

Winners and Losers? The Effect of Gaining and Losing Access to Selective Colleges on Education and Labor Market Outcomes

Sandra E. Black
Jeffrey T. Denning
Jesse Rothstein

Online Appendix

LASSO estimator

Our primary analyses rely on a somewhat ad hoc rule for defining the Pulled In, Pushed Out, and control groups. As an alternative, we also present results that use a LASSO procedure (James et al. 2013, pp. 219-228) to select groups. This appendix describes that procedure.

We begin by assigning each observation to an s-q cell. Let y_{sq} represent the change in the share of students in cell s-q who attend UT Austin between 1996-1997 and 1998-2002, and let X^{sq} be an indicator for cell s-q. Our baseline algorithm is based on the un-smoothed y_{sq} surface. However, the individual cell values of y_{sq} are noisily estimated, producing the volatile patterns seen in Figure 5C and in, presented differently, in Panel A of Appendix Figure 1. We use the LASSO estimator to smooth this surface.

Our starting point is a simple saturated regression of y_{sq} on a full set of s-q indicators:

$$y_{sq} = \alpha + \sum_{(s',q') \neq (s_0,q_0)} X_{sq}^{s'q'} \beta^{s'q'} + e_{sq}. \quad (A1)$$

Here, (s_0, q_0) represents a base category, which we define as $(s_0, q_0) = (5, 25)$. Because (A1) is saturated, the coefficient estimates are simply $\hat{\alpha} = y_{s_0q_0}$ and $\hat{\beta}^{sq} = y_{sq} - y_{s_0q_0}$, and the fitted values exactly equal the unsmoothed y_{sq} :

$$\hat{y}_{sq} = \hat{\alpha} + \hat{\beta}^{sq} = y_{sq}. \quad (A2)$$

We use a LASSO penalty to smooth this. Specifically, the LASSO coefficients are the solution to the following problem:

$$\min_{\alpha, \{\beta^{sq}\}} \sum_{s,q} \left[(y_{sq} - \alpha - \beta^{sq})^2 + \lambda |\beta^{sq}| \right]. \quad (A3)$$

With $\lambda = 0$, this is merely the least squares criterion, and the solutions are the coefficients of the un-smoothed regression (A1). When $\lambda > 0$, however, the β^{sq} coefficients are smoothed toward zero. As James et al. (2013) note, this form of penalty forces some coefficients to be exactly zero, while others are shrunken relative to their OLS values.

A drawback of specification (A3) is that the fitted values will remain highly “spiky” – when a coefficient is set to zero, the fitted value is simply $\hat{\alpha} = y_{s_0q_0}$, which may be quite different from those of adjacent cells. To avoid this problem, we reparameterize the model to permit a smooth fitted surface. Specifically, we define a new set of variables Z^{sq} as follows:

$$Z^{sq} = \begin{cases} \sum_{s',q'} \mathbf{1}(s' \geq s, q' \geq q) X^{s'q'} & \text{if } s \geq s_0, q \geq q_0 \\ \sum_{s',q'} \mathbf{1}(s' \geq s, q' \leq q) X^{s'q'} & \text{if } s \geq s_0, q < q_0 \\ \sum_{s',q'} \mathbf{1}(s' \leq s, q' \geq q) X^{s'q'} & \text{if } s < s_0, q \geq q_0 \\ \sum_{s',q'} \mathbf{1}(s' \leq s, q' \leq q) X^{s'q'} & \text{if } s < s_0, q < q_0 \end{cases} \quad (A4)$$

These saturate the s-q space just as do the X^{sq} variables, but their coefficients have a different interpretation: Where β^{sq} represented the fitted value for cell s-q relative to the omitted

category, the γ^{sq} coefficients represent “steps” upward or downward as one moves away from the omitted category, and the fitted value for cell s-q is the sum of all steps on this route. When estimated without penalization, the surface will be identical to that obtained via (A1), but the penalized estimates are different: Here, when LASSO sets a particular γ_{sq} to zero, this eliminates a step, making the predicted value for that cell similar to those for the adjacent cells closer to the omitted category.ⁱ

We use a two-step process for generating the smoothed surface. First, we fit the LASSO regression of y_{sq} on the Z^{sq} variables. Second, we identify the subset of non-zero γ^{sq} coefficients in the LASSO fit, and we estimate a simple OLS regression of y_{sq} on the corresponding subset of Z^{sq} s. This OLS-after-LASSO approach means that we use the LASSO procedure only for variable selection, and that the coefficients that are estimated are not shrunk. We use the fitted values from this regression to define treatment groups: The Pulled In group is the set of cells with fitted values greater than 0.003, and the Pushed Out group is the set of cells with fitted values less than -0.003. The control group consists of all cells with $\hat{q} \geq 25$ that are not included in either of the two treatment groups. We also present some analyses that use a continuous treatment measure; these use the fitted values from the OLS-after-LASSO model, rescaled to range from -1 to +1.

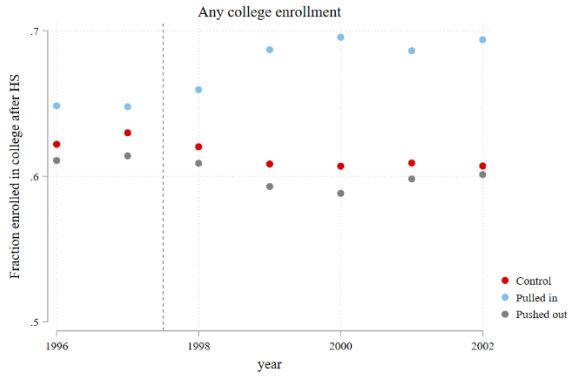
The final issue is the choice of the smoothing parameter λ . A larger λ will mean fewer non-zero coefficients and a smoother surface. We use a cross-validation approach to this. We construct two measures of the change in UT enrollment in each cell: y_{sq}^A is the difference between the 1996 and 1998 cohorts, and y_{sq}^B is the difference between the 1997 and 1999 cohorts. We choose the λ that minimizes the mean squared error when a LASSO model fit to y_{sq}^A is used to out of sample to predict y_{sq}^B , then fit the LASSO model on the full sample (where y_{sq} represents the change from 1996-1997 to 1998-2002) using the selected λ .

Panel B of Appendix Figure 1 shows the fitted values from our OLS-after-LASSO model, using the cross-validation choice of λ . Comparison to Panel A makes clear that the LASSO model substantially smooths the raw data while retaining the basic features of the data.

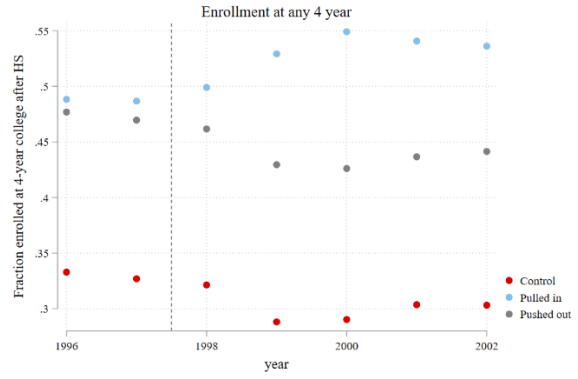
ⁱ Specifically, for $s > s_0$ and $q > q_0$ one can write $\hat{y}_{sq} = \hat{y}_{s-1,q-1} + \gamma_{s-1,q} + \gamma_{s,q-1} + \gamma_{sq}$ (with similar expressions in other quadrants). Thus, if LASSO sets $\gamma_{sq} = 0$ then the fitted value for cell s-q will be that for cell s-1,q-1 plus the steps associated with (s-1,q) and (s,q-1).

Appendix Figure 1: Event Studies in Levels

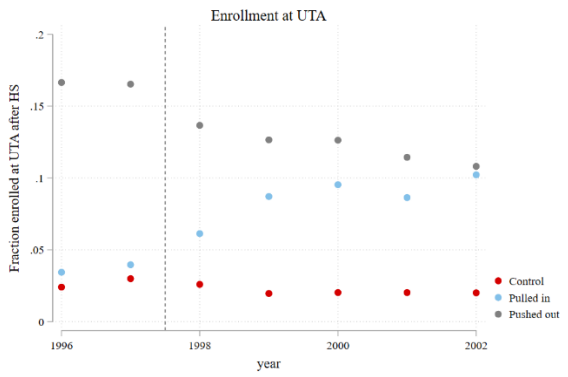
A. Any College Enrollment



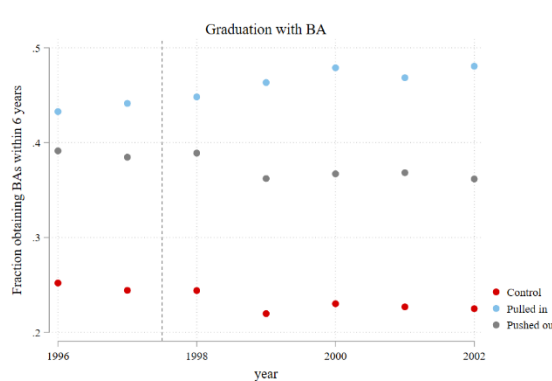
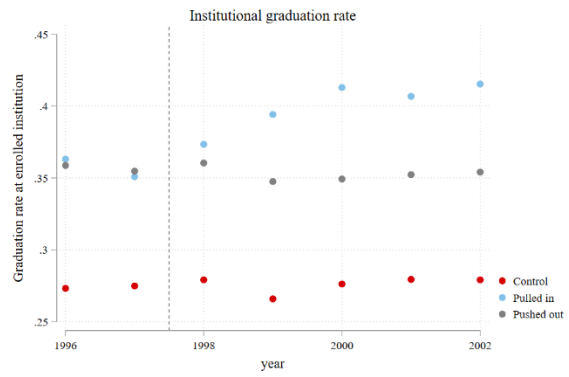
B. Any 4 year Enrollment

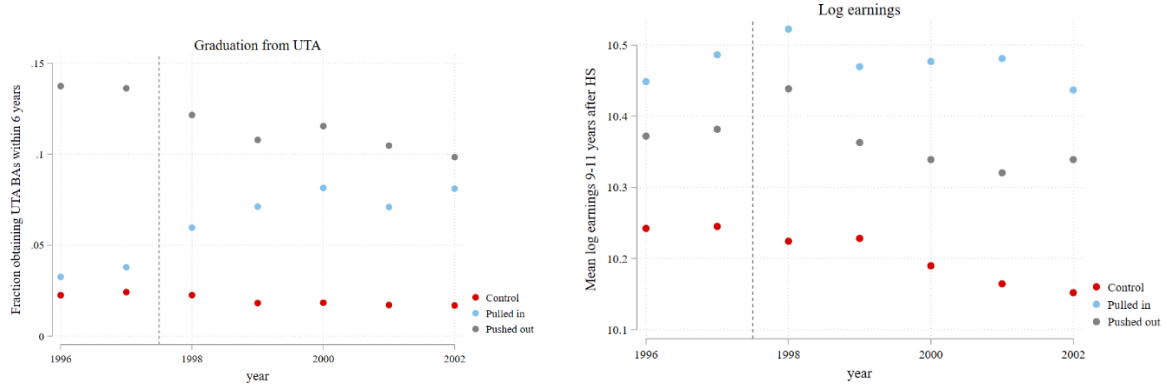


C. UTA Enrollment

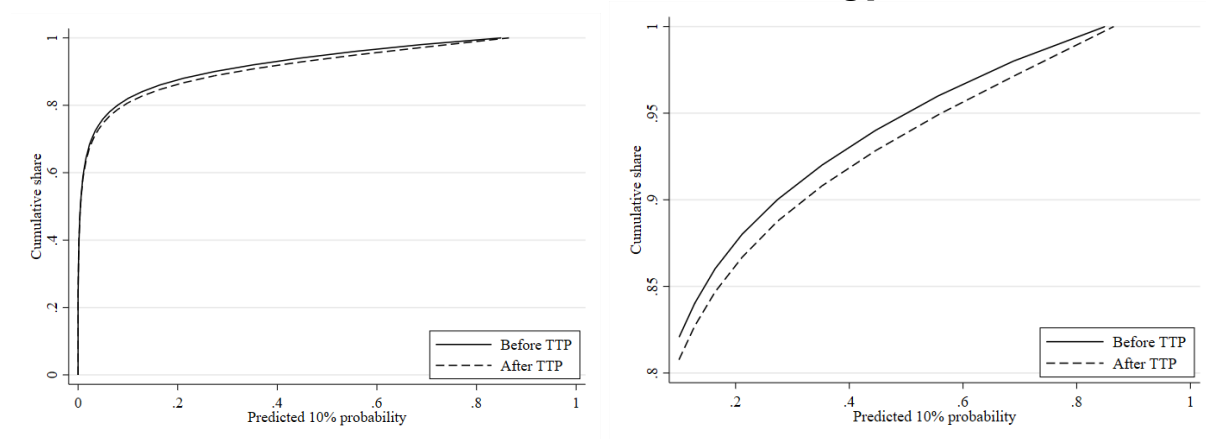


D. Institutional Graduation Rate





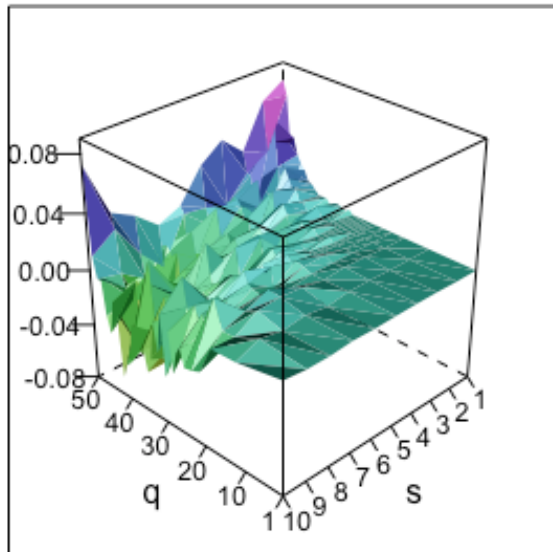
Appendix Figure 2. Distribution of predicted top ten probability (\hat{p}) before and after TTP
A. Full distribution **B. Excluding $\hat{p} < 0.1$**



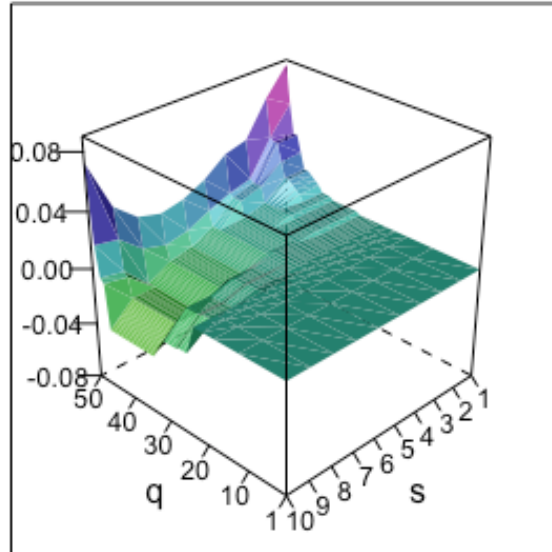
Notes: Figures show the CDF of estimated top-ten-percent probability (\hat{p}) across all students in the sample, separately for 1996-1997 (Before TTP) and 1998-2002 (After TTP). Predicted probabilities of being in the top ten percent are derived from a random forest model fit to 1999-2002 data. Right panel shows the portion of the CDF for $\hat{p} > 0.1$.

Appendix Figure 3. LASSO-smoothed changes in UT Austin enrollment shares

A. Raw

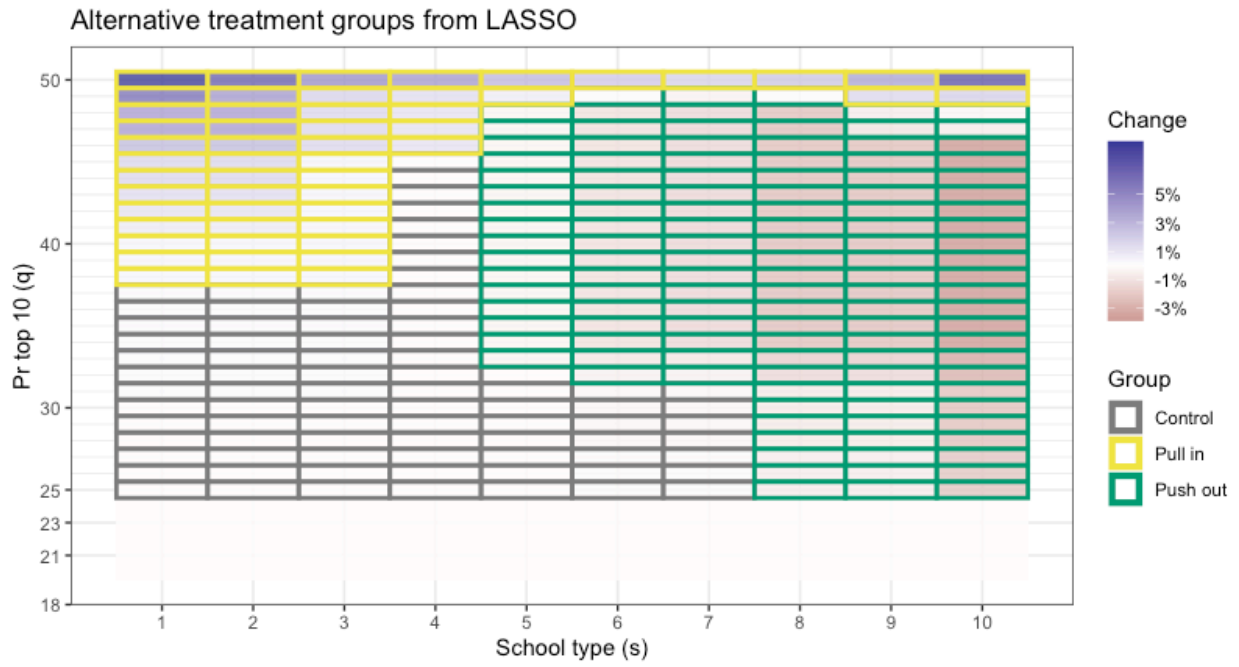


B. With LASSO regularization



Notes: Panel A shows the same data that is displayed in Figure 5A; the vertical axis is the change in the share of students in the cell who attended UT Austin between 1996-1997 and 1998-2002. Panel B presents estimates obtained by fitting a LASSO model to the data in Panel A, as described in the Appendix, then fitting OLS using only the explanatory variables whose estimated LASSO coefficients are not zero.

Appendix Figure 4. Pulled in, pushed out, and control groups using LASSO



Notes: Axes are defined as in Figure 5, though only the upper half ($\hat{q} \geq 18$) is shown. Outlines indicate treatment and control groups selected using the LASSO method described in the text.

Appendix Table 1. Predicting the Top 10%

	Summary statistics		Logit			Random forest predictions (OLS)	
	Mean	SD	Coeff.	SE	Marg. effect	Coeff.	SE
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Math score (statewide percentile/100)	0.51	[0.29]	0.35	(0.23)	0.025	0.067	(0.027)
Reading score (statewide percentile/100)	0.51	[0.29]	-0.02	(0.27)	-0.001	0.023	(0.020)
Writing score (statewide percentile/100)	0.51	[0.29]	0.08	(0.28)	0.005	-0.043	(0.016)
Math score (percentile within school/100)	0.46	[0.26]	3.05	(0.23)	0.220	0.222	(0.025)
Reading score (percentile within school/100)	0.46	[0.26]	2.32	(0.28)	0.167	0.271	(0.015)
Writing score (percentile within school)/100	0.46	[0.26]	2.09	(0.26)	0.151	0.162	(0.020)
Indicator for taking regular math in 9th grade	0.35	[0.48]	-0.47	(0.07)	-0.032	-0.056	(0.008)
Indicator for taking regular math in 10th grade	0.51	[0.50]	0.08	(0.04)	0.006	0.012	(0.005)
Indicator for taking regular math in 11th grade	0.52	[0.50]	0.06	(0.05)	0.004	-0.006	(0.004)
Indicator for taking regular math in 12th grade	0.27	[0.44]	0.39	(0.05)	0.029	0.025	(0.005)
Indicator for taking advanced math in 9th grade	0.14	[0.34]	-0.10	(0.07)	-0.007	-0.023	(0.011)
Indicator for taking advanced math in 10th grade	0.15	[0.36]	0.32	(0.06)	0.025	0.023	(0.007)
Indicator for taking advanced math in 11th grade	0.14	[0.34]	-0.13	(0.07)	-0.009	-0.013	(0.006)
Indicator for taking advanced math in 12th grade	0.07	[0.26]	0.84	(0.07)	0.078	0.266	(0.008)
Indicator for taking any math in 9th grade	0.51	[0.50]	-0.47	(0.07)	-0.036	-0.037	(0.009)
Indicator for taking any math in 10th grade	0.60	[0.49]	-0.48	(0.05)	-0.038	-0.036	(0.005)
Indicator for taking any math in 11th grade	0.61	[0.49]	0.13	(0.05)	0.009	0.015	(0.003)
Indicator for taking any math in 12th grade	0.35	[0.48]	0.19	(0.06)	0.014	0.017	(0.003)
Indicator for taking biology	0.86	[0.35]	0.08	(0.09)	0.006	0.025	(0.004)
Indicator for taking chemistry	0.57	[0.50]	0.13	(0.05)	0.009	-0.001	(0.003)
Indicator for taking physics	0.24	[0.43]	0.07	(0.04)	0.005	-0.014	(0.002)
Indicator for taking remedial algebra	0.17	[0.38]	-1.12	(0.11)	-0.056	-0.037	(0.005)
Number of foreign language courses taken	3.40	[1.99]	0.084	(0.009)	0.006	0.003	(0.000)
Number of days absent	8.35	[8.75]	-0.009	(0.018)	-0.001	-0.003	(0.000)
Percent of schools days absent (/100)	0.05	[0.05]	-8.90	(3.18)	-0.642	-0.260	(0.058)
Percent of school Black (/100)	0.12	[0.17]	1.13	(0.23)	0.082	0.074	(0.015)
Percent of school Hispanic (/100)	0.29	[0.30]	0.11	(0.24)	0.008	-0.005	(0.011)
Percent of school Asian (/100)	0.03	[0.05]	-3.47	(0.69)	-0.250	-0.263	(0.063)
Percent of school other minority race (/100)	0.00	[0.00]	12.00	(4.59)	0.866	0.944	(0.340)
Percent of school on Free/Reduced Lunch (/100)	0.22	[0.21]	0.79	(0.34)	0.057	0.075	(0.019)
Percent of school female (/100)	0.52	[0.03]	1.69	(1.40)	0.122	0.059	(0.090)
Percent of school English Lang. Learner (/100)	0.04	[0.07]	0.93	(0.78)	0.067	0.036	(0.041)
Percent of school Special Education (/100)	0.04	[0.02]	0.99	(0.98)	0.072	0.061	(0.077)
Indicator for being 18 years old	0.16	[0.37]	-0.10	(0.04)	-0.007	0.014	(0.002)
Number of courses failed	0.95	[1.70]	-0.59	(0.09)	-0.042	-0.006	(0.001)

Notes: N=195,710. Columns 6 and 7 present coefficients of an linear (OLS) regression of the predicted values from the nonlinear random forest model on the covariates.

Appendix Table 2. Stability of predicted top 10% probability across specifications

	RF	Logit	RF 1999- 2000	RF 2001- 2002	RF 25	RF 100
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. Full sample						
Random Forest	1					
Logit	0.94	1				
Random Forest (fit to 1999-2000 data)	0.98	0.93	1			
Random Forest (fit to 2001-2002 data)	0.99	0.95	0.96	1		
Random Forest (Leaf=25 obs.)	0.99	0.94	0.97	0.97	1	
Random Forest (Leaf=100 obs.)	0.99	0.93	0.98	0.98	0.97	1
Panel B. 1996-1997 (pre TTP)						
Random Forest	1					
Logit	0.95	1				
Random Forest (fit to 1999-2000 data)	0.99	0.94	1			
Random Forest (fit to 2001-2002 data)	0.99	0.95	0.97	1		
Random Forest (Leaf=25 obs.)	0.99	0.96	0.98	0.98	1	
Random Forest (Leaf=100 obs.)	0.99	0.93	0.98	0.99	0.97	1
Panel C. 1998-2002 (post TTP)						
Random Forest	1					
Logit	0.94	1				
Random Forest (fit to 1999-2000 data)	0.98	0.93	1			
Random Forest (fit to 2001-2002 data)	0.99	0.95	0.96	1		
Random Forest (Leaf=25 obs.)	0.99	0.93	0.97	0.97	1	
Random Forest (Leaf=100 obs.)	0.99	0.92	0.98	0.98	0.97	1

Notes: The baseline random forest model discussed in the text, fit to 1999-2002 data and using a "leaf" size of 50 observations, is labeled "random forest" here. Table shows correlations of predicted values across alternative models. Correlations are estimated at the individual level.

Appendix Table 3. Sensitivity to prediction model

Group: Prediction model:	Pulled In			Pushed Out		
	RF	RF	Logit	RF	RF	Logit
	1999-00 (1)	2001-02 (2)	(3)	1999-00 (4)	2001-02 (5)	(6)
Enrollment outcomes						
UT Austin	0.030 (0.007)	0.047 (0.006)	0.049 (0.004)	-0.027 (0.008)	-0.034 (0.006)	-0.036 (0.003)
Any college	0.017 (0.012)	0.042 (0.014)	0.040 (0.010)	0.010 (0.011)	-0.006 (0.009)	0.007 (0.007)
Any 4-year	0.018 (0.012)	0.065 (0.014)	0.049 (0.011)	0.005 (0.013)	-0.009 (0.011)	-0.006 (0.008)
Degree attainment within 6 years						
Bachelors from UT Austin	0.024 (0.006)	0.035 (0.006)	0.034 (0.003)	-0.013 (0.007)	-0.019 (0.007)	-0.023 (0.003)
Bachelors from any institution	0.012 (0.011)	0.044 (0.013)	0.036 (0.009)	0.007 (0.011)	0.002 (0.010)	0.002 (0.006)
Bachelors with STEM major	-0.011 (0.006)	-0.003 (0.008)	-0.004 (0.006)	0.010 (0.006)	-0.007 (0.004)	-0.006 (0.002)
Labor market outcomes 9-11 years after HS graduation						
Employment (0/1)	-0.011 (0.011)	0.004 (0.012)	0.003 (0.007)	0.024 (0.010)	0.007 (0.008)	0.016 (0.005)
Log average annual earnings (excluding 0s)	0.051 (0.024)	0.045 (0.027)	0.050 (0.018)	0.050 (0.021)	0.040 (0.025)	0.030 (0.013)
Labor market outcomes 13-15 years after HS graduation						
Employment (0/1)	-0.013 (0.011)	-0.008 (0.013)	-0.006 (0.008)	0.024 (0.009)	0.019 (0.011)	0.016 (0.006)
Log average annual earnings (excluding 0s)	0.003 (0.023)	0.004 (0.029)	0.020 (0.020)	0.022 (0.026)	-0.006 (0.024)	0.006 (0.014)

Notes: In columns 1 and 4, predictions of top ten percent status are from a random forest model fit to 1999-2000 data only, while only 1997, 2001, and 2002 data are used to fit the difference-in-differences model. In columns 2 and 5, the random forest model is fit to 2001-2002 data, and 1996, 1999, and 2000 data are used for the difference-in-differences. In columns 3 and 6, predictions are based on a logit model using 1999-2002 data, and all years are used in the difference-in-differences model.

Appendix Table 4. Summary statistics by decile of high school pre-TTP UT sending rate (s)

s	Demographics				Mean statewide percentile			Send to UT	
	Black	Hispanic	Free lunch	ELL	Math	Reading	Writing	Pre	Post
1	18%	28%	31%	3%	46.0	45.7	46.3	NR	0.7%
2	19%	44%	38%	9%	43.7	42.5	44.2	NR	1.1%
3	10%	38%	29%	4%	48.1	47.4	48.2	0.8%	1.2%
4	9%	37%	28%	5%	48.1	47.7	48.3	1.3%	1.5%
5	14%	39%	29%	4%	48.0	48.1	48.8	1.8%	1.9%
6	11%	27%	21%	3%	50.6	50.4	50.3	2.4%	2.2%
7	11%	30%	18%	3%	51.0	51.5	51.3	3.2%	3.0%
8	10%	21%	12%	2%	53.4	54.1	53.5	4.3%	3.7%
9	10%	24%	13%	3%	55.4	55.5	54.4	6.6%	5.8%
10	8%	12%	6%	2%	60.0	60.5	59.1	12.8%	10.9%

Notes: All columns present student-weighted means of school average characteristics, based (except in final column) on pre-TTP data. S=1 represents the decile of high schools with the lowest pre-TTP UT sending rates, and s=10 the decile with the highest rates. "ELL" represents English Language Learners.

Appendix Table 5. Difference-in-Differences Analysis, Including P=50, S=9 and P=50, S=10 Cells in Pulled-In Group

	Pulled In	Pushed Out
	(1)	(2)
Enrollment outcomes		
UT Austin	0.050 (0.004)	-0.036 (0.004)
Texas A&M	-0.006 (0.006)	0.008 (0.004)
Any college	0.056 (0.010)	0.000 (0.007)
Any 4-year	0.069 (0.010)	-0.006 (0.008)
Any Community College	-0.003 (0.008)	0.011 (0.005)
Any other 4-year	0.026 (0.008)	0.022 (0.006)
Characteristics of institution attended (fixed at pre-policy levels)		
Graduation rate (conditional on enrollment)	0.019 (0.005)	-0.006 (0.004)
Graduation rate (with non-enrollment as institution)	0.040 (0.006)	-0.006 (0.005)
Math state %ile (conditional on enrollment)	1.29 (0.31)	-1.15 (0.26)
Math state %ile (with non-enrollment as institution)	1.91 (0.30)	-0.84 (0.24)
Instructional expenditures per student (conditional on enrollment)	584 (105)	-409 (82)
Average log earnings in years 9-11 (conditional on enrollment)	0.028 (0.005)	0.000 (0.003)
Average log earnings in years 9-11 (with non-enrollment as an institution)	0.039 (0.004)	-0.001 (0.004)
Degree attainment within 6 years		
Bachelors from UT Austin	0.039 (0.004)	-0.021 (0.003)
Bachelors from any institution	0.043 (0.009)	-0.001 (0.006)
Associates or better	0.036 (0.009)	-0.006 (0.007)
Bachelors with STEM major	-0.003 (0.006)	-0.001 (0.003)
Labor market outcomes 9-11 years after HS graduation		
Employment (0/1)	-0.001 (0.009)	0.008 (0.006)
Average annual earnings (excluding 0s)	1004 (503)	-119 (478)
Average annual earnings (including 0s)	507 (567)	305 (384)
Log average annual earnings (excluding 0s)	0.063 (0.018)	0.036 (0.017)
Labor market outcomes 13-15 years after HS graduation		
Employment (0/1)	-0.006 (0.009)	0.012 (0.006)
Average annual earnings (excluding 0s)	-557 (798)	-58 (675)
Average annual earnings (including 0s)	-758 (814)	721 (506)
Log average annual earnings (excluding 0s)	0.000 (0.019)	-0.004 (0.017)

Notes: N=205,693.

Appendix Table 6. Baseline difference-in-differences analysis of alternative outcomes

Graduation outcomes, by year

	DD estimates		Pre-policy means [SDs]		
	Pulled In	Pushed Out	Control	Pulled In	Pushed Out
	(1)	(2)	(3)	(4)	(5)
UT, year 4	0.019 (0.003)	-0.005 (0.003)	0.01 [0.10]	0.02 [0.14]	0.06 [0.24]
UT, year 5	0.036 (0.004)	-0.017 (0.003)	0.02 [0.14]	0.03 [0.17]	0.12 [0.32]
UT, year 6	0.039 (0.004)	-0.021 (0.003)	0.02 [0.15]	0.04 [0.19]	0.14 [0.34]
UT, year 7	0.038 (0.004)	-0.021 (0.004)	0.03 [0.16]	0.04 [0.19]	0.15 [0.35]
UT, year 8	0.037 (0.004)	-0.019 (0.004)	0.03 [0.16]	0.04 [0.20]	0.15 [0.36]
University, year 4	0.022 (0.007)	0.006 (0.005)	0.07 [0.26]	0.20 [0.40]	0.15 [0.36]
University, year 5	0.031 (0.009)	0.003 (0.006)	0.19 [0.39]	0.37 [0.48]	0.32 [0.47]
University, year 6	0.037 (0.010)	-0.001 (0.006)	0.25 [0.43]	0.44 [0.50]	0.39 [0.49]
University, year 7	0.035 (0.010)	0.001 (0.007)	0.28 [0.45]	0.47 [0.50]	0.42 [0.49]
University, year 8	0.038 (0.011)	0.001 (0.007)	0.30 [0.46]	0.49 [0.50]	0.44 [0.50]
University, STEM, year 4	0.002 (0.004)	0.004 (0.001)	0.01 [0.09]	0.05 [0.22]	0.02 [0.15]
University, STEM, year 5	-0.006 (0.006)	0.002 (0.002)	0.02 [0.14]	0.11 [0.31]	0.06 [0.24]
University, STEM, year 6	-0.007 (0.006)	-0.001 (0.003)	0.03 [0.16]	0.13 [0.33]	0.07 [0.26]
University, STEM, year 7	-0.005 (0.006)	-0.002 (0.003)	0.03 [0.17]	0.13 [0.34]	0.08 [0.27]
University, STEM, year 8	-0.002 (0.007)	-0.003 (0.003)	0.03 [0.17]	0.13 [0.34]	0.08 [0.27]

Notes: Each row represents a separate difference-in-differences regression. Standard errors, clustered at the school district, in parentheses; standard deviations in square brackets. N=201,167 for DD specifications, with smaller samples for years 7 and 8, where we do not have graduation outcomes for all cohorts).

	Pulled In						Pushed Out					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Controls												
Individual demographics		y	y	y	y	y		y	y	y	y	y
School demographics			y						y			
S group dummies				y	y	y				y	y	y
Predicted top-ten probability				y	y	y (cubic)				y	y	y (cubic)
s-by-p dummies					y	y					y	y
Enrollment outcomes												
UT Austin	0.057 (0.005)	0.057 (0.005)	0.057 (0.005)	0.050 (0.005)	0.054 (0.005)	0.053 (0.005)	-0.038 (0.004)	-0.038 (0.004)	-0.036 (0.004)	-0.036 (0.004)	-0.036 (0.004)	-0.036 (0.004)
Any college	0.054 (0.011)	0.052 (0.011)	0.052 (0.011)	0.053 (0.011)	0.051 (0.011)	0.052 (0.011)	0.001 (0.007)	0.001 (0.007)	-0.001 (0.007)	-0.000 (0.007)	-0.000 (0.007)	0.000 (0.007)
Any 4-year	0.074 (0.011)	0.075 (0.011)	0.074 (0.011)	0.065 (0.011)	0.066 (0.011)	0.066 (0.011)	-0.006 (0.008)	-0.006 (0.008)	-0.005 (0.008)	-0.007 (0.008)	-0.007 (0.008)	-0.006 (0.008)
Degree attainment within 6 years												
Bachelors from UT Austin	0.043 (0.004)	0.043 (0.004)	0.043 (0.004)	0.037 (0.004)	0.041 (0.004)	0.039 (0.004)	-0.023 (0.003)	-0.023 (0.003)	-0.021 (0.003)	-0.021 (0.003)	-0.021 (0.003)	-0.021 (0.003)
Bachelors from any institution	0.051 (0.010)	0.048 (0.010)	0.048 (0.009)	0.037 (0.010)	0.038 (0.010)	0.037 (0.010)	0.001 (0.007)	-0.000 (0.006)	0.002 (0.006)	-0.001 (0.006)	-0.001 (0.006)	-0.001 (0.006)
Associates or better	0.044 (0.010)	0.040 (0.010)	0.039 (0.009)	0.031 (0.010)	0.032 (0.009)	0.032 (0.010)	-0.004 (0.007)	-0.005 (0.007)	-0.005 (0.006)	-0.006 (0.007)	-0.006 (0.007)	-0.006 (0.007)
Labor market outcomes 9-11 years after HS graduation												
Employment (0/1)	-0.001 (0.010)	-0.001 (0.010)	-0.001 (0.009)	0.002 (0.009)	0.000 (0.009)	-0.000 (0.009)	0.009 (0.006)	0.009 (0.006)	0.006 (0.006)	0.008 (0.006)	0.008 (0.006)	0.008 (0.006)
Log average annual earnings (excluding 0s)	0.060 (0.019)	0.065 (0.018)	0.064 (0.018)	0.051 (0.019)	0.053 (0.019)	0.055 (0.019)	0.035 (0.017)	0.036 (0.017)	0.039 (0.017)	0.035 (0.018)	0.035 (0.018)	0.036 (0.017)
Labor market outcomes 13-15 years after HS graduation												
Employment (0/1)	-0.009 (0.010)	-0.009 (0.010)	-0.008 (0.010)	-0.006 (0.010)	-0.007 (0.009)	-0.007 (0.010)	0.012 (0.007)	0.013 (0.006)	0.010 (0.006)	0.012 (0.006)	0.012 (0.006)	0.012 (0.006)
Log average annual earnings (excluding 0s)	0.002 (0.022)	0.011 (0.020)	0.012 (0.020)	-0.006 (0.020)	-0.002 (0.020)	-0.004 (0.020)	-0.005 (0.017)	-0.003 (0.017)	0.001 (0.016)	-0.004 (0.017)	-0.005 (0.017)	-0.004 (0.017)

Notes: Each row presents estimates from six separate difference-in-differences regressions: One model is presented in columns 1 and 7, another in 2 and 8, and so on. Standard errors, clustered at the school district, in parentheses. All specifications include calendar year and group (Pulled In, Pushed Out) indicators. Individual demographics are indicators for race (Hispanic, Black, Asian, other), gender, free or reduced price lunch, and immigrant status. School demographics are the school fraction Black, Hispanic, Asian, free or reduced price lunch, English language learner, or special education.

Appendix Table 8. Difference-in-Differences Analysis, Excluding High Schools That Participate in Longhorn Opportunity and Century Scholar Programs

	Pulled In	Pushed Out
	(1)	(2)
Enrollment outcomes		
UT Austin	0.047 (0.005)	-0.035 (0.004)
Texas A&M	-0.011 (0.008)	0.008 (0.004)
Any college	0.043 (0.011)	0.003 (0.008)
Any 4-year	0.062 (0.012)	-0.003 (0.008)
Any Community College	-0.008 (0.009)	0.011 (0.006)
Any other 4-year	0.026 (0.010)	0.024 (0.006)
Characteristics of institution attended (fixed at pre-policy levels)		
Graduation rate (conditional on enrollment)	0.017 (0.005)	-0.006 (0.004)
Graduation rate (with non-enrollment as institution)	0.034 (0.007)	-0.005 (0.005)
Math state %ile (conditional on enrollment)	1.42 (0.36)	-1.18 (0.28)
Math state %ile (with non-enrollment as institution)	1.75 (0.35)	-0.81 (0.26)
Instructional expenditures per student (conditional on enrollment)	583 (119)	-399 (86)
Average log earnings in years 9-11 (conditional on enrollment)	0.024 (0.006)	0.000 (0.003)
Average log earnings in years 9-11 (with non-enrollment as an institution)	0.036 (0.005)	-0.000 (0.004)
Degree attainment within 6 years		
Bachelors from UT Austin	0.036 (0.005)	-0.020 (0.003)
Bachelors from any institution	0.029 (0.012)	0.000 (0.006)
Associates or better	0.026 (0.012)	-0.005 (0.007)
Bachelors with STEM major	-0.013 (0.007)	-0.001 (0.002)
Labor market outcomes 9-11 years after HS graduation		
Employment (0/1)	-0.001 (0.011)	0.008 (0.007)
Average annual earnings (excluding 0s)	669 (605)	-262 (491)
Average annual earnings (including 0s)	311 (699)	225 (403)
Log average annual earnings (excluding 0s)	0.044 (0.021)	0.032 (0.018)
Labor market outcomes 13-15 years after HS graduation		
Employment (0/1)	-0.007 (0.011)	0.012 (0.007)
Average annual earnings (excluding 0s)	-923 (948)	-313 (665)
Average annual earnings (including 0s)	-1161 (998)	613 (538)
Log average annual earnings (excluding 0s)	-0.009 (0.023)	-0.007 (0.018)

Notes: N=180,101.

Appendix Table 9. Bootstrap vs. clustered standard errors

	Pulled In				Pushed Out			
	Baseline		Lasso		Baseline		Lasso	
	Cluster	Bootstrap	Cluster	Bootstrap	Cluster	Bootstrap	Cluster	Bootstrap
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Enrollment outcomes								
UT Austin	0.053 (0.005)	(0.004)	0.016 (0.002)	(0.004)	-0.036 (0.004)	(0.003)	-0.016 (0.002)	(0.002)
Any college	0.052 (0.011)	(0.011)	0.030 (0.005)	(0.008)	0.000 (0.007)	(0.008)	0.010 (0.005)	(0.005)
Any 4-year	0.066 (0.011)	(0.011)	0.017 (0.005)	(0.008)	-0.006 (0.008)	(0.008)	-0.013 (0.005)	(0.006)
Degree attainment within 6 years								
Bachelors from UT Austin	0.039 (0.004)	(0.004)	0.011 (0.002)	(0.003)	-0.021 (0.003)	(0.004)	-0.010 (0.002)	(0.002)
Bachelors from any institution	0.037 (0.010)	(0.010)	0.015 (0.004)	(0.005)	-0.001 (0.006)	(0.007)	-0.004 (0.004)	(0.004)
Associates or better	0.032 (0.010)	(0.010)	0.016 (0.005)	(0.005)	-0.006 (0.007)	(0.007)	-0.002 (0.004)	(0.004)
Labor market outcomes 9-11 years after HS graduation								
Employment (0/1)	0.000 (0.009)	(0.008)	0.002 (0.004)	(0.005)	0.008 (0.006)	(0.007)	0.011 (0.004)	(0.004)
Log average annual earnings (excluding 0s)	0.055 (0.019)	(0.019)	0.016 (0.009)	(0.012)	0.036 (0.017)	(0.016)	-0.008 (0.009)	(0.011)
Labor market outcomes 13-15 years after HS graduation								
Employment (0/1)	-0.007 (0.010)	(0.009)	-0.004 (0.004)	(0.005)	0.012 (0.006)	(0.007)	0.009 (0.004)	(0.005)
Log average annual earnings (excluding 0s)	-0.004 (0.020)	(0.021)	0.014 (0.011)	(0.015)	-0.004 (0.017)	(0.018)	0.009 (0.009)	(0.010)

Notes: All specifications use predicted probabilities of being in the top 10% from the random forest model. In odd numbered columns, standard errors of the DD regression are clustered at the school district level. In even numbered columns, standard errors are computed from bootstrap replications of the entire estimation procedure, from estimation of predicted probabilities to formation of treatment groups and estimation of the DD regression. Bootstrap samples are clustered at the school level; standard errors are the standard deviations across 250 bootstrap replications.

Appendix Table 10. Estimates based on 1996-1999 cohorts only

	Pulled In		Push
	Using 1999 to train predictions (1)	Using 1999. 2002 to train predictions (2)	Using 1999 to train predictio ns (3)
Enrollment outcomes			
UT Austin	0.046 (0.006)	0.042 (0.005)	-0.037 (0.005)
Any college	0.046 (0.011)	0.041 (0.011)	-0.017 (0.008)
Any 4-year	0.057 (0.011)	0.054 (0.011)	-0.021 (0.008)
Degree attainment within 6 years			
Bachelors from UT Austin	0.030 (0.005)	0.033 (0.005)	-0.026 (0.004)
Bachelors from any institution	0.036 (0.011)	0.033 (0.010)	-0.013 (0.006)
Bachelors with STEM major	-0.001 (0.007)	-0.003 (0.007)	-0.008 (0.003)
Labor market outcomes 9-11 years after HS graduation			
Employment (0/1)	0.002 (0.009)	-0.004 (0.009)	-0.007 (0.008)
Log average annual earnings (excludi	0.018 (0.023)	0.042 (0.022)	0.011 (0.017)
Labor market outcomes 13-15 years after HS graduation			
Employment (0/1)	-0.006 (0.010)	-0.014 (0.010)	-0.002 (0.007)
Log average annual earnings (excludi	0.014 (0.026)	-0.007 (0.024)	0.025 (0.019)

Notes: Each row represents a separate difference-in-differences regression. Standard errors, clustered at the school district, in parentheses; standard deviations in brackets. N=201,167 for DD specifications, with smaller samples for years 7-11 (where we do not have graduation outcomes for all cohorts).