Appendix A: Data

A. Wellesley College applicants, 1999-2013

There were 62,805 applicants for Fall admission between 1999 and 2013 (see Figure 2). Of these, 42,326 were U.S. citizens and applied through regular admissions or early evaluation. In both groups, admission does not obligate the applicant to enroll at Wellesley College. We exclude all other applicants, including early decision applicants who are required to enroll if admitted, as well as other applicants who may have already completed some college elsewhere (e.g., transfer and Davis Scholar applicants). We further limit this paper's sample to the 17,217 applicants who received offers of admission. The estimation sample includes 15,390 applicants for whom we could also identify degree and major data from the National Student Clearinghouse.

B. National Student Clearinghouse data

In September 2017, we submitted the full applicant dataset for matching to the StudentTracker service of the National Student Clearinghouse. To identify applicants' degree(s) and major(s), we used the following procedure in the full applicant sample:

- We kept all observations that recorded a graduation event. This excluded records with privacy-related blocks, with matching errors, or for whom only enrollment episodes were reported.
- We kept observations with non-missing string values of the degree title, which we coded into bachelor's degrees, graduate degrees (of any kind), and other degrees or certificates.
- In the remaining sample, we identified the first (and usually only) college or university that awarded a bachelor's degree to a student. We identified the six-digit Classification of Instructional Program (CIP) code(s) for the major(s) in this college or university.
- We imputed missing CIP codes using college-reported string descriptions of the majors. First, we identified non-missing, exactly-matching string major descriptions within colleges, and imputed missing CIP codes using non-missing CIP codes in the same string group. Second, we identified the modal CIP code in sample-wide groups defined by non-missing, exactly-matching major strings, and imputed this CIP for missing observations in the same string group. Note that all applicants were used to impute missing CIP codes, even if they were not included in the estimation sample.
- We used CIP codes to generate the dependent variables, as described in Table A1.
- Finally, we used non-missing major strings and keyword matches (e.g., "Economics") to impute remaining missing values of the dependent variables. We used keywords from descriptions of the CIP codes at https://nces.ed.gov/ipeds/cipcode/.

C. NSF Graduate Fellowships

We downloaded spreadsheets of NSF Graduate Research Fellowship awardees and honorable mentions (from 2002 to 2020).[1] We matched by the strings of first name, last name, and undergraduate institution using the Stata command matchit.[2] We hand-verified all exact and close matches.

[1] https://www.research.gov/grfp/AwardeeList.do?method=loadAwardeeList
[2] https://ideas.repec.org/c/boc/bocode/s457992.html

D. High School Data

We extracted public high school variables from the 2013 Common Core of Data (https://nces.ed.gov/ccd/), and private school variables from the Private School Universe Surveys in 2005, 2007, 2009, 2011, 2013, and 2015 (https://nces.ed.gov/surveys/pss/), using multiple rounds because of non-reporting. In the latter case, we used the most recent non-missing observation for each private school code (with the exception of free-and-reduced price lunch, which was only available in 2015). We linked the variables to the Wellesley dataset using the College Board's high school CEEB codes. Because CEEB codes are not available in the Common Core or Private School Surveys, we created a cross-walk between the CEEB codes in our admitted applicant dataset and the high school codes used in the national datasets. Specifically, we matched school names and cities using the Stata command matchit, and then hand-verified exact and close matches.

D. College and University Data

We created a college-level panel of variables from the Integrated Postsecondary Education Data System (https://nces.ed.gov/ipeds/) for each year from 2000 and 2017, including the proportion of female full-time instructional faculty. For each Spring graduation year—and being attentive to the timing of Fall and Spring IPEDs surveys—we constructed an average of non-missing observations in that academic year and the three prior ones, in order to proxy exposure for a typical graduate. We merged these to Wellesley College applicants with non-missing bachelor's degree data, using the graduation year and college codes that are shared across NSC and IPEDS datasets. We also created a college-level panel of a single variable—proportion of tenure-track and non-tenure-track female faculty in Economics—using the restricted-use panel dataset of the Committee on the Status of Women in the Economics Profession (CSWEP) (https://www.icpsr.umich.edu/web/ICPSR/studies/37118). We constructed 4-year averages as with the IPEDs data, and merged the exposure proxy to Wellesley College applicants.

Table A1: Mathematically-intensive and/or STEM majors

Binary dependent variable	Classification of Instructional Program (CIP) code(s)
Economics[a]	4506
Business (math-intensive)[b]	5206 (Business/Managerial Economics), 5208 (Finance and Financial Management Services), 5213 (Management Sciences and Quantitative Methods)
Mathematics, Statistics, Physical Sciences, Computing[c]	Mathematics and Statistics (27), Physical Sciences (40), Computing[c]
Engineering	14
Biological and Biomedical Sciences	26
Other STEM	Other officially designated CIP codes not referenced above.

Notes: A full list of 2010 CIP codes are available at https://nces.ed.gov/ipeds/cipcode/browse.aspx?y=56. DHS (2016) lists officially-designated STEM codes.

[a] Within the four-digit economics CIP code (4506), only 450603 (Econometrics and Quantitative Economics) is officially designated as STEM.
[b] Among four-digit business CIP codes, only 5213 is officially designated as STEM.
[c] "Computing" includes 24 six-digit CIP codes (all with a two-digit code of 11) that are officially designated as STEM, including Computer Science. One six-digit code is excluded (110899), corresponding to the Media Arts and Sciences major at Wellesley College.

Table A2: Descriptive statistics for applicant and high school variables

	Degree and major data *available* in NSC						Degree and major data *missing* in NSC					
	Enrollees			Non-enrollees			Enrollees			Non-enrollees		
	Mean	S.D.	N	Mean	S.D.	N	Mean	S.D.	N	Mean	S.D.	N
<u>Panel A: Applicant variables</u>												
Mathematics SAT	678.9	66.50	5,586	701.8	64.63	8,454	664	88.49	5	698.6	67.14	1,620
Verbal SAT	694.8	67.20	5,587	716.3	63.36	8,454	706	53.67	5	715.7	65.42	1,620
Writing SAT	702.2	63.25	2,686	723.0	61.53	4,240	736.7	101.2	3	717.1	64.77	954
ACT	29.86	3.035	2,042	30.82	2.873	2,826	32.50	3.536	2	30.78	2.922	583
First generation (1/0)	0.126	0.331	6,192	0.106	0.308	9,198	0.286	0.488	7	0.125	0.330	1,820
Under-represented minority (1/0)	0.152	0.359	6,192	0.170	0.376	9,198	0.286	0.488	7	0.198	0.399	1,820
Alumna(e) relative(s):												
Sister (1/0)	0.034	0.181	6,192	0.022	0.148	9,198	0.143	0.378	7	0.015	0.121	1,820
Mother (1/0)	0.046	0.210	6,192	0.040	0.196	9,198	0.143	0.378	7	0.037	0.188	1,820
Aunt (1/0)	0.030	0.171	6,192	0.025	0.154	9,198	0	0	7	0.021	0.145	1,820
Grandmother or higher (1/0)	0.028	0.165	6,192	0.024	0.154	9,198	0.143	0.378	7	0.023	0.148	1,820
Application major preferences												
Economics (1/0)	0.076	0.265	6,192	0.087	0.282	9,198	0	0	7	0.069	0.254	1,820
Biological sciences (1/0)	0.225	0.418	6,192	0.267	0.442	9,198	0.143	0.378	7	0.280	0.449	1,820
Math, physical, computer (1/0)	0.124	0.329	6,192	0.160	0.367	9,198	0	0	7	0.157	0.364	1,820
Other STEM (1/0)	0.050	0.218	6,192	0.055	0.227	9,198	0.143	0.378	7	0.065	0.247	1,820
<u>Panel B: High school variables</u>												
Public charter (1/0)	0.009	0.094	5,622	0.011	0.106	8,393	0	0	6	0.014	0.118	1,618
Public magnet (1/0)	0.060	0.237	5,622	0.069	0.253	8,393	0	0	6	0.063	0.243	1,618
Private non-religious (1/0)	0.219	0.414	5,622	0.235	0.424	8,393	0.167	0.408	6	0.256	0.436	1,618
Private Catholic (1/0)	0.056	0.229	5,622	0.058	0.233	8,393	0	0	6	0.064	0.245	1,618
Private other religious (1/0)	0.047	0.211	5,622	0.047	0.212	8,393	0.167	0.408	6	0.046	0.210	1,618
Single-sex (1/0)	0.046	0.210	5,622	0.051	0.220	8,393	0	0	6	0.059	0.235	1,618
Proportion free-and-reduced lunch	0.208	0.231	5,070	0.203	0.234	7,558	0.312	0.375	6	0.213	0.247	1,433

Appendix B

Table B1: Heterogeneity of effects on Economics major choice

	<700 math	≥700 math	Not URM	URM
Economics	0.049**	0.093**	0.077**	0.038**
	(0.005)	(0.011)	(0.008)	(0.006)
R^2	0.13	0.16	0.15	0.13
p-value [R,U]	[0.000,0.000]	[0.019,0.000]	[0.000,0.000]	[0.004,0.000]
Non-enrollee mean	0.048	0.100	0.082	0.052
Bound	0.050	0.089	0.083	0.045
N	6,828	7,212	12,882	2,508

	Economics preference	Any STEM preference	Neither Economics nor STEM preference
Economics	0.277**	0.053**	0.062**
	(0.032)	(0.008)	(0.007)
R^2	0.20	0.12	0.08
p-value [R,U]	[0.003,0.000]	[0.019,0.000]	[0.003,0.000]
Non-enrollee mean	0.296	0.068	0.056
Bound	0.294	0.058	0.065
N	1,270	6,157	8,281

	Single-sex high school	Non-single-sex high school	Any Wellesley College relative	No Wellesley College relative
Economics	0.105**	0.071**	0.077**	0.072**
	(0.021)	(0.007)	(0.013)	(0.007)
R^2	0.33	0.14	0.23	0.14
p-value [R,U]	[0.127,0.000]	[0.001,0.000]	[0.055,0.000]	[0.001,0.000]
Non-enrollee mean	0.047	0.078	0.059	0.078
Bound	0.138	0.077	0.090	0.077
N	688	14,702	1,532	13,858

	Applicant group		Years of Fall enrollment	
	Early evaluation	Regular admission	1999-2002	2003-2013
Economics	0.071**	0.072**	0.062**	0.074**
	(0.009)	(0.006)	(0.009)	(0.007)
R^2	0.14	0.16	0.17	0.14
p-value [R,U]	[0.005,0.000]	[0.000,0.000]	[0.004,0.000]	[0.001,0.000]
Non-enrollee mean	0.083	0.070	0.070	0.079
Bound	0.078	0.079	0.062	0.082
N	7,911	7,479	4,083	11,307

Notes: Each cell reports the coefficient on enrollment from a separate regression in the sample denoted in the header; all regressions include dummy variables for δ_j and θ_k, and application and high school variables. Standard errors are in parentheses, clustered by the college or university where the degree was received. ** indicates significance at 1%, and * at 5%. The p-values are from the restricted (with null imposed) and unrestricted wild cluster bootstrap-t, respectively. The bounds are calculated relative to a specification that only controls for enrollment; see text for a description of other assumptions.