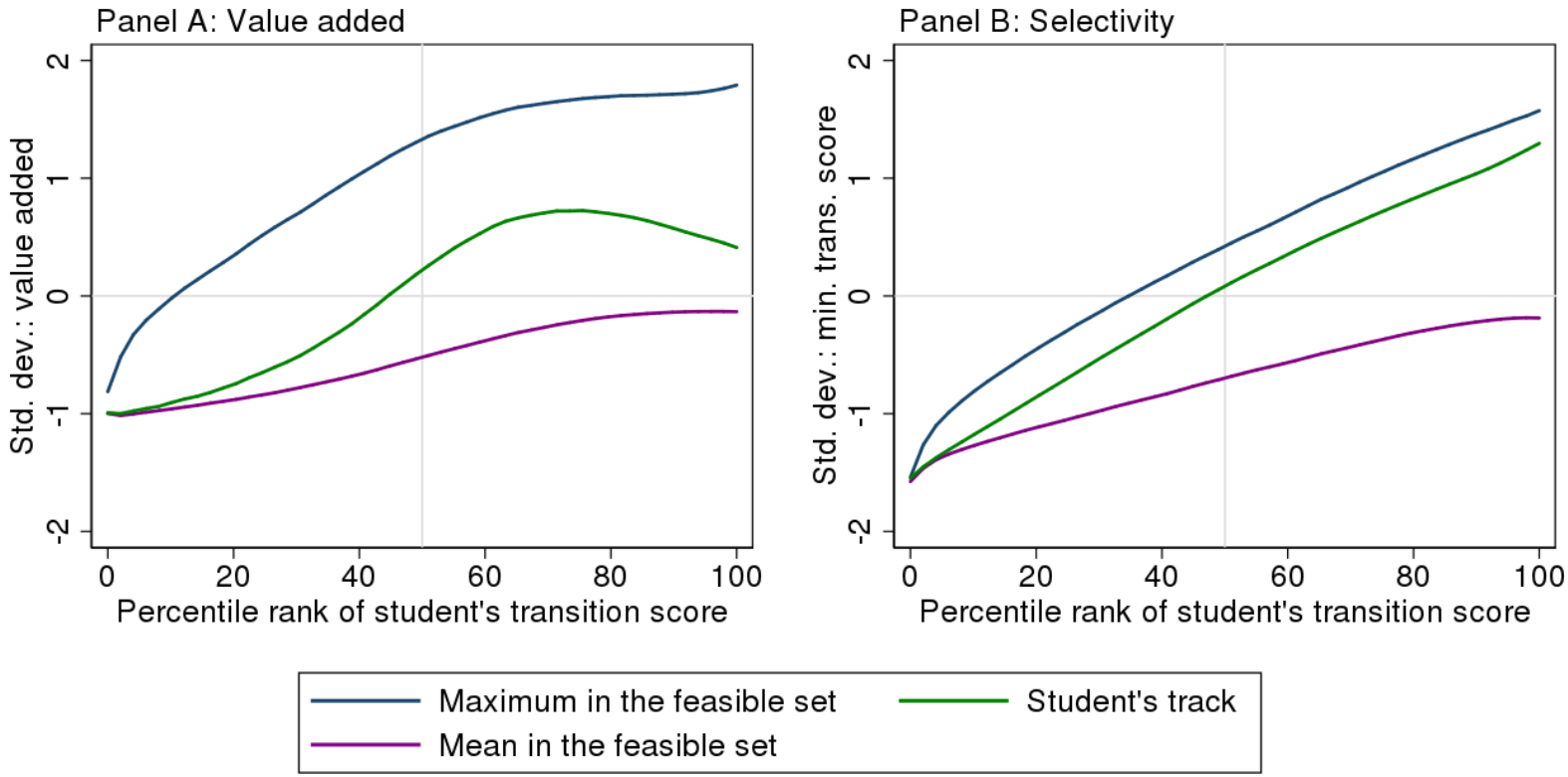
→ Households choose above-average tracks by VA

→ But can substantially increase VA by switching tracks

- Less room for increase in selectivity

1. Heterogeneity in choice outcomes by student achievement

- Does the amount by which households can ↑ VA or selectivity vary by transition score?



→ Large and stable potential increase in VA across the distribution

- Small and stable potential increase in selectivity

Outline

- Preliminaries

1. Do households choose high-value added tracks?
2. **Do households have accurate beliefs regarding value added?**
3. Does providing information on value added change their choices?
4. Do household preferences enhance/limit the impact of such policy?

2. Do hhlds. have accurate beliefs on value added?

- Use households' quality scores from the survey data
- Compare measured values of VA with households' scores for VA
  - Same for measured selectivity and hhlds.' scores for peer quality
- Quality scores on 1-5 scale

  $\Rightarrow$ compare with within-town *quintiles* of true values

2. Regressions of quality scores on true values

- Regress scores on quintiles of measured traits:

$$s_{ij}^{V} = \alpha_{0,V} + \alpha_{1,V} \cdot \text{quintile}(V_{jt}^{P}) + \alpha_{ij,V}$$
$$s_{ij}^{PQ} = \alpha_{0,PQ} + \alpha_{1,PQ} \cdot \text{quintile}(MTS_{jt-1}) + \alpha_{ij,PQ}$$

	All students		**Low-achieving**		**High-achieving**	
	Score: pass	Score: peers	Score: pass	Score: peers	Score: pass	Score: peers
Value added, V_{jt}^{P} (quintile)	0.427***		0.358***		0.473***	
	(0.0172)		(0.0209)		(0.0188)	
Prior-year selectivity, MTS_{jt-1} (quintile)		0.573***		0.465***		0.643***
		(0.0162)		(0.0223)		(0.0167)
R-sq.	0.19	0.33	0.15	0.23	0.21	0.39
Clusters	188	188	171	171	177	177
Students	2,370	2,370	891	891	1,479	1,479
Student-tracks	17,460	17,460	6,483	6,483	10,977	10,977

→ Hhlds. have significant but imperfect information on VA

- Substantially more on selectivity

2. Do hhlds. have accurate beliefs on value added?

- Further analysis: link
  - → Hhlds. know if track is good (top-3/5^{ths}) or bad (bot.-2/5^{ths}) by VA
    - Right 75% of time
  - → But do not know track's exact quintile
    - Right 31% of time

⇒ Significant room to influence households' beliefs

Outline

- Preliminaries

1. Do households choose high-value added tracks?
2. Do households have accurate beliefs regarding value added?
3. **Does providing information on VA change their choices?**
4. Do household preferences enhance/limit the impact of such policy?

3. Does information on value added change hhlds.' choices?

Overview

- Use the experiment to examine the effect of information
- Main results are on track assignments – "real-world" outcome
  - Use endline beliefs and preference rankings to probe mechanisms

Estimation

- Obtain main results using simple treatment effects regression:

$$Y_i = \eta_0 + \eta_1 \cdot T_i + \eta_X{}' \cdot X_i + \eta_i$$

  - Y_i: an attribute of the track i *actually* attends
  - T_i: an indicator for treatment
  - X_i: a vector of controls

$\rightarrow$ η_1 is the average treatment effect

3. ATEs on the VA of students' assigned tracks

- Treatment $\Rightarrow$ low-achieving students attend tracks with higher VA
  - 0.125 s.d. $\Leftrightarrow$ 1.1 %-point $\uparrow$ in probability of passing the bacc.

	Value added (s.d.)		
	All students	Low-achieving	High-achieving
Treated	0.046*	0.125**	-0.005
	(0.0249)	(0.0529)	(0.0213)
Effect in percentage points	0.41	1.13	-0.04
Predicted pass rate (%)	65.2	31.0	83.3
Clusters	81	80	80
Students	2,692	932	1,760

- No effect for high-achieving students
- Tradeoffs? – no effects for any students on: link
  1. track selectivity
  2. household's location score

3. Breaking down the ATE

- In follow-up survey, find treatment
  - ↑ accuracy of beliefs regarding VA
  - ↑ the correlation between VA and preference rankings

  → But only for tracks *other than a household's top two baseline choices*

⇒ TE driven by hhlds. who were unable to access top choices:

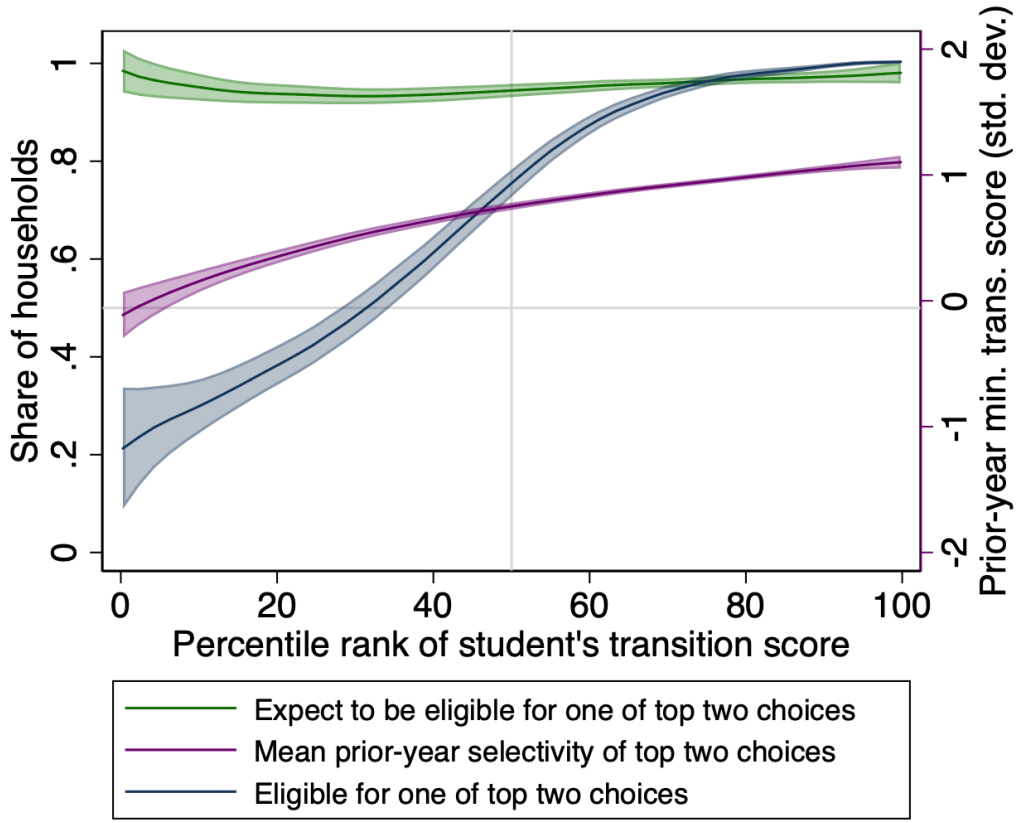
	Eligible for x^{th} most-preferred track in the baseline					
	Most-preferred	2nd-most-preferred	≥ 3rd-most-preferred	≥ 4th-most-preferred	≥ 5th-most-preferred	≥ 6th-most-preferred
Treated	-0.020	0.016	0.257***	0.251***	0.226**	0.296***
	(0.022)	(0.069)	(0.074)	(0.080)	(0.091)	(0.105)
Effect in %-points	-0.18	0.15	2.31	2.26	2.03	2.67
Predicted pass rate	75.8	52.0	38.4	37.4	37.2	37.0
Clusters	80	72	78	77	75	74
Students	1,824	280	588	465	388	338

→ Serial dictatorship ⇒ low-achieving students eligible for fewer tracks

⇒ more likely to be ineligible for top choices ⇒ larger TE

3. Why did hhlds.' beliefs change only for their less preferred tracks?

- $>$ 95% expected to be eligible for one of top two baseline choices
  - $\rightarrow$ more willing to change beliefs on tracks that expect to be irrelevant?



- Many households were overconfident $\Rightarrow$ TE

3. TE heterogeneity by student characteristics

- Students ineligible for top choices: TE is robust across sub-groups

	Ineligible for top two baseline choices						
	All	*Gender*		*Transition score*		*Mother's schooling*	
		Female	Male	Bot. $1/3^{rd}$	Top $2/3^{rds}$	$\leq$ 12 years	$>$ 12 years
Treated	0.257***	0.355***	0.173*	0.262***	0.223**	0.231**	0.308***
	(0.074)	(0.095)	(0.101)	(0.088)	(0.110)	(0.095)	(0.109)
Clusters	78	72	69	72	63	76	65
Students	588	278	310	278	310	380	208

- Students who were eligible: TE is always 0

	Eligible for at least one of top two baseline choices						
	All	*Gender*		*Transition score*		*Mother's schooling*	
		Female	Male	Bot. $2/3^{rds}$	Top $1/3^{rd}$	$\leq$ 12 years	$>$ 12 years
Treated	-0.016	-0.010	-0.025	-0.015	-0.014	-0.008	-0.022
	(0.024)	(0.023)	(0.036)	(0.039)	(0.020)	(0.034)	(0.024)
Clusters	81	81	80	81	78	81	80
Students	2,104	1,157	947	908	1,196	1,000	1,104

Outline

- Preliminaries

1. Do households choose high-value added tracks?
2. Do households have accurate beliefs regarding value added?
3. Does providing information on value added change their choices?
4. **How much do preferences limit the response to information?**

4. How much do preferences limit the response to information?

- Treatment only partially influenced beliefs
  - → How would households' choices change if they had fully accurate beliefs about VA & selectivity?

- Two steps:
  1. Use baseline survey to explore hhlds.' preferences for track attributes
     - Explain track preference rankings using track quality scores
  2. Simulate track choices under accurate beliefs.[2]
     - Replace hhlds.' quality scores with w/in-town quintiles of true values

[2] Keep feasible sets constant – ignore effects on capacity constraints

4. Explaining hhlds.' preference rankings using their quality scores

- i's utility from track j is a function of its scores on dimensions q:

$$U_{ij} = \sum_q \beta_q \cdot s_{ij}^q + \epsilon_{ij}$$

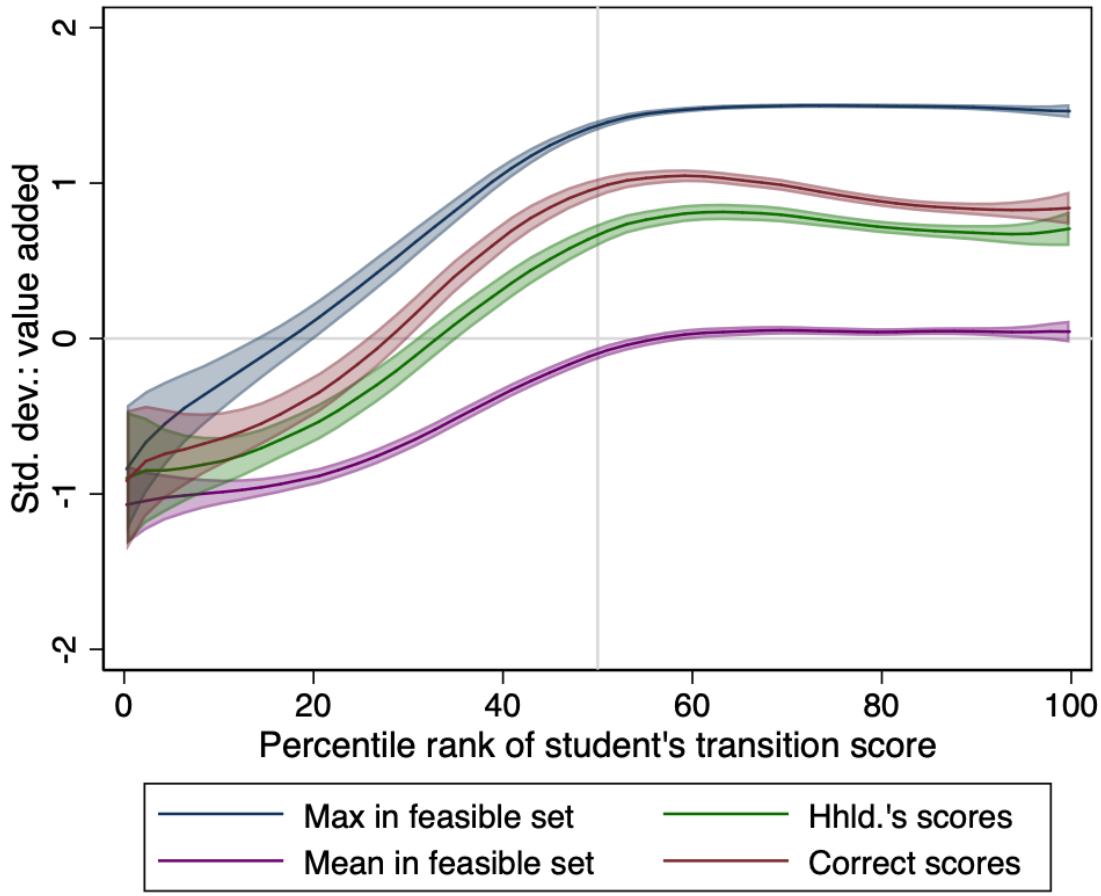
  - Estimate using a rank-ordered logit:

	All students	**Low-achieving**	**High-achieving**
Location	0.529***	0.399***	0.725***
	(0.094)	(0.131)	(0.126)
Peer quality	0.407***	0.012	0.901***
	(0.105)	(0.112)	(0.140)
Curriculum	1.20***	1.02***	1.38***
	(0.134)	(0.182)	(0.158)
VA: pass the bacc.	0.452***	0.503***	0.326*
	(0.140)	(0.163)	(0.167)
R-sq.	0.37	0.27	0.46
Clusters	129	98	116
Students	849	332	517
Student-tracks	10,911	4,327	6,584

→ Households care about multiple track characteristics

4. Simulating track choices under accurate beliefs

- Given preferences, how would track choices change if hhlds. had correct scores for VA and selectivity?



→ Students attend tracks with higher VA:

- Low-achieving: 0.27 s.d. High-achieving: 0.18 s.d.
- But they also leave a lot on the table

Conclusion

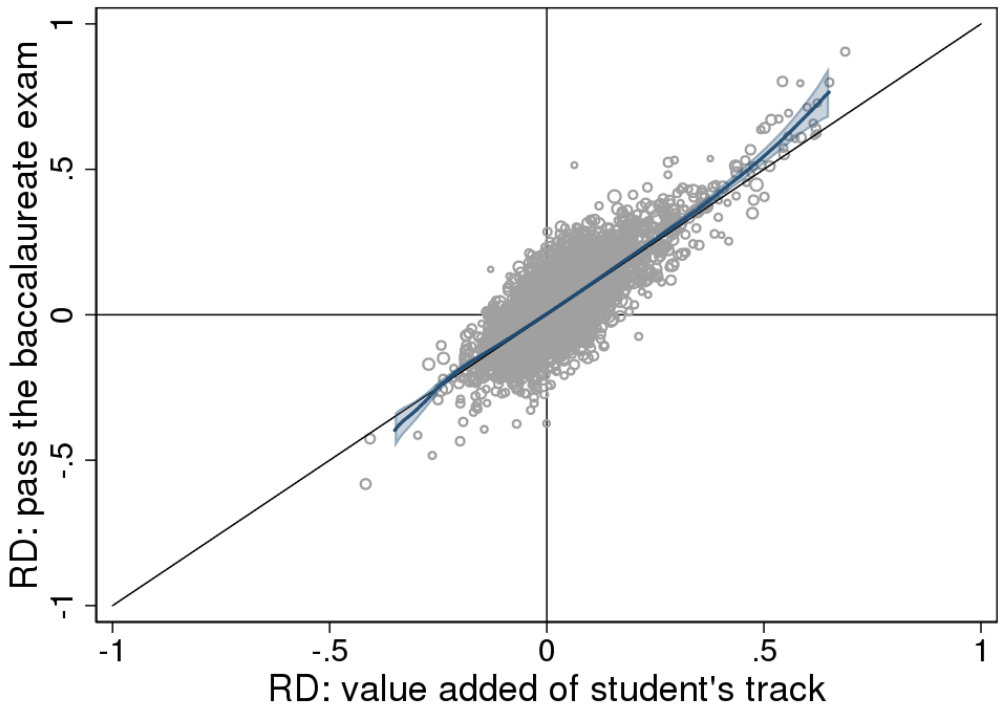
- Why do households not always favor providers that economists deem productive?
  - They might **prefer** other dimensions of service quality
  - They might lack **information** about productivity

- We explore this distinction in the context of school choice
- Our findings suggest:
  - Providing information has a positive but overall modest effect
  - Preferences somewhat limit the demand for productive options
  - Households have strongly held beliefs about their top options

Validating value added

- Calculate admissions-cutoff RDs for:
  1. whether the student passes the bacc.
  2. the value added of the student's track

→ If V_{jt} is a track's constant causal effect, then RDs should be equal



- Best fit line: slope = 1.01 (s.e. = 0.012); raw correlation = 0.80

Back

Forecasting value added, V_{jt}^P

- In experiment, aim to inform households about track VA *for their child's admissions cohort*
  - But not observed until students finish high school

$\Rightarrow$ Forecast track VA based on past values & other track covariates

- Use a local linear forest: Athey, Friedberg, Tibshirani, Wager 2019

Years	**Standard deviation**		**R-sq.**	**Towns**	**Tracks**	**Students**
	V_{jt}	V_{jt}^P				
2004-2007	0.165	-	-	1,600	13,734	603,458
2008-2014	0.128	0.112	0.827	2,976	30,132	1,106,572
2015-2019	-	0.091	-	1,735	16,784	558,091

$\rightarrow$ For 2008-2014, V_{jt}^P predicts 83% of variation in V_{jt}

- (75% if don't adjust for measurement error)
- Much more predictive than using the mean of prior VA

Back

Experiment balance tests

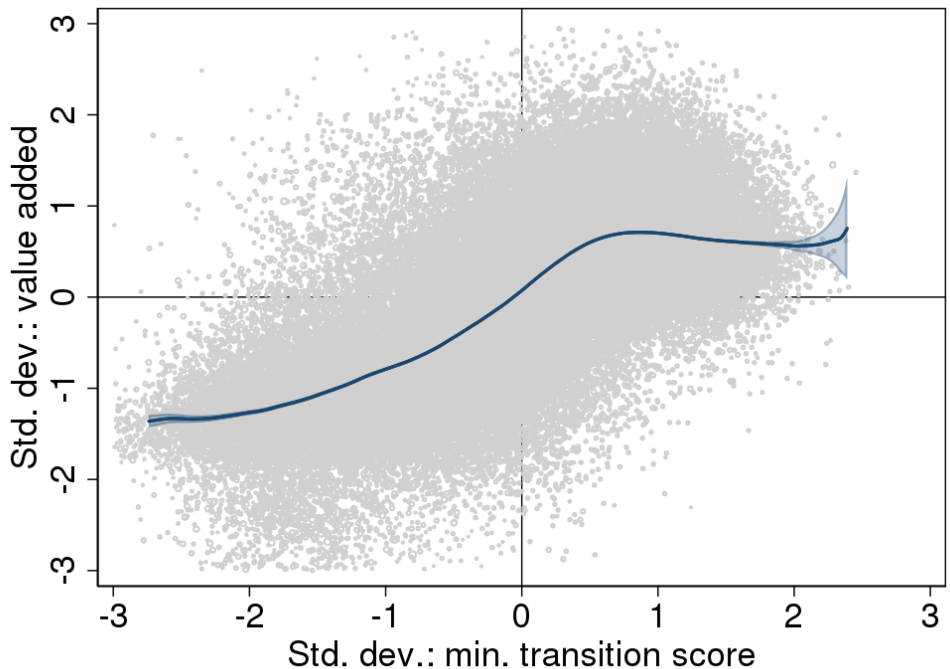
- Experiment is well balanced:

Covariate	Summary statistics		Balance tests			
	Mean	Std. dev.	Coef.	Std. error	Clusters	N
Matched with the administrative data	0.829	0.377	0.034	0.021	81	3,540
Assigned to a track	0.816	0.388	0.036*	0.021	81	3,540
In the follow-up survey	0.556	0.497	-0.023	0.025	81	2,933
Student demographics:						
Female	0.528	0.499	0.016	0.020	81	2,933
Mother's years of schooling	12.3	2.0	0.079	0.102	81	2,856
Parents not married	0.136	0.343	-0.011	0.015	81	2,741
High school application process:						
Num. of tracks in the town	13.1	4.7	0.097	0.353	81	2,933
Share of tracks ranked	0.464	0.318	-0.017	0.028	81	2,933
Share of tracks scored on peer quality	0.408	0.425	-0.010	0.031	81	2,933
Share of tracks scored on passing the bacc.	0.396	0.421	-0.018	0.031	81	2,933
Very certain of preference ranking	0.442	0.497	0.038	0.026	81	2,812
Somewhat certain of preference ranking	0.497	0.500	-0.024	0.022	81	2,812
Administrative data:						
Transition score	7.83	1.36	0.112	0.096	81	2,933
Middle school GPA	9.19	0.69	0.032	0.050	81	2,933
Transition exam score: math	6.86	1.84	0.096	0.127	81	2,933
Transition exam score: language	8.07	1.56	0.135	0.111	81	2,933

Back

1. The aggregate reln. between value added and selectivity

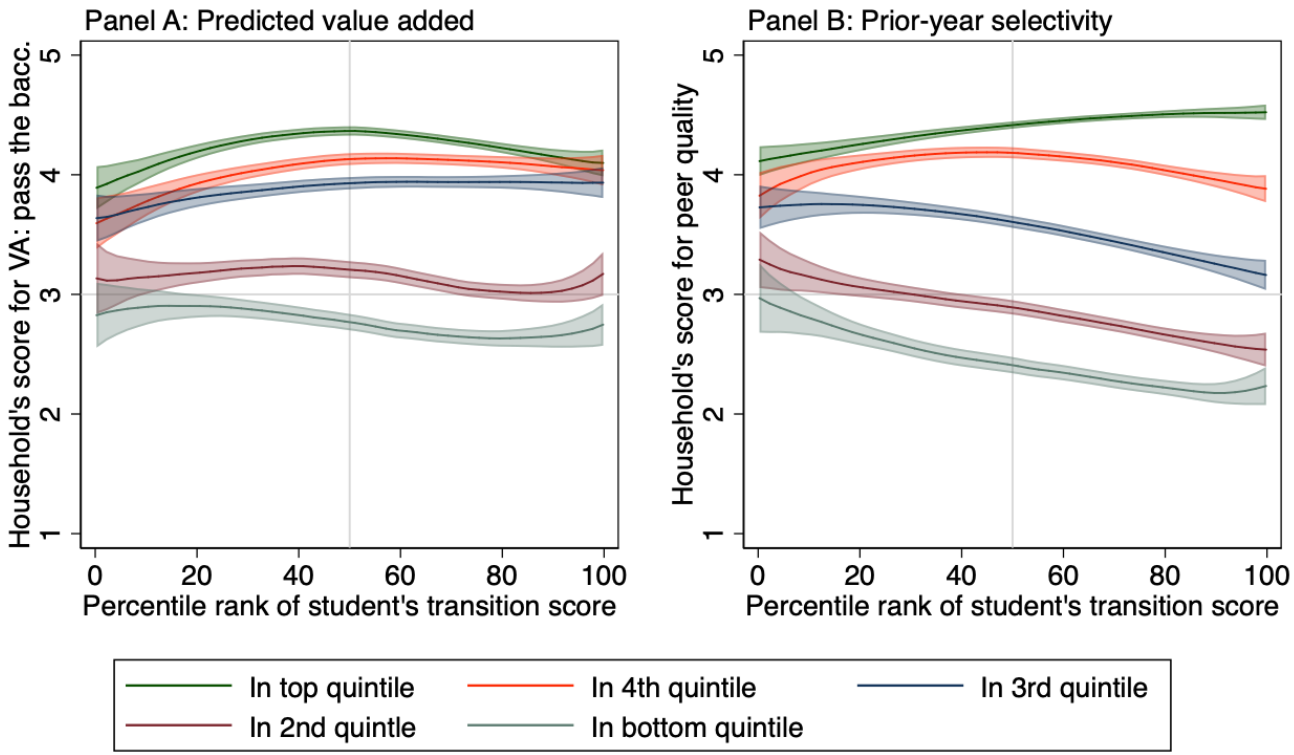
- In serial dictatorship, selectivity reflects demand
- Positive relationship between V_{jt}, and min. trans. score, MTS_{jt}
  - Over full sample, correl. of ~ 0.7



- But effect entirely driven by tracks in bottom two terciles of MTS_{jt}

2. Accuracy by track quintile and student achievement

- Households' quality scores by quintiles of the track's true values
  - Heterogeneity by student achievement



→ Hhlds. have less information on VA than on selectivity

- High-achieving hhlds. have substantial info. on selectivity

Back

3. ATEs on the characteristics of students' assigned tracks

- Treatment doesn't cause students to attend tracks with different selectivity or location quality:

	Selectivity (s.d.)			**Location (1-5)**		
	All students	Low-achieving	High-achieving	All students	Low-achieving	High-achieving
Treated	0.004	0.027	-0.014	0.020	0.055	0.011
	(0.019)	(0.037)	(0.016)	(0.023)	(0.068)	(0.017)
Clusters	81	80	80	81	73	79
Students	2,692	932	1,760	1,679	459	1,220

$\rightarrow$ No evidence of tradeoffs

Back