Targeting the Gender Placement Gap: Marks versus Money
Online Appendix

Pelin Akyol Kala Krishna Sergey Lychagin

October 17, 2024

A Additional Tables and Figures

Table A.1: Type-Specific Demand Coefficients and Type Shares: Science Track, Female

Placement score used:								
SAY	11.13	-14.36***	-8.28**	-7.51***	-3.41	-6.21***	-2.22	-3.10**
EA	7.70	-3.97	-4.55*	-2.25	-2.64	-6.89**	1.86**	2.53***
Program major:								
Agriculture	4.92***	-0.06	-2.79***	-3.13***	-4.45***	-4.84***	-4.93***	-7.75***
Architecture	7.50***	-2.11**	0.56	-0.00	-3.27***	0.52	-5.12***	-3.83**
Business	1.21	-0.35	-0.17	2.95**	-1.73*	4.07***	-1.00	-6.68***
Economics	1.14	-4.22***	0.46	2.81*	2.50	3.41**	-1.47	-7.55***
Engineering	-2.05***	3.30***	0.28	0.30	-0.75	0.89**	-4.69***	-5.65***
Health Service	6.10***	-0.25	-3.95***	-3.24***	4.50***	-4.73***	-3.07*	-2.32*
Mathematics	3.93***	-1.20	0.22	0.06	-0.92	-0.24	0.08	-6.45***
Medicine	5.32***	3.33*	2.17	-0.74	7.18***	-3.98***	2.50	2.33**
Science	2.79***	-1.46	0.78	-4.60***	0.99	-0.28	-0.08	-5.42***
Other majors	1.19	-1.27	-0.50	0.36	-3.63	0.80*	-7.61***	-10.59***
Non-placement×								
predicted score	5.18	2.63**	2.09*	1.02	3.71***	-0.59	0.66	0.57
Type share	0.02	0.08	0.32***	0.09	0.06	0.08	0.09	0.26***

Notes: Significance levels (* — 10%, ** — 5%, *** — 1%) are obtained using 1,000 bootstrap samples. Placement score dummies indicate programs accepting SAY, EA or SÖZ scores. The interaction of non-placement dummy and predicted score captures the value of the outside option depending on the student's expected score in the main field (SAY for the Science track, EA for the Turkish-Math track and SÖZ for the Social Studies track) predicted using demographic variables and past performance.

Table A.2: Type-Specific Demand Coefficients and Type Shares: Science Track, Male

Placement score used:								
SAY	38.53***	-2.61	-5.55	-10.90***	-7.08***	-7.16***	-6.37***	-3.65***
EA	24.35***	-0.64	-3.24	-7.37***	-5.07*	0.30*	-2.06	1.05**
Program major:								
Agriculture	2.03	0.72	0.59	0.08	-1.48	-1.60	-3.19***	-6.86***
Architecture	-1.00	3.57***	-5.25***	0.14	-2.39	-1.54	-3.98***	-4.56***
Business	1.58	3.29***	-0.95	3.89***	2.17	-0.23	3.42***	-4.89***
Economics	2.28**	-2.48***	-0.55	2.88***	2.19	0.66	3.43***	-7.56***
Engineering	-0.55	2.49**	-0.71*	2.38	2.48	4.50***	0.21	-4.35***
Health Service	-3.64***	-3.71***	0.06	-2.03	-3.81***	1.77**	-3.95***	-3.53*
Mathematics	-2.92***	-0.37	-1.05	1.27	-3.34***	6.10***	-0.87	-3.29**
Medicine	17.96***	-1.95***	3.44	2.21	0.46	5.39***	0.05	1.58***
Science	-1.50	-0.32	0.22	0.02	-4.27***	5.39***	-1.26	-5.13***
Technical Science	-2.33*	-2.24	-1.78	2.33*	-3.88***	3.48***	-3.91***	-4.86***
Technical Services	-0.23	-1.51	-0.10	-2.40***	-1.99***	7.28***	-3.13***	-2.74**
Veterinary	6.48***	2.58	1.96	0.58	-0.23	4.17***	-1.89*	-1.12
Other	3.34***	-1.94	-0.20	-5.11***	-3.97***	-2.73**	-1.34	-4.94***
Non-placement× predicted score	23.08***	-0.23	1.87	2.33***	0.44	0.89	0.03	-0.48
Type share	0.01	0.05	0.16*	0.29***	0.22***	0.07	0.06	0.13***

Notes: Significance levels (* — 10%, ** — 5%, *** — 1%) are obtained using 1,000 bootstrap samples. Placement score dummies indicate programs accepting SAY, EA or SÖZ scores. The interaction of the non-placement dummy and predicted score captures the value of the outside option depending on the student's expected score in the main field (SAY for the Science track, EA for the Turkish-Math track, and SÖZ for the Social Studies track) predicted using demographic variables and past performance.

Table A.3: Type-Specific Demand Coefficients and Type Shares: Turkish-Math Track, Female

Placement score used:								
EA	5.46*	-4.43**	-3.71	-13.38***	-5.73	-3.15	0.22	-0.94**
SOZ	-9.57**	-9.00	-13.51***	-14.11***	-13.28***	-10.52*	-9.60**	-5.91
Program major:								
Arts	5.79***	5.20***	2.92***	-0.02	-0.72	-1.09	-3.80***	-5.44***
Business	2.71***	-4.73***	-3.37	-1.32	1.09	-3.47	-7.96***	-7.49***
Economics	2.03**	-1.57	-6.00***	0.14	1.19	-3.39	-7.32***	-6.97***
Humanities	-3.05*	-1.13	-2.23	4.34***	-2.41	-7.16***	-11.24***	-9.95***
Journalism	-0.78	-2.12	-2.88	6.23***	-4.71*	-8.57***	-12.20***	-10.42***
Language and Literature	-0.00	2.92***	-4.23***	1.04	-1.36	-5.35***	1.64	-3.21*
Law	-0.99**	6.96***	-0.41	11.05***	1.95	3.72**	-7.16***	-6.12***
Personal services	-1.22	-0.62	-3.65**	-0.12	0.36	-6.55***	-7.48***	-7.68***
Public Administration	3.06***	-3.84***	-0.59	10.74***	1.33	-3.27	-9.71***	-5.09
Social sciences	0.71	-4.44***	-0.00	6.59***	-1.05	-6.36***	-7.54***	-8.89***
Other	3.91***	-7.50***	-8.67***	-9.24***	-10.05***	-9.16***	-3.79	-8.85***
Non-placement×								
predicted score	9.33***	-0.29	2.67*	-0.67	0.37	3.09**	0.83	2.44***
Type share	0.03	0.02	0.13*	0.05	0.13	0.29***	0.14	0.22***

Notes: Significance levels (* — 10%, ** — 5%, *** — 1%) are obtained using 1,000 bootstrap samples. Placement score dummies indicate programs accepting SAY, EA or SÖZ scores. The interaction of the non-placement dummy and predicted score captures the value of the outside option depending on the student's expected score in the main field (SAY for the Science track, EA for the Turkish-Math track, and SÖZ for the Social Studies track) predicted using demographic variables and past performance.

Table A.4: Type-Specific Demand Coefficients and Type Shares: Turkish-Math Track, Male

Placement score used:								
EA	7.20*	-6.32***	0.70	-7.87**	-9.66***	-4.23**	-2.72**	-1.09*
SOZ	-3.02	-18.11***	0.60	-10.98**	-9.76***	-8.99***	-6.01**	-12.02***
Program major:								
Business	3.80***	2.03*	1.64***	1.32***	-1.70	-3.62***	-4.97***	-8.89***
Economics	2.94**	2.15**	2.85***	1.57	3.84***	-6.82***	-5.90***	-6.84***
Journalism	-2.16***	-5.18***	7.64***	-3.31**	3.05	-6.07***	-7.87***	-6.15***
Language and Literature	7.97***	-0.13	-2.30	-3.93***	-1.34	-5.32***	-1.83	2.63***
Law	5.26***	2.20	-0.02	3.95	6.78***	3.18**	-5.85***	-5.05***
Personal services	-1.95***	0.47	10.22***	-2.91**	-1.47	-3.97	-7.54***	-8.15***
Public Administration	4.11***	1.47	-1.39	-3.20***	7.16***	-2.02	-4.95***	-5.09***
Social sciences	-2.52***	-0.69	8.52***	-4.72***	-2.17	-3.36	-8.77***	-4.33
Other	4.71***	-10.45***	-2.12	-8.86***	-8.73***	-9.82***	-9.92***	-2.60**
Non-placement×								
predicted score	10.34**	0.57	-23.28	4.42**	1.63	3.11***	2.26***	2.62***
Type share	0.02**	0.19***	0.00	0.11*	0.07	0.28***	0.29***	0.04

Notes: Significance levels (* — 10%, ** — 5%, *** — 1%) are obtained using 1,000 bootstrap samples. Placement score dummies indicate programs accepting SAY, EA, or SÖZ scores. The interaction of the non-placement dummy and predicted score captures the value of the outside option depending on the student's expected score in the main field (SAY for the Science track, EA for the Turkish-Math track, and SÖZ for the Social Studies track) predicted using demographic variables and past performance.

Table A.5: Type-Specific Demand Coefficients and Type Shares: Social Science Track, Female

Placement score used:								
EA	-42.01***	-7.78	-9.23	-1.65	44.36***	17.22	-14.36**	-12.86***
SOZ	-41.67***	-13.25	-10.19	-8.98*	0.00	14.50***	-11.76*	-5.11
Program major:								
Arts	3.62*	2.07	1.42**	0.05	-0.00	-0.55	-2.13	-5.12***
Business	1.77**	-3.54*	0.66	-7.57***	4.18***	-3.23	-0.71	-2.41
Humanities	-4.83**	-3.24	-6.79***	-5.88***	-0.01	-2.31	-3.34	-2.98
Journalism	1.39	0.01	1.48**	-5.86***	-0.00	0.33	-0.04	-7.77***
Language and Literature	-3.87*	0.67	-5.81**	-4.90***	-0.00	2.66***	2.35***	-4.95**
Public Administration	4.17**	-2.82*	4.30**	-4.89***	4.24***	-5.23***	0.01	-0.05
Other	-4.11**	0.74	-7.27***	-9.11***	2.64***	-2.27	-5.22	-5.47*
Non-placement×								
predicted score	-62.96***	-3.04	-1.05	0.01	78.83***	72.40***	-1.58	-0.73
Type share	0.02	0.06	0.31***	0.16*	0.03	0.01	0.13	0.28***

Notes: Significance levels (* — 10%, ** — 5%, *** — 1%) are obtained using 1,000 bootstrap samples. Placement score dummies indicate programs accepting SAY, EA or SÖZ scores. The interaction of the non-placement dummy and predicted score captures the value of the outside option depending on the student's expected score in the main field (SAY for the Science track, EA for the Turkish-Math track, and SÖZ for the Social Studies track) predicted using demographic variables and past performance.

Table A.6: Type-Specific Demand Coefficients and Type Shares: Social Science Track, Male

Placement score used:							
EA	-6.46	1.47	5.00	-2.10	-9.91**	-0.02	-9.19**
SOZ	-2.27	5.85	0.75	8.93**	0.85	-4.62**	2.41
Program major:							
Arts	3.38**	-0.55**	-0.93	-1.85	-2.26	-5.13***	-5.61**
Business	-0.90	1.34	3.94***	2.62**	-0.43	-7.13***	-0.35
Humanities	-1.27	-1.17	3.23***	-5.88***	0.63	-4.43	-6.48**
Journalism	3.61***	2.78*	-1.29	-9.45***	-2.43	-1.08	-6.36*
Public Administration	6.72***	4.34**	3.20*	3.70***	-0.00	-1.69	-0.01
Other	-5.46***	2.28*	3.23***	-5.11**	-5.68**	-5.08**	-2.98
Non-placement×							
predicted score	-0.53	14.16	5.18	-7.34	-1.67	-0.96	2.12
Type share	0.16***	0.03	0.04	0.02	0.13	0.22	0.41***

Notes: Significance levels (* — 10%, ** — 5%, *** — 1%) are obtained using 1,000 bootstrap samples. Placement score dummies indicate programs accepting SAY, EA or SÖZ scores. The interaction of the non-placement dummy and predicted score captures the value of the outside option depending on the student's expected score in the main field (SAY for the Science track, EA for the Turkish-Math track, and SÖZ for the Social Studies track) predicted using demographic variables and past performance.

Table A.7: Average Monthly Earnings in TL and Employment Probability by Field of Study in 2009

	25-30 years-old				40-50 years-old			
	Earnings		Employment		Earnings		Employment	
Field of Study	Female	Male	Female	Male	Female	Male	Female	Male
Teacher training and education science	1281.24	1405.21	0.74	0.81	1572.70	1686.67	0.73	0.90
Arts	1139.93	1144.00	0.51	0.67	1965.35	1665.00	0.67	0.82
Humanities	1040.10	1350.76	0.65	0.81	1647.96	1619.64	0.77	0.92
Social and behavioral science	1324.96	1575.46	0.56	0.74	1836.75	1823.84	0.62	0.87
Journalism and information	1158.46	1337.50	0.65	1.00	1575.00	2350.00	0.55	1.00
Business and administration	1074.64	1227.87	0.58	0.79	1701.64	1863.27	0.59	0.83
Law	1998.49	2031.44	0.75	0.92	2400.00	2767.08	0.91	0.97
Life science	1046.83	1069.44	0.63	0.66	1461.09	1743.56	0.79	0.88
Physical science	1327.31	1472.16	0.69	0.71	2157.74	2088.06	0.69	0.90
Mathematics and statistics	1042.57	1288.38	0.75	0.82	1583.32	1803.50	0.79	0.97
Computing	1450.17	1239.94	0.59	0.79	2000.00	2045.56	0.25	0.83
Engineering and engineering trades	1419.92	1238.02	0.62	0.83	2052.05	2001.92	0.69	0.92
Manufacturing and processing	1074.75	1287.87	0.55	0.81	1630.00	1741.71	0.53	0.87
Architecture and building	1226.24	1425.72	0.70	0.79	1814.29	2081.39	0.74	0.91
Agriculture, forestry and fishery	980.69	1205.58	0.55	0.75	1747.24	1878.02	0.74	0.93
Veterinary	1561.29	1304.81	0.89	0.79	1798.50	2034.94	0.92	1.00
Health	1592.14	2156.33	0.86	0.88	4031.55	5497.93	0.77	0.95
Personal services	1024.21	1031.26	0.59	0.69	1454.10	1585.42	0.52	0.84
Security services	1895.00	1882.24	0.75	1.00		2166.33		0.75

Note: The Average Dollar-Turkish Lira exchange rate in 2009 is 1.65 TL

Figure A.1: Minimum Cutoff Scores

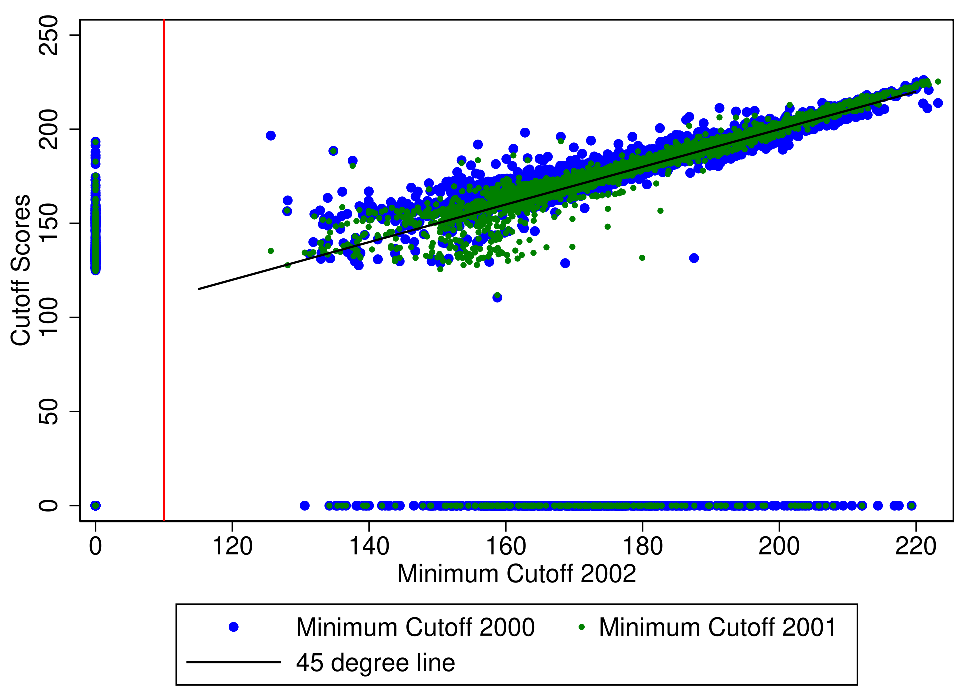


Figure A.2: Gender Differences in Major Choice (Science Track)

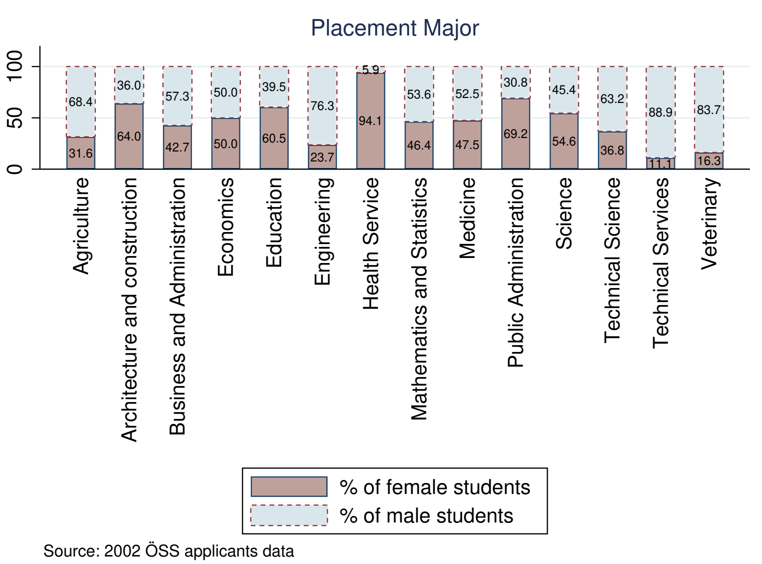


Figure A.3: Gender Differences in Major Choice (Turkish-Math Track)

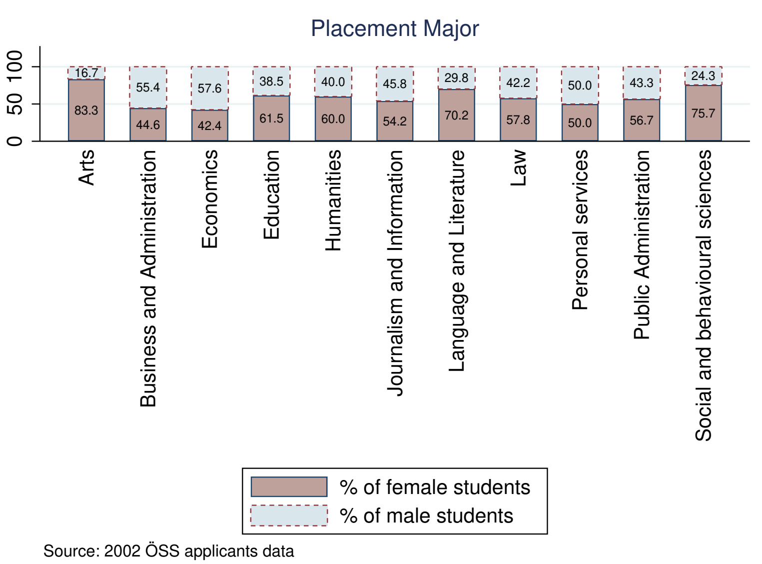


Figure A.4: Gender Differences in Major Choice (Social Science Track)

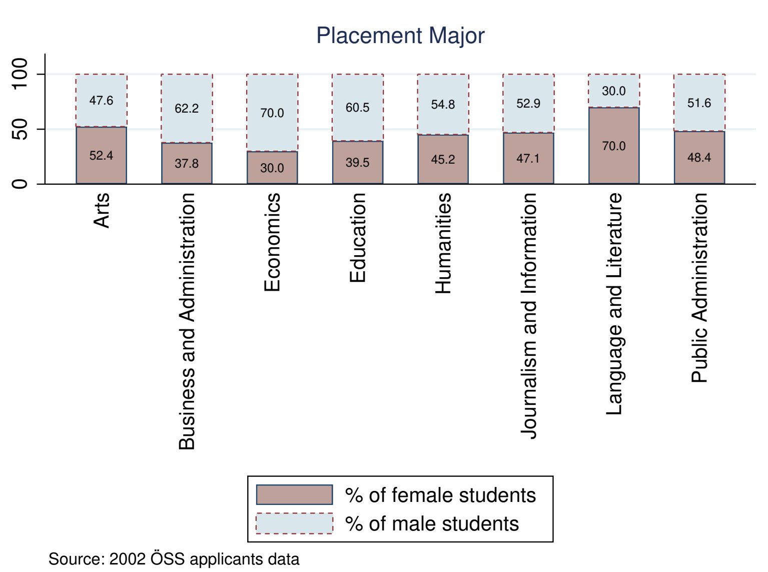


Figure A.5: 1st Preference Major (Turkish-Math Track)

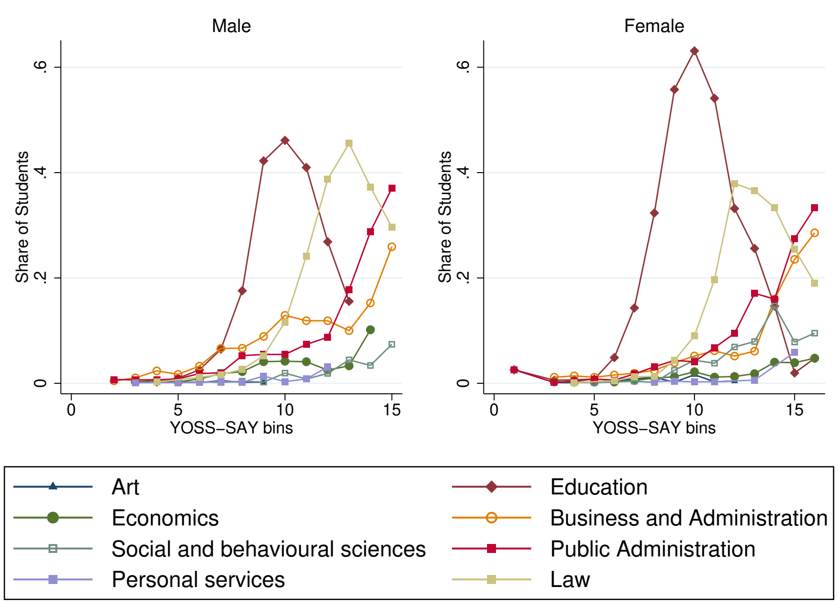


Figure A.6: 1st Preference Major (Social Science Track)

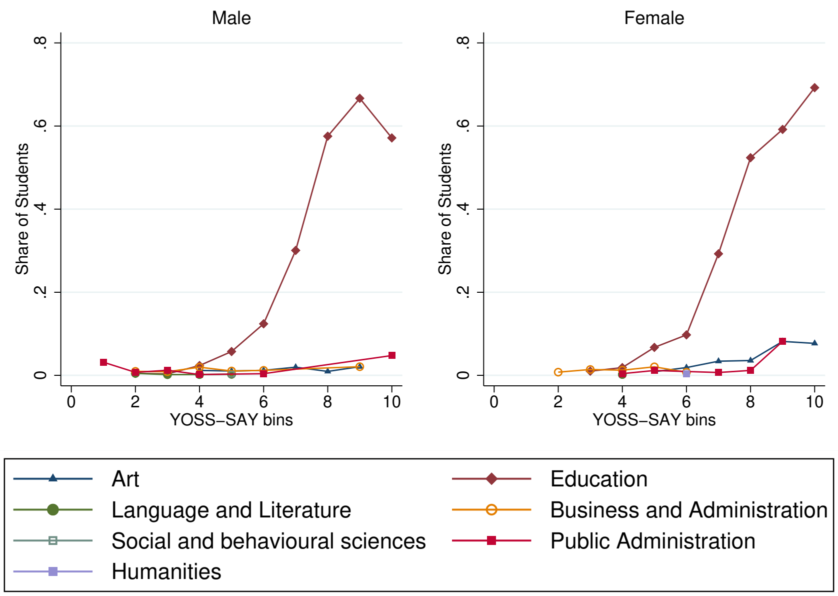


Figure A.7: Distribution of Students According to the Share of Dominant Major in Their Preference List

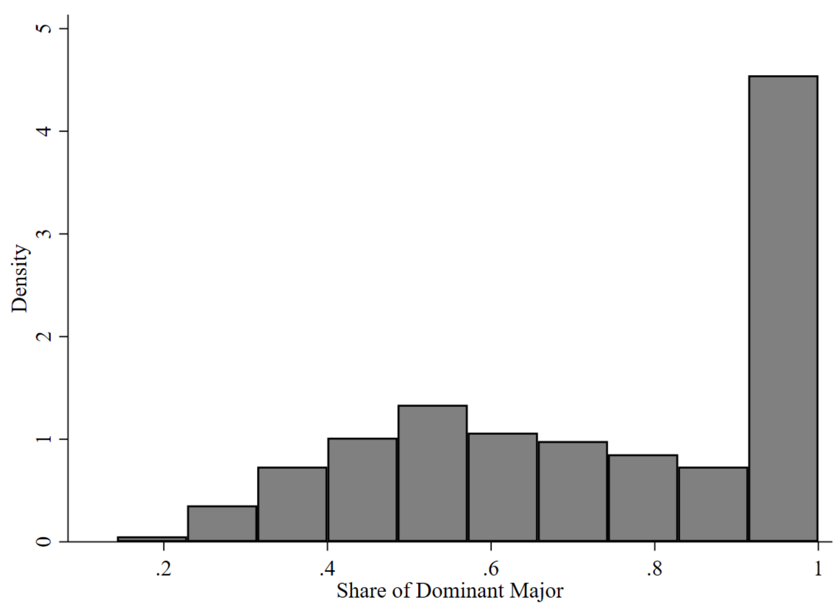
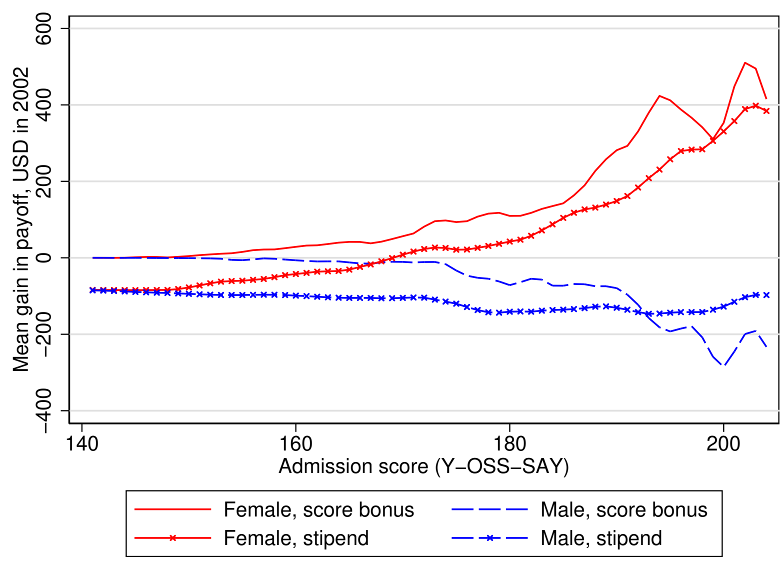


Figure A.8: Stipend vs Score Bonus: Expected Changes in Payoffs Net of Tax by Score



Notes: This graph compares two bonus policies: (a) a stipend for female engineering students, (b) a score bonus granting an admission priority for female applicants in engineering programs. Both policies are calibrated to achieve gender equality in admission probabilities to the engineering major and restricted to first-time takers from the academic track of high school. The payoffs are expressed as annual stipend equivalents in the 2002 US dollars net of tax. To finance the stipend policy, a uniform tuition charge is applied to all the admitted students irrespective of gender or major of study.

B Deriving the Likelihood Function

For each student i, we observe the program of placement in 2002, j_{i2}, and the predicted program of placement under the cutoff scores in 2001, j_{i1}, given i's scores and preference list submitted in 2002, s_i and $\mathcal{L}_i$. We also observe whether j_{i1} is ranked above j_{i2} in the student's list L_i.

The likelihood function is defined as the probability of j_{i1} and j_{i2} being ranked in the order given by L_i and being the best choices in the sets of programs ex-post feasible for i in 2001 and 2002, C_{i1} and C_{i2}. Denoting the vector of all parameters as θ, one can express the likelihood function for observation i via a likelihood function conditional on unobserved types:

$$\mathcal{L}_i(\theta; j_{i1}, j_{i2}, L_i, C_{i1}, C_{i2}) = \sum_{t=1}^{T} \sigma_t \mathcal{L}_{it}(\theta; j_{i1}, j_{i2}, L_i, C_{i1}, C_{i2}, t) \tag{1}$$

In what follows, we omit the indices i and t whenever this does not cause confusion. We also use the following notation for the parts of the choice sets: $A_{i1} = C_{i1} \backslash C_{i2}$, $A_{i2} = C_{i2} \backslash C_{i1}$.

Case 1: $j_1 \neq j_2$, $j_1 \succeq j_2$

First, we consider the case in which the choices j_1 and j_2 are different and j_1 is ranked above j_2. This implies that j_1 is the best choice not only in the set C_1, but also in the union of C_1 and C_2. Note that $j_1 \neq j_2$ implies $j_1 \in A_1$ by revealed preference — otherwise, j_1 would be feasible in C_2 and the agent would prefer it to j_2. One can find a closed-form solution for the type- and student-specific likelihood as follows:

$$\begin{aligned}
&\mathcal{L}_t(\theta; j_1, j_2, L, C_1, C_2) = \Pr\{c(C_1 \cup C_2) = j_1, c(C_2) = j_2\} = \\
&\quad = \Pr\{u_{j_1} \geq u_k, u_{j_2} \geq u_l, k \in A_1 \cup j_2 \setminus j_1, l \in C_2\} \\
&\quad = \int \cdots \int I[\varepsilon_k \leq \varepsilon_{j_1} + \delta_{j_1} - \delta_k, k \in A_1 \cup j_2 \setminus j_1] I[\varepsilon_l \leq \varepsilon_{j_2} + \delta_{j_2} - \delta_l, l \in C_2] \prod_j f(\varepsilon_j) d\varepsilon_1 \ldots d\varepsilon_J \\
&\quad = \int \left[\int_{-\infty}^{\varepsilon_{j_1} + \delta_{j_1} - \delta_{j_2}} \prod_{k \in A_1 \backslash j_1} F\left(\varepsilon_{j_1} + \delta_{j_1} - \delta_k\right) \prod_{l \in C_2 \backslash j_2} F\left(\varepsilon_{j_2} + \delta_{j_2} - \delta_l\right) f(\varepsilon_{j_2}) d\varepsilon_{j_2} \right] f(\varepsilon_{j_1}) d\varepsilon_{j_1}
\end{aligned}$$

$$= \int \left[\int_{-\infty}^{\varepsilon_{j_1}+\delta_{j_1}-\delta_{j_2}} \prod_{l \in C_2 \setminus j_2} \exp(-\exp(-\varepsilon_{j_2} - \delta_{j_2} + \delta_l)) \exp(-\varepsilon_{j_2} - \exp(-\varepsilon_{j_2})) d\varepsilon_{j_2} \right]$$
$$\times \prod_{k \in A_1 \setminus j_1} \exp(-\exp(-\varepsilon_{j_1} - \delta_{j_1} + \delta_k)) \exp(-\varepsilon_{j_1} - \exp(-\varepsilon_{j_1})) d\varepsilon_{j_1}$$
$$= \int \left[\int_{-\infty}^{\varepsilon_{j_1}+\delta_{j_1}-\delta_{j_2}} \exp\left(-e^{-\varepsilon_{j_2}} \sum_{l \in C_2 \setminus j_2} e^{\delta_l - \delta_{j_2}} \right) \exp\left(-e^{-\varepsilon_{j_2}}\right) e^{-\varepsilon_{j_2}} d\varepsilon_{j_2} \right]$$
$$\times \exp\left(-e^{-\varepsilon_{j_1}} \sum_{k \in A_1 \setminus j_1} e^{\delta_k - \delta_{j_1}} \right) \exp\left(-e^{-\varepsilon_{j_1}}\right) e^{-\varepsilon_{j_1}} d\varepsilon_{j_1}$$

One can calculate the inner integral by substituting $z = -e^{-\varepsilon_{j_2}}$:

$$\int_{-\infty}^{\varepsilon_{j_1}+\delta_{j_1}-\delta_{j_2}} \exp\left(-e^{-\varepsilon_{j_2}} \sum_{l \in C_2 \setminus j_2} e^{\delta_l - \delta_{j_2}} \right) \exp\left(-e^{-\varepsilon_{j_2}}\right) e^{-\varepsilon_{j_2}} d\varepsilon_{j_2}$$
$$= \int_{-\infty}^{-\exp(-\varepsilon_{j_1}-\delta_{j_1}+\delta_{j_2})} \exp\left(z \sum_{l \in C_2} e^{\delta_l - \delta_{j_2}} \right) dz$$
$$= \frac{e^{\delta_{j_2}}}{\sum_{l \in C_2} e^{\delta_l}} \exp\left(-\exp(-\varepsilon_{j_1} - \delta_{j_1} + \delta_{j_2}) \sum_{l \in C_2} e^{\delta_l - \delta_{j_2}} \right)$$
$$= \frac{e^{\delta_{j_2}}}{\sum_{l \in C_2} e^{\delta_l}} \exp\left(-e^{-\varepsilon_{j_1}} \sum_{l \in C_2} e^{\delta_l - \delta_{j_1}} \right)$$

Substituting the last line back into the expression for the joint probability yields

$$\mathcal{L}_t(\theta; j_1, j_2, L, C_1, C_2) =$$
$$= \int \left[\int_{-\infty}^{\varepsilon_{j_1}+\delta_{j_1}-\delta_{j_2}} \exp\left(-e^{-\varepsilon_{j_2}} \sum_{l \in C_2 \setminus j_2} e^{\delta_l - \delta_{j_2}} \right) \exp\left(-e^{-\varepsilon_{j_2}}\right) e^{-\varepsilon_{j_2}} d\varepsilon_{j_2} \right]$$
$$\times \exp\left(-e^{-\varepsilon_{j_1}} \sum_{k \in A_1 \setminus j_1} e^{\delta_k - \delta_{j_1}} \right) \exp\left(-e^{-\varepsilon_{j_1}}\right) e^{-\varepsilon_{j_1}} d\varepsilon_{j_1}$$
$$= \frac{e^{\delta_{j_2}}}{\sum_{l \in C_2} e^{\delta_l}} \int \exp\left(-e^{-\varepsilon_{j_1}} \sum_{k \in C_1 \cup C_2 \setminus j_1} e^{\delta_k - \delta_{j_1}} \right) \exp\left(-e^{-\varepsilon_{j_1}}\right) e^{-\varepsilon_{j_1}} d\varepsilon_{j_1}$$

$$= \frac{e^{\delta_{j_2}}}{\sum_{l \in C_2} e^{\delta_l}} \frac{e^{\delta_{j_1}}}{\sum_{k \in C_1 \cup C_2} e^{\delta_k}}$$

The last line is obtained by following the same steps as we used to compute the inner integral.

Case 2: $j_1 \neq j_2$, $j_2 \succeq j_1$

This case is symmetric to the previous one. The conditional likelihood function is obtained from the above formula by changing indices:

$$\mathcal{L}_t(\theta; j_1, j_2, L, C_1, C_2) = \frac{e^{\delta_{j_2}}}{\sum_{l \in C_1 \cup C_2} e^{\delta_l}} \frac{e^{\delta_{j_1}}}{\sum_{k \in C_1} e^{\delta_k}}$$

Case 3: $j_1 = j_2$

In this case, $j_1, j_2 \in C_1 \cup C_2$. Also, j_1 is optimal in C_1 and C_2 if and only if it is optimal in $C_1 \cup C_2$. Thus, the formula boils down to the standard multinomial logit probability:

$$\mathcal{L}_t(\theta; j_1, j_2, L, C_1, C_2) = \Pr\{c(C_1) = c(C_2) = j_1\} = \Pr\{c(C_1 \cup C_2) = j_1\} = \frac{e^{\delta_{j_1}}}{\sum_{k \in C_1 \cup C_2} e^{\delta_k}}$$

C Estimation Details

We estimate the parameters of the model in six sub-populations, defined by gender and three high school tracks: Science, Turkish-Math, and Social Science. Preferences for broad categories of subjects (science vs. humanities) tend to correlate with one's high school track. Preferences may also vary between genders if, for example, certain career paths are incompatible with commonly accepted gender roles.

The set of choice characteristics with common valuation across unobserved types, X_{ij}, includes the following variables:

1. The highway distance between the student's high school and the program's campus.[1] A dummy for the high school and the campus being in the same province.

[1] Obtained from the Directorate of Highways at https://www.kgm.gov.tr/

2. A full set of university dummies and program ranking by the cutoff score in the preceding admission cycle in 2001. These variables control for program quality.

3. Dummies for the type of admission score accepted by the program.

4. Interactions of net tuition with student income dummies capture preference heterogeneity associated with one's income.

The coefficients on the following choice characteristics, Z_{ij}, are allowed to vary across the unobserved student types:

1. A set of dummies for program majors.

2. A dummy variable for the option of not being placed and its interaction with the student's predicted exam score. These terms are meant to serve as a reduced form for the value of retaking the exam in the following year or not attending university at all.

When we implement the maximum likelihood estimator, we are confronted by two practical issues. First, the log-likelihood function in latent class logit models is well-known to have multiple local maxima. Second, latent classes tend to separate in terms of preference for majors. For instance, the estimation algorithm may split the population of students into a latent class that favors medical degrees and never applies for economics and a class that favors economics and never applies to medical programs. This means that the coefficient γ_t on the economics major is nearly minus infinity for the former class, and so is the coefficient on medical majors for the latter one. Moreover, the log-likelihood function is nearly flat for these coefficients, which makes the numerical maximization procedure stop prematurely and produce noisy results. Perfect separation is a well-known issue in estimating latent class discrete choice models.

We tackle the multiple maxima problem in three steps. First, we use the simple multinomial logit instead of the latent class logit to give us the first starting value for the parameter vector β. Second, we set the number of latent classes to the number of majors popular among the students from the sample. The initial values for γ are estimated using simple multinomial logit on the subsample of students who are placed in the respective major; for instance, we run the multinomial logit with no heterogeneity in γ using the subsample of students who are placed in economics, estimate the coefficients on Z_{ij} and use these estimates as a starting value for one of the γ_t's. Third, once we have the starting value for the vector $(\beta, \gamma_1, \ldots, \gamma_T, \sigma_1, \ldots, \sigma_T)$, we generate 100 perturbations of this vector by adding small random shocks to it. We then maximize the log-likelihood function for the fully specified latent class logit model using these 100 random starting values and pick the solution that corresponds to the highest value of log-likelihood. Although we did find that the log-likelihood function has multiple local optima, we could not find visible differences between them in terms of the demand substitution patterns they produce.

To address the preference separation problem, we impose a quadratic penalty on the coefficients β and γ_t:

$$\mathcal{L}_{penalized}(\beta, \gamma) = \mathcal{L}(\beta, \gamma) - \sum_k w_{penalty,\beta_k} \beta_k^2 - \sum_{t,l} w_{penalty,\gamma_{tl}} \gamma_{tl}^2$$

The penalty parameters $w_{penalty}$ are set at 0.01 for the coefficients on universities and majors, the main culprits behind the preference separation issue. For all the other coefficients, $w_{penalty} = 0.0001$. One way to view penalized maximum likelihood is that it represents a Bayesian estimator with a vague normal prior. The variance of the prior for a coefficient is inversely related to the penalty placed on this coefficient. This estimator has the usual large-sample asymptotic properties (consistency and normality), and the choice of the weights has vanishing impact on the estimates since the likelihood term $\mathcal{L}(\beta, \gamma)$ becomes dominant on the right hand side as the estimation sample grows in size.[2]

D Predicting Substitution Patterns

We present a "heat map" to illustrate the performance of our model and its competitors relative to the benchmark in Column 5. We first create transition matrices. For each student in a given track of a given gender, we use the associated model to simulate placement. Then, we average the results across all students to produce the transition matrices, which are shown

[2]See Gelman et al. (2014) for more details on large-sample frequentist properties of Bayesian estimators.

in Figures D.1 to D.3.

In Figure D.1, we depict the substitution patterns from the data. The vertical axis depicts the actual major of placement under the 2002 admission cutoffs, while the horizontal axis corresponds to the placements predicted using the preference list of the student but under the cutoff scores in 2001. The programs are ordered in terms of their popularity with the most popular ones at the top. Each colored cell depicts the conditional probability of switching majors, with darker colors representing higher probabilities. The substitution patterns predicted by our preferred model are shown in Figure D.2, while the patterns predicted by the models in Columns 2, 3, and 4 of Table 5 are analogously depicted in Figures D.3b, D.3c and D.3d.

Note that our preferred model reproduces the transition matrix for majors quite well. In most cases, students seem to have a strong preference for a specific major, as evidenced by the dark colors on the diagonal: the predicted probability of not switching majors is 91.4% whether we use the fitted model or predict placements using preference lists as given. Programs in education seem to be a backup option for many students, and this is reflected in the fact that whatever major the student was placed in 2002, there is a movement to education with 2001 cutoffs. When our preferred model or its alternatives predict non-negligible switching rates, this usually involves related majors. For instance, economics seems to be a substitute for education, engineering, business, and public administration, subjects that either deal with similar domains or require similar skills.

A feature of the transition matrices that may be puzzling is that they are darker below the diagonal. This comes from the fact that if you are going to switch from one major to another, you are more likely to switch to a popular major than an unpopular one. To draw an analogy to demand for colas, if you were to switch from Coke, you would most likely switch to Pepsi, not RC Cola.

It is hard to see how Figures D.1 to D.3d differ from one another. To make the differences between the predictions and the data more salient, Figure D.4 present a heat map of the differences between Figures D.3a to D.3d and Figure D.1. The solid lines drawn delineate the programs that account for 90% of the placements. The dotted line drawn does the same but for 95% of the placements. Each colored cell represents differences in the transition matrices. White means the differences are close to zero, red shows the difference is positive, and blue shows the difference is negative. Our preferred model (in the upper left panel) overall performs better, as its colors are lighter everywhere than any of the others. More importantly, it does particularly well inside the box delineated by the solid and dashed lines, where most of the action occurs.

E Calculating Placement Score (Y-ÖSS)

The University Entrance Exam placement score (Y-ÖSS) of student i is a function of his ÖSS scores and the weighted normalized high school grade points (AOBP).

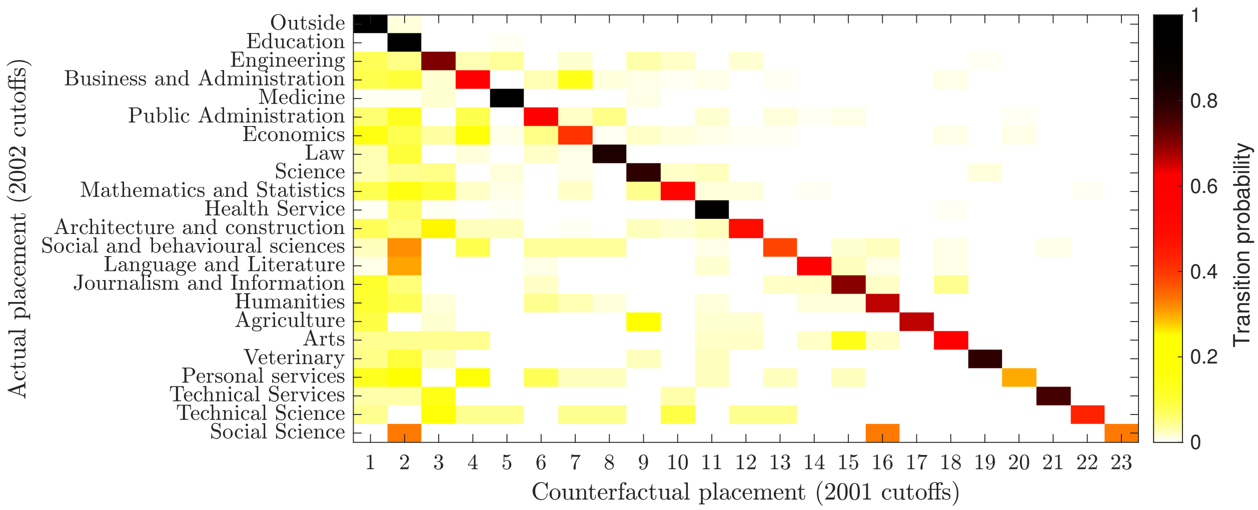
$$Y\_OSS\_X_i = OSS\_X_i + \alpha AOBP\_X_i$$

where $X \in \{SAY, SOZ, EA\}$, and α is a pre-determined constant that changes according to the students' track, preferred department, and whether the student was placed in a regular program in the previous year. The Student Selection and Placement Center (ÖSYM) publishes the lists of departments open to students according to their tracks. When students choose a program from this "open" list, α equals to 0.5. If it is outside the open list, α equals 0.2. If a student graduated from a vocational high school and prefers a department that is compatible with his high school field, α equals 0.65. If a student was placed in a regular university program in the previous year, the student is punished with a 50% penalty to his GPA, so that α equals 0.25, 0.1, and 0.375, respectively in the three above cases.

In turn, the AOBP score of student i from a given track in school j is a function of normalized high school GPA, OBP_j, the minimum and the maximum GPA of the same-school peers, $\min_{i' \in j} OBP_{i'}, \max_{i' \in j} OBP_{i'}$, and the mean ÖSS score in in the respective subject, OSS_X_j, $X \in \{SAY, SOZ, EA\}$, among graduating seniors in that school as in equation (2). Students keep their AOBP over attempts made.

$$AOBP\_X_{ij} = \left[\left(\frac{OSS\_X_j}{80} \times \min_{i' \in j} OBP_{i'} \right) - \left(\frac{OSS\_X_j - 80}{10} \right) \right]$$

Figure D.1: Transition Matrix for Majors of Placement, Predicted Using the Preference Data



Notes: Actual major — major of placement in 2002. Counterfactual major — major of placement if the admission cutoffs are the same as in 2001. "Outside" corresponds to not being placed. Counterfactual majors follow the same order as the actual ones (e.g., the label 3 corresponds to engineering). The value in each cell is the mean probability of placement into the counterfactual major conditional on the actual placement. The probabilities are predicted using the preference lists submitted by the students in 2002 and the admission cutoffs from 2001 and 2002.

Figure D.2: Transition Matrix for Majors of Placement, Predicted Using the Estimated Model

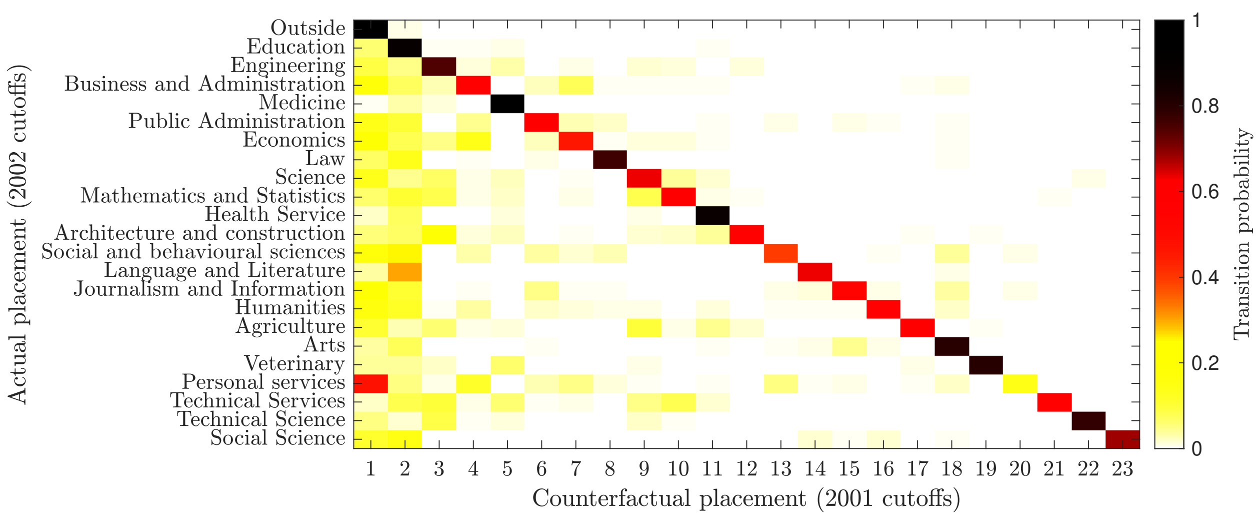
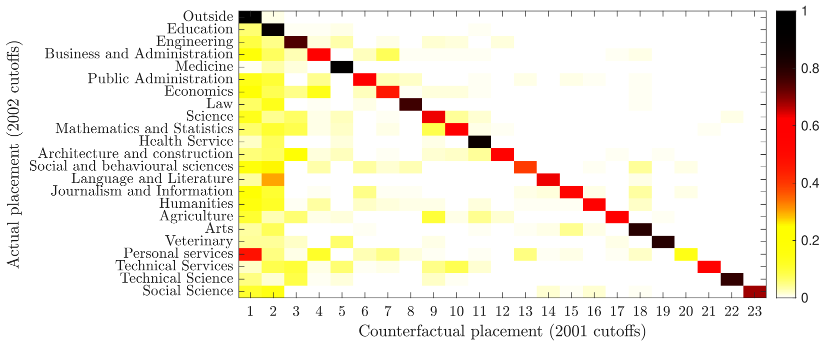
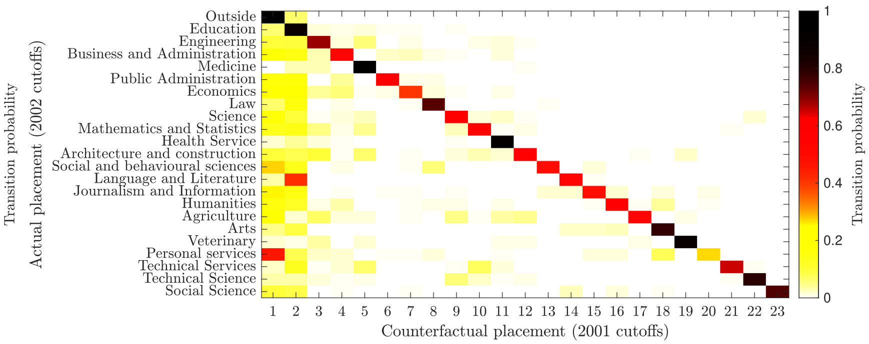


Figure D.3: Transition Probabilities for Different Model Specifications

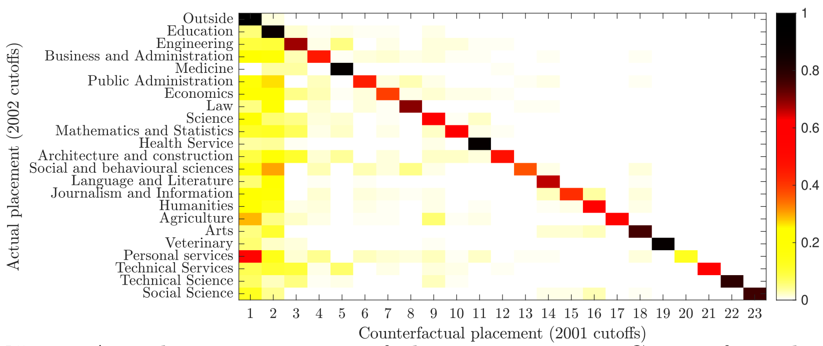
(a) Main Model



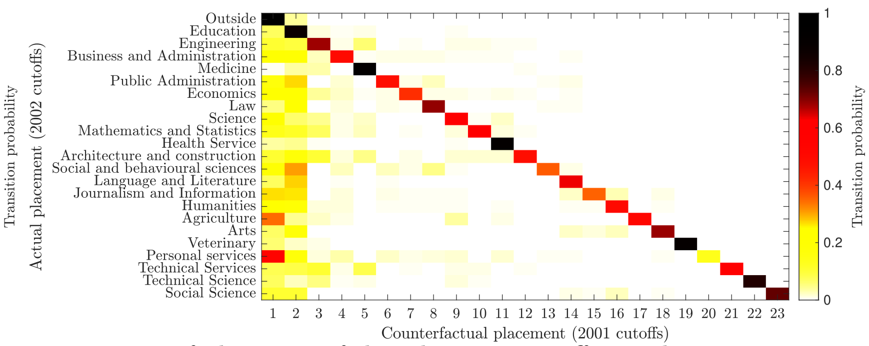
(b) Alternative 1



(c) Alternative 2

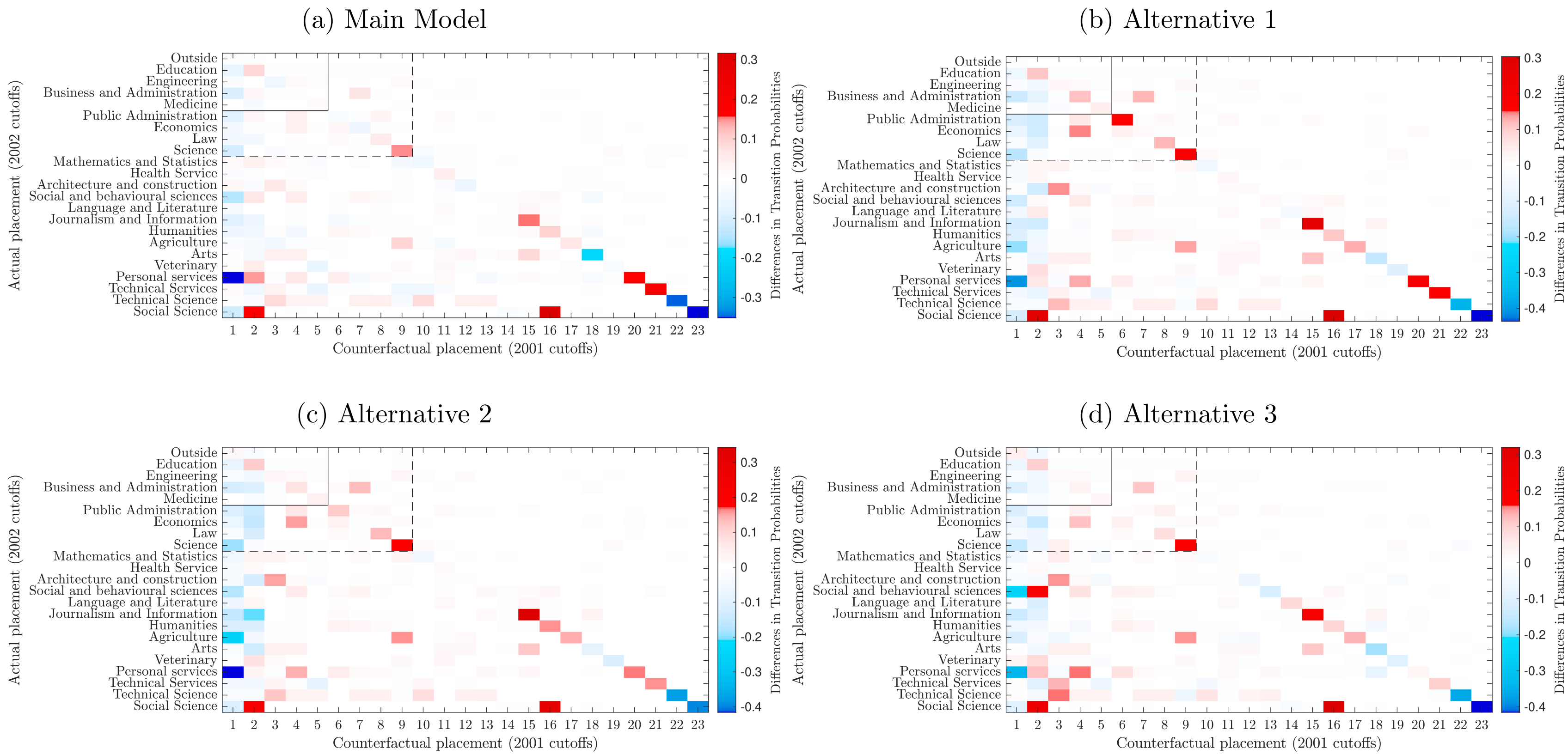


(d) Alternative 3



Notes: Actual major — major of placement in 2002. Counterfactual major — major of placement if the admission cutoffs are the same as in 2001. "Outside" corresponds to not being placed. Counterfactual majors follow the same order as the actual ones (e.g., label 3 corresponds to engineering). The value in each cell is the mean probability of placement into the counterfactual major conditional on the actual placement.

Figure D.4: Differences in Transition Probabilities between Model and Data for Different Model Specifications



(a) Main Model

(b) Alternative 1

(c) Alternative 2

(d) Alternative 3

$$
+ \left[\left(OBP_i \times \frac{OSS\_X_j}{80} \right) - \left(\frac{OSS\_X_j}{80} \times \min_{i' \in j} OBP_{i'} \right) \right]
$$

$$
\times \left[\frac{80 - \left[\left(\frac{OSS\_X_j}{80} \times \min_{i' \in j} OBP_{i'} \right) - \left(\frac{OSS\_X_j - 80}{10} \right) \right]}{\left(\frac{OSS\_X_j}{80} \times \max_{i' \in j} OBP_{i'} \right) - \left(\frac{OSS\_X_j}{80} \times \min_{i' \in j} OBP_{i'} \right)} \right] \tag{2}
$$

We do not observe student AOBP scores in our data set, but we do observe the inputs on the right hand side in (2) other than the minimum and maximum OBP scores in the school.[3] In our sample, we observe the normalized high school GPA (OBP_i) and the raw GPA for all students. ÖSYM calculates OBP as follows:

$$
OBP_i = \max \left\{ 30, \min \left[80, 10 \frac{GPA_i - \mu_{GPA,j}}{\sigma_{GPA,j}} + 50 \right] \right\} \tag{3}
$$

where GPA_i is the students' own GPA, while $\mu_{GPA,j}$ and $\sigma_{GPA,j}$ are the average and the standard deviation of GPA within i's school j. The student's own GPA and OBP are observed in the data. As long as we have at least two students from a given school, we can use equation (3) to solve for $\mu_{GPA,j}$ and $\sigma_{GPA,j}$ for almost almost all the high schools in Turkey.

Since our data set only includes a sample of students, we cannot observe the minimum or the maximum OBP within school for the entire population. To pin down the maximum OBP, we first look at the schools where we have their first-ranking student in our sample (there is a variable that identifies whether the student ranked first or not). In the data set, we observe 445 first-ranked students. This gives us the maximum OBP for 445 schools.

For the remaining schools, we resort to simulations. First, note that the raw GPA is bounded from above by 5. If follows for equation (3) that OBP has an upper bound $\overline{OBP}_j = [80, 10(5 - \mu_{GPA,j})/\sigma_{GPA,j} + 50]$. We assume that OBP scores in each school have a beta distribution with the mean equal to 50, the standard deviation of 10,[4] and the support on $[30, \overline{OBP}_j]$. We find the parameters of this distribution independently for each school. Since the mean and the standard deviation are the same in all schools, parameters differ in

[3] We obtained each school's mean ÖSS scores in each field for the 2002 high school graduates from the OSYM website.

[4] Recall that by definition, OBP is normalized so that its mean within a school is 50 points and the standard deviation is 10 points.

each school only because of the differences in the support of the distribution.

In the second step, we draw from the estimated beta distribution the number of simulated OBP realizations equal to the class size in the school, which is known from the official statistics. We do this S times for each school and then find the average minimum and average maximum OBP over the S draws. We use these averages as our approximation for the minimum and maximum OBP scores:

$$\min_{i \in j} OBP_i = \frac{1}{S} \sum_{k=1}^{S} \min_{i \in j} OBP_i^k$$

$$\max_{i \in j} OBP_i = \frac{1}{S} \sum_{k=1}^{S} \max_{i \in j} OBP_i^k$$

Finally, we match these estimated minimum and maximum OBP scores with our data set. If we observe a lower bound for OBP in our data set than what was simulated, we use it as the minimum OBP for this school. If we observe a higher maximum OBP in the data, we use it as the maximum OBP for this school. Otherwise, we use the simulated minimum and maximum OBP scores.

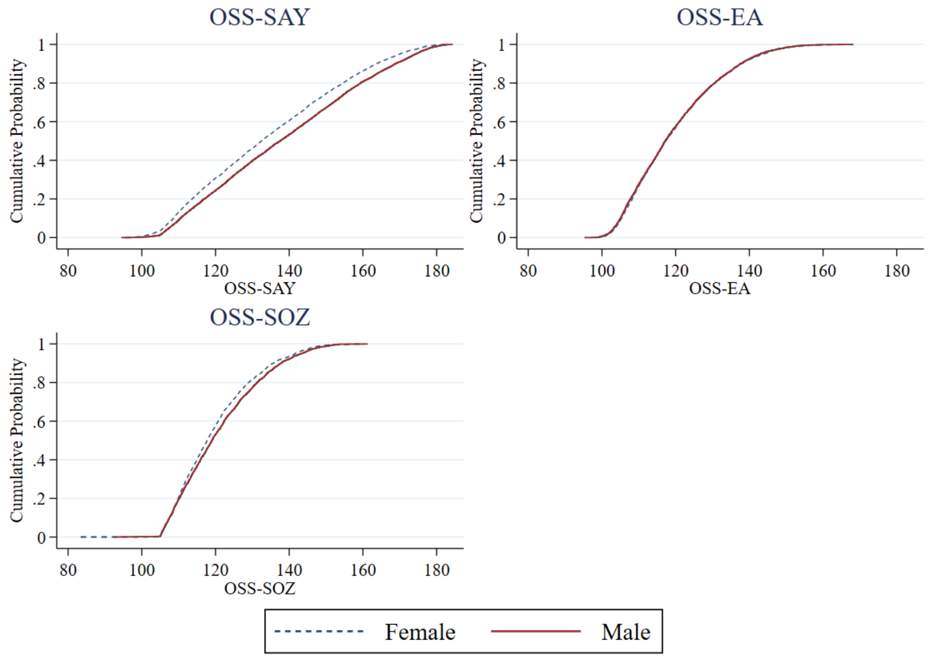
F Score and GPA Distributions

Figure F.1 presents the cumulative distribution of exam scores (ÖSS) by gender and high school track. For the score used in the science track programs (ÖSS-SAY) the male students' score distribution (in red) first-order stochastically dominates that of female students. The Kolmogorov-Smirnov test shows this difference is significant. The same pattern holds for ÖSS-SÖZ. On the other hand, for ÖSS-EA, the score usually relevant for students in the Turkish-Math Track, the difference is not as obvious, and the difference in the distributions is not significant (p-value 0.215).

The distributions of high school GPA (AOBP)[5] follow the opposite pattern: women tend

[5]Since different schools could differ in their grading standards, AOBP score is normalized as explained in Appendix E.

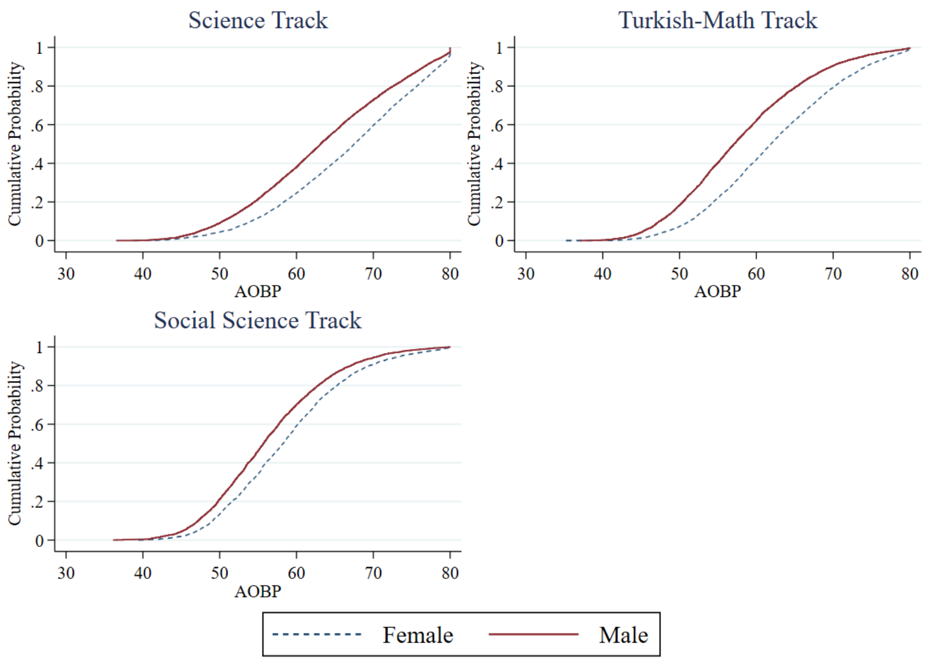
Figure F.1: ÖSS Score Distributions by Gender



to perform better in school than men do[6] (see Figure F.2). Since the placement score (Y-ÖSS) is a mix of the exam score (ÖSS) and the GPA (AOBP), the gap in placement scores is less than that in exam scores.

[6]This pattern, where males do better in high-stakes exams has also been observed in other settings. In a meta-analysis, Voyer and Voyer (2014) show that girls do better than men in high school and have been doing so for quite a while. This pattern is often attributed to women maturing earlier than men.

Figure F.2: AOBP Distributions by Gender



References

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., and Rubin, D. B. (2014). *Bayesian Data Analysis*. Chapman and Hall/CRC, 3rd edition.

Voyer, D. and Voyer, S. D. (2014). Gender differences in scholastic achievement: A meta-analysis. *Psychological Bulletin*, 140(4):1174–1204.