A Data Appendix [Online]

This section provides additional details on the construction of the dataset.

ACT scores: Some students take the ACT instead of the SAT; if a student took both exams, I keep that student's SAT scores. Otherwise, I convert ACT composite scores to SAT combined scores using the College board's 1999 ACT-SAT concordance tables.[31]

Average SAT scores: Raw SAT scores range from 0 (all questions answered incorrectly) to 1600. After converting ACT scores to SAT scores, I rescale SAT scores by dividing by 1600. The standard deviation of the rescaled measure is 0.119.

To construct the ratio of student *i*'s SAT scores to their high school's average score, I use school-level average scores reported by THEOP. I found and corrected one outlier in the school-level score data. Between 10 and 20 students at a single high school have a reported school-average score of 29 points out of 1600. Because students can obtain a score of 200 by leaving the exam blank, this value seems implausible. The next-lowest school average score is greater than 690 points out of 1600. I replace the (presumably erroneous) 29-point school average with the mean score of surveyed students at this high school, which is 933 points out of 1600. At other schools, I have confirmed that the means of surveyed students' scores are close to the reported school mean scores.[32]

Income: The data do not provide *i*'s household income; I do, however, observe the education and occupational category of each of *i*'s parents. I draw incomes from the March CPS. To do so, for each household I construct the household head's education and occupational category. For each simulation draw, I then draw from incomes in the 2002 and 2003 March CPS samples of Texas residents with the same occupational category and education. Because the THEOP dataset uses a different encoding of occupations, I convert the CPS sample occupation codes to 1990 CPS codes as described in the data appendix. If either parents' occupation category or education is missing in the THEOP data, but one variable is present, I draw from the March CPS conditional on the variable that is present.

[31] http://research.collegeboard.org/sites/default/files/publications/2012/7/researchnote-1999-7-concordance-sat-act-students.pdf

[32] The numbers may differ because of survey nonparticipation and because of the use of converted ACT scores for students who took only the ACT. The mean and variance of the difference between the within-school mean student score and the mean school score are 17 and 52 points out of 1600, respectively.

Table A1: Administrative dataset

	Application Data		College Transcript Data	
Institution	N	Years	N	Years
Texas A&M	163,027	1992-2002	637,018	1992-2007
Texas A&M Kingsville*	18,872	1992-2002	91,106	1992-2004
UT Arlington	29,844	1994-2002	51,315	1994-2002
UT Austin	210,006	1991-2003	659,102	1991-2004
UT Pan American**	44,747	1995-2002	115,812	1995-2005
UT San Antonio#	61,221	1990-2004	160,604	1990-2004
Texas Tech	81,153	1995-2003	211,771	1995-2004
Rice	36,190	2000-2004	18,149	2000-2005
SMU	45,549	1998-2005	60,607	1998-2005

* Applicant data for enrollees only: 1992-1994
** Limited variables provided
\# Applicant data for enrollees only, 1990-1997

http://www.texastop10.princeton.edu/admin_overview.html

Expected family contribution: In estimation, I use the family's expected family contribution (EFC) as a measure of the amount of financial aid the student is likely to receive. I use the federal government's "simplified EFC formula worksheet A" from the 2002-2003 FAFSA, using parents' income, the number of parents/guardians who live with the applicant, and the number who work full-time, if the applicant lives with two parents.

In estimation, I draw a large number of income draws for each student from the CPS. For each income draw, I calculate the EFC using the formula. I integrate over these draws when calculating the likelihood and moments. Details are provided in the computational appendix.

Administrative dataset: Table A1 provides a list of all institutions for which THEOP provided administrative data, together with the years covered and number of applications (in application data), or semester-individual observations (in transcript data).

The THEOP survey is based on a stratified sampling design that placed unequal weights on different types of high schools. Using the population weights, however, one obtains patterns very similar to those in the administrative data. A sequence of papers by Marta Tienda and coauthors describes the survey design and findings. I have confirmed that population-weighted mean SAT scores, class rank, and high school characteristics (poverty, SAT scores) of students in the survey who attend flagship universities closely match the actual numbers in the administrative data.

B Model Appendix [Online]

B.1 Optimality of cutoff rules

This section shows that cutoff rules are optimal for colleges.

As before, each college j maximizes the expected quality of its entering class subject to the constraint that its expected enrollment is less than its capacity. The quality of a class $C(j) \subset I$, denoted $\Pi_j(C_j)$, is the sum of the quality of enrolled students:

$$\Pi_j(C(j)) = \sum_{i \in C(j)} \pi_{ij}.$$

Each college j chooses an offer probability for each applicant, $B(j) \in [0,1]^{A(j)}$ where $A(j)$ denotes the set of applicants to j, in order to solve

$$\max_{B(j) \in [0,1]^{A(j)}} E\Pi_j(C(j)|A(j)) \text{ s.t. } E(C(j)|A(j)) \leq K_j.$$

college j observes (z_i, q_i, μ_{ij}) for all applicants before choosing who to admit. College j does not observe $\mu_{ij'}$ for $j' \neq j$. Colleges cannot commit to an admission rule for non-automatically-admitted students that is not optimal given applications.

A college is said to be ***nonselective*** (given students' and other colleges' decisions) if its capacity constraint does not bind when it admits all students with $\pi_{ij} > 0$. Nonselective colleges admit all students for whom $\pi_{ij} > 0$.

Colleges whose capacity constraints bind are said to be *selective*. It is also optimal for selective colleges to employ cutoff rules, a result stated in the following Lemma:

Lemma 1. *Colleges' optimal admissions rules consist of cutoffs $\{\underline{\pi}_j\}_{j \in \mathscr{J}}$ such that an applicant is admitted to j if and only if $\pi_{ij} > \underline{\pi}_j$, and either $\sum_{i:\pi_{ij} > \underline{\pi}_j, j \in A_i} Pr(C_i = j) = k_j$ or $\underline{\pi}_j = 0$.*

Proof. Consider a selective college. Let $B(j)$ be an admissions rule that satisfies the expected capacity constraint, and suppose it is not a cutoff rule. Then there are two applicants i and i' such that $\pi_{ij} < \pi_{i'j}$ but $B(j)(i) > 0$ and $B(j)(i') < 1$. If admitted, i attends j with probability P_{ij} and i' attends with probability $P_{i'j}$, for some $P_{ij}, P_{i'j} \in (0,1)$. It is feasible and profitable for j to reduce $B(j)(i)$ by $\frac{\varepsilon}{P_{ij}}$ and increase $B(j)(i')$ by $\frac{\varepsilon}{P_{i'j}}$ for some ε. □

B.2 Rational-Expectations Equilibrium

A rational-expectations equilibrium is a tuple

$$\left\{\{A_i\}_{i\in I}, \{Admit(j)\}_{j\in J}\right\}$$

that satisfies the following properties:

1. $A_i \in \mathscr{A}$ solves i's application problem

$$\max_{A\in\mathscr{A}} V_i(A)$$

   given admissions rules $Admit(j)$and i's characteristics, where $\mathscr{A}$ is the set of feasible portfolios.

2. For each college j, $Admit(j) \subset I$ maximizes

$$\mathbb{E}\left(\Pi_j | x_i, z_i, q_i, \mu_i, Admit_{-j}\right)$$

   subject to

$$\sum_{i\in Admit(j)} Pr(C_i = j | x_i, z_i, q_i, Admit_{-j}) \leq k_j,$$

   where $Pr(C_i = j|x_i, z_i, q_i, \mu_i; Admit_{-j})$ are the matriculation probabilities induced by A, integrating over the distribution of the student's other applications and admissions offers given the characteristics x_i, z_i, q_i, μ_i observed by college j and the other colleges' policies $Admit_{-j}$.

3. $i \in Admit_j$ if and only $j \in B_i$.

Because optimal admissions rules are cutoffs, I identify admissions rules with the corresponding cutoffs and consider equilibria of the form:

$$\left\{\{A_i\}_{i\in I}, \{\underline{\pi}_j\}_{j\in J}\right\}.$$

B.3 Existence of Equilibrium

Suppose there is a continuum of students with type $\omega \in \Omega \subseteq R^N$ and measure $F(\omega)$ with density f with respect to Lebesgue measure. (In the model, an individual is defined by his observables z_i, his preference terms and random coefficients, and his caliber and signal (q,s).) Each cutoff vector $\underline{\pi}$ induces a joint distribution of application portfolios $A \in \mathscr{A}$ and types ω. In particular, for almost all ω there is a unique portfolio $A(\underline{\pi}, \omega) \in \mathscr{A}$ that is optimal.

To prove existence, we need to show that there is a cutoff $\underline{\pi}$ such that A_i is a best response to $\underline{\pi}$ for all i, and that $\underline{\pi}$ maximizes quality subject to capacity constraints given applications $A(\underline{\pi}, \omega)_{\omega \in \Omega}$.

Let $\tilde{A}(\underline{\pi})$ denote the joint distribution of $(A \in \mathscr{A}, \omega \in \Omega)$ induced by students' best responses to $\underline{\pi}$. Because ω admits a density and the indirect utility $U(\mathscr{A}; \underline{\pi})$ is continuous in $\underline{\pi}$, the share of applicants to each portfolio $A \in \mathscr{A}$ changes continuously in $\underline{\pi}$. Moreover, the share of students with admission set B given application set A changes continuously in $\underline{\pi}$, for all $A, B \in \mathscr{A}$. Conditional on admission set B, application set A, and j's information about ω, the probability of attending $j \in B$ is continuous in $\underline{\pi}$. As a result the best-response cutoff function $h(\underline{\pi}) \equiv \underline{\pi}^*(\tilde{A}(\underline{\pi})) : R^J \to R^J$ is continuous in $\underline{\pi}$.

We now show that $h(\cdot)$ is bounded, i.e. we can restrict attention to the map $h(\cdot)$ on a box in R^J. For each j, taking the distribution of applications A that has every student apply to $\{j\}$ with probability 1 gives an upper bound $\underline{\pi}_j^{high} \leq \underline{\pi}^*(E(A))$. There exists a cutoff $\underline{\pi}_j^{low}$ such that j strictly prefers empty seats to students with caliber less than $\underline{\pi}_j^{low}$.

From Brouwer's fixed point theorem, h has a fixed point $\underline{\pi}^*$. By construction $\underline{\pi}^*$ is a set of market-clearing cutoffs given applications $\tilde{A}(\underline{\pi})$, and applications $\tilde{A}$ are optimal given $\underline{\pi}^*$.

B.4 Limited multiplicity of equilibria

In general, there is no guarantee that equilibrium is unique.[33] In this section, I show that given fixed distribution of application portfolio choices, there is a unique mutual best response by colleges. That is, there is precisely one vector of cutoffs $\underline{\pi}$ such that each $\underline{\pi}_j$ is optimal given $\underline{\pi}_{-j}$ and applications.

To show these results, I show that holding applications fixed, students' matriculation probabilities are isotone in $\underline{\pi}$. As a result, market-clearing cutoffs $\underline{\pi}$ form a complete lattice. As a corollary, for any set of applications, cutoffs exist that satisfy the colleges' problems. Letting $\underline{\pi}_L$ denote the lowest cutoff and $\underline{\pi}_H$ denote the highest, When students' choice probabilities satisfy a substitutes property, I show that $\underline{\pi}_L \neq \underline{\pi}_H$ leads to a contradiction,

[33] See Chade, Lewis, and Smith (2014).

as the share of students attending the outside option and/or nonselective colleges must be strictly higher under $\underline{\pi}_H$ than under $\underline{\pi}_L$, but the share at each selective college cannot decrease.

Let $Pr(C_i = j|B)$ denote the probability that student i enrolls in college j given admission to a set of colleges $B \subseteq \mathscr{A}$. Let $Pr(C_i = 0|B)$ denote the probability that i chooses the outside option.

Condition (Substitutes)**.** For each individual, the following condition holds:

$$B \subseteq B' \implies Pr(C_i = 0|B') < Pr(C_i = 0|B).$$

This condition holds generally whenever the outside option is a possible substitute for each college.[34] In particular it is satisfied for the nested logit specification of utility in this model conditional on random coefficients, and hence for the mixed nested logit specification.

Proposition 2. *If the substitutes condition holds, then conditional on applications there is a unique cutoff vector* $\underline{\pi}$ *that satisfies equilibrium condition (3).*

Proof. For each set $B \subseteq \mathscr{A}$ and and $B' \subseteq A$ with $B \subseteq B'$, each agent's choice probabilities must satisfy[35]

$$Pr(C_i = j|B') \leq Pr(C_i = j|B). \qquad (*)$$

Define $\underline{\pi}^*(\underline{\pi}) : R^{|J|} \rightarrow R^{|J|}$ by $\underline{\pi}^*_j(\underline{\pi}) = \underline{\pi}^*_j(\underline{\pi}_{-j})$. That is, the jth component of $\underline{\pi}^*$ is the cutoff that college j optimally chooses given (fixed) applications and student preferences, and the cutoffs of other colleges. By (*), the function $\underline{\pi}^*$ is isotone in $R^{|J|}$. Its fixed points therefore form a complete lattice.

Let $\underline{\pi}^*_L$ be the lowest equilibrium cutoffs, and $\underline{\pi}^*_H$ be the highest. Let $\mathscr{J}_u \subseteq \mathscr{J}$ denote the set of schools that are not selective at $\underline{\pi}_L$, i.e. the union of the outside option and the set of schools for which $\underline{\pi}^*_{Lj} = \pi^0_j$. For a contradiction, suppose $B_i(\underline{\pi}^*_H) \neq B_i(\underline{\pi}^*_L)$ for some student i.

The share of students attending the outside option must strictly increase. Moreover, in any random utility model the the share of students attending other colleges in $\mathscr{J}_u$ must also

[34]The condition is related to the "connected substitutes" condition of Berry, Gandhi and Haile with continuous prices. In my setting, what is needed is that when a capacity-constrained college is removed from the choice set, the probability of not attending a capacity-constrained college strictly increases.

[35]This result is due to Block and Marshak (1960) and Falmagne (1978). See also Haile, Hortacsu and Kosenok (2008). In particular, a direct application of Theorem 1 of Falmagne (1978) gives that for any sets $B_0, B_1 \subseteq \mathscr{A}$, we have $Pr(C_i = j|B_0) - Pr(C_i = j|B_0 \cup B_1) \geq 0$.

weakly increase.[36] Therefore the share of students attending colleges in $\mathscr{J}_u$ must strictly increase:

$$\sum_i Pr(C_i \in \mathscr{J}_u | \underline{\pi}_H^*) > \sum_i Pr(C_i \in \mathscr{J}_u | \underline{\pi}_L^*).$$

Cutoffs of college in $\mathscr{J} \setminus \mathscr{J}_u$ are weakly higher under $\underline{\pi}_H^*$, however, implying that they are at full capacity in both $\underline{\pi}_L^*$ and $\underline{\pi}_H^*$, i.e.

$$\sum_i Pr(C_i \notin \mathscr{J}_u | \underline{\pi}_H^*) = \sum_i Pr(C_i \notin \mathscr{J}_u | \underline{\pi}_L^*),$$

which is a contradiction. □

C Estimation Appendix [Online]

C.1 Likelihood

Let $\theta \in \Theta \subset R^N$ be a vector of parameters, where Θ is the set of allowed parameter values. Recall that A_{ij}, B_{ij}, C_{ij} denote applications, admission, and matriculation respectively for student i and college j. $Aware_{ij}^{obs} \in \{0,1\}$ is an indicator for completing an application for financial aid at college j and is observed only if $A_{ij} > 0$. $Aware_{ij}^{obs} = 1(\text{i applied for financial aid at j})$ is generally not the full vector of financial-aid awareness, because we do not observe whether i would have completed an application for aid at colleges outside his application set.

Let ω_i denote the vector of random coefficients β_i of individual i as well as income y_i, expected family contribution EFC_i, random terms in aid, and shocks ν_i.

The likelihood of applications ℓ_i^{Ap} is the measure of the set of preference and signal draws such that the observed application portfolio maximizes expected utility among all application portfolios net of costs:

$$P(A_i) = \frac{\exp \lambda^{\text{app}} \left(V_{iA} - c_{z_i^{\text{app}}}^{fixed} - c_{z_i^{\text{app}}}^{var} |A| \right)}{\sum_{A \in \mathscr{A}} \exp \lambda^{\text{app}} \left(V_{iA} - c_{z_i^{\text{app}}}^{fixed} - c_{z_i^{\text{app}}}^{var} |A| \right)}.$$

[36] For $j \in \mathscr{J}_u$, if $\underline{\pi}_{Hj}^* = \underline{\pi}_{Lj}^*$ then the share must weakly increase by (*). Alternatively, $\underline{\pi}_{Hj}^* > \underline{\pi}_{Lj}^*$ if and only if college j has filled its capacity under $\underline{\pi}_H^*$.

The following expression is the likelihood of admissions conditional on a signal s:

$$\ell_i^{B|A}(\theta,s) = \int \prod_{j\in B} \left(\Phi(z_{ij}\gamma_j + q_i - \underline{\pi}_j)\right) \prod_{j\in A\setminus B} \left(1-\Phi(z_{ij'}\gamma_{j'} + q_i - \underline{\pi}_{j'})\right) dF_i(q|s;\sigma_{q|s}).$$

The likelihood of financial-aid awareness $\ell_i^{\text{Aware}|A}$ is analogous to the term above, except that a double integral is taken over the distribution of income y and random coefficient α_i^{aware} rather than an integral over q, and coefficients on y, x, z_i^{admit}, and program indicators are estimated.

The likelihood of matriculation is the nested-logit choice probability of school j conditional on i's characteristics:

$$\ell_i^{C|B}(\theta,\omega_i) = \frac{\exp(u_{ij}(\omega_i)/\lambda)\left(\sum_{j'\in B}\exp(u_{ij'}(\omega_i)/\lambda\right)^{\lambda-1}}{1+\left(\sum_{j'\in B}\exp(u_{ij'}(\omega_i)/\lambda\right)^{\lambda}}.$$

The dependence on aid awareness is implicit, via the impact of awareness at j on u_{ij}. For students who appear in wave 1 of the survey, I set $\ell_i^{C|B} = 1$, as matriculation decisions are not observed. The likelihood of all observables in the data is given by

$$\ell_i(\theta) = \int_s \int_{\omega_i} \ell_i^A(\theta,\omega_i,s)\ell_i^{Aware|A}(\theta,\omega_i)\ell_i^{B|A}(\theta,s)\ell_i^{C|B}(\theta,\omega_i)dF_i(s;\theta)dG_i(\omega_i;\theta).$$

C.2 Calculating the objective

Consider the optimization problem

$$\max_\theta \frac{1}{N}\sum_{i=1}^{N}\log\ell_i(\theta) + g(\theta)W_g g(\theta)^T, \tag{8}$$

where

$$g(\theta) = \frac{1}{N}\sum_i g_i(\theta)$$

are non-likelihood moments, $\log\ell_i(\cdot)$ is the log-likelihood, and W_g is a weighting matrix.

Suppose that the objective in 8 is maximized at θ_0. Then the FOC of objective (8) is given by

$$\frac{1}{N}\sum_i \nabla_\theta \log\ell_i(\theta_0) + 2g(\theta)W_g(\nabla_\theta g(\theta_0))^T = 0. \tag{FOC1}$$

Define the matrix W by

$$W = \begin{pmatrix} \frac{1}{2}\left[\frac{1}{N}\sum_i \left((\nabla_\theta \log \ell_i(\theta_0))^T * (\nabla_\theta \log \ell_i(\theta_0))\right)\right]^{-1} & 0 \\ 0 & W_g \end{pmatrix}$$
$$= \begin{pmatrix} \frac{1}{2}\cdot \mathscr{I}_{\theta_0}^{-1} & 0 \\ 0 & W_g \end{pmatrix}$$
$$= \begin{pmatrix} -\frac{1}{2}\left[\frac{1}{N}\sum_i \frac{\partial^2}{\partial\theta\partial\theta'} log\ell_i(\theta_0)\right]^{-1} & 0 \\ 0 & W_g \end{pmatrix},$$

where $\mathscr{I}_{\theta_0}$ is the Fisher information matrix evaluated at θ_0. Define $loglik(\theta) = \sum_i \log \ell_i(\theta)$. Then $FOC1$ is also the first-order condition of the following GMM objective function:

$$\min_\theta [\frac{1}{N}\nabla_\theta \log \ell(\theta), g(\theta)] W [\frac{1}{N}\nabla_\theta \log \ell(\theta), g(\theta)]^T, \tag{9}$$

Hence the solution to 8 will also satisfy 9. The advantage of 8 for my purposes is that it is possible to use gradient-based optimization methods without needing to calculate second derivatives of $\log \ell(\theta)$. Therefore, 8 is a relatively computationally inexpensive way to optimize the GMM objective 9.

The weight matrix is not known in advance; however it is a fixed function of the parameters θ_0 that optimize the penalized-likelihood objective.

C.3 Standard errors

I estimate standard errors via the bootstrap. I resample applicants with replacement from the survey data, and draw new draws of financial-aid awareness, random coefficients, shocks, and other latent variables in order to allow my procedure to account for simulation error as well as sampling error. Using this new resampled dataset, and taking the point estimates as starting values, I re-estimate the "first step" parameters θ.

Population weights are needed for second-step estimation and counterfactuals. I construct new population weights proportional to the weights of each resampled individual, but summing to one.

Next, I redraw auxiliary-model parameters using their asymptotic mean and covariance matrix as estimated on the administrative datasets, then re-estimate the indirect-inference outcome specifications. Finally, I re-solve the model in counterfactuals and compute outcomes.

D Computation [Online]

D.1 Computing the value of application sets via the inclusion-exclusion principle

Computing the value of an application set $A \subset \mathscr{A}$ requires integrating over all possible outcomes $B \subseteq A$, as the value depends on the probabilities and utilities of each admissions set B that is possible given application to A.

In principle, and dropping i subscripts, computing $V(A)$ for all A requires computing $|\mathscr{A}|$ utility terms $\{U(B)\}_{B \in \mathscr{A}}$, and $\mathscr{O}(|\mathscr{A}|^2)$ multivariate normal CDF evaluations $\{P(B|A)\}_{B \subseteq A, A \in \mathscr{A}}$. It would be expensive to compute all of these probabilities directly for each draw for each individual at each trial parameter value. In what follows, I show that one needs only evaluate $|\mathscr{A}|$ multivariate normal CDFs and perform some matrix multiplication.

The argument does not rely on the presence of a single common correlating factor, q. Therefore, this method may be useful in cases beyond this paper where it is necessary to compute all conditional admissions probabilities under a complicated correlation structure.

Define

$$P_B = \int_q \prod_{j \in B} \Phi(q_i + z_{ij}^{\text{admit}} \gamma_j - \underline{\pi}_j) dF(q|s)$$

as the probability of admission to every school in B given application set B and a realization of the applicant's information q^s. Let X_B be the event that i is admitted to all schools in B. Let $X_{A;B}$ denote the event that i is admitted to all schools in B and rejected from all schools in $A \backslash B$, and $P_{A;B}$ the probability of this event. If $B \not\subseteq A$ then let $X_{A;B}$ be empty and $P_{A;B} = 0$. Let $\mathbb{P}$ be the probability measure associated with i's characteristics and signal s, so that $P_{A;B} = \mathbb{P}(X_{A;B})$.

Proposition. *(inclusion-exclusion formula) Given the above definitions, the following result holds:*

$$P_{A;B} = \sum_{B' : B \subseteq B' \subseteq A} P_{B'} \cdot (-1)^{|A|-|B'|} \textit{ for all sets } A, B \in \mathscr{A} \textit{ with } B \subseteq A.$$

Proof.

$$\begin{aligned}
P_{A;B} &= \mathbb{P}\left(X_B \backslash \left(\cup_{j\in A\backslash B}\left(X_{B\cup\{j\}}\right)\right)\right) \\
&= \mathbb{P}(X_B) - \mathbb{P}\left(\cup_{j\in A\backslash B}\left(X_{B\cup\{j\}}\right)\right) \\
&= P_B - \sum_{j\in B\backslash A} \mathbb{P}(X_{B\cup\{j\}}) + \sum_{j_1,j_2\in B\backslash A} \mathbb{P}(X_{B\cup\{j_1\}} \cap X_{B\cup\{j_2\}}) - \ldots + \mathbb{P}(\cap_{j\in A\backslash B} X_{B\cup\{j\}}) \\
&= P_B - \sum_{B':B\subseteq B'\subseteq A} P_{B'} \cdot (-1)^{|A|-|B'|+1}.
\end{aligned}$$

The second line follows because $X_{B\cup\{j\}} \subseteq X_B$ and the third line is the standard inclusion-exclusion formula. □

List all the portfolios $A_0, A_{1,}, \ldots, A_{|\mathscr{A}|}$ in some order, and define the matrix T by

$$T_{kl} = 1_{A_k\subseteq A_l} \prod_{j\in A_k} \left(\begin{array}{c} A_{lj} \\ A_{kj} \end{array} \right),$$

where A_{lj} is the number of applications to j in portfolio A_l. Similarly, define the matrix $S = T^{-1}$. We have

$$S_{kl} = (-1)^{|A_l|-|A_k|} T_{kl}.$$

It follows that $P_{A_k;A_l} = \sum_{A_k\subseteq A_r\subseteq A_l} T_{kr} P_r$

Corollary. *Let P be the diagonal matrix with kth entry P_{A_k}. The vector $V = \{V_A\}_{A\in\mathscr{A}}$ is given by*

$$V = T * Diag(P) * S * U.$$

Computing only the admissions probabilities P_B rather than all $P_{A;B}$ simplifies computation. As an example, if $\mathscr{A} = \{\{1\},\{2\},\{1,2\}\}$ we have the following calculation:

$$\left(\begin{array}{c} V_{\{1\}} \\ V_{\{2\}} \\ V_{\{1,2\}} \end{array} \right) = \left(\begin{array}{ccc} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 1 & 1 \end{array} \right) * \left(\begin{array}{ccc} P_{\{1\}} & 0 & 0 \\ 0 & P_{\{2\}} & 0 \\ 0 & 0 & P_{\{1,2\}} \end{array} \right) * \left(\begin{array}{ccc} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -1 & -1 & 1 \end{array} \right) * \left(\begin{array}{c} U_{\{1\}} \\ U_{\{2\}} \\ U_{\{1,2\}} \end{array} \right).$$

By using the inclusion-exclusion principle, we avoid having to compute $P_{\{1,2\};\{1\}}$ or $P_{\{1,2\};\{2\}}$ by integrating over multivariate normal densities.

D.2 Discretization

To compute integrals over s, and over the residual uncertainty $q|s$, I use Gauss-Hermite quadrature with 9 nodes.

I draw $M = 384$ draws $\omega_{i,m}$ for each individual i, which contain random coefficients, income y_{im}, EFC efc_{im}, and preference shocks ν. The construction of income and EFC draws is described in the data appendix. To draw random coefficients and shocks ν, I fix a set of iid standard Normal draws. I use an importance-sampling procedure, which I describe below to integrate over financial-aid awareness draws.

D.3 Simulation of financial-aid awareness

I observe financial-aid application or failure to apply for aid only at schools to which i submits an application. In order to compute the likelihood that i chooses his observed application portfolio, however, we need to calculate the value of all application portfolios he could have chosen. Therefore calculating the likelihood requires integrating over financial-aid awareness at colleges to which i did not apply.

In order to obtain an estimator that is smooth in θ, I draw financial-aid awareness for student i using an importance sampling procedure. Before searching over parameter values, I draw a vector of outcomes $aware_{ijm} \in \{0, 1\}$ once for each simulation draw, for all colleges j to which i did not submit an application. I estimate weights on each vector for each parameter vector θ. These weights are then smooth functions of θ.

Initially, I estimate the parameters of equation (4), taking the set of observed applications as given, ignoring selection, and integrating over the distribution of the random effect α_i^{aware}.

for each individual i and simulation draw m, I then draw a vector of financial-aid awareness outcomes for the schools to which i did not apply, according to their distribution given the parameters estimated above, and the agent's observed financial-aid awareness, subject to the constraint that, at colleges to which i applied, awareness must take its observed value.

Let P_{0i}^{aware} denote the probability of the vector of financial aid awareness draws under $\left\{Pr_{0ijm}\right\}_{j\in\mathscr{J}}$ for simulation draws $m = 1, \ldots, M$.

For each trial parameter value θ in estimation, I calculate the probability of financial-aid awareness,

$$P_{im}^{aware}(\theta) = \int_\alpha \prod_{j\in\mathscr{J}} \left[\left(Pr(Aware_{ij} = 1|\theta, \alpha_i^{\text{aware}})\right)^{aware_{ijm}} \left(Pr(Aware_{ij} = 0|\theta, \alpha_i^{\text{aware}})\right)^{1-aware_{ijm}} dF(\alpha_i^{\text{aware}}|\theta)\right].$$

Note that the probability of the financial aid awareness draw under parameters θ includes the probability of $aware_{ijm}$ for $j \in A_i$.

The likelihood of financial aid awareness is then approximated by

$$\hat{\ell}_i^{F^{aware}|A,B}(\theta, m) \approx \frac{P_{im}^{aware}(\theta)}{P_{0im}^{aware}}.$$

In estimation, I obtain a starting guess for financial-aid awareness parameters as described above, which forms part of the starting values θ^0. I draw financial-aid awareness according to these parameters. I then run the estimation procedure, arriving at estimates θ^1, and use the estimated values as new initial parameters for constructing p_{0im}^{aware}. I then run the estimation procedure again to obtain estimates θ^2. Finally, I redraw financial-aid awareness again, and rerun the estimation procedure, to obtain final point estimate θ^*.

D.4 Implementation

The first-stage estimation procedure uses an implementation of the L-BFGS algorithm provided by the NLopt solver package. I provide efficient analytic first derivatives to the solver for first-stage estimation.

The second-stage uses the L-BFGS algorithm as well. I use automatic differentiation to obtain gradients for the outcome equations in the second stage.

Counterfactuals require solving for a vector of new cutoffs at three programs. I use the baseline cutoffs, recovered during estimation, as starting values. I use the L-BFGS algorithm to minimize the sum of squared differences between total enrollments in the counterfactual and at baseline, using automatic differentiation to obtain the gradient. In practice this procedure converges rapidly.

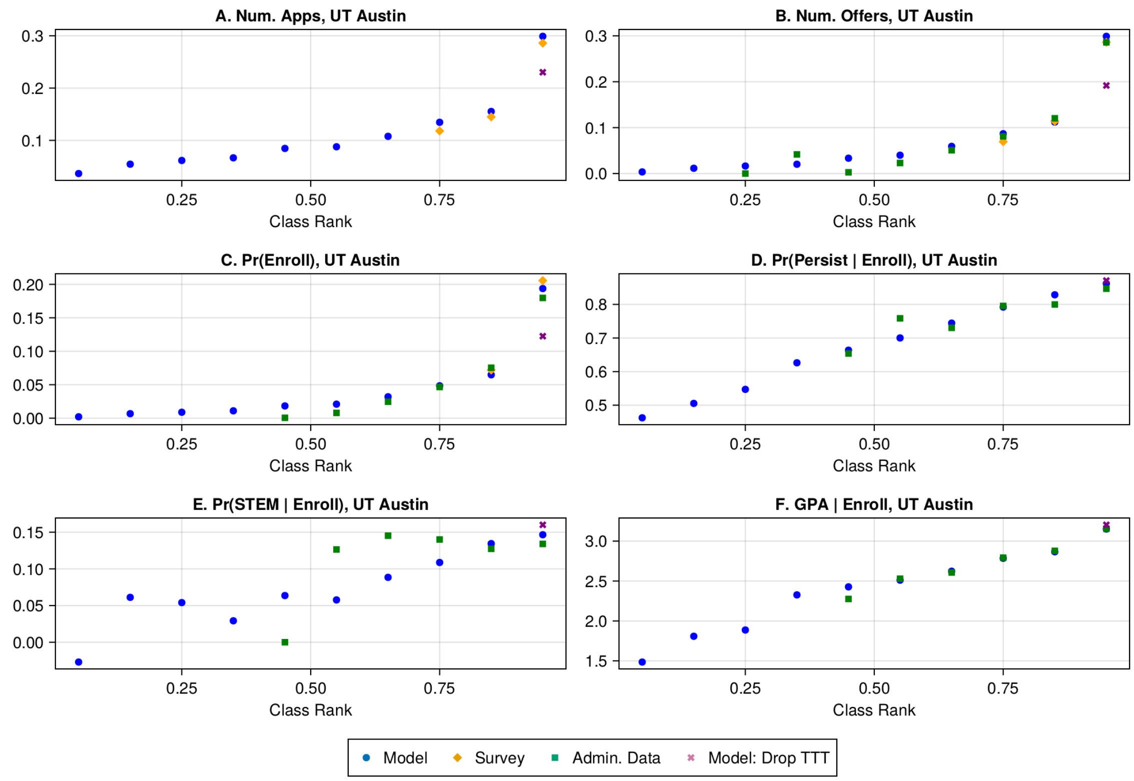
E Additional Estimates and Results [Online]

This section provides additional model-fit results, results on the importance of accounting for selection, and additional counterfactuals and extensions.

E.1 Additional Model-Fit Figures

Figures A1 and A2 show the analogues of Figure 1, separately for each flagship institution. Figure A3 shows fit of applications, admissions, and matriculation by class rank decile for all public institutions, and for all four-year institutions. Results are similar to those in the main text.

Figure A1: Model Fit: UT Austin



E.2 Parameter Estimates

This section provides additional information on parameters. A full set of estimates is given in the supplementary material. Table A2 shows cost and information parameters. While estimates of information parameters γ^s are noisy, we can reject large values of σ^s. Table A3 shows the estimated change in cutoffs across counterfactual scenarios. Table A4 shows probit versions of the linear specifications reported in Table 2 for binary outcomes.

E.3 Additional Results

This section provides additional results and context. I first consider the importance of accounting for selection. I then decompose post-enrollment outcomes at flagship universities.

Figure A2: Model Fit: Texas A&M

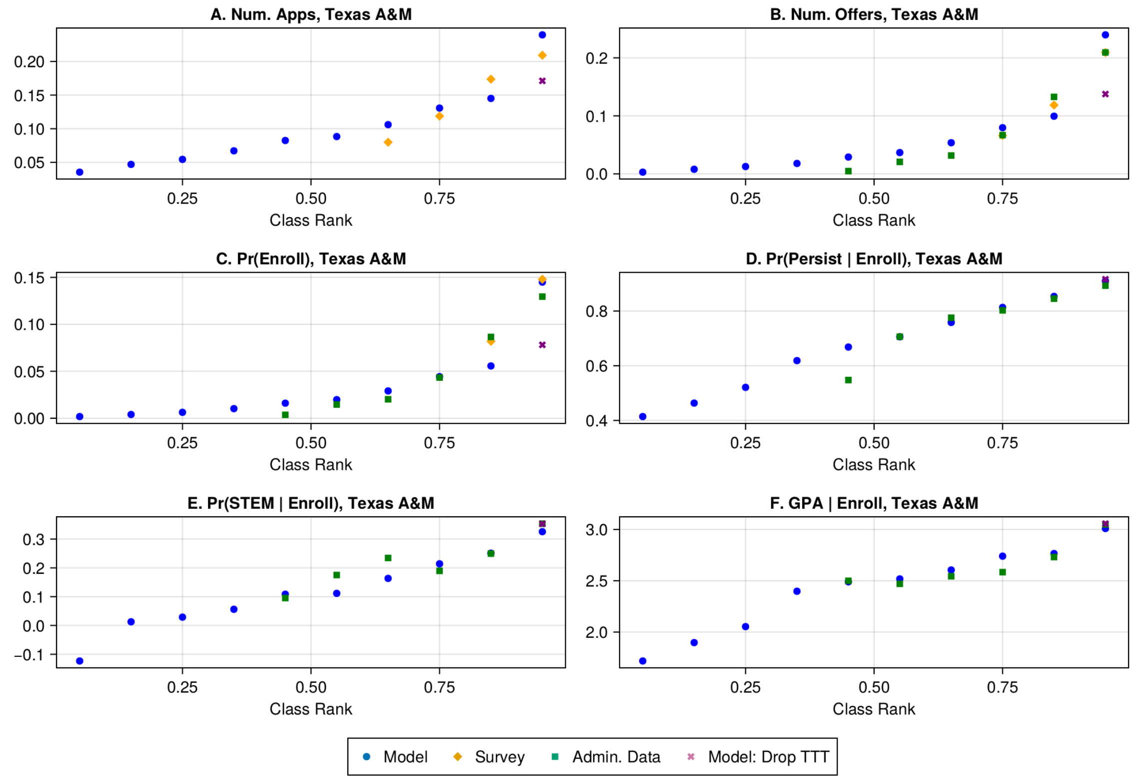


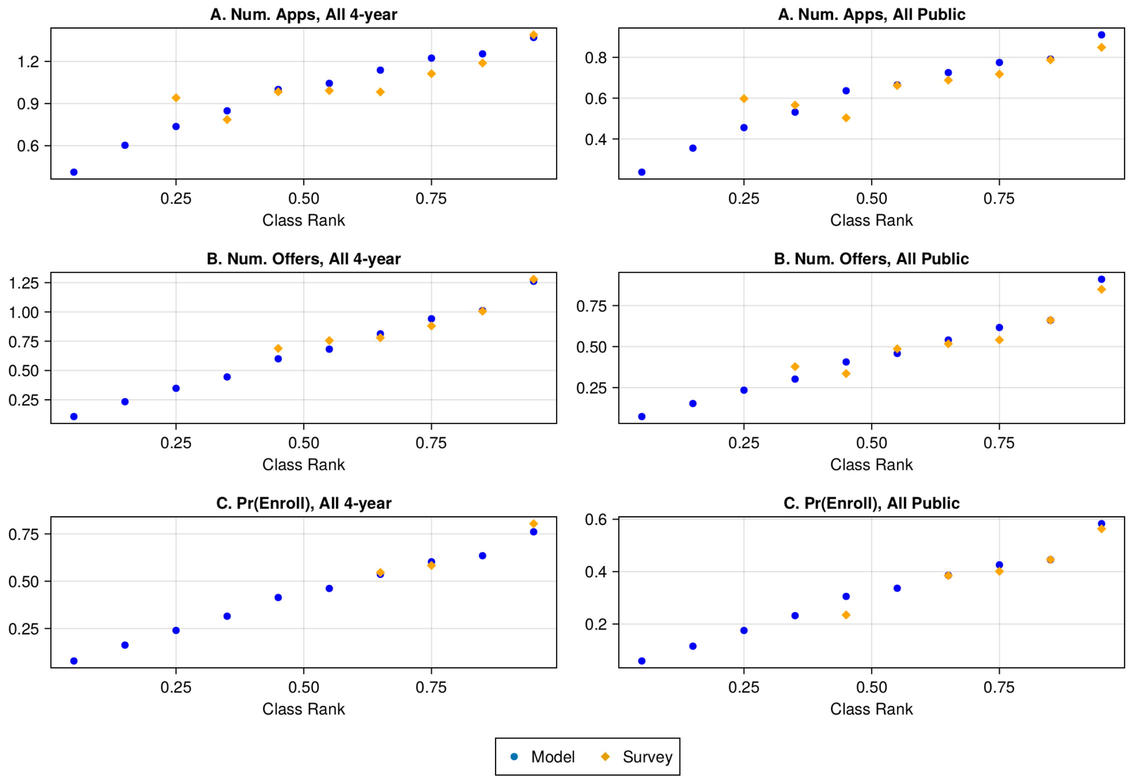
Table A2: Information Parameters

	Student Info. γ_s	Uncertainty $\gamma_{q\|s}$	Fixed Cost γ_0^c	Var. Cost γ_1^c	Var. Shocks $\gamma^{\varepsilon_{app}}$
Constant	-10.0	-0.692	0.31	1.038	-0.88
	(0.556)	(0.353)	(0.091)	(0.053)	(0.094)
Poverty	1.563	1.625	-0.239	-0.219	-0.592
	(8.376)	(0.763)	(0.111)	(0.119)	(0.21)
URM	5.748	0.545	0.047	-0.176	-0.075
	(6.26)	(0.305)	(0.06)	(0.045)	(0.079)

Note: this table shows information and cost parameter estimates.

Figure A4 shows the impact of accounting for selection on applications and admissions. Estimates using matriculation choices only, taking choice sets as given, would have underestimated demand for non-flagship institutions and overstated the popularity of flagships.

Figure A3: Model Fit: Public Institutions and All Four-Year Institutions



While I estimate that top-decile applicants are generally well-informed about aid availability, Panel B of A4 shows that naive estimates ignoring selection would have slightly overestimated aid awareness among non-top-decile students at elite private institutions, at out-of-state institutions, and at Texas A&M.

In Figure A5 I show means of students enrolled at baseline, of students "pulled in" [Black et al., 2023] by mechanical and informative effects, and of students "pushed out" by equilibrium responses. I show decompositions of the percent plan as implemented (blue bars), of the percent plan combined with aid information (orange bars), and of the "top 20%" aligned percent plan (green bars). The leftmost column, "baseline," shows means of students enrolled in flagships at baseline. "Aid info" shows means among students pulled in by aid information without the percent plan (e.g. 3.051 GPA, top panel) and means among students "pushed out" by equilibrium cutoff changes (2.926 GPA, top panel). The next columns, "Mechanical" and "informative," show means of students pulled in by mechanical and informative effects respectively. Finally, the last column, "Eqbm," shows means of

Table A3: Change in Cutoffs Relative to Baseline

Policy	UT Austin	Texas A&M	Other Public
TTP	0.404	0.382	0.124
	(0.03)	(0.03)	(0.01)
Aid Awareness Only	0.092	0.021	0.062
	(0.03)	(0.05)	(0.06)
Aid Awareness + TTP	0.537	0.387	0.185
	(0.06)	(0.09)	(0.06)
Top 5% zg	0.029	0.019	0.007
	(0.12)	(0.12)	(0.04)
Top 10% zg	0.08	0.061	0.018
	(0.11)	(0.11)	(0.04)
Top 15% zg	0.138	0.109	0.03
	(0.09)	(0.09)	(0.03)
Top 20% zg	0.221	0.181	0.046
	(0.07)	(0.06)	(0.03)

Note: This table shows estimates of changes in program cutoffs relative to baseline. Lower cutoffs denote lower admissions standards.

Table A4: Outcome Probit Models

	Probit UTA Persist	Probit UTA STEM	Probit TAMU Persist	Probit TAMU STEM
Constant	1.403	-3.098	1.755	-2.542
	(0.218)	(0.313)	(0.34)	(0.305)
SAT	0.314	1.088	0.214	1.155
	(0.112)	(0.177)	(0.108)	(0.492)
Class Rank	-0.714	-1.661	-0.845	-0.923
	(0.518)	(1.188)	(0.569)	(0.962)
SAT Ratio	-0.208	1.35	-0.25	1.41
	(0.141)	(0.207)	(0.212)	(0.369)
Poverty	-0.953	-0.855	-0.908	-1.118
	(0.109)	(0.229)	(0.179)	(0.173)
URM	0.067	0.252	0.095	0.238
	(0.065)	(0.073)	(0.08)	(0.061)
Scholarship	0.333	0.104	-0.227	0.703
	(0.101)	(0.228)	(0.236)	(0.317)
Caliber (q)	-0.773	1.137	-1.314	-0.383
	(0.346)	(1.002)	(0.581)	(0.593)

Note: This table shows probit specifications for outcome equations (Equation (6)).

Figure A4: Preferences and Awareness, Estimates vs. Naive Models

Note: This figure shows average choice probabilities when B_i contains one copy of each institution (Panel A), and when elite private schools are excluded (Panels C and D). Panel B shows the share of students aware of aid at each institution. Panels A and C: all students fully aware of aid. Panel D: aid awareness is set to zero for all students. "Ignore selection": In panels A, C and D, use estimates from a random-coefficients nested-logit specification estimated on final enrollment decisions only, taking choice sets as given; in Panel B, use observed aid awareness only, ignoring selection into applications. Top-decile class rank and other students shown separately. Institutions: (1) Non-Flagship In-State Public; (2) Private Secular; (3) Religious; (4) Out-of-State Public; (5) Elite Private; (6) Texas A&M; (7) UT Austin.

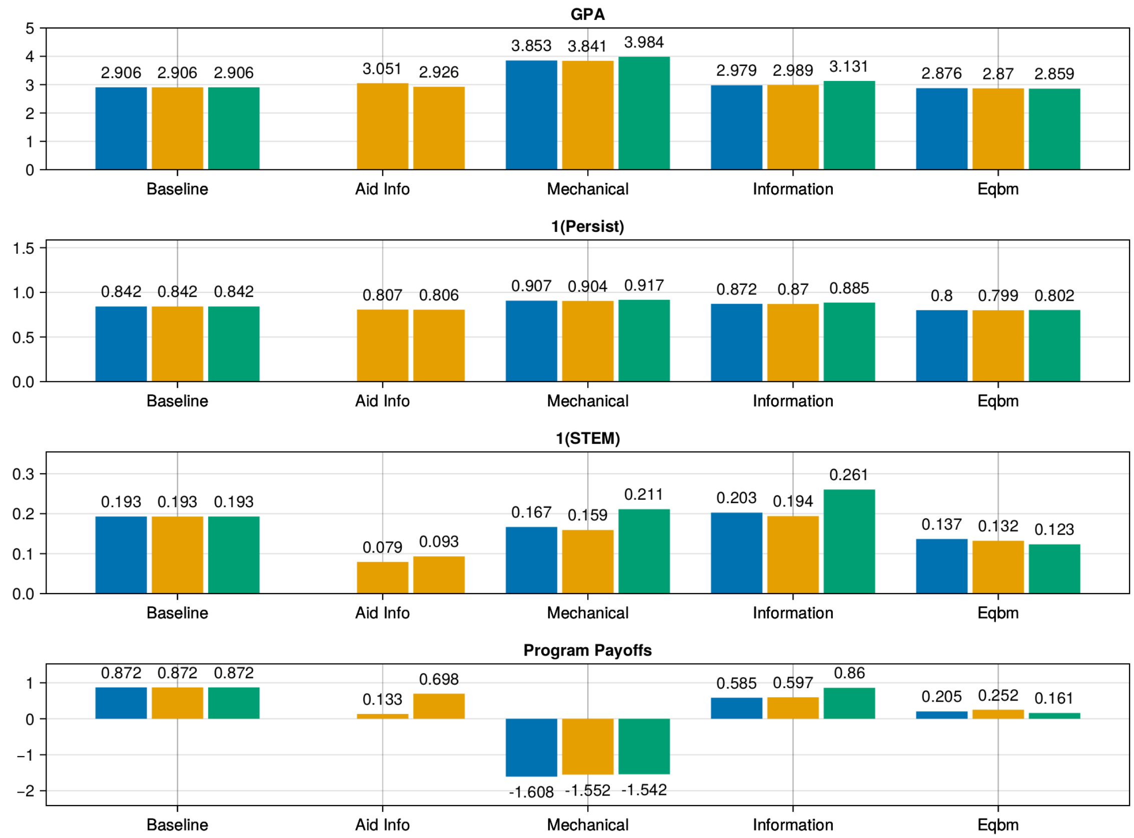
students "pushed out" by cutoff changes due to the percent plan.

In all academic measures, students pulled in by either the mechanical or informative channels outperform those pushed out.

Those induced to enroll by information are comparable to those enrolled at baseline in propensity to enroll in STEM, and are slightly more likely to persist and to earn high grades. Those pulled in mechanically earn even higher GPAs, but are less likely to major in STEM. Unsurprisingly, given that they were below the bar, those pulled in by mechanical effects provide very low payoffs to colleges.

Interestingly, when only aid information is provided, students "pulled in" by this information earn slightly higher GPAs but are less likely to persist and to major in STEM than the average student or the students they displace.

Figure A5: Decomposition of Post-Enrollment Outcomes



Note: This table shows mean post-enrollment outcomes of students enrolled in flagships, and of students shifted into or out of flagships by mechanical, informative, and equilibrium effects. "Baseline": means of students enrolled in flagships at baseline. "Aid info": left column shows means among students pulled in by aid information without the percent plan; right column shows means among students in turn "pushed out" by equililbrium cutoff changes. "Mechanical" and "informative": means of students pulled in by mechanical and informative effects respectively. "Eqbm": means of students "pushed out" by cutoff changes due to the percent plan.

E.4 Enrollment, Match Effects, and Student-Level Impacts

So far, I have focused on outcomes at flagship institutions. These are of interest in evaluating transparency. For instance, if the students shifted into flagships by the percent plan were to drop out at high rates, this would provide an argument against the percent plan. To quantify impacts on students, however, one needs to understand what would have happened had they not attended flagship institutions. This in turn depends on two factors: (1) where students would counterfactually enroll, and (2) the presence or absence of "match effects". In this section I first show impacts on enrollment, then discuss the possibility of match effects. Finally, I present estimated impacts of policy changes on the number of students persisting two years across all institutions, not just flagships. I find that the gains for top-decile students were larger than the losses for those displaced, leading to an overall increase in persistence.

Enrollment impacts: Figure A6 describes counterfactual enrollment, comparing equilibrium under the percent plan to the baseline scenario, in equilibrium, in which there is no such plan. Blue bars show the gains in enrollment at each institution in equilibrium under the percent plan, relative to baseline, among top-decile students, as a percentage of the total measure of top-decile students. Orange bars show *decreases* in enrollment at each institution in equilibrium under the percent plan, relative to baseline, among non-top-decile students. To make the scale comparable, changes are divided by the measure of top-decile students. This figure shows that the probability of enrollment at each flagship institution among top-decile students increases by roughly 4.5 points. At UT Austin, this is comparable to the 5.3-point increase among "pulled in" students estimated in Black et al. [2023]. In contrast, top-decile enrollment at other institutions falls. In particular, despite the presence of an admissions guarantee at non-flagship in-state public institutions, in fact the probability of enrollment there among top-decile students falls slightly with the percent plan, as some of these students substitute to flagships.[37]

Importantly, Under my counterfactuals, total four-year college enrollment essentially does not change. Because capacities are held fixed, necessarily an equal measure of non-top-decile students is displaced from flagships and from public non-flagship institutions. Moreover, while I do not hold capacities at the remaining four options fixed, aggregate changes in enrollment there are negligible. The total increase in four-year college enrollment among top-decile students is equal to 5.26% of top-decile students. The total decrease in four-year college enrollment among non-decile students is equal to 2.38% of non-top-decile students. In total these changes essentially offset, leading to a 0.01% increase in total four-year enrollment.

My results here differ from those of Black et al. [2023] because, in addition to the direct effects of admissions changes at flagships, I am including indirect effects in which students who are displaced from flagships in turn displace other students, who then displace others, and so on. By solving for equilibrium, I am accounting for the possibility that non-top-

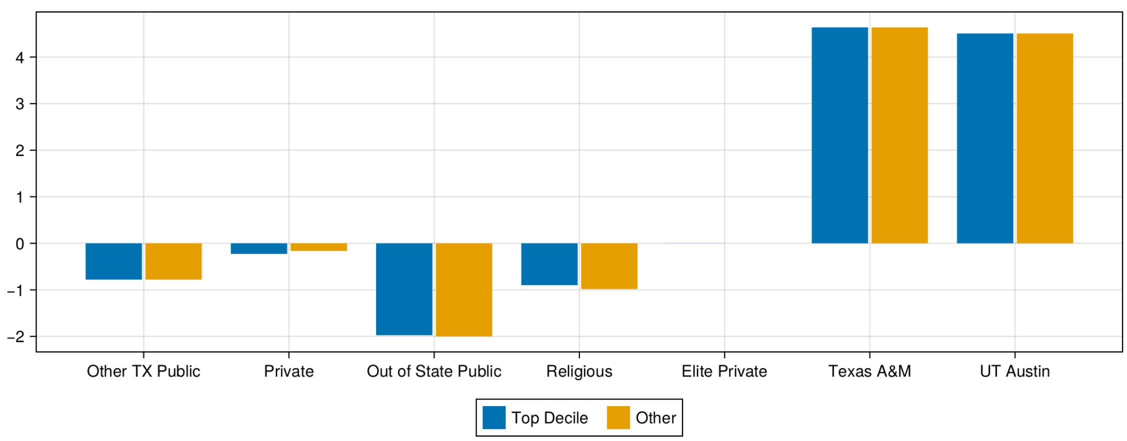
[37] I estimate a 5.26 point increase in overall four-year college enrollment for this group, comprising decreases in out-of-state and private institution enrollment and a 8.36 point increase in in-state public university enrollment. This latter number is larger than the 6.6 point increase in public four-year college enrollment estimated by Black et al. [2023] (p.49), and is driven by larger estimated enrollment increases at Texas A&M University, which may be due to differences in the population being considered.

decile students whom the percent plan pulled out of flagships set off chains of displacement.

Alternatively, one may wish to focus on specific groups of students defined by their potential assignments, as in Black et al. [2023]. There, "Pulled In" students are those who attend a flagship if the percent plan is in place and otherwise do not. Conversely, "Pushed Out" students attend a flagship if and only if the percent plan is not in place. Most students who will in turn be displaced from non-flagship institutions by the "Pushed Out" group would not have attended flagships under either scenario. They are therefore not part of either the "pulled in" or "pushed out" group.

To make my analysis comparable to Black et al. [2023], I compute a counterfactual in which flagships (only) provide automatic admission, and flagship programs' cutoffs adjust to their equilibrium values, but there is no automatic admission at non-flagship institutions, and cutoffs at non-flagship institutions are held fixed at their baseline values. This counterfactual captures the direct effect of the guarantee at flagships, holding the probability of having an option to attend a non-flagship institution fixed at its baseline value for every student.

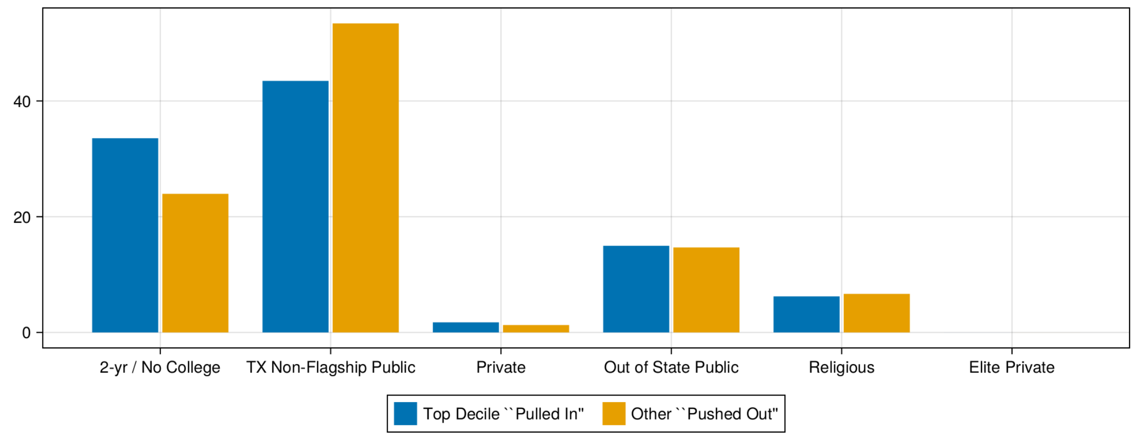
Figure A6: Equilibrium Changes in Enrollment



Note: This figure shows gains in enrollment shares in equilibrium under the percent plan, relative to baseline, for top-decile students (blue bars), and decreases in enrollment shares in equilibrium, relative to baseline, for non-top-decile students (orange bars).

Figure A7 provides results, showing the institutions that top-decile students who are "pulled in" to flagship institutions come from (blue bars), and the institutions that non-top-decile students "pushed out" of flagships attend (orange bars). I find that 34 percent of students "pulled in" substitute into flagships from the outside option, while only 24% of students "pushed out" substitute into the outside option (leftmost column). Instead, non-top-decile students displaced by the percent plan are more likely to substitute to non-flagship in-state institutions. These substitution patterns are broadly consistent with Black et al. [2023]. They find that about two thirds of "pushed out" students substitute into non-flagship institutions, while one third substitute into community colleges, while "pulled in" students are more likely to have been pulled into the public four-year sector from outside. Differences between Figures A7 and A6 are due to the subsequent adjustments in programs' cutoffs, and chains of rejections, set off by non-top-decile students' moves.

Figure A7: Where do Students Substitute From/To?



Note: This figure considers a counterfactual in which automatic enrollment is provided to top-decile students at flagships only. It shows enrollment shares, under this counterfactual, for top-decile students who attend flagship universities under this counterfactual but not at baseline (blue bars), and enrollment shares under the percent plan of students who attend flagships at baseline but are displaced when flagships' cutoffs adjust after the percent plan is introduced (orange bars).

Match Effects: The previous results suggest that, if there are gains or losses in aggregate human capital production, they will have to be driven by match effects. While students "pushed out" of flagships are more likely to attend other four-year institutions than those "pulled in" in the event they do not attend a flagship, the "pushed out" students in turn displace other students, so that overall effects on enrollment at any institution type are close to zero in equilibrium.

Estimating these effects requires estimating outcomes at non-flagship universities. To do so, I use THEOP public data from the non-flagship universities listed in Table A1. I restrict to Texas-resident students from public schools entering in Fall 2002, and drop institutions that do not have transcript data through at least 2004. As an outcome, I focus on two-year persistence.

I pool Rice University and SMU, and ascribe potential outcomes estimated at those universities to students enrolling in options 2 ("Secular Private"), 4 ("Religious"), and 5 ("Elite Private"). The remaining institutions, which are public, are used to estimate persistence at college 1 ("In-State Non-Flagship Public"). In addition, as I lack data on out-of-state institutions, I use the outcome equation estimated there to predict persistence for students matched to out-of-state public institutions.

Estimation is as described in Section 5.4. However, the auxiliary-model covariates w_{ij}^{outcome} in Equation (7) differ. I observe high schools' poverty quartile, but not exact FRPL rate, mean SAT scores, or the availability of the LOS or CS programs. Hence, for non-flagship options j, the auxiliary-model covariates include indicators for poverty quartiles, and do not include "SAT ratio":

$$w_{ij}^{\text{outcome}} = \left(1, \text{SAT}_i, \text{classrank}_i, 1\left((\text{poverty quartile})_{h(i)}\right), \text{classrank}_i^2, \text{classrank}_i^3, \text{topdecile}_i, \text{urm}_i\right).$$

A second difference is that I do not include scholarships among the outcome covariates, as there were no targeted scholarship programs for these institutions comparable to the LOS or CS programs.

Table A5: Persistence Parameters, Non-Flagship Institutions

	Non-Flagship Public	Private
Constant	0.22	0.578
SAT	-2.754	1.433
Class Rank	0.14	-0.311
SAT Ratio	2.193	-0.681
Poverty	-0.142	0.015
URM	0.152	0.005
Caliber (q)	-0.89	0.032

Note: This table shows outcome indices (Equation (6)) for non-flagship institutions. As these institutions do not have targeted (LOS or Century Scholars) scholarships, "Scholarship" is omitted.

Table A5 gives estimates of persistence parameters. While the coefficient on SAT is negative at public institutions, the coefficient on SAT ratio is also highly positive, leading to positive effects of increases in SAT scores all else equal (for the mean student enrolling in non-flagship public institutions, the peer-average SAT score at baseline is equal to 0.68.) Caliber is negatively associated with persistence at public institutions but not relevant at private institutions.

Figure A8 shows the share of students who enroll and persist at least two years in some four-year institution under my counterfactuals of interest overall (Panel A) and by subgroups (Panels B-F).

Overall persistence is roughly half a point higher under the percent plan than at baseline, and would be a further half-point higher if accompanied by aid information. Persistence under the "aligned" percent plan is similar to baseline, however.

Turning to subgroups, as the percent plan raises top-decile enrollment, the fraction of top-decile students enrolling and persisting grows, from 50% at baseline to 57.5% with the percent plan and aid-information intervention. The percent plan, with or without the aid intervention, would also raise the share of URM students (Panel C) and students from lowest-quartile schools (Panel D) who enroll and persist.

In sum, this evidence suggests that the percent plan raised total persistence, and improved outcomes for disadvantaged subgroups, by assigning students to colleges who were more likely to benefit. However, in equilibrium, tracing chains of rejection, some non-top-decile students were crowded out.

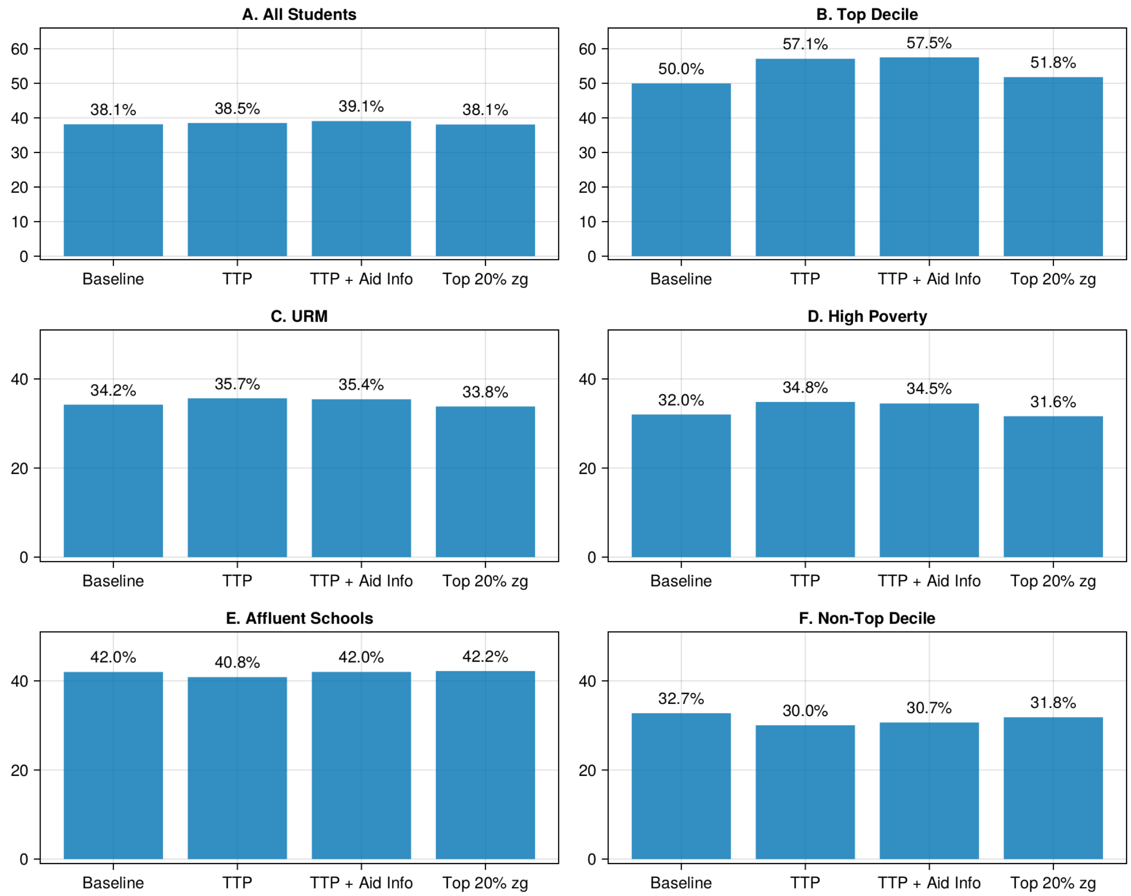
A caveat is that I am holding capacity at non-flagship in-state institutions fixed. If there is slack at some of those institutions, then total enrollment could increase with the percent plan (See Figure A7). This in turn would mechanically lead to further increases in the number of students persisting.

Black et al. [2023] conjecture that "Pushed Out students are likely to come from families with more support for college success, so may be less dependent on inputs received from the college itself."(p.30) My results are consistent with this conjecture. Conversely, Bleemer [2024] argues that gains from selectivity may be larger for students who are less well prepared. In contrast to the setting of Bleemer [2024], "Pulled In" students in my setting earn higher grades, and are more likely to persist, than the average student at the institutions they are pulled into.

E.5 Effects of the Longhorn Opportunity Scholars (LOS) Program

This section provides an analysis of the effects of the LOS program, under the assumption that the assignment of the program to high schools was "as good as random" conditional on schools' poverty rates, average SAT scores, and other observables. I first show that estimated effects are consistent with results from the literature. I then provide further evidence of complementarities between admissions transparency and other forms of support—in this case, the availability of scholarship and mentorship programs.

Figure A8: Two-Year Persistence, All Institutions

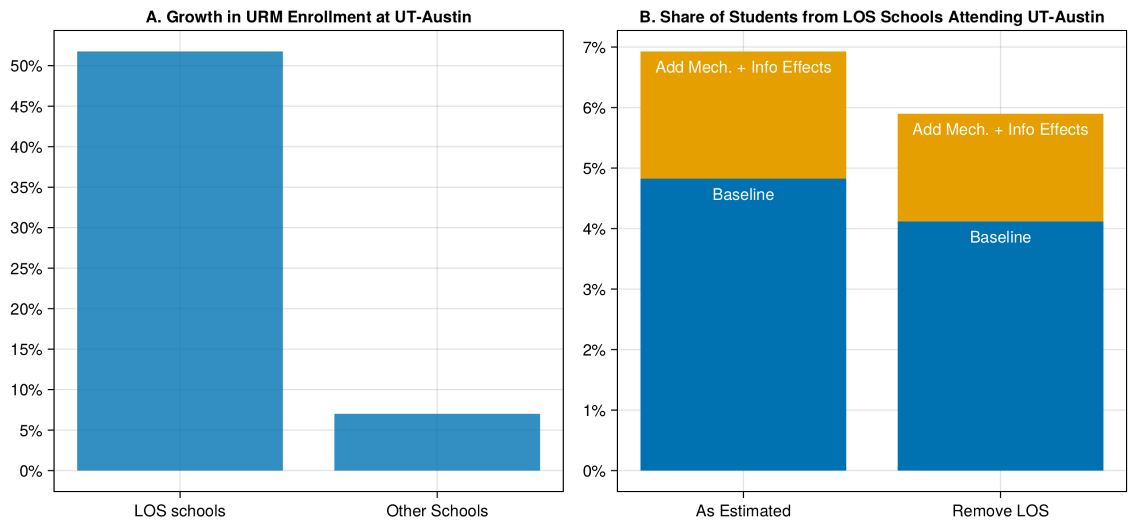


Panel A: Percentage of all students (population-weighted) who enroll in a four-year college and persist at least two years. Panels B through F: share of students in given subgroup who enroll in a four-year college and persist at least two years. “Top Decile”: top 10% by class rank; “URM”: Black or Hispanic; “High Poverty”: top quartile of HS ever-received-FRPL; “Affluent Schools”: bottom quartile of HS ever-received-FRPL; “Non-Top Decile”: class rank outside of top decile.

Panel A of Figure A9 decomposes differences between my equilibrium and baseline counterfactuals, showing percentage growth in URM enrollment at UT Austin, relative to baseline enrollment numbers, separately by LOS schools—defined as those schools that participated in the LOS program in 2002—and other schools. At the baseline admissions cutoffs that I estimate, if there were no LOS program, 13.0% of URM students at UT Austin would have come from LOS schools. I estimate that the percent plan and the LOS program combined caused this share to grow by 51.8%. In contrast, the share of minority students from other high schools would have grown by 7.0%. As such, most of the growth in URM enrollment at UT Austin was driven by LOS schools.

This result is consistent with the argument and descriptive evidence in Kain et al. [2005] that effects were driven by LOS schools. Those authors argue that apparent increases in minority enrollment at UT Austin elsewhere may be due to other demographic changes, such as increases in the share of minority students at suburban high schools over the years 1998-2002, which were not caused by the percent plan. An advantage of my model-based approach is that it is able to evaluate policy changes while holding demographics fixed.

Figure A9: Effects of LOS Program



Note: Panel A shows growth in URM enrollment, relative to baseline total among students from LOS schools (left column) and other schools (right column), when the percent plan is introduced. Results come from model simulations, comparing "baseline" and "equilibrium" (with TTT) scenarios. In both cases, the LOS program is present at LOS schools. Panel B shows the impacts of the LOS program. Left column: baseline (no TTP) and mechanical+information effects of percent plan, taking LOS as given, among students from LOS schools. Right column: baseline (no TTP) and mechanical+information effects of percent plan on UT Austin enrollment among students from LOS schools, if there were no LOS program.

Panel B of Figure A9 shows baseline shares of students from LOS schools attending UT Austin (blue), and baseline shares + mechanical and informative effects (orange). The left column, "as estimated," reports mean estimated propensities to enroll at UT Austin from the main specification, restricting to students from LOS schools. In the right column, "Remove LOS", I consider the same set of students, but counterfactually remove the LOS program. I find that the LOS program would have increases UT Austin enrollment even without the percent plan, with 4.8% of students from LOS schools attending UT Austin, as opposed to 4.1% if there were no LOS program. Moreover, enrollment effects of the percent plan and the LOS program are complementary. UT-Austin enrollment would have increased by 2.1 points, in this population, if LOS and the percent plan were both introduced, as they were in the data (left column). If only the percent plan had been introduced, the increase in enrollment would have been 1.8 points, which is smaller. Hence, in the same way that admissions transparency and aid information are complements, admissions transparency and targeted scholarship/mentorship programs are complementary as well.

There are two important caveats to this analysis. First, for simplicity, and to isolate complementarities holding all else fixed, I am using previously-estimated baseline and equilibrium cutoffs which held the presence of the LOS program fixed. It would be possible to solve for equilibrium without the LOS program. Second, I am not modeling school-level unobservables with which LOS participation might be correlated. If LOS schools were negatively selected on baseline share of students sent to UT Austin, then my estimates of these complementarities may be lower bounds.

E.6 Extension: Class balancing

In this section I consider robustness to a form of class balancing. Indices for GPA and persistence load heavily on class rank and exam scores. Because the Top Ten Percent plan forces colleges to admit students with high class rank, a college that wants some students with strong personal scores might place additional weight on personal characteristics for the remaining seats when the percent plan is in effect. In the absence of the percent plan, it might stop doing so [Antonovics and Backes, 2013b].

Accordingly, I allow admissions offices to place additional weight on academic characteristics in the baseline scenario in which there is no percent plan. To operationalize this idea, I take $\gamma_{UTA}^{\text{persist}}$, the coefficients on (z,q) in equation (6), as the academic index (results are similar for Texas A&M's indices, and for GPA indices). I define $\gamma^{\text{personal}} \equiv (\gamma^{\text{admit}}, \gamma_q^{\text{admit}}) - (\gamma_{UTA,z}^{\text{persist}}, \gamma_{UTA,q}^{\text{persist}})$, where $\gamma_q^{\text{admit}} = 1$. Table A6 provides results. In the first row, at baseline, the admissions index is given by $\gamma^{\text{personal}} + 4\gamma_{UTA}^{\text{persist}}$. (I also consider a $2\times$ weight). Once top-decile students are automatically admitted, colleges return to their estimated preferences for the remaining applicants.

In this case, mechanical effects on top decile students' enrollment are not distinguishable from zero, and (unsurprisingly) impacts on persistence and other outcomes are attenuated. However, informative effects remain large and would raise enrollment of top-decile students by up to eight percentage points.

Table A6: Effects of TTP with Class Balancing

	Pr(Enroll in Flagship), Treated Group				Average Effect on Outcomes		
	Base	+ Mech.	+ Info	% Info	ΔGPA	ΔPersist	ΔSTEM
Class Balance: Additional Weight on Acad. Index							
(7): 4×	26.13	1.41	6.33	81.78	0.52	7.51	1.88
	(2.74)	(2.73)	(0.4)	(13.75)	(0.33)	(6.5)	(8.43)
(8): 2×	25.22	2.32	6.33	73.21	0.66	13.08	7.68
	(2.04)	(1.93)	(0.4)	(9.87)	(0.12)	(5.72)	(7.32)

Note: This table shows average impacts on academic outcomes, allowing for increased weight on an academic index in the baseline scenario. "Base": Share of "treated" group enrolling at baseline. "Mech": mechanical effect. "Info": information effect. "% Info": Share of total increase in treated-group enrollment due to information. Outcomes: total change in outcome (GPA, 1(Persist), 1(Persist and major in STEM), college payoff) / measure of top-decile students induced to enroll in flagships by mechanical and informative effects; this is equivalent to the average difference in outcomes between these students and the students they displaced.

E.7 Extension: Selection on student preferences and caliber

In this section I extend the model of outcomes to include selection on students' preferences. Let $\tilde{\nu}_{ij} = \nu_{ij} + w_j \beta_i^w$ denote the value of preference unobservables—application-time shocks plus terms arising from random coefficients—that i will receive if she enrolls in j. I assume:

$$\text{outcome}_{ij} = z_i^{\text{outcome}} \gamma_{z,j}^{\text{outcome}} + q_i \gamma_{q,j}^{\text{outcome}} + \tilde{\nu}_{ij} \gamma_{\nu,j}^{\text{outcome}} + \mu_{ij}^{\text{outcome}}. \tag{10}$$

Thus outcomes may be selected both on "caliber," which is observed by colleges, and on students' preferences.[38] For instance, the model allows students with a stronger preference for Texas A&M University to be more likely to persist if enrolled there, conditional on caliber. Estimation is as before. Implicitly, the indirect-inference procedure is making use of the "opposite" scholarship, e.g. the availability of the Longhorn Scholarship for students who enroll at Texas A&M. Intuitively, because the LOS program makes UT Austin observably more appealing, if a student from a LOS school attends Texas A&M this student must have had a higher preference shock for Texas A&M, in expectation, all else equal.

I hold the application-admissions-matriculation model parameters fixed, so that patterns of enrollment are as in the main specification. Table A7 presents parameters of the

[38] I have also estimated specifications of the form

$$\text{outcome}_{ij} = z_i^{\text{outcome}} \gamma_{z,j}^{\text{outcome}} + q_i \gamma_{q,j}^{\text{outcome}} + \nu_{ij} \gamma_{\nu,j}^{\text{outcome}} + \mu_{ij}^{\text{outcome}}$$

in which selection on preferences occurs only through the match-level shock ν. Results are similar.

outcome equations, and Table A8 presents post-enrollment outcomes under counterfactuals, analogous to the main results in Table 3.

Table A7: Outcome Parameters Allowing Selection on Preferences

	UTA GPA	LPM UTA Persist	LPM UTA STEM	Probit UTA Persist	Probit UTA STEM	TAMU GPA	LPM TAMU Persist	LPM TAMU STEM	Probit TAMU Persist	Probit TAMU STEM
Constant	2.646	0.598	0.964	-0.095	-1.586	2.718	0.962	1.635	-0.299	-2.327
	(0.382)	(0.194)	(0.071)	(0.263)	(0.739)	(0.492)	(0.098)	(0.146)	(0.372)	(0.535)
SAT	4.748	0.251	0.541	0.191	0.522	3.866	0.132	0.215	0.34	1.586
	(0.357)	(0.536)	(0.03)	(0.117)	(0.205)	(0.856)	(0.026)	(0.248)	(0.307)	(0.822)
Class Rank	-0.787	-0.394	-1.202	-0.267	-0.424	-0.38	-0.454	-0.805	-0.18	-1.799
	(0.325)	(0.152)	(0.164)	(0.502)	(1.809)	(0.493)	(0.178)	(0.208)	(0.536)	(0.797)
SAT Ratio	-2.714	0.005	-0.121	0.215	0.951	-2.223	-0.088	-0.244	0.436	1.161
	(0.239)	(0.393)	(0.032)	(0.135)	(0.3)	(0.459)	(0.044)	(0.229)	(0.317)	(0.501)
Poverty	-0.22	-0.342	-1.124	-0.111	-0.346	-0.176	-0.262	-0.974	-0.296	-0.961
	(0.157)	(0.145)	(0.031)	(0.148)	(0.264)	(0.511)	(0.085)	(0.105)	(0.224)	(0.273)
URM	0.053	0.028	0.08	0.045	0.184	-0.009	0.002	0.098	0.085	0.227
	(0.038)	(0.023)	(0.014)	(0.064)	(0.087)	(0.071)	(0.022)	(0.024)	(0.087)	(0.094)
Scholarship	0.192	0.094	0.345	0.018	0.073	0.077	0.016	-0.217	0.15	0.586
	(0.054)	(0.03)	(0.029)	(0.1)	(0.197)	(0.137)	(0.072)	(0.1)	(2.207)	(0.275)
Caliber (q)	-0.991	0.042	-0.298	0.129	-0.129	-1.097	-0.062	-1.25	-0.199	0.128
	(0.349)	(0.124)	(0.126)	(0.367)	(1.113)	(0.479)	(0.141)	(0.211)	(0.493)	(0.565)
Pref. Shock	-0.043	0.389	0.543	-0.202	-1.837	-0.51	0.21	0.489	-0.096	-0.827
	(0.71)	(0.253)	(0.059)	(0.243)	(0.579)	(1.499)	(0.149)	(0.489)	(0.326)	(0.835)

Note: This table shows outcome indices (Equation (10)), extending the model to allow selection on unobserved components of utility.

GPA parameters are nearly unchanged. Persistence and STEM parameters are similar to the main estimates, but there is evidence of selection on match quality, although the preference coefficients (final row) are noisily estimated. However, even though preferences appear to matter for persistence in this specification—i.e. students who like their matched programs more indeed are more likely to persist—the counterfactual results are statistically indistinguishable from those of the main specification.

Turning to counterfactuals in Table A8, estimated impacts on GPA are nearly identical to those of the main specification, reported in Table 3. Point estimates of impacts on persistence and STEM major are slightly smaller and slightly larger, respectively, than in the main specification, but estimates here are not statistically distinguishable from the main estimates.

An implication of outcomes' dependence on preference shocks is that, to the extent that colleges value those outcomes, they would now face a selection problem in their admissions decisions. If colleges were to maximize persistence, and this in turn depended on ν_{ij}, then the conditional expectation of outcome-relevant unobservables, conditional on

Table A8: Changes in Outcomes, Allowing Selection on Preferences

	ΔGPA	ΔPersist	ΔSTEM
A. Texas Top Ten			
(1): TTP	0.72	13.1	11.96
	(0.08)	(2.77)	(3.25)
(2) +Aware	0.73	13.32	11.01
	(0.08)	(2.81)	(3.43)
B. Automatic Admission By $z^{admit}\gamma$			
(3): Top 5%	0.21	2.02	5.15
	(0.12)	(6.98)	(13.04)
(4): Top 10%	0.39	4.36	9.58
	(0.05)	(4.81)	(8.08)
(5): Top 15%	0.54	6.34	12.97
	(0.06)	(3.82)	(5.62)
(6): Top 20%	0.69	8.5	16.51
	(0.07)	(2.75)	(3.49)

Note: This table shows a decomposition of changes in flagship enrollment for treated households, and average impacts on academic outcomes and programs' payoffs, allowing academic outcomes to be selected on preference shocks. "Base": Share of "treated" group enrolling at baseline. "Mech": mechanical effect. "Info": information effect. "% Info": Share of total increase in treated-group enrollment due to information. Outcomes: total change in outcome (GPA, 1(Persist), 1(Persist and major in STEM), college payoff) / measure of top-decile students induced to enroll in flagships by mechanical and informative effects; this is equivalent to the average difference in outcomes between these students and the students they displaced.

students' observables and accepting the offer, could differ across counterfactuals or policy changes. For instance, top-decile students might be more positively selected on preference shocks if there were no top-ten-percent plan. Put differently, if colleges did not also observe preference shocks, they would no longer have private values. For this reason, in counterfactuals the score π_{ij} or relative ranking of students could change in principle, in addition to changes in "cutoffs".

As one obtains essentially the same impacts on counterfactual persistence, grades, and STEM majors abstracting from this source of selection, I abstract from it in the main specification.

F Supplementary Materials [Not For Publication]

F.1 Additional Descriptives

Table S1 reports THEOP sample means, using the population weights provided by the survey. According to these weights, about 7% of students attend a LOS school, and 15% attend a school with a poverty rate greater than 60%. Table S2 shows sample means at the student-by-college level. Tables S3 through S5 study how students' demographic covariates and application behavior differ by survey wave 2 participation, top-decile status, and participation in the Longhorn Opportunity Scholars program, respectively. Wave-2 students are disproportionately Black and Asian. These differences reflect an intentional decision by survey designers to oversample Black and Asian students. They are similar to wave-1 students in SAT scores and high school poverty rates. Top-decile applicants are about 6 percentage points more female than non-top-decile students. They are more likely to be Asian, and less likely to be Black, than non-top-decile students, but are equally likely to be White and/or Hispanic. They are much more likely to apply to UT-Austin, to apply for aid there, and to be admitted there. LOS participants are more likely to be Black, equally likely to be Hispanic, and less likely to be White or Asian, than the general population. They come from poorer schools than average and have lower SAT scores.

Table S6 provides statistics on admissions and matriculations by in-state and out-of-state applicants at UT Austin over the years 1991 through 2003 as given in the administrative data. The fraction of the enrolled students whom the university classified as in-state applicants remained slightly above 90% in each year with the exception of 1997, where 89.9% of first-time freshmen were in-state. The share of students from in-state public high schools is also nearly constant.

F.2 Parameter Estimates

Tables S7 through S9 provide estimates from the main specification.

Table S1: Sample Means: Students (Pop. Weighted)

Variable	Mean	Std. Dev.	Min.	Max.
Participates in Century Scholars Program	0.025	0.157	0	1
Participates in Longhorn Opp. Scholars Program	0.069	0.253	0	1
White	0.626	0.484	0	1
Black	0.077	0.266	0	1
Latino	0.192	0.394	0	1
Asian	0.049	0.215	0	1
Female	0.567	0.496	0	1
Num. Guardians	1.714	0.546	0	2
Class Rank	0.268	0.181	0.1	1
SAT*	0.67	0.117	0.006	1
HS mean SAT	0.682	0.055	0.431	0.806
SAT / HS mean SAT	0.984	0.161	0.008	1.553
HS Poverty (share ever FRPL)	0.323	0.244	0.023	1
High Poverty HS	0.147	0.354	0	1
Affluent HS	0.283	0.45	0	1
Applied Anywhere	0.861	0.346	0	1
Applications	1.417	1.142	0	5
Admissions Offers	1.085	0.998	0	5
Aid Applications	1.041	1.067	0	5
Enrolled	0.326	0.469	0	1
Did Not Enroll in 4-year Institution	0.14	0.347	0	1
Not in Wave 2	0.534	0.499	0	1
N		4143		

Table S2: Sample Means: Applications

	All Potential Applications	Applications	Offers	Matriculations
Distance (100mi)	188.32	104.07	97.79	86.39
Cost of Attendance ($10000)	1.83	1.53	1.49	1.39
Frac. minority	0.20	0.28	0.28	0.29
Frac. own race	0.55	0.69	0.69	0.70
1(Top Ten)	0.12	0.23	0.28	0.30
Applied for aid	0.13	0.86	0.91	0.93
Applications	0.18	1.20	1.22	1.20
Admitted	0.26	0.89	1.16	1.01
In Wave 2	0.48	0.49	0.49	1.00
Enrolled	0.06	0.27	0.31	1.00
N	33,144	4,945	3,780	1,382

Table S3: Sample Balance, Wave 2 Participation

Variable	Mean, Not wv2	Increase if wv2	P-value	N
Female	0.57	0.01	0.54	4,143
Black	0.04	0.09***	0.00	4,143
Hispanic	0.20	-0.02	0.15	4,143
White	0.67	-0.12***	0.00	4,143
Asian	0.02	0.07***	0.00	4,143
SAT	0.67	0.01**	0.05	4,143
SAT / HS mean SAT	0.98	0.01***	0.01	4,143
HS poverty rate	0.32	0.01*	0.06	4,143
Apply Anywhere	0.85	0.03***	0.00	4,143
Num. Apps	2.37	0.20***	0.00	4,143
Apply to UT-A	0.14	0.04***	0.00	4,143
Apply for aid at UT-A	0.09	0.04***	0.00	4,143
Admitted to UT-A	0.12	0.04***	0.00	4,143

Note: this table shows means among surveyed students not participating in Survey Wave 2, and differences in means between Wave 2 and non-Wave 2 students. Wave 1 population weights are used. Oversampling of Black students in Wave 2 reflects an intentional choice on the part of the survey designers. *** indicates $p < 0.01$. ** indicates $p < 0.05$. * indicates $p < 0.1$.

Table S4: Sample Balance, Top-Decile Status

Variable	Mean, Not ttt	Increase if ttt	P-value	N
Female	0.54	0.06***	0.00	4,143
Black	0.09	-0.05***	0.00	4,143
Hispanic	0.19	0.01	0.43	4,143
White	0.63	-0.01	0.45	4,143
Asian	0.03	0.06***	0.00	4,143
SAT	0.64	0.10***	0.00	4,143
SAT / HS mean SAT	0.94	0.16***	0.00	4,143
HS poverty rate	0.31	0.05***	0.00	4,143
Apply Anywhere	0.82	0.11***	0.00	4,143
Num. Apps	2.34	0.32***	0.00	4,143
Apply to UT-A	0.11	0.19***	0.00	4,143
Apply for aid at UT-A	0.06	0.17***	0.00	4,143
Admitted to UT-A	0.07	0.22***	0.00	4,143

Note: this table shows means among surveyed students not placing in the top decile of their class, and differences in means between top-decile and non-top-decile students. Wave 1 population weights are used. *** indicates $p < 0.01$. ** indicates $p < 0.05$. * indicates $p < 0.1$.

F.3 Data Supplement

Calculating EFC

Matching THEOP and CPS datasets In order to draw incomes from the CPS, I recode parents' education in the THEOP survey to match the CPS dataset. I then draw incomes for each student in THEOP from the distribution of income given parents' occupation and the education of the most-educated parent.

1. THEOP senior survey

   - Mother's education: q71
   - Father's education: q67
   - `label list q71`
     ```
     q71:
     1 No Schooling
     2 Elementary School
     ```

Table S5: Sample Balance, LOS Participation

Variable	Mean, Not longhorn	Increase if longhorn	P-value	N
Female	0.56	-0.01	0.63	4,143
Black	0.06	0.28***	0.00	4,143
Hispanic	0.17	0.29***	0.00	4,143
White	0.67	-0.55***	0.00	4,143
Asian	0.05	-0.05***	0.00	4,143
SAT	0.68	-0.09***	0.00	4,143
SAT / HS mean SAT	0.98	0.02**	0.03	4,143
HS poverty rate	0.29	0.51***	0.00	4,143
Apply Anywhere	0.86	-0.07***	0.00	4,143
Num. Apps	2.45	-0.11	0.34	4,143
Apply to UT-A	0.16	-0.06***	0.00	4,143
Apply for aid at UT-A	0.11	-0.03	0.12	4,143
Admitted to UT-A	0.13	-0.05***	0.00	4,143

Note: this table shows means among surveyed students whose high schools do not participate in the Longhorn Opportunity Scholars program, and differences in means between students whose schools participate and those who don't. Wave 1 population weights are used. *** indicates $p < 0.01$. ** indicates $p < 0.05$. * indicates $p < 0.1$.

Table S6: Fraction of admissions and matriculations from in-state applicants, UT Austin

year	admitted (total)	enrolled (total)	% admitted (in-state)	(TX public)	% enrolled (in-state)	(TX public)
1991	10403	5818	85.0	71.8	90.4	84.1
1992	9726	5613	85.0	71.5	90.9	84.2
1993	10085	5861	86.0	72.3	91.6	84.3
1994	10278	6156	84.0	74.3	89.6	81.7
1995	10506	6346	84.3	74.3	90.9	82.8
1996	11041	6381	81.7	71.7	90.1	81.7
1997	11352	7138	82.2	72.8	89.9	82.2
1998	10693	6833	84.9	75.7	92.0	83.2
1999	10990	7288	89.1	79.4	93.5	84.6
2000	13061	8118	87.5	77.8	93.1	84.7
2001	12564	7243	85.0	76.7	91.1	83.8
2002	14138	7868	86.3	77.7	90.9	83.6
2003	10820	6493	87.1	80.0	92.9	86.9

Note: this table shows the fraction of admissions and matriculations from in-state applicants and from in-state applicants who attended public schools at the University of Texas at Austin.

```
3 Some High School
4 High School Graduate
```

Table S7: Parameters 1/3

	Parameter	Estimate	SE	p5	p95
1	$\underline{\pi}$ Eqbm. (In-State Public)	0.373	0.354	-0.039	1.104
2	$\underline{\pi}$ Eqbm. (Private)	1.626	0.44	1.182	2.405
3	$\underline{\pi}$ Eqbm. (Out-of-State Public)	0.717	0.363	0.328	1.54
4	$\underline{\pi}$ Eqbm. (Relig.)	0.901	0.377	0.433	1.654
5	$\underline{\pi}$ Eqbm. (Elite Private)	4.529	0.418	3.987	5.354
6	$\underline{\pi}$ Eqbm. (Texas A&M)	1.669	0.372	1.243	2.526
7	$\underline{\pi}$ Eqbm. (UT Austin)	1.573	0.38	1.153	2.423
8	γ (SAT)	5.624	0.948	4.318	7.43
9	γ (Class Rank)	-2.435	0.197	-2.73	-2.11
10	γ (SAT/HS mean SAT)	-1.2	0.486	-1.929	-0.318
11	γ (Poverty)	0.278	0.244	-0.123	0.683
12	γ (URM)	0.011	0.074	-0.108	0.104
13	γ^s (Const)	-10.0	0.556	-10.0	-8.715
14	γ^s (Poverty)	1.563	8.376	-10.0	10.0
15	γ^s (URM)	5.748	6.26	-10.0	6.596
16	$\gamma^{a\|s}$ (Const)	-0.692	0.353	-1.217	-0.31
17	$\gamma^{a\|s}$ (Poverty)	1.625	0.763	0.442	2.705
18	$\gamma^{a\|s}$ (URM)	0.545	0.305	0.145	1.045
19	λ (Matric. shock scale)	10.0	0.0	10.0	10.0
20	γ^{fixed} (Const)	0.31	0.091	0.19	0.44
21	γ^{fixed} (Poverty)	-0.239	0.111	-0.361	-0.062
22	γ^{fixed} (URM)	0.047	0.06	-0.077	0.122
23	γ^{var} (Const)	1.038	0.053	0.95	1.099
24	γ^{var} (Poverty)	-0.219	0.119	-0.46	-0.064
25	γ^{var} (URM)	-0.176	0.045	-0.225	-0.098
26	γ^{shock} (Const)	-0.88	0.094	-1.044	-0.795
27	γ^{shock} (Poverty)	-0.592	0.21	-0.965	-0.272
28	γ^{shock} (URM)	-0.075	0.079	-0.168	0.088
29	β^p (Const)	2.808	0.555	2.57	4.309
30	β^p (income)	-0.422	0.055	-0.534	-0.369
31	$\beta^{(w,x,z)}$ (In-State Public)	3.058	0.398	2.876	4.034
32	$\beta^{(w,x,z)}$ (Private)	-0.898	1.03	-1.442	0.798

Note: This table shows all estimated parameters (Part 1 of 3).

```
5 Some College
6 Two-Year College
7 Four-Year College
8 Master's Degree
9 Professional Degree
```

2. 2002 March CPS

- Education of household head
- `label list EDUC_HEAD`

Table S8: Parameters 2/3

	Parameter	Estimate	SE	p5	p95
33	$\beta^{(w,x,z)}$ (Out-of-State Public)	0.937	0.583	0.493	2.112
34	$\beta^{(w,x,z)}$ (Relig.)	2.822	0.613	2.076	4.049
35	$\beta^{(w,x,z)}$ (Elite Private)	1.798	2.688	-0.611	5.286
36	$\beta^{(w,x,z)}$ (Texas A&M)	2.888	0.605	2.362	4.175
37	$\beta^{(w,x,z)}$ (UT Austin)	0.208	0.707	-0.422	1.898
38	$\beta^{(w,x,z)}$ (Dist < 25)	0.17	0.07	0.053	0.27
39	$\beta^{(w,x,z)}$ (Distance)	-0.184	0.028	-0.229	-0.14
40	$\beta^{(w,x,z)}$ (URM)	-0.108	0.133	-0.346	0.101
41	$\beta^{(w,x,z)}$ (URM X UTA)	0.127	0.095	-0.011	0.268
42	$\beta^{(w,x,z)}$ (URM X TAMU)	-0.367	0.19	-0.732	-0.148
43	$\beta^{(w,x,z)}$ (Poverty)	-0.7	0.397	-1.374	-0.147
44	$\beta^{(w,x,z)}$ (Poverty X UTA)	-0.354	0.188	-0.745	-0.174
45	$\beta^{(w,x,z)}$ (Poverty X TAMU)	-0.708	0.353	-1.401	-0.423
46	$\beta^{(w,x,z)}$ (LOS X UTA)	0.099	0.175	-0.217	0.289
47	$\beta^{(w,x,z)}$ (Century X TAMU)	-0.203	0.3	-0.763	0.1
48	$\beta^{(w,x,z)}$ (SAT)	-1.141	0.647	-2.047	-0.277
49	$\beta^{(w,x,z)}$ (Class Rank)	-0.098	0.545	-0.926	0.683
50	$\beta^{(w,x,z)}$ (SAT / HS Mean SAT)	-0.222	0.377	-0.702	0.338
51	$\beta^{(w,x,z)}$ (SAT X TAMU)	1.296	0.532	0.094	1.782
52	$\beta^{(w,x,z)}$ (SAT X UTA)	4.485	0.637	3.186	5.174
53	$\beta^{(w,x,z)}$ (SAT X Private)	2.502	0.533	1.617	3.251
54	$\log(\sigma^{rc})$ (Distance)	-0.61	3.104	-10.0	-0.249
55	$\log(\sigma^{rc})$ (S/F Ratio)	-10.0	0.005	-10.0	-10.0
56	$\log(\sigma^{rc})$ (UTA vs TAMU)	-10.0	1.529	-10.0	-10.0
57	β^{aware} (In-State Public)	4.014	2.686	-1.403	5.895
58	β^{aware} (Private)	-0.261	3.455	-7.282	0.251
59	β^{aware} (Out-of-State Public)	0.832	2.873	-5.287	2.077
60	β^{aware} (Relig.)	3.542	2.549	-1.723	5.399
61	β^{aware} (Elite Private)	-1.675	4.038	-10.0	0.208
62	β^{aware} (Texas A&M)	-3.733	3.42	-10.0	-0.113
63	β^{aware} (UT Austin)	1.023	3.602	-9.453	0.791
64	β^{aware} (Dist < 25)	-0.421	0.706	-0.884	0.886

Note: This table shows all estimated parameters (Part 2 of 3).

```
EDUC_HEAD:
0 NIU or no schooling
1 niu
2 None or preschool
10 Grades 1, 2, 3, or 4
11 Grade 1
12 Grade 2
13 Grade 3
14 Grade 4
```

Table S9: Parameters 3/3

	Parameter	Estimate	SE	p5	p95
65	β^{aware} (Distance)	-0.05	0.194	-0.334	0.157
66	β^{aware} (URM)	0.779	0.55	0.327	1.797
67	β^{aware} (URM X UTA)	1.168	1.15	-0.342	3.156
68	β^{aware} (URM X TAMU)	-0.404	1.806	-1.535	3.772
69	β^{aware} (Poverty)	2.319	1.111	1.683	5.004
70	β^{aware} (Poverty X UTA)	0.749	3.027	-1.792	7.566
71	β^{aware} (Poverty X TAMU)	7.591	2.649	3.327	10.0
72	β^{aware} (LOS X UTA)	10.0	1.853	5.345	10.0
73	β^{aware} (Century X TAMU)	-3.785	3.652	-9.571	1.994
74	β^{aware} (SAT)	3.394	3.351	0.643	10.0
75	β^{aware} (Class Rank)	-10.0	0.0	-10.0	-10.0
76	β^{aware} (SAT / HS Mean SAT)	-2.208	1.664	-3.99	0.722
77	β^{aware} (SAT X TAMU)	7.162	3.461	0.963	10.0
78	β^{aware} (SAT X UTA)	1.775	4.757	-2.407	10.0
79	β^{aware} (SAT X Private)	6.043	2.662	2.119	10.0
80	β^{aware} (income)	-0.066	0.112	-0.155	0.233
81	$\log(\sigma)(\beta_i^0)$	0.667	0.219	0.711	1.492
82	α^{aid} (In-State Public)	-0.505	0.128	-0.604	-0.207
83	α^{aid} (Private)	-0.332	0.077	-0.423	-0.187
84	α^{aid} (Out-of-State Public)	-0.175	0.039	-0.181	-0.079
85	α^{aid} (Relig.)	-0.461	0.084	-0.497	-0.252
86	α^{aid} (Elite Private)	-0.55	0.195	-0.817	-0.297
87	α^{aid} (Texas A&M)	-1.044	0.245	-1.246	-0.514
88	α^{aid} (UT Austin)	-0.405	0.123	-0.437	-0.094
89	$\log(\sigma^e)$ (In-State Public)	-10.0	1.646	-10.0	-10.0
90	$\log(\sigma^e)$ (Private)	0.732	0.242	0.274	1.087
91	$\log(\sigma^e)$ (Out-of-State Public)	-0.309	2.665	-8.246	-0.046
92	$\log(\sigma^e)$ (Relig.)	-0.262	2.763	-8.744	0.288
93	$\log(\sigma^e)$ (Elite Private)	1.699	0.503	0.864	2.321
94	$\log(\sigma^e)$ (Texas A&M)	-1.075	4.125	-10.0	-0.13
95	$\log(\sigma^e)$ (UT Austin)	-0.578	3.074	-10.0	-0.187

Note: This table shows all estimated parameters (Part 3 of 3).

```
20 Grades 5 or 6
21 Grade 5
22 Grade 6
30 Grades 7 or 8
31 Grade 7
32 Grade 8
40 Grade 9
50 Grade 10
60 Grade 11
70 Grade 12
```

```
71 12th grade, no diploma
72 12th grade, diploma unclear
73 High school diploma or equivalent
80 1 year of college
81 Some college but no degree
90 2 years of college
91 Associate's degree, occupational/vocational program
92 Associate's degree, academic program
100 3 years of college
110 4 years of college
111 Bachelor's degree
120 5+ years of college
121 5 years of college
122 6+ years of college
123 Master's degree
124 Professional school degree
125 Doctorate degree
999 Missing/Unknown
```

3. Matching procedure:

   - EDUC_HEAD $\in \{0,1\}$: $Ed = 1$
   - EDUC_HEAD $\in \{10,...,32\}$: $Ed = 2$
   - EDUC_HEAD $\in \{40,...,71\} \mapsto Ed = 3$.
     - Note that in the 2002 March CPS, TX subsample there are no household heads with unclear graduation status EDUC_HEAD=72.

   EDUC_HEAD $\in \{73\} \mapsto Ed = 4$.

   - EDUC_HEAD $\in \{81\} \mapsto Ed = 5$.
   - EDUC_HEAD $\in \{91,92\} \mapsto Ed = 6$.
   - EDUC_HEAD $\in \{111\} \mapsto Ed = 7$.
   - EDUC_HEAD $\in \{123\} \mapsto Ed = 8$.
   - EDUC_HEAD $\in \{124\} \mapsto Ed = 9$.

4. Administrative data: UT Austin codes parents' education similarly to the THEOP survey.

Aggregate colleges

In this section I provide lists of the colleges and universities that make up each aggregate college. The data use agreement requires me to aggregate each college that has fewer than ten applicants in the data.

Institution	Sum	
	apply	admit
ANGELO STATE UNIVERSITY	148	119
Institutions with ten or fewer apps	10	9
LAMAR UNIVERSITY-BEAUMONT	80	61
MIDWESTERN STATE UNIVERSITY	30	24
PRAIRIE VIEW A & M UNIVERSITY	95	49
SAM HOUSTON STATE UNIVERSITY	187	143
SOUTHWEST TEXAS STATE UNIVERSITY	329	266
STEPHEN F AUSTIN STATE UNIVERSITY	282	227
SUL ROSS STATE UNIVERSITY	27	15
TARLETON STATE UNIVERSITY	71	49
TEXAS A & M INTERNATIONAL UNIVERSITY	56	37
TEXAS A & M UNIVERSITY-CORPUS CHRISTI	55	43
TEXAS A & M UNIVERSITY-GALVESTON	26	21
TEXAS A & M UNIVERSITY-KINGSVILLE	77	61
TEXAS A&M UNIVERSITY-COMMERCE	47	30
TEXAS SOUTHERN UNIVERSITY	125	79
TEXAS TECH UNIVERSITY	425	360
TEXAS WOMAN'S UNIVERSITY	38	27
THE UNIVERSITY OF TEXAS AT ARLINGTON	170	120
THE UNIVERSITY OF TEXAS AT BROWNSVILLE	62	40
THE UNIVERSITY OF TEXAS AT DALLAS	86	61
THE UNIVERSITY OF TEXAS AT EL PASO	275	183
THE UNIVERSITY OF TEXAS AT SAN ANTONIO	166	121
THE UNIVERSITY OF TEXAS AT TYLER	27	17
THE UNIVERSITY OF TEXAS OF THE PERMIAN BASIN	30	22
THE UNIVERSITY OF TEXAS-PAN AMERICAN	194	165
UNIVERSITY OF HOUSTON-DOWNTOWN	37	18
UNIVERSITY OF HOUSTON-UNIVERSITY PARK	470	305
UNIVERSITY OF NORTH TEXAS	317	231
WEST TEXAS A & M UNIVERSITY	42	37
Total	3,984	2,940

Colleges comprising OTHER TX PUBLIC 4-YEAR

Institution	Sum	
	apply	admit
EMBRY-RIDDLE AERONAUTICAL UNIVERSITY	12	10
Institutions with ten or fewer apps	235	190
NEW YORK UNIVERSITY	29	21
NORTHWOOD UNIVERSITY	21	17
TULANE UNIVERSITY OF LOUISIANA	17	16
UNIVERSITY OF SOUTHERN CALIFORNIA	23	22
VANDERBILT UNIVERSITY	12	11
Total	349	287

Colleges comprising PRIVATE NONRELIGIOUS

Institution	Sum	
	apply	admit
ABILENE CHRISTIAN UNIVERSITY	60	53
AUSTIN COLLEGE	16	15
BAYLOR UNIVERSITY	242	213
BRIGHAM YOUNG UNIVERSITY	28	24
DILLARD UNIVERSITY	18	14
EAST TEXAS BAPTIST UNIVERSITY	29	21
HARDIN-SIMMONS UNIVERSITY	21	19
HOUSTON BAPTIST UNIVERSITY	41	29
HOWARD PAYNE UNIVERSITY	27	19
Institutions with 15 or fewer apps	383	310
LUBBOCK CHRISTIAN UNIVERSITY	17	14
MCMURRY UNIVERSITY	19	17
OUR LADY OF THE LAKE UNIVERSITY-SAN ANTONIO	53	42
SAINT EDWARDS UNIVERSITY	35	29
SOUTHERN METHODIST UNIVERSITY	58	50
SOUTHWESTERN UNIVERSITY	33	32
ST MARYS UNIVERSITY	59	47
TEXAS CHRISTIAN UNIVERSITY	126	106
TRINITY UNIVERSITY	40	39
UNIVERSITY OF MARY HARDIN BAYLOR	19	13
UNIVERSITY OF SAINT THOMAS	28	23
UNIVERSITY OF THE INCARNATE WORD	25	23
XAVIER UNIVERSITY OF LOUISIANA	20	16
Total	1,397	1,168

Colleges comprising RELIGIOUS

Institution	Sum	
	apply	admit
ARIZONA STATE UNIVERSITY-MAIN CAMPUS	30	27
COLORADO STATE UNIVERSITY	14	12
FLORIDA AGRICULTURAL AND MECHANICAL UNIVERSITY	11	10
FLORIDA STATE UNIVERSITY	14	12
GRAMBLING STATE UNIVERSITY	11	6
Institutions with ten or fewer apps	421	346
KANSAS STATE UNIVERSITY OF AGRICULTURE AND APP SCI	11	9
LOUISIANA ST UNIV & AGRL & MECH & HEBERT LAWS CTR	64	51
NEW MEXICO STATE UNIVERSITY-MAIN CAMPUS	102	76
OKLAHOMA STATE UNIVERSITY-MAIN CAMPUS	45	42
PURDUE UNIVERSITY-MAIN CAMPUS	12	12
SOUTHERN UNIVERSITY-NEW ORLEANS	10	3
THE UNIVERSITY OF ALABAMA	13	12
UNITED STATES AIR FORCE ACADEMY	13	12
UNITED STATES NAVAL ACADEMY	11	8
UNIVERSITY OF CALIFORNIA-BERKELEY	16	14
UNIVERSITY OF COLORADO AT BOULDER	15	15
UNIVERSITY OF GEORGIA	12	10
UNIVERSITY OF MISSISSIPPI MAIN CAMPUS	11	10
UNIVERSITY OF NEW MEXICO-MAIN CAMPUS	17	9
UNIVERSITY OF OKLAHOMA NORMAN CAMPUS	36	32
Total	889	728

Colleges comprising NON-TX PUBLIC

Institution	Sum	
	apply	admit
HARVARD UNIVERSITY	16	12
Institutions with ten or fewer apps	55	41
MASSACHUSETTS INSTITUTE OF TECHNOLOGY	12	11
NORTHWESTERN UNIVERSITY	13	12
PRINCETON UNIVERSITY	14	10
RICE UNIVERSITY	52	44
STANFORD UNIVERSITY	19	17
WASHINGTON UNIVERSITY	19	17
YALE UNIVERSITY	11	7
Total	211	171

Colleges comprising SELECTIVE PRIVATE