Students' Heterogeneous Preferences and the Uneven Spatial Distribution of Colleges: Online Appendix

Admissions Probabilities

Using data on admissions outcomes from the ELS survey, we define the binary dependent variable $Admit_{ij}$ to be an indicator for whether student i was admitted to college j (conditional on applying), and estimate the probit model

$$Prob(Admit_{ij} = 1) = \Phi(Z'_{ij}\beta)$$

where the covariates Z_{ij} include a constant and the following:

- $Female_i$, an indicator for whether student i is female
- $Black_i$, an indicator for whether student i is Black
- $Hispanic_i$, an indicator for whether student i is Hispanic
- $TookAP_i$, an indicator for whether student i took any AP courses in high school
- $ParentsTogether_i$, an indicator for whether student i has two parents living at home
- $ParentsNoCollege_i$, an indicator equal to one if student i's parents do not have college degrees
- GPA_i, student i's high school grade point average
- SAT_i, student i's SAT score
- SAT_j, the median SAT score of incoming freshmen at college j
- $LowRelSAT_{ij}$, an indicator equal to one if SAT_i is lower than SAT_j by more than the interquartile range of SAT scores at college j
- $InState_{ij}$, an indicator equal to one if college j is in student i's home state
- Indicators for seven categories of student income

Including the $LowRelSAT$ variable improves the model's fit in cases where students apply to "reach" schools; in the data, students are very rarely admitted to colleges where their SAT scores would be abnormally low.

In the interest of flexibility, we estimate separate β coefficients for each of six categories defined by (Public vs. Private university)$\times$(college's SAT tercile). In practice we do this by running one probit regression with many interaction terms. The detailed results of these regressions are extensive, and are available from the authors on request. Below we summarize the predicted admissions probabilities by showing how they depend on the applicant's SAT score and GPA for four sample

universities: one at the median of the SAT distribution (e.g. University of South Alabama), one at the 75th percentile (e.g. Michigan State), one at the 90th percentile (e.g. University of Texas at Austin), and one at the very top (e.g. Princeton). In general the levels and slopes are fairly reasonable, though perhaps the admissions probabilities do not increase steeply enough at very high SAT scores. The relatively low admission probabilities for students with extremely high SAT scores likely reflects the fact that the graphs were drawn for a student with the average high school GPA of 3.1. The discrete jumps in the SAT graph reflect the role of the *LowRelSAT* indicator; strangely, for the top tier universities, the coefficient on this variable is actually positive.

Figure 1: Admissions probabilities and SAT scores

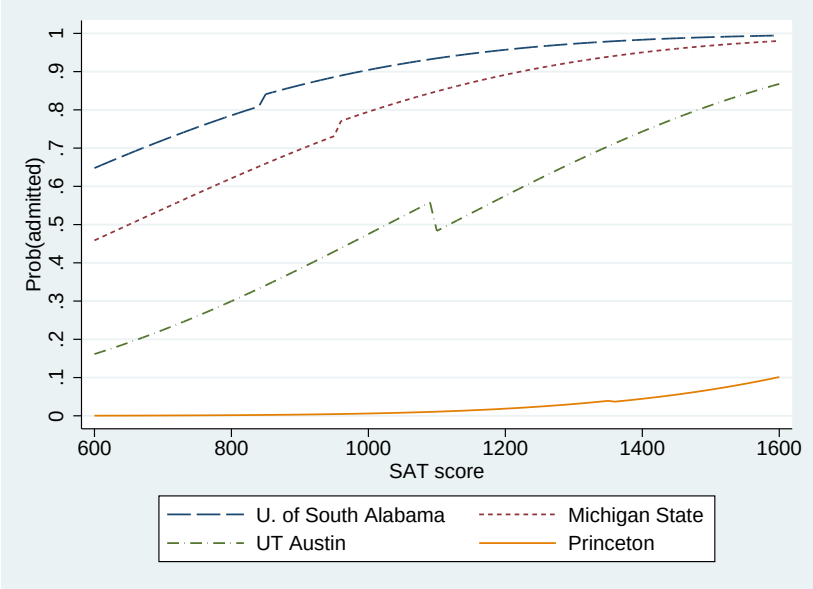
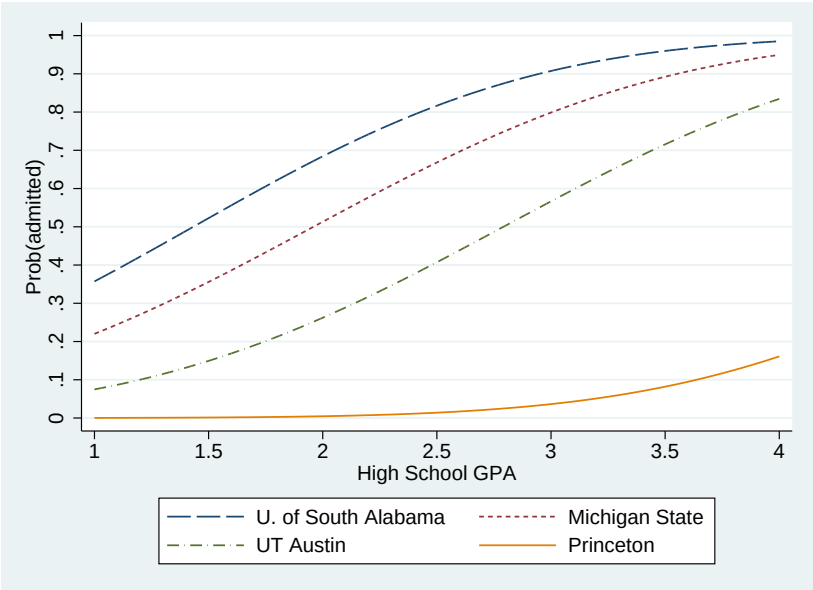


Figure 2: Admissions probabilities and high school GPA



Financial Aid

To estimate the probability that student i receives any financial aid at college j, which we denote with the binary dependent variable Aid_{ij}, we again run a probit regression using the outcomes observed in the ELS survey. The covariates include all of those listed above for the admissions model, along with:

- $Tuition_{ij}$, tuition that student i would pay at college j, including fees
- EFC_i, student i's expected family contribution, calculated from the standard formula

As with the admissions model, we estimate separate coefficients for the six categories corresponding to (Public vs. Private university)×(college's SAT tercile). The full table of estimates is available from the authors upon request. The table below summarizes marginal effects for three main variables of interest (EFC, SAT, and GPA) for different types of colleges. The first row shows the average predicted probability of getting aid.

Table 1: Probability of getting any aid: selected marginal effects

	Public			Private		
	Low SAT	Mid SAT	High SAT	Low SAT	Mid SAT	High SAT
Avg. Prob.	0.459	0.431	0.520	0.693	0.615	0.780
EFC	-0.0038	-0.00008	-0.0023	0.0011	0.00053	-0.0013
SAT	0.0002	0.0006	0.0005	0.0004	0.0006	0.0004
GPA	0.1491	0.1700	0.1418	0.1031	0.1781	0.1137

To model the *amount* of aid received, conditional on receiving any, we use data on aid amounts from the NPSAS survey to estimate a truncated regression model. Letting $AidAmount_{ij}$ denote the amount of financial aid received by student i at college j, we assume

$$ln(\text{AidAmount}_{ij}) \sim TN(Z'_{ij}\beta, \sigma, -\infty, max\_amount)$$

that is, the log of aid received is distributed as a truncated normal with a mean equal to $Z'_{ij}\beta$ and an upper truncation point of max_amount. We set max_amount to be 20% above the highest amount we observe in the data. The reason for this is that if left untruncated, the long right tail of the lognormal distribution introduces implausibly large values into our simulations. The covariates Z_{ij} include the following:

- $Tuition_{ij}$, tuition student i would pay at college j (including fees)
- $Tuition^2_{ij}$, squared tuition
- EFC_i, student i's expected family contribution

- EFC_i^2, EFC squared
- $ZeroEFC_i$, a binary indicator equal to one if student i's EFC is zero
- $RelativeSAT_{ij}$, the difference between student i's SAT score and college j's median SAT score, divided by the difference between the 75th and 25th percentiles of college j students' SAT scores
- $Female_i$, an indicator for whether student i is female
- $Black_i$, an indicator for whether student i is Black
- $Hispanic_i$, an indicator for whether student i is Hispanic
- SAT_i, student i's SAT score
- SAT_i^2, student SAT squared
- SAT_j, the median SAT score of incoming freshmen at college j
- $InState_{ij}$, an indicator equal to one if college j is in student i's home state
- $SAT_j \times EFC_i$, college SAT interacted with student EFC
- $SAT_i \times EFC_i$, student SAT interacted with student EFC

We estimate separate models for Private and Public colleges; a detailed table of coefficient estimates is available from the authors upon request. The table below summarizes marginal effects for EFC and student SAT at Public and Private colleges. Note that the dependent variable is log(Aid), so the marginal effects can be interpreted as approximate percentage increases in predicted aid.

Table 2: Aid amount, conditional on receiving aid: selected marginal effects

	Public	Private
EFC	-0.0038	0.0673
SAT	0.0022	0.0016