Online Appendix to "Old Boys' Clubs and Upward Mobility Among the Educational Elite"

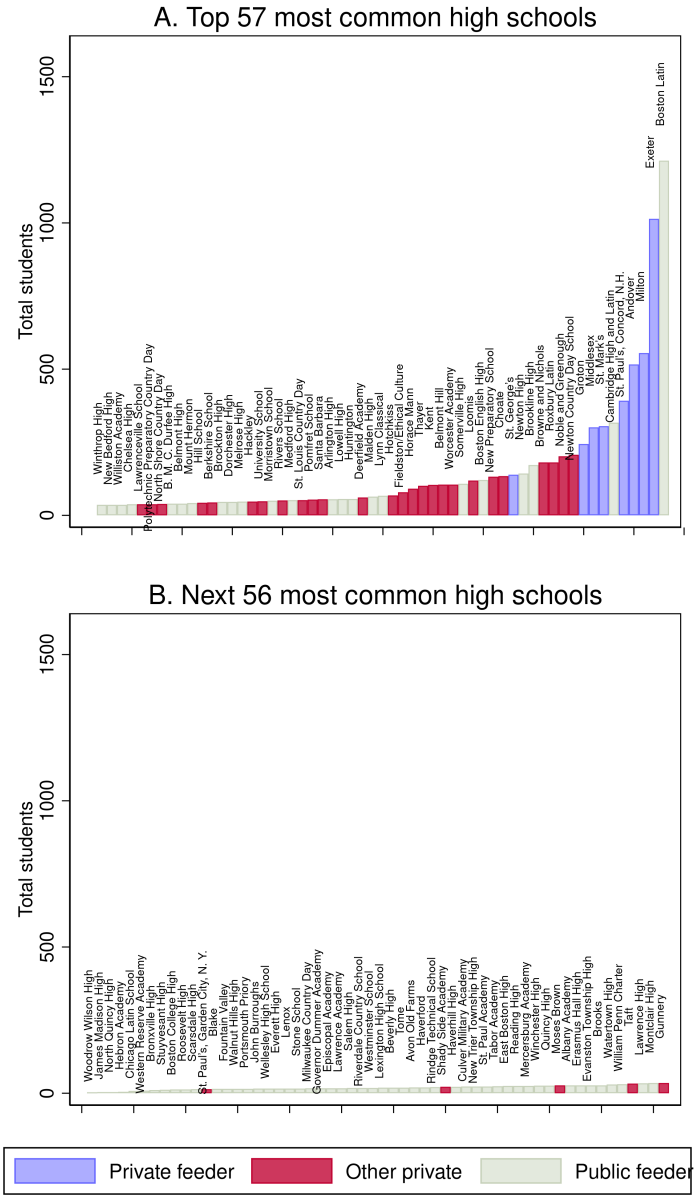
Valerie Michelman
Joseph Price
Seth Zimmerman

March 12, 2021

A Additional tables and figures

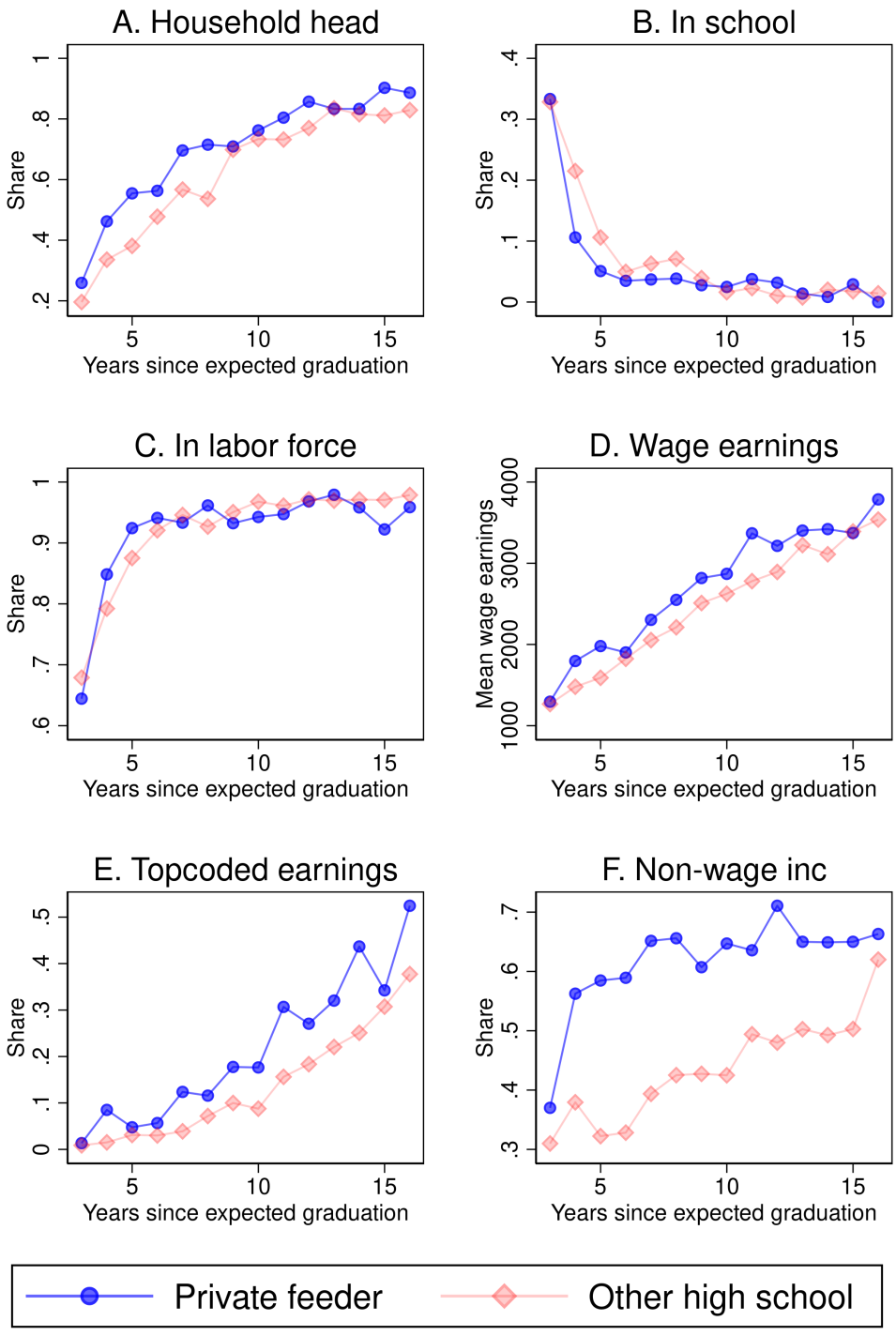
A.1 Figures

Figure A.1: Counts of Harvard students by high school



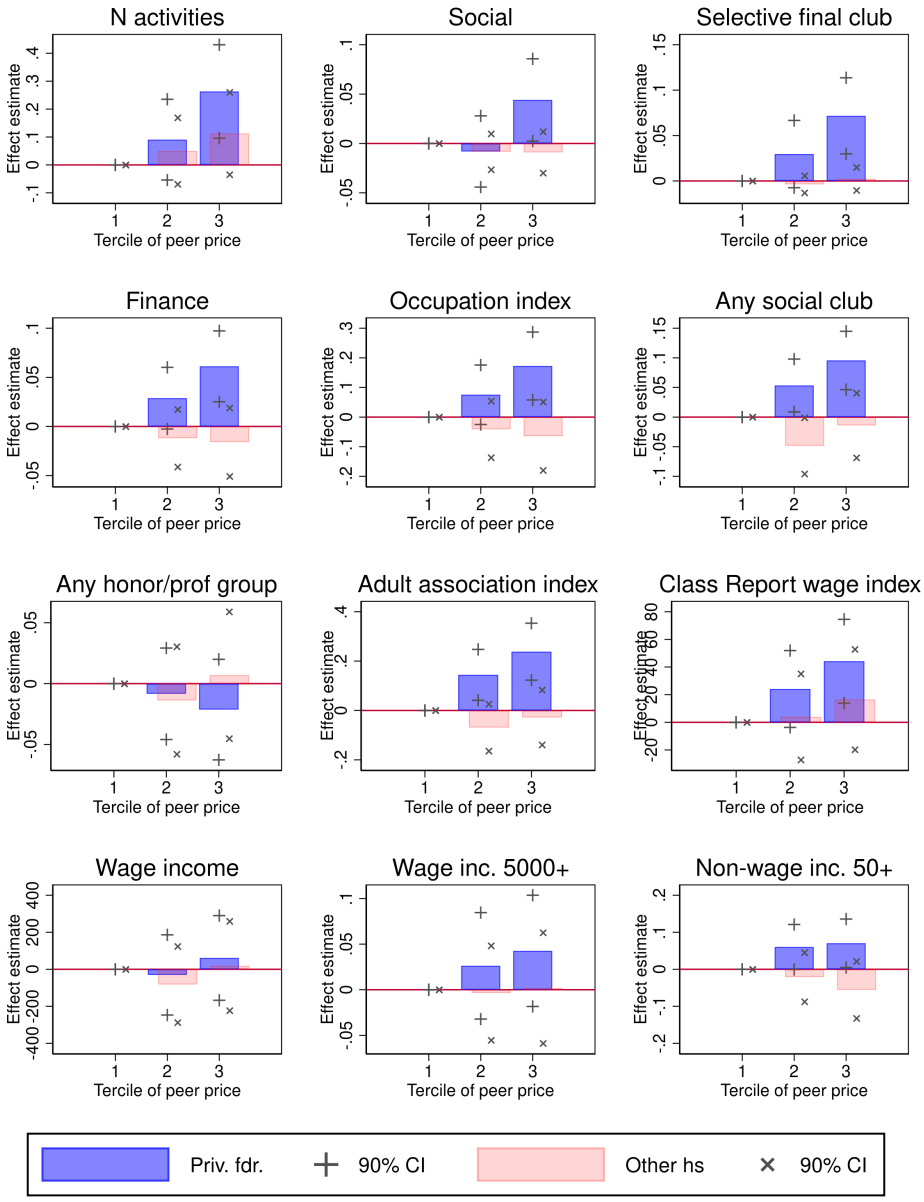
Count of students in Harvard sample universe by high school for each high school that we classify as private feeder, other private, or public feeder. Bar colors denote high school classification. See Section 3.4.1 for school classification details.

Figure A.2: Census outcome profiles by cohort and high school type



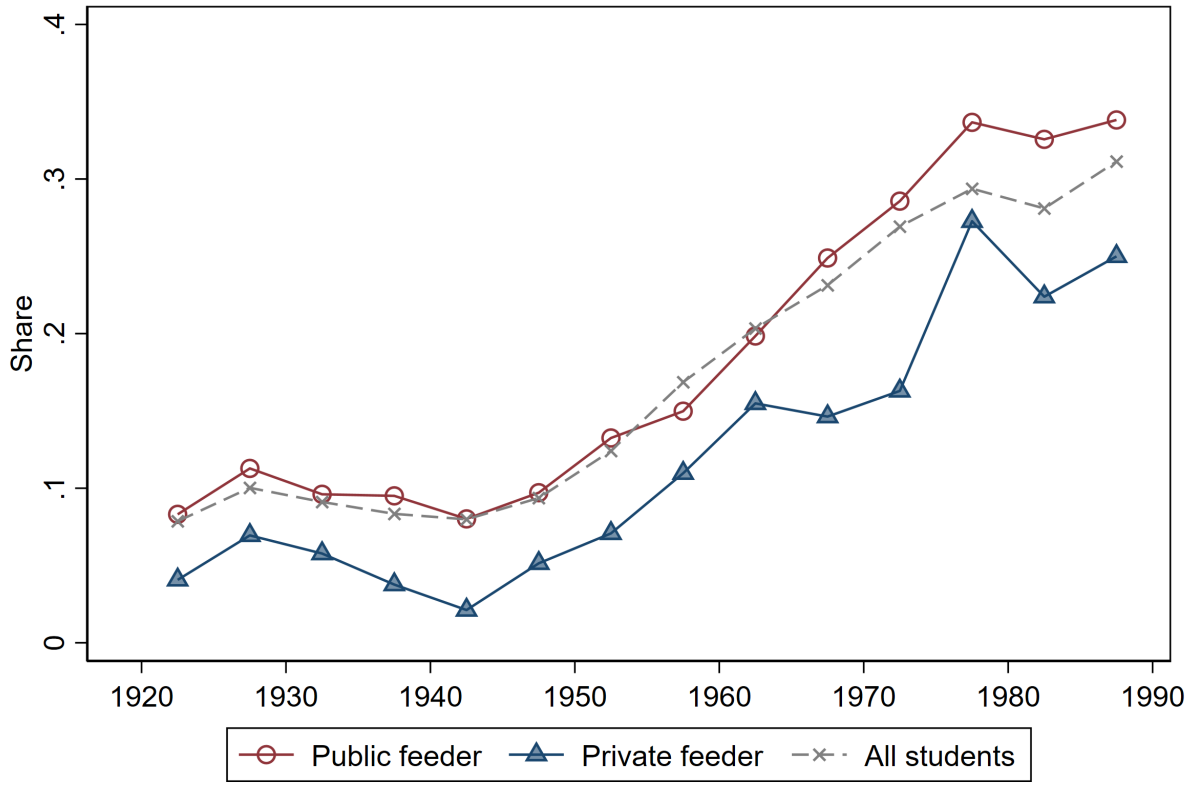
Census outcome profiles presented by cohort. The horizontal axis is years since expected graduation, set equal to 1940 − (entry year + 4). Sample: Harvard students matched to 1940 Census data. See section 3.3 for discussion.

Figure A.3: Peer effect estimates by tercile



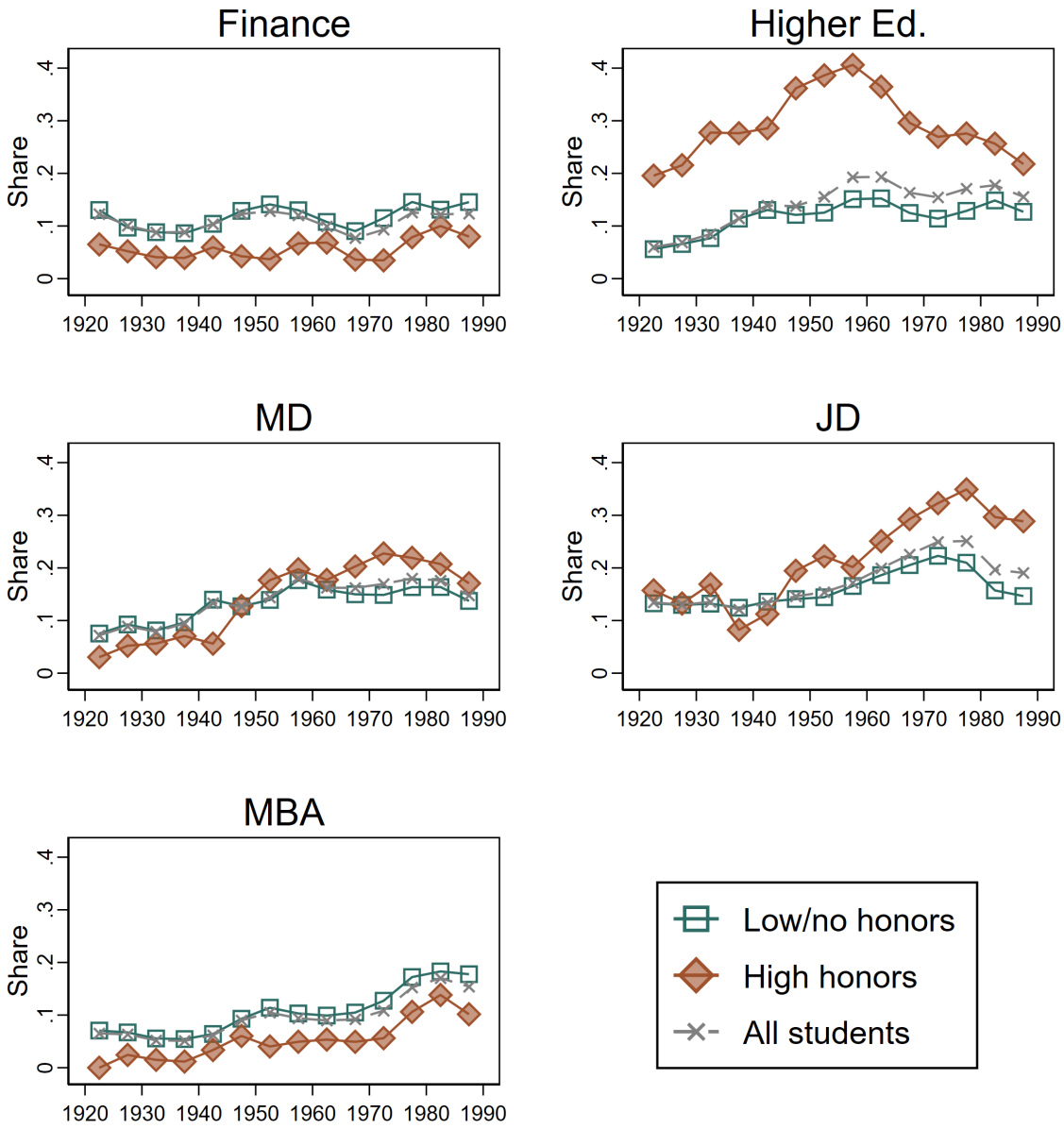
Estimated effects and 90% confidence intervals of assignment to rooms in the top and middle terciles of the peer neighborhood price distribution, relative to omitted bottom tercile category, for outcomes listed in graph title. Specifications are as in equation 2 but use tercile dummies rather than continuous measures of neighborhood price. See text for outcome definitions and section 5.7 for more discussion.

Figure A.4: Share of high honors students by high school type– men and women



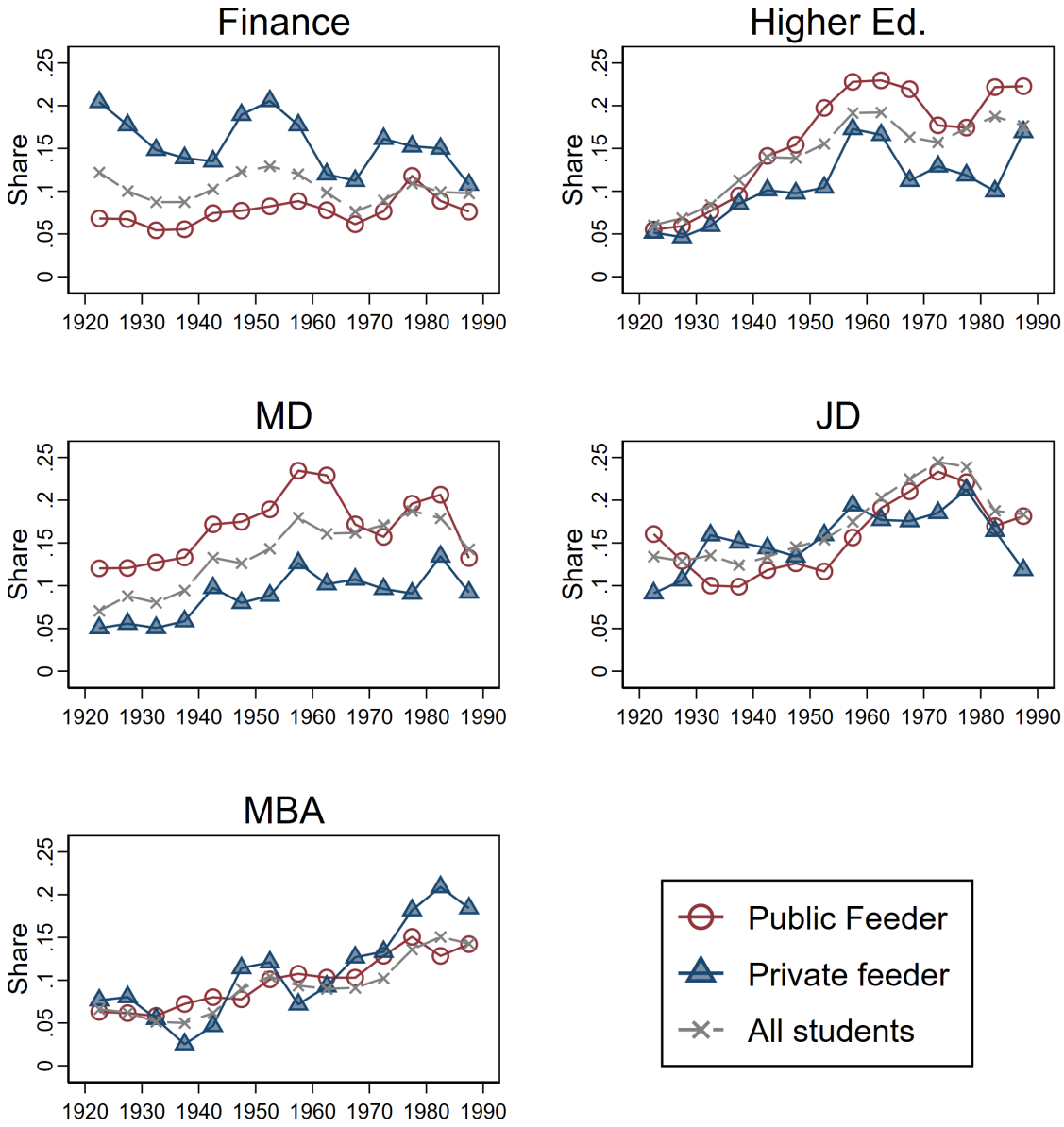
Share of students with high honors pooled for all students as well as separately for private feeder and public feeder students by graduating class year. High honors is defined as graduating with a *magna* or *summa cum laude* degree. Sample includes male and female students. Figure uses data from Class Reports that is available for each graduating class from 1924 through 1939 and then at five year intervals starting in 1940. Points display means over all years within 2.5 years on either side of the centered value. For example, the 1982.5 datapoint is an average of 1980 and 1985 class years, and the 1987.5 datapoint is an average of 1985 and 1990 class years. See section 6.1 for details.

Figure A.5: Long-run trends in career outcomes by academic achievement



Share of students in graduating class in listed occupation or with listed degree type. Horizontal axis in all panels is graduating class year. All panels present rates for all students as well as split by undergraduate academic achievement at Harvard, with "high honors" corresponding to students who graduate *magna* or *summa* and "low/no honors" corresponding to all other students. Sample is male students. "JD" panel includes both LLB and JD degrees. All use data from Class Reports that is available for each graduating class from 1924 through 1939 and then at five year intervals starting in 1940. Points are means over all years within 2.5 years on either side of the centered value. For example, the 1982.5 datapoint is an average of 1980 and 1985 class years, and the 1987.5 datapoint is an average of 1985 and 1990 class years. See section 6.1 for details.

Figure A.6: Long-run trends in career outcomes by high school background – men and women



Share of students in graduating class in listed occupation or with listed degree type. Horizontal axis in all panels is graduating class year. All panels present rates for all students as well as split by high school type. Sample includes both male and female students. "JD" panel includes both LLB and JD degrees. All use data from Class Reports that is available for each graduating class from 1924 through 1939 and then at five year intervals starting in 1940. Points are means over all years within 2.5 years on either side of the centered value. For example, the 1982.5 datapoint is an average of 1980 and 1985 class years, and the 1987.5 datapoint is an average of 1985 and 1990 class years. See section 6.1 for details.

A.2 Tables

Table A.1: Peer effects on short-run outcomes without occupancy controls

	All	Private	Non-private	Test
A. First-year activities				
Have any activity	0.083	0.100	0.063	0.566
	(0.031)	(0.053)	(0.039)	
N activities	0.303	0.656	0.123	0.013
	(0.098)	(0.191)	(0.110)	
Activity leadership position	0.036	0.089	0.007	0.042
	(0.017)	(0.037)	(0.018)	
Social	0.063	0.200	-0.016	0.000
	(0.021)	(0.048)	(0.019)	
Sports	0.041	0.105	-0.005	0.115
	(0.032)	(0.058)	(0.039)	
Music	0.033	0.054	0.026	0.596
	(0.024)	(0.046)	(0.029)	
Other activities	0.038	-0.007	0.071	0.162
	(0.028)	(0.049)	(0.032)	
First-year activity index	0.224	0.528	0.033	0.003
	(0.073)	(0.151)	(0.079)	
N	9356	2856	6387	
B. Upper-year social clubs				
Selective final club	0.073	0.182	0.018	0.002
	(0.020)	(0.051)	(0.016)	
Less selective final club	-0.016	-0.060	-0.002	0.184
	(0.018)	(0.040)	(0.020)	
Hasty Pudding Inst. 1770	0.053	0.136	0.016	0.048
	(0.026)	(0.056)	(0.027)	
Upper-year club index	0.192	0.441	0.058	0.025
	(0.073)	(0.160)	(0.072)	
N	8601	2633	5862	
C. First-year academic rank				
Rank group 1	0.000	-0.010	0.005	0.382
	(0.009)	(0.012)	(0.012)	
Rank group 2	0.010	0.021	0.004	0.555
	(0.015)	(0.021)	(0.020)	
Rank group 3	-0.003	0.006	-0.010	0.715
	(0.022)	(0.036)	(0.027)	
Rank group 4	-0.008	0.076	-0.031	0.032
	(0.026)	(0.041)	(0.032)	
Rank group 5	0.032	-0.024	0.073	0.131
	(0.029)	(0.054)	(0.035)	
Rank group 6	0.029	0.012	0.017	0.915
	(0.020)	(0.039)	(0.024)	
Rank listed year 1	0.060	0.081	0.056	0.657
	(0.026)	(0.046)	(0.032)	
Class rank year 1	-0.055	0.089	-0.078	0.376
	(0.093)	(0.154)	(0.116)	
N	7048	2123	4821	

Coefficients on peer neighborhood price rank from regressions of form given in equation 2. Specifications are identical to those reported in Table 6 but with randomization blocks defined only by interactions between per-student price and entry year. Occupancy interactions and controls for high school identifiers are dropped. The first three columns denote samples. Rows are outcome variables. "Test" column reports the p-value from a test of the null that the coefficients reported in the private and non-private columns are equal. Panels A and B report first-year activity outcomes and upper-year club outcomes respectively. "First-year activity index" and "upper-year club index" are the standardized indices of the association between activities and private high school background described in section 3.4.3. Other variables are indicators for a given activity type unless stated otherwise. Panel C describes academic outcomes in the first year. Class rank is a continuous variable from one through six, with one the best and six the worst. The other outcomes are dummies for having grades in the listed rank group and being listed at all. Standard errors clustered at peer neighborhood level.

Table A.2: Peer effects on long-run outcomes without occupancy controls

	All	Private	Non-private	Test
A. Adult associations				
Any social club	0.073	0.203	0.020	0.012
	(0.033)	(0.061)	(0.041)	
Country club	0.083	0.209	0.033	0.010
	(0.030)	(0.058)	(0.037)	
Gentleman's club	0.021	0.055	0.001	0.334
	(0.024)	(0.052)	(0.026)	
Fraternal order	0.008	0.008	0.009	0.984
	(0.020)	(0.032)	(0.026)	
Any honor/prof group	-0.033	0.022	-0.057	0.247
	(0.029)	(0.055)	(0.037)	
Prof. Association	-0.043	0.010	-0.071	0.225
	(0.029)	(0.055)	(0.037)	
Honor society	-0.020	0.013	-0.026	0.313
	(0.017)	(0.032)	(0.022)	
Adult association index	0.177	0.349	0.099	0.149
	(0.073)	(0.154)	(0.081)	
N	8061	2478	5473	
B. Occupation choice				
Finance	0.025	0.145	-0.035	0.002
	(0.023)	(0.052)	(0.025)	
Accounting	0.021	0.056	-0.003	0.217
	(0.021)	(0.041)	(0.026)	
Doctor	0.008	-0.014	0.027	0.309
	(0.020)	(0.031)	(0.026)	
Law	-0.011	0.005	-0.014	0.686
	(0.024)	(0.038)	(0.032)	
Higher ed.	-0.018	-0.023	-0.023	0.995
	(0.019)	(0.028)	(0.026)	
Teach	-0.016	-0.033	-0.007	0.524
	(0.019)	(0.033)	(0.025)	
Government	-0.002	0.014	-0.007	0.486
	(0.014)	(0.026)	(0.016)	
Art/pub	-0.016	-0.040	-0.002	0.363
	(0.020)	(0.035)	(0.024)	
Senior management	0.023	0.038	0.007	0.642
	(0.030)	(0.056)	(0.037)	
Low management	0.059	0.031	0.073	0.395
	(0.024)	(0.039)	(0.031)	
Retail	-0.004	0.035	-0.024	0.267
	(0.025)	(0.041)	(0.033)	
Occupation index	0.067	0.439	-0.130	0.002
	(0.077)	(0.167)	(0.087)	
N	6958	2107	4742	
C. Adult income				
Wage income	53.7	-251.2	212.0	0.246
	(182.7)	(351.9)	(222.0)	
Wage inc. 5000+	0.052	0.009	0.091	0.459
	(0.050)	(0.097)	(0.060)	
Non-wage inc. 50+	0.055	0.093	0.032	0.589
	(0.051)	(0.095)	(0.062)	
Class Report wage index	43.6	143.9	-12.8	0.005
	(26.2)	(46.2)	(33.2)	
N	2392	710	1619	

Coefficients on peer neighborhood price rank from regressions of form given in equation 2. The first three columns denote samples. Specifications are identical to those reported in Table 7 but with randomization blocks defined only by interactions between per-student price and entry year. Occupancy interactions and controls for high school identifiers are dropped. Rows are outcome variables."Test" column reports the p-value from a test of the null that the coefficients reported in the private and non-private columns are equal. Panels A and B report adult social club and occupation outcomes, respectively. "Adult association index" and "Occupation index" are standardized indices of the association between adult outcomes and private high school background. See section 3.4.3. Other variables are indicators for outcomes of the listed type. The sample in Panels A and B is students matched class reports; Panel B further restricts to students with occupation outcomes. Panel C reports labor market outcomes from the 1940 census. Sample is students matched to Census wage records in the 1920-30 entering cohorts. Standard errors clustered at peer neighborhood level.

Table A.3: Peer effects on short-run outcomes without large high school controls

	All	Private	Non-private	Test
A. First-year activities				
Have any activity	0.068	0.088	0.061	0.703
	(0.036)	(0.057)	(0.043)	
N activities	0.277	0.723	0.087	0.013
	(0.117)	(0.226)	(0.127)	
Activity leadership position	0.042	0.100	0.011	0.063
	(0.020)	(0.044)	(0.021)	
Social	0.048	0.193	-0.031	0.000
	(0.026)	(0.058)	(0.023)	
Sports	0.022	0.087	-0.010	0.207
	(0.037)	(0.064)	(0.043)	
Music	0.050	0.113	0.029	0.173
	(0.027)	(0.053)	(0.033)	
Other activities	0.023	-0.015	0.057	0.262
	(0.031)	(0.058)	(0.035)	
First-year activity index	0.195	0.510	0.013	0.011
	(0.089)	(0.173)	(0.090)	
N	9343	2828	6367	
B. Upper-year social clubs				
Selective final club	0.067	0.182	-0.001	0.006
	(0.026)	(0.065)	(0.018)	
Less selective final club	-0.031	-0.062	-0.013	0.353
	(0.021)	(0.050)	(0.022)	
Hasty Pudding Inst. 1770	0.021	0.137	-0.024	0.036
	(0.034)	(0.072)	(0.031)	
Upper-year club index	0.126	0.443	-0.040	0.030
	(0.098)	(0.211)	(0.082)	
N	8589	2606	5845	
C. First-year academic rank				
Rank group 1	-0.001	-0.011	0.003	0.499
	(0.011)	(0.015)	(0.015)	
Rank group 2	0.019	0.015	0.015	0.990
	(0.016)	(0.025)	(0.021)	
Rank group 3	0.025	0.006	0.033	0.588
	(0.024)	(0.039)	(0.030)	
Rank group 4	-0.007	0.091	-0.052	0.011
	(0.029)	(0.045)	(0.036)	
Rank group 5	0.020	-0.063	0.071	0.059
	(0.032)	(0.060)	(0.039)	
Rank group 6	0.020	0.020	0.003	0.741
	(0.023)	(0.046)	(0.028)	
Rank listed year 1	0.076	0.058	0.072	0.823
	(0.029)	(0.050)	(0.036)	
Class rank year 1	0.029	0.089	0.039	0.816
	(0.106)	(0.179)	(0.129)	
N	7020	2085	4790	

Coefficients on peer neighborhood price rank from regressions of form given in equation 2. Specifications are identical to those reported in Table 6 but with controls for high school identifiers dropped. The first three columns denote samples. Rows are outcome variables. "Test" column reports the p-value from a test of the null that the coefficients reported in the private and non-private columns are equal. Panels A and B report first-year activity outcomes and upper-year club outcomes respectively. "First-year activity index" and "upper-year club index" are the standardized indices of the association between activities and private high school background described in section 3.4.3. Other variables are indicators for a given activity type unless stated otherwise. Panel C describes academic outcomes in the first year. Class rank is a continuous variable from one through six, with one the best and six the worst. The other outcomes are dummies for having grades in the listed rank group and being listed at all. Standard errors clustered at peer neighborhood level.

Table A.4: Peer effects on long-run outcomes without large high school controls

	All	Private	Non-private	Test
A. Adult associations				
Any social club	0.067	0.246	-0.004	0.002
	(0.037)	(0.066)	(0.046)	
Country club	0.071	0.223	0.004	0.004
	(0.033)	(0.064)	(0.040)	
Gentleman's club	0.021	0.070	-0.002	0.254
	(0.027)	(0.058)	(0.028)	
Fraternal order	0.001	0.018	-0.001	0.689
	(0.023)	(0.037)	(0.030)	
Any honor/prof group	-0.010	0.038	-0.023	0.401
	(0.032)	(0.059)	(0.040)	
Prof. Association	-0.021	0.029	-0.039	0.351
	(0.031)	(0.059)	(0.040)	
Honor society	-0.010	0.034	-0.017	0.225
	(0.018)	(0.035)	(0.023)	
Adult association index	0.139	0.367	0.024	0.077
	(0.079)	(0.173)	(0.087)	
N	8046	2450	5449	
B. Occupation choice				
Finance	0.025	0.164	-0.047	0.001
	(0.026)	(0.059)	(0.028)	
Accounting	0.027	0.062	0.006	0.312
	(0.024)	(0.048)	(0.029)	
Doctor	0.008	-0.024	0.029	0.224
	(0.022)	(0.033)	(0.029)	
Law	-0.026	-0.020	-0.025	0.936
	(0.027)	(0.044)	(0.035)	
Higher ed.	-0.019	-0.032	-0.022	0.823
	(0.021)	(0.033)	(0.028)	
Teach	-0.019	-0.025	-0.011	0.771
	(0.022)	(0.037)	(0.028)	
Government	0.005	0.021	0.004	0.634
	(0.015)	(0.031)	(0.017)	
Art/pub	-0.013	-0.019	-0.003	0.694
	(0.022)	(0.034)	(0.027)	
Senior management	0.031	0.065	0.011	0.463
	(0.033)	(0.061)	(0.041)	
Low management	0.065	0.042	0.072	0.565
	(0.026)	(0.041)	(0.033)	
Retail	0.003	0.025	-0.009	0.582
	(0.028)	(0.049)	(0.037)	
Occupation index	0.071	0.525	-0.171	0.001
	(0.084)	(0.185)	(0.095)	
N	6928	2075	4712	
C. Adult income				
Wage income	35.8	-306.6	229.5	0.204
	(190.0)	(365.4)	(234.7)	
Wage inc. 5000+	0.031	-0.006	0.073	0.493
	(0.050)	(0.099)	(0.062)	
Non-wage inc. 50+	0.060	0.124	0.028	0.417
	(0.053)	(0.099)	(0.065)	
Class Report wage index	38.0	139.0	-14.6	0.008
	(27.2)	(49.4)	(34.3)	
N	2361	685	1587	

Coefficients on peer neighborhood price rank from regressions of form given in equation 2. Specifications are identical to those reported in Table 7 but with controls for high school identifiers dropped. "Test" column reports the p-value from a test of the null that the coefficients reported in the private and non-private columns are equal. Panels A and B report adult social club and occupation outcomes, respectively. "Adult association index" and "Occupation index" are standardized indices of the association between adult outcomes and private high school background. See section 3.4.3. Other variables are indicators for outcomes of the listed type. The sample in Panels A and B is students matched class reports; Panel B further restricts to students with occupation outcomes. Panel C reports labor market outcomes from the 1940 census. Sample is students matched to Census wage records in the 1920-30 entering cohorts. Standard errors clustered at peer neighborhood level.

Table A.5: Peer effects on short-run outcomes with alternate private high school classifications

	Private feeder	More expansive	Less expansive	Private non-feeder
A. First-year activities				
Have any activity	0.080	0.087	0.054	0.142
	(0.057)	(0.053)	(0.060)	(0.081)
N activities	0.656	0.576	0.639	0.082
	(0.222)	(0.206)	(0.231)	(0.253)
Activity leadership position	0.099	0.088	0.099	-0.023
	(0.044)	(0.039)	(0.045)	(0.048)
Social	0.184	0.124	0.177	-0.002
	(0.057)	(0.053)	(0.059)	(0.054)
Sports	0.085	0.071	0.073	0.004
	(0.064)	(0.058)	(0.066)	(0.080)
Music	0.094	0.097	0.100	0.027
	(0.052)	(0.047)	(0.054)	(0.061)
Other activities	-0.027	-0.048	-0.039	-0.012
	(0.058)	(0.053)	(0.059)	(0.073)
First-year activity index	0.480	0.387	0.467	0.049
	(0.170)	(0.157)	(0.179)	(0.190)
N	2828	3472	2705	1858
B. Academic outcomes				
Rank group 1	-0.012	-0.002	-0.013	0.026
	(0.015)	(0.013)	(0.016)	(0.022)
Rank group 2	0.017	0.032	0.017	0.021
	(0.025)	(0.022)	(0.025)	(0.032)
Rank group 3	0.006	0.016	-0.005	-0.010
	(0.039)	(0.036)	(0.040)	(0.048)
Rank group 4	0.093	0.051	0.101	-0.064
	(0.046)	(0.044)	(0.048)	(0.069)
Rank group 5	-0.068	-0.038	-0.077	0.079
	(0.061)	(0.057)	(0.062)	(0.076)
Rank group 6	0.016	0.018	0.006	0.018
	(0.046)	(0.043)	(0.047)	(0.056)
Rank listed year 1	0.052	0.077	0.030	0.070
	(0.051)	(0.046)	(0.051)	(0.073)
Class rank year 1	0.098	0.140	0.094	0.102
	(0.180)	(0.165)	(0.182)	(0.253)
N	2085	2579	2000	1359
C. Upper-year social clubs				
Selective final club	0.168	0.135	0.169	-0.005
	(0.056)	(0.049)	(0.059)	(0.049)
Less selective final club	-0.074	-0.055	-0.079	0.023
	(0.050)	(0.045)	(0.051)	(0.055)
Hasty Pudding Inst. 1770	0.106	0.088	0.069	0.011
	(0.062)	(0.056)	(0.063)	(0.074)
Upper-year club index	0.358	0.300	0.304	0.057
	(0.176)	(0.158)	(0.181)	(0.204)
N	2606	3210	2494	1706

Coefficients on peer neighborhood price rank from regressions of form given in equation 2. Specifications are identical to those reported in Table 6 but for alternate high school classifications. "Private feeder" column repeats estimates from Table 6 for private feeder sample. "More expansive" column reports results for the sample of the 13 private schools that sent the most students to Harvard over the period. "Less expansive" reports results for the sample of the seven private schools that sent the most students. "Private non-feeder" reports results for all private schools not part of the main feeder sample. Panel A reports first-year activity outcomes. "First year activity index" and "Upper-year club index" are the standardized private school indices described in section 3.4.3. See text for other variable descriptions. Standard errors clustered at peer neighborhood level.

Table A.6: Peer effects on long-run outcomes with alternate private high school classifications

	Private feeder	More expansive	Less expansive	Private non-feeder
A. Adult associations				
Any social club	0.238	0.188	0.241	0.035
	(0.067)	(0.060)	(0.071)	(0.085)
Country club	0.216	0.199	0.220	0.097
	(0.066)	(0.056)	(0.068)	(0.078)
Gentleman's club	0.060	0.047	0.054	0.024
	(0.057)	(0.050)	(0.059)	(0.058)
Fraternal order	0.024	0.006	0.011	0.017
	(0.037)	(0.035)	(0.039)	(0.048)
Any honor/prof group	0.037	0.010	0.031	-0.083
	(0.059)	(0.053)	(0.061)	(0.075)
Prof. Association	0.030	0.005	0.017	-0.081
	(0.059)	(0.053)	(0.061)	(0.074)
Honor society	0.025	0.023	0.030	-0.007
	(0.035)	(0.030)	(0.036)	(0.041)
Adult association index	0.337	0.321	0.343	0.246
	(0.173)	(0.147)	(0.177)	(0.172)
N	2450	3009	2346	1610
B. Occupation choice				
Finance	0.150	0.110	0.156	-0.080
	(0.059)	(0.052)	(0.060)	(0.061)
Accounting	0.065	0.050	0.059	0.095
	(0.048)	(0.043)	(0.050)	(0.058)
Doctor	-0.023	-0.032	-0.019	0.019
	(0.034)	(0.031)	(0.035)	(0.040)
Law	-0.009	-0.014	-0.026	-0.009
	(0.044)	(0.038)	(0.046)	(0.062)
Higher ed.	-0.026	-0.005	-0.009	0.001
	(0.034)	(0.031)	(0.034)	(0.054)
Teach	-0.018	-0.008	-0.020	-0.007
	(0.038)	(0.034)	(0.040)	(0.052)
Government	0.020	0.020	0.017	0.021
	(0.031)	(0.027)	(0.032)	(0.033)
Art/pub	-0.017	-0.038	-0.015	-0.016
	(0.035)	(0.031)	(0.036)	(0.045)
Senior management	0.064	0.038	0.055	-0.099
	(0.061)	(0.055)	(0.063)	(0.085)
Low management	0.041	0.089	0.047	0.107
	(0.042)	(0.039)	(0.043)	(0.063)
Retail	0.030	0.062	0.027	0.060
	(0.049)	(0.044)	(0.050)	(0.079)
Occupation index	0.470	0.348	0.485	-0.269
	(0.183)	(0.161)	(0.186)	(0.195)
N	2075	2576	1987	1373
C. Adult income				
Wage income	-368.1	-92.7	-367.5	753.2
	(370.5)	(307.5)	(381.0)	(476.9)
Wage inc. 5000+	-0.020	0.019	-0.045	0.240
	(0.099)	(0.088)	(0.101)	(0.139)
Non-wage inc. 50+	0.094	0.085	0.110	0.261
	(0.099)	(0.097)	(0.101)	(0.126)
Class Report wage index	130.2	99.5	116.1	-50.9
	(49.2)	(45.1)	(51.8)	(61.8)
N	685	868	652	428

Coefficients on peer neighborhood price rank from regressions of form given in equation 2. Specifications are identical to those reported in Table 7 but for alternate high school classifications. "Private feeder" column repeats estimates from Table 6 for private feeder sample. "More expansive" column reports results for the sample of the 13 private schools that sent the most students to Harvard over the period. "Less expansive" reports results for the sample of the seven private schools that sent the most students. "Private non-feeder" reports results for all private schools not part of the main feeder sample. "Adult association index" and "Occupation index" are standardized private school indices described in section 3.4.3. See text for other variable descriptions. Standard errors clustered at peer neighborhood level.

Table A.7: Peer neighborhood effects on short-run outcomes excluding cohorts 1919-1921

	All	Private	Non-private	Test
A. First-year activities				
Have any activity	0.074	0.081	0.066	0.850
	(0.037)	(0.063)	(0.046)	
N activities	0.283	0.408	0.205	0.431
	(0.114)	(0.230)	(0.128)	
Activity leadership position	0.036	0.105	0.003	0.045
	(0.021)	(0.047)	(0.022)	
Social	0.049	0.158	-0.010	0.006
	(0.025)	(0.058)	(0.022)	
Sports	0.039	0.064	0.015	0.559
	(0.037)	(0.071)	(0.045)	
Music	0.029	0.026	0.030	0.957
	(0.029)	(0.054)	(0.035)	
Other activities	0.046	0.008	0.073	0.342
	(0.033)	(0.061)	(0.038)	
First-year activity index	0.226	0.454	0.078	0.070
	(0.089)	(0.188)	(0.094)	
N	8018	2418	5470	
B. Upper-year social clubs				
Selective final club	0.069	0.145	0.016	0.032
	(0.022)	(0.058)	(0.017)	
Less selective final club	-0.024	-0.055	-0.008	0.427
	(0.022)	(0.055)	(0.023)	
Hasty Pudding Inst. 1770	0.010	0.058	-0.004	0.391
	(0.030)	(0.068)	(0.030)	
Upper-year club index	0.129	0.308	0.019	0.162
	(0.085)	(0.194)	(0.083)	
N	7264	2196	4948	
C. First-year academic rank				
Rank group 1	-0.001	-0.010	0.004	0.519
	(0.011)	(0.015)	(0.016)	
Rank group 2	0.019	0.019	0.013	0.864
	(0.016)	(0.027)	(0.021)	
Rank group 3	0.020	0.008	0.027	0.717
	(0.026)	(0.042)	(0.032)	
Rank group 4	-0.021	0.089	-0.075	0.006
	(0.031)	(0.049)	(0.039)	
Rank group 5	0.021	-0.071	0.074	0.058
	(0.034)	(0.064)	(0.042)	
Rank group 6	0.031	0.001	0.027	0.623
	(0.022)	(0.045)	(0.027)	
Rank listed year 1	0.068	0.037	0.070	0.610
	(0.031)	(0.055)	(0.038)	
Class rank year 1	-0.017	0.129	-0.038	0.462
	(0.107)	(0.188)	(0.132)	
N	6238	1862	4253	

Coefficients on peer neighborhood price rank from regressions of form given in equation 2. Cohorts 1919, 1920, and 1921 are excluded. The first three columns denote samples. Rows are outcome variables. All specifications include randomization block and dummies for large feeder high schools; see section 5.1 for details."Test" column reports the p-value from a test of the null that the coefficients reported in the private and non-private columns are equal. Panels A and B report first-year activity outcomes and upper-year club outcomes respectively. "First-year activity index" and "upper-year club index" are the standardized indices of the association between activities and private high school background described in section 3.4.3. Other variables are indicators for a given activity type unless stated otherwise. Panel C describes academic outcomes in the first year. Class rank is a continuous variable from one through six, with one the best and six the worst. The other outcomes are dummies for having grades in the listed rank group and being listed at all. Private and non-private column sample sizes add up to slightly less than "all" column sample size because there is no variation in peer neighborhood assignment for private-feeder or non-private-feeder students only within some randomization blocks; students in these blocks are omitted from split sample regressions. Standard errors clustered at peer neighborhood level.

Table A.8: Peer neighborhood effects on long-run outcomes excluding cohorts 1919-1921

	All	Private	Non-private	Test
A. Adult associations				
Any social club	0.064	0.257	-0.008	0.002
	(0.038)	(0.071)	(0.048)	
Country club	0.079	0.235	0.017	0.007
	(0.034)	(0.069)	(0.042)	
Gentleman's club	0.015	0.045	0.002	0.516
	(0.028)	(0.060)	(0.029)	
Fraternal order	-0.000	0.027	-0.011	0.458
	(0.023)	(0.040)	(0.031)	
Any honor/prof group	-0.014	0.046	-0.024	0.389
	(0.035)	(0.066)	(0.044)	
Prof. Association	-0.025	0.044	-0.042	0.289
	(0.034)	(0.066)	(0.043)	
Honor society	-0.010	0.032	-0.015	0.312
	(0.020)	(0.039)	(0.025)	
Adult association index	0.135	0.310	0.053	0.210
	(0.082)	(0.184)	(0.092)	
N	7223	2214	4878	
B. Occupation choice				
Finance	0.010	0.138	-0.050	0.005
	(0.026)	(0.061)	(0.029)	
Accounting	0.054	0.106	0.028	0.176
	(0.025)	(0.049)	(0.031)	
Doctor	0.014	-0.016	0.034	0.291
	(0.024)	(0.036)	(0.031)	
Law	-0.039	-0.031	-0.039	0.897
	(0.028)	(0.046)	(0.037)	
Higher ed.	-0.017	-0.030	-0.021	0.853
	(0.022)	(0.037)	(0.030)	
Teach	-0.023	-0.019	-0.017	0.957
	(0.023)	(0.041)	(0.029)	
Government	0.006	0.014	0.006	0.834
	(0.016)	(0.034)	(0.019)	
Art/pub	-0.007	-0.008	0.005	0.771
	(0.023)	(0.036)	(0.029)	
Senior management	0.034	0.075	0.010	0.414
	(0.035)	(0.066)	(0.044)	
Low management	0.058	0.026	0.065	0.481
	(0.028)	(0.044)	(0.035)	
Retail	-0.003	-0.013	-0.002	0.861
	(0.030)	(0.051)	(0.038)	
Occupation index	0.032	0.452	-0.175	0.003
	(0.086)	(0.193)	(0.099)	
N	6190	1860	4207	
C. Adult income				
Wage income	-44.9	-494.3	161.6	0.141
	(201.7)	(378.0)	(253.9)	
Wage inc. 5000+	0.030	-0.014	0.077	0.446
	(0.052)	(0.104)	(0.063)	
Non-wage inc. 50+	0.037	0.014	0.068	0.663
	(0.055)	(0.103)	(0.069)	
Class Report wage index	22.0	126.2	-25.6	0.017
	(29.4)	(53.5)	(37.1)	
N	1961	582	1300	

Coefficients on peer neighborhood price rank from regressions of form given in equation 2. Cohorts 1919, 1920, and 1921 are excluded. The first three columns denote samples. Rows are outcome variables. All specifications include randomization block and dummies for large feeder high schools; see section 5.1 for details."Test" column reports the p-value from a test of the null that the coefficients reported in the private and non-private columns are equal. Panels A and B report adult social club and occupation outcomes, respectively. "Adult association index" and "Occupation index" are standardized indices of the association between adult outcomes and private high school background. See section 3.4.3. Other variables are indicators for outcomes of the listed type. The sample in Panels A and B is students matched class reports; Panel B further restricts to students with occupation outcomes. Panel C reports labor market outcomes from the 1940 census. Sample is students matched to Census wage records in the 1920-30 entering cohorts. Standard errors clustered at peer neighborhood level.

Table A.9: Peer neighborhood effects on short-run outcomes with randomization inference

	All	Private	Non-private	Test
A. First-year activities				
Have any activity	0.070	0.080	0.064	0.829
	[0.042]	[0.150]	[0.164]	
N activities	0.282	0.656	0.105	0.024
	[0.008]	[0.003]	[0.384]	
Activity leadership position	0.043	0.099	0.012	0.063
	[0.035]	[0.021]	[0.567]	
Social	0.049	0.184	-0.028	0.001
	[0.021]	[0.000]	[0.213]	
Sports	0.028	0.085	-0.007	0.240
	[0.437]	[0.175]	[0.865]	
Music	0.047	0.094	0.030	0.326
	[0.093]	[0.078]	[0.364]	
Other activities	0.021	-0.027	0.059	0.191
	[0.491]	[0.663]	[0.103]	
First-year activity index	0.205	0.480	0.027	0.008
	[0.006]	[0.005]	[0.748]	
N	9343	2828	6367	
B. Upper-year social clubs				
Selective final club	0.065	0.168	0.001	0.008
	[0.003]	[0.006]	[0.960]	
Less selective final club	-0.028	-0.074	-0.007	0.207
	[0.198]	[0.126]	[0.782]	
Hasty Pudding Inst. 1770	0.019	0.106	-0.018	0.072
	[0.486]	[0.092]	[0.567]	
Upper-year club index	0.122	0.358	-0.021	0.070
	[0.124]	[0.054]	[0.815]	
N	8589	2606	5845	
C. First-year academic rank				
Rank group 1	-0.002	-0.012	0.002	0.380
	[0.877]	[0.358]	[0.845]	
Rank group 2	0.018	0.017	0.014	0.927
	[0.270]	[0.441]	[0.540]	
Rank group 3	0.022	0.006	0.029	0.663
	[0.349]	[0.855]	[0.326]	
Rank group 4	-0.007	0.093	-0.053	0.011
	[0.786]	[0.051]	[0.132]	
Rank group 5	0.023	-0.068	0.075	0.046
	[0.509]	[0.265]	[0.062]	
Rank group 6	0.021	0.016	0.006	0.885
	[0.368]	[0.752]	[0.840]	
Rank listed year 1	0.075	0.052	0.073	0.745
	[0.008]	[0.341]	[0.038]	
Class rank year 1	0.016	0.098	0.017	0.664
	[0.885]	[0.541]	[0.903]	
N	7020	2085	4790	

Coefficients on peer neighborhood price rank from regressions of form given in equation 2. The first three columns denote samples. Rows are outcome variables. All specifications include randomization block and dummies for large feeder high schools; see section 5.1 for details."Test" column reports the p-value from a test of the null that the coefficients reported in the private and non-private columns are equal. Panels A and B report first-year activity outcomes and upper-year club outcomes respectively. "First-year activity index" and "upper-year club index" are the standardized indices of the association between activities and private high school background described in section 3.4.3. Other variables are indicators for a given activity type unless stated otherwise. Panel C describes academic outcomes in the first year. Class rank is a continuous variable from one through six, with one the best and six the worst. The other outcomes are dummies for having grades in the listed rank group and being listed at all. Private and non-private column sample sizes add up to slightly less than "all" column sample size because there is no variation in peer neighborhood assignment for private-feeder or non-private-feeder students only within some randomization blocks; students in these blocks are omitted from split sample regressions. P-values are presented in brackets below each coefficient. Two-sided p-values computed using randomization inference with 2000 re-randomizations of roommate groups to rooms within blocks. Statistic reported is $\frac{1}{N}\sum_{j=1}^{N} 1[\mid \hat{\beta}_1^j \mid > \mid \hat{\beta}_1 \mid]$, where $N = 2000$ is the number of rerandomizations, $\hat{\beta}_1^j$ is the estimate of β_1 in the jth randomization, and $\hat{\beta}_1$ is the estimate of β_1 observed in the data.

Table A.10: Peer neighborhood effects on long-run outcomes with randomization inference

	All	Private	Non-private	Test
A. Adult associations				
Any social club	0.063	0.238	-0.006	0.004
	[0.082]	[0.003]	[0.937]	
Country club	0.069	0.216	0.004	0.010
	[0.051]	[0.002]	[0.888]	
Gentleman's club	0.015	0.060	-0.005	0.310
	[0.542]	[0.306]	[0.839]	
Fraternal order	0.002	0.024	-0.003	0.571
	[0.896]	[0.521]	[0.920]	
Any honor/prof group	-0.012	0.037	-0.023	0.423
	[0.719]	[0.525]	[0.601]	
Prof. Association	-0.022	0.030	-0.039	0.348
	[0.489]	[0.608]	[0.333]	
Honor society	-0.014	0.025	-0.019	0.242
	[0.435]	[0.427]	[0.417]	
Adult association index	0.126	0.337	0.021	0.092
	[0.095]	[0.035]	[0.804]	
N	8046	2450	5449	
B. Occupation choice				
Finance	0.020	0.150	-0.047	0.004
	[0.511]	[0.014]	[0.098]	
Accounting	0.029	0.065	0.008	0.357
	[0.263]	[0.213]	[0.817]	
Doctor	0.010	-0.023	0.031	0.238
	[0.662]	[0.514]	[0.294]	
Law	-0.026	-0.009	-0.029	0.695
	[0.333]	[0.827]	[0.419]	
Higher ed.	-0.017	-0.026	-0.021	0.930
	[0.408]	[0.423]	[0.424]	
Teach	-0.017	-0.018	-0.011	0.863
	[0.458]	[0.625]	[0.709]	
Government	0.005	0.020	0.005	0.678
	[0.746]	[0.503]	[0.759]	
Art/pub	-0.012	-0.017	-0.001	0.720
	[0.576]	[0.626]	[0.980]	
Senior management	0.031	0.064	0.011	0.509
	[0.354]	[0.352]	[0.785]	
Low management	0.065	0.041	0.070	0.607
	[0.009]	[0.328]	[0.029]	
Retail	0.005	0.030	-0.010	0.495
	[0.853]	[0.515]	[0.826]	
Occupation index	0.050	0.470	-0.173	0.002
	[0.576]	[0.011]	[0.069]	
N	6928	2075	4712	
C. Adult income				
Wage income	6.7	-368.1	217.4	0.183
	[1.0]	[0.3]	[0.4]	
Wage inc. 5000+	0.022	-0.020	0.073	0.409
	[0.663]	[0.820]	[0.227]	
Non-wage inc. 50+	0.044	0.094	0.025	0.574
	[0.430]	[0.362]	[0.759]	
Class Report wage index	31.1	130.2	-17.6	0.017
	[0.3]	[0.0]	[0.5]	
N	2361	685	1587	

Coefficients on peer neighborhood price rank from regressions of form given in equation 2. The first three columns denote samples. Rows are outcome variables. All specifications include randomization block and dummies for large feeder high schools; see section 5.1 for details."Test" column reports the p-value from a test of the null that the coefficients reported in the private and non-private columns are equal. Panels A and B report adult social club and occupation outcomes, respectively. "Adult association index" and "Occupation index" are standardized indices of the association between adult outcomes and private high school background. See section 3.4.3. Other variables are indicators for outcomes of the listed type. The sample in Panels A and B is students matched class reports; Panel B further restricts to students with occupation outcomes. Panel C reports labor market outcomes from the 1940 census. Sample is students matched to Census wage records in the 1920-30 entering cohorts. P-values are presented in brackets below each coefficient. Two-sided p-values computed using randomization inference with 2000 re-randomizations of roommate groups to rooms within blocks. Statistic reported is $\frac{1}{N}\sum_{j=1}^{N} 1[|\ \hat{\beta}_1^j\ | > |\ \hat{\beta}_1\ |]$, where $N = 2000$ is the number of rerandomizations, $\hat{\beta}_1^j$ is the estimate of β_1 in the jth randomization, and $\hat{\beta}_1$ is the estimate of β_1 observed in the data.

Table A.11: Harvard sample compared to similarly aged men

	Men 27-37	Men 27-37 w col. 1+	Men 27-37 w col. 4+	Harvard	Private feeder
A. Census: individual					
Yrs. of col. 4+	0.072	0.532	1.000	0.851	0.834
Yrs. of col. 5+	0.024	0.176	0.332	0.339	0.280
In school	0.012	0.037	0.040	0.030	0.026
Cen. Occ.: Doc.	0.007	0.048	0.086	0.073	0.052
Cen. Occ.: Law.	0.006	0.041	0.070	0.124	0.094
In labor force	0.949	0.962	0.963	0.953	0.951
Non-farm self emp.	0.084	0.160	0.191	0.207	0.195
Farm	0.191	0.064	0.042	0.038	0.075
Non-wage inc.50+	0.253	0.301	0.331	0.503	0.647
Has wage income	0.758	0.774	0.757	0.726	0.709
Wage income	1167	1841	2043	2747	2958
Wage inc. 5000+	0.008	0.034	0.047	0.175	0.247
Household head	0.692	0.704	0.699	0.718	0.775
Own home	0.188	0.209	0.201	0.210	0.272
Home value	2719	5296	6284	13612	19062
Monthly rent	60.1	89.6	99.2	103.6	109.5
B. Census: enumeration district					
Central city	0.193	0.226	0.245	0.514	0.546
Dist. share farm	0.200	0.078	0.058	0.039	0.063
Dist. share non-farm self emp.	0.106	0.146	0.151	0.162	0.159
Dist. share non-wage inc. 50+	0.304	0.290	0.287	0.325	0.357
Dist. 50p wage income	1050	1408	1492	1800	1732
Dist. share wage 5000+	0.014	0.036	0.045	0.109	0.122
Dist. 50p home value	2946	4375	4865	8538	9856
Dist. share college 4+	0.060	0.128	0.151	0.234	0.258
Dist. share Harvard	0.001	0.002	0.003	0.045	0.062
Dist. N men 27-37	129.3	135.1	135.8	114.9	110.4
N	11160105	1479483	787453	5623	1377

Demographic and labor market comparison of Harvard students to the broader Census population of similarly-aged men. Panel A: individual Census records. Panel B: attributes of Census enumeration districts. Columns are samples. Men 27-37: all Census men of listed age. Men 27-37 w col. 1+: men with at least one year of college. Men 27-37 w col. 4+: men with at least four years of college. Harvard: all Harvard students. Private feeder: Harvard students from private feeder schools. Degree attainment is not available in the 1940, so years of schooling serves as a proxy. Mean wage income and share with top-coded wages are conditional on reporting positive wage income. Home ownership, home value, and monthly rent are conditional on the individual being the head of their household.

B Data construction

B.1 Data construction details

This appendix describes data sources and match procedures we use to construct the Harvard archival dataset. Unless noted, all materials were accessed through the Harvard Archives.

B.1.1 Freshman Red Book / Freshman Register

The Freshman Red Books present information about each Harvard class's freshman year in a manner analogous to a yearbook. For the Harvard classes of 1913 through 1952, Freshman Red Books were published at the end of the spring semester of a class's freshmen year.[B.1] In 1953 the Freshmen Red Book was renamed the Freshman Register and publication was moved to the beginning of the fall semester. Starting in 1962, the Freshmen Registers no longer contain campus address or activities. Note that the Freshman Registers are distinct from the Student Council Registers and Official Registers.

In the Red Books, there is a "Class Directory" section with entries for each student containing the student's picture, name, home address, college address, high school, age, and activities the student participated in during his freshman year. See Figure B.1 for an example. Following the student records, the Red Books contain club and activity pages which list freshman membership in campus activities. See Figure B.2 for a club page example. We use the Red Books as our primary source for constructing cohorts of entering Harvard classes and for information on high schools, campus addresses, home addresses, and freshman activities. We have this information for all cohorts in our sample – the entering classes of 1919 through 1935 – with the exception of the entering class of 1926 (the graduating class of 1930) for whom we do not have college addresses because they were not included in the class of 1930 Red Book. We digitized the Red Books using hand transcription.

We clean student home address records into standard address format. We process student campus addresses by cleaning them and then matching to college dorm records so that, in addition to knowing the exact street address or dormitory and room number, we also categorize whether students were living at home, at another address off campus, in a dorm primarily housing freshmen in that year, or in a dorm primarily housing upper classmen or graduate students in that year. The small share of records where students are housed in non-freshman dorms are primarily due to demand for freshman dorms exceeding capacity or the student sharing a room with an older brother.

[B.1] The classes of 1940-1945 have a fall Freshmen Red Book in addition to the standard spring Red Book. The fall Red Books do not contain student activity information.

We categorize student high school records as follows. First we clean and standardize the free text entries by constructing a key that maps regular expressions description of common string patterns to standardized high school names. We do this for the 67 high schools listed identically by at least 15 students. We supplement this list with 46 high schools needed to distinguish records of the most common schools (e.g., we standardize reports of attending St. Paul's, Garden City, N. Y. in order to distinguish its graduates from students who attended St. Paul's School in New Hampshire). We categorize all standardized school names as either public or private. Among the private schools, we also flag the most elite New England boarding schools as "private feeder" schools. See Table B.1 for a complete list of school codes. We use that key to assign a high school code to each matching student record. 9813 students in our sample attended one of the schools we standardized. 4342 students report a high school other than our set of standardized schools. The majority of these students attended a school that only sent a handful of students to Harvard across our cohorts (e.g., Zanesville High). We include these students as non-private feeder, non-public feeder, and non-private school students.

As discussed in Section 3.4.1, the choice of Andover, Exeter, Groton, Middlesex, Milton, St. George's, St. Paul's, and St. Mark's as our private feeder school population was driven by the high shares of incoming Harvard students from each school and by historical accounts which describe these schools as a signature of high social status. In addition, these schools were characterized by high tuition, making them only accessible to wealthy families. For example, tuition and expenses at Andover and Groton in 1920 ranged from $500 - $1,200 (Levine, 1980), a significant cost compared to average incomes estimated at $1,283 in 1919 (Bowden, 1946) and Harvard tuition and expenses estimated at minimum $650 (Official Register, 1921). We discuss alternative private school classifications in sections 2 and 5.7 of the main text.

Finally, we process student activities records as follows. We clean and standardize activities listed in free text entries by constructing a key that maps regular expressions descriptions of identifiable string patterns to standardized activity names (identifiers). We group activities into a straightforward taxonomy under the guiding principle that activities whose participants may have interacted in a shared activity setting are grouped in the same category. See Table B.2 for the complete code list. Our categorization was exhaustive, meaning that all activites were categorized except for a few rare cases (e.g., activities that were only listed by a couple of students). We also flag if students were described as having a leadership role (e.g., manager, captain, president). Students may participate in multiple activities.

B.1.2 Student Council Registers

To supplement the freshman student activity participation information we extract from the Red Books, we also digitize the "Registered Clubs" chapter of the Harvard University Registers published by the Student Council of Harvard College. We refer to these as the Student Council Reg-

isters to avoid confusion with the Freshman Registers (a later name for the Red Books) and the Harvard University Official Registers published by the University (described below under 'Official Registers'). We obtained Student Council Registers for school years 1919-1920 through 1923-1924 from the Hathi Trust and Google Books and accessed physical copies of Student Council Registers for school years 1924-1925, 1925-1926, and 1927-1928 in the Harvard Archives.[B.2] Each club record in Student Council Registers lists club officers and members by name and class year. See Figure B.3 for an example. From the "Registered Clubs" section, we only extract the clubs from the "Miscellaneous – Social" subsection which contains records for fraternities and final clubs.[B.3] Club and fraternity membership records extracted from the Student Council Registers lists students in all classes in contrast to the activity membership from the Red Books which only lists Freshmen activity participation. As the typical final club does not admit members until at least sophomore year, we need these additional records in order to track final club membership.

Membership records were digitized by the firm Suntec Digital whose work we then audited for quality. We categorize the clubs into three categories – fraternities, most elite final clubs, and other final clubs. To distinguish the most elite final clubs, we relied on contemporary records naming the most inclusive clubs such as Amory's The Proper Bostonian. See main text for details.

B.1.3 Class Albums

The Harvard Class Albums edited by the Senior Album Committee are yearbooks produced at the time of graduation. In the Class Albums, each student entry contains the student's photo, name, and brief description of the student's background and time at Harvard. Figure B.4 presents an example. We use these records to extend our series of final club membership from 1928 to 1935 following the end of the Student Council Register series.

B.1.4 Rank Lists

Rank Lists report class rank based on course grades discretized into six groups.[B.4] These lists were published for freshmen, sophomore, and junior students from 1921-1932 and for freshmen

[B.2] As best we and the librarians at the Harvard Archives can tell, Student Council Registers were either discontinued or not preserved following 1928.

[B.3] It is somewhat surprising and fortuitous for us that final clubs are included in the "Registered Clubs" section as they were not, in fact, formally recognized by the administration of the University. Amory (1947) writes that "[f]or the most part official Harvard prefers to ignore the clubs in the same manner that the columns of the Crimson, undergraduate newspaper, take no notice of their existence. A Harvard professor recently declared that having been a faithful attendance at faculty meetings for twenty-five years he has yet to hear the clubs brought up for discussion."

[B.4] The rank lists were distributed to distinguish academic achievement with more granularity - especially in the middle of the distribution - than other previously available measures such as probation and receiving academic prizes (Annual Report of 1920-21). Lists were distributed publicly to students and their families.

only begining in 1933.[B.5] Students are divided into six groups "according to the average of the grades attained" (Annual Report of 1921-22). See Figure B.5 for the description of the ranks and which students were excluded from the rank list. Students' rank is reported alongside their names and class. See Figure B.6 for an example. Rank Lists were hand transcribed with quality audits.

B.1.5 Official Register

The Official Registers of Harvard University were a means by which the administration communicated information to entering freshmen. We refer to these as 'Official Registers' to prevent confusion with the Student Council Registers and the Freshman Registers. In these Official Registers, we find descriptions of how freshman students are assigned to dorms. Official Registers from 1922 through 1944 contain near identical language describing how students should submit an application indicating the price range they are willing to pay, if they have special financial hardship, and an optional choice of roommate. The text further describes that, with the exception of trying to reserve the very cheapest rooms for the poorest students and an attempt to make sure that students from the same high school are somewhat spread out, students are assigned to rooms "by lot" within the requested price range. Figure B.7 provides an example of this text.

In addition, we use floor plans and room price menus from the Official Registers of 1920 and 1932. We use the floor plans in order to inform our definition of 'neighborhoods' by the architectural features of the dorm. Price menus allow us to stratify rooms by their per-student price. As we do not have price lists available for every year, we rely on the prices from 1920 and 1932. In 1931 freshmen were relocated from the dorms by the river to dorms in the yard, so we rely on the 1920 records for information on dorms occupied by entering classes 1919-1930 and on the 1932 records for information on dorms occupied by entering classes 1931-1935. See Figures B.8 and B.9 for examples of price lists and floor plans. We describe the way we use these data to identify neighborhoods in section B.6

B.1.6 Miscellaneous Student Housing Records

In the archived records of the Bursar, we located a blank copy of the housing application form entering freshman students were asked to submit in 1922. Figure B.10 displays the application. Students could fill out the form individually or in pairs if they were requesting to live together.

[B.5] In 1933, publication of rank lists for Sophomores and Juniors was discontinued due to concern that producing a metric of course grades alone distorted upperclass students' incentives towards focusing on course grades when departments were defining students' academic performance more broadly to also encompass students' tutorial work, which they preferred to leave ungraded. Freshmen had no tutorial duties, so course grades were considered an appropriate measure and the publication of Freshmen Rank Lists continued (Annual Report of 1933-1934, p. 107-108).

B.1.7 Class Reports

The 25th Reunion Class Reports were compiled by Harvard twenty-five years following graduation. Nearly all graduates self-reported the information for the Class Reports. Of the minority of graduates who did not respond, Harvard entered the information it had on file (such as last known address or date and location of death). For graduating classes of 1923 through 1937, these records usually include the graduates' home address, office address, birth place, parents' names, high school, years at Harvard, marriages and children, occupation, offices held, and clubs and associations of which the graduate is a member. This more standardized information is followed by narrative text usually containing some combination of Harvard memories and updates on the graduate's life since graduation. Figure B.11 shows an example of a Class Report record. Starting in with the class of 1938 (Red Book year 1934), some previously common fields were frequently omitted, including those relating to parents and place of birth.

We digitize the class reports by photographing the pages with a high resolution camera, running an OCR program to collect all text from the images, and parsing the raw text file to extract standardized information from each record (e.g., the text following "PREPARED AT:" was assigned as a student's high school). A team of BYU research assistants went over each of the OCRed records to fill in missing names and match the individuals to Census records on FamilySearch. Observations that were incomplete or indecipherable due to imperfect OCR were sent to the firm SunTec Digital to be entered by hand.

We categorize occupations listed in the Class Reports by searching for key words associated with occupation categories. Our occupation categories are not mutually exclusive. We do not distinguish between occupation and industry. For example, an indiviual who reports "Lawyer, secretary and resident legal counsel, Brookhav National Laboratory, Associated Universities, Inc" would be categorized in both the "law" category and the "academics/research" category. Starting with common identifying words, we continue to add identifiers until all reported occupations are assigned to at least one category. Table B.3 provides the full list of codes. When we construct private school indices, we include all categories with at least 100 students as potential predictors. This amounts to adding the science and engineering categories to the eleven we discuss in the main text.

We categorize associations and clubs listed in the Class Reports as follows. We clean and standardize clubs and associations listed in free text entries by constructing a key that maps regular expressions descriptions of identfitiable string patterns to standardized club and association names (identifiers). We categorize clubs and associations as social (including subcategories for country clubs, gentlemen's clubs, and fraternal orders), professional, or honorary/political. Table B.4 provides the full list of club codes. We categorize any club listed identically by at least five students across the Class Reports, excluding Harvard-specific clubs.

B.2 Record linking procedures

B.2.1 Linking red books and class reports

We linked student records from the Red Books to records from the Class Reports via the following process:

1. First, last, and full (combined first, last, and possibly middle) names, years, and high schools are extracted for each person in both the red books and the class reports.

2. Red book-class report pairs with similar first OR last names (within 3 or 2 positions in an alphabetically sorted list, respectively) AND similar years (± 2) are identified as candidate matches.

3. The following features are assembled for each candidate match:

   - A first name score that is the Jaro-Winkler distance between the candidates' first names
   - A last name score that is the Jaro-Winkler distance between the candidates' last names
   - A full name score that is the Jaro-Winkler distance between the candidates' full names
   - A year score based on the difference between the birth date in the class report and the estimated birth date from the red book.
   - A high school score that is a normalized Levenshtein distance between the candidates' listed high schools

4. Matches with very bad first name or last name scores are removed.

5. The features for the remaining candidate matches are plugged into a random forest classifier (50 trees, max depth 5, class weights normalized to be balanced), which outputs a match score for each candidate match. These scores are the proportion of trees in the random forest that classify a given candidate pair as a match. The random forest classifier was trained on a data set 690 candidate matches which were marked as a match only if a research assistant was highly confident that the records were for the same person. By being trained on this data set, the classifier learns to estimate the probability that a human would confidently say that the two records belong to the same person.

6. Candidate matches are retained if their score exceeds a threshold (0.4). If a class report entry has more than one Red Book match, the match with the highest score is retained and all others are dropped. It is possible that a given red book entry may still have multiple potential matches indicated in the output file, but this is rare, and the match with a higher score should be preferred.

We assess the performance of links between Red Books and Class Reports using k-fold cross-validation on the manually labeled training data, with k set to 10. We split the data into k folds, then withold one fold at a time and re-fit the model on the other k-1 folds. We compute accuracy, precision, and recall on the witheld fold. Figure B.12 displays precision-recall curves generated from this analysis. At the threshold currently used (0.4) the average precision (percent of candidates predicted to be matches that are actually matches) is about 0.95, and the average recall (percent of true matches correctly predicted to be matches) is essentially 1.0.

B.2.2 Linking clubs and honor lists to student records within the Red Books

1. First and last names are extracted from the clubs lists and the main Red Book data. First names are transformed to just first initials because the clubs lists often only include first initials.

2. For each entry in the clubs lists:

   (a) Candidate matches are selected from the set of those with exactly matching years.

   (b) Candidate matches are filtered down to those with exactly matching last names. The match with the closest first name is kept.

   (c) If there are no exactly matching last names, candidate matches are reduced to those with exactly matching first names, and Jaro Winkler scores are computed for the last names of each candidate match. The candidate with the highest score above a threshold (0.85) is kept; if no candidates meet that threshold, there will be no match.

We evaluate match quality by calculating precision on a random sample of 200 records. See section B.2.4 for a description of this process and Table B.5 for results. We find precision exceeding 98% on our quality check sample.

B.2.3 Linking of class rank data to red books and class reports

Entries in the class ranks data with missing information or some invalid value (for example, some entries do not report class years) are dropped.

1. First and last names are extracted from class rank, red book and class report data. First and last initials are also extracted.

2. For each entry in the class rank data:

   (a) Entries missing either a first or a last name are dropped (no match)

(b) Look for matches with exact last name match and exact first initial (or full name if available) match in the class report for the same year. If more than one match is found, it is ambiguous which one is correct, so we drop the record (no match)

(c) If still no match found, repeat the above looking in surrounding years (± 1).

(d) If still no match found, try again in the original year, but allow inexact first and last name matches (within a damerau-levenshtein distance of 2, i.e. 2 or fewer string edits away)

(e) Repeat the above 4 steps, this time looking for matches in red book data.

3. We assign each candidate a rough confidence score which we use to select a single match for entries with multiple candidate matches. These scores were constructed to reflect the amount and quality of information used to form the match. Looking for inexact matches or matches in surrounding years decreases the confidence score, and making use of the full name (rather than just initials) increases the confidence score. The confidence score will also be higher if a match is found both in the Red Books and the Class Reports and those matches agree (as per the Red Book-class report linking described above). If a given Class Report entry or Red Book entry is given as a match to more than one class ranking entry, the match with a lower confidence score is dropped.

We evaluate match quality by calculating precision on a random sample of 200 records. See section B.2.4 for a description of this process and Table B.5 for results. We find precision of 100% on our quality check sample.

B.2.4 Report linking accuracy for class rank, Student Council Registers, and Class Albums

When we link records such as class rank data from the Rank Lists or social club members from the Student Council Registers, we only have name and year to rely on. This means that the random forest approach we used to link the Red Books and Class Reports is not feasible given the paucity of features and that we have many ambiguous candidate matches. We err on the side of false negatives rather than false positives when designing the matching schemes. Manually checking a small random subsamples of 200 class rank records, 200 Student Council Register records, and 200 Senior Album records, we report precision for matches with Red Books and Class Reports in Table B.5. All quality checks yield precisions exceeding 98%.

B.2.5 Linking class reports and Census records

The approach we use to link the Harvard Class Reports to US census records uses the same types of family history tools used by Costa et al. (2020) to link the children of Union army soldiers

to census records. They created profiles for each family on Ancestry.com and then used the record hints that are generated by the machine learning algorithms employed by Ancestry to find additional records for these individuals. The hinting algorithm used by Ancestry is described in Folkman et al. (2018) and is similar to the types of supervised machine learning algorithms described in Feigenbaum (2018b) and Abramitzky et al. (2020). The search tools on Ancestry also make to possible to look for potential records by widening the search parameters to include a broader set of possible matches based on similarity of name, birth year, and birth place.

We chose to use FamilySearch as our genealogical platform for this project because it provides a wiki-style platform in which 12 million registered users contribute to a shared Family Tree that includes the profiles of over 1.2 billion deceased individuals. The Family Tree has very good coverage for the United States and we were able to identify a profile for 59.9% of the Harvard students in our sample. FamilySearch provides an API that allows us to query the Family Tree for individuals based on their name, birth place, birth year, and names of family members. The Class Reports provides the information we need to do this query.

For the other 40.1% of our sample, we had research assistants create a profile for the individual on the Family Tree using the information from the class reports. While most people added to the Family Tree are added as a relative to someone already on the Family Tree, there is an option to add an "unconnected person" to the Family Tree. All profiles of deceased individuals on the Family Tree are publicly available and accessible through the API. The detailed records attached to each person in our sample, their vital information, and links to each of their family members can be accessed from the authors or via the FamilySearch API.

We use the FamilySearch API to check which records are already attached to each of the profiles that were previously on the Family Tree. Of those students who already had a profile, we found that 69.5% of the students in our sample were already attached to the 1940 census and 70.5% to at least one census prior to when they entered Harvard. We had research assistants search for the census records that were not already attached to the individual. FamilySearch has a built-in hinting algorithm and other search tools that allow trained research assistants to find census records. FamilySearch has a search feature that also makes it easy to check for hints on Ancestry as well, and we use that search tool to find potential matches.

B.3 Comparing occupation classifications and alternative wage measures

B.3.1 Occupations classifications in Class Reports vs. the Census

We use occupation categories constructed from the Class Reports as our primary measure of occupation. As described in section B.1.7, we construct these by applying a categorization scheme to the free text describing occupations in the Class Reports. For students linked to the 1940 Census, the occupations reported in the 1940 Census are available as an alternative measure of

occupation. These occupation measures differ in terms of data availability, the number of years post-college when occupation is observed, and the granularity of occupation category definition.

We consider our Class Report occupations to be preferable across each of these dimensions. Class Report occupations capture a student's occupation twenty-five years after graduation for cohorts 1920-1935 and are available for 78% of students in those cohorts, while the 1940 Census reports a students' occupation sixteen to six years after graduation for cohorts 1920-1930[B.6] respectively and is available for 60.1% of students in those cohorts. This advantages Class Report occupations in two ways. First, we observe Class Report occupation for twice as many students (10,752 versus 5,399). Second, we observe Class Report occupations at the same time in students' career trajectories and well into their working years (approximately age 47) while the 1940 census occupations provides a different early-career snapshot for each cohort.

Turning to differences in the granularity of occupation reports between the Class Reports and the Census, Table B.6 presents the modal Census occupation for four illustrative Class Report occupations: finance, senior management, medicine, and law. The Census provides a reasonable approximation of Class Report data for the professions. Medicine and law map closely to the Census occupation codes for, respectively, "Physicians and Surgeons" and "Lawyers and judges." Despite the ten to fifteen year gap in reporting most doctors in our data also show up as doctors in the Census (54%), and most lawyers in our data show up as lawyers in the Census (59%). The weaknesses of Census data are more pronounced for business-oriented careers. The modal occupation for students reporting both finance and senior manager positions is "Managers, officials, and proprietors, not elsewhere classified." The essential issue here is that the Census occupation categories do not offer a very detailed taxonomy of distinctions across elite business and management careers, which are highly relevant for Harvard students.

A corrollary to the point that Census occupation measures do not describe elite business careers very well is that measures of income based on Census occupations do not perform well in our setting. Some studies of historical Census data use occupation wage scores or occupation prestige scores from the 1950 Census to describe individual outcomes. Table B.6 compares the scores for the modal Census occupation for each Class Report occupation to observed earnings for Harvard students in the Class Report occupation who also report the modal Census occupation. What we see is that students reporting finance and senior management occupations in the Class Reports (and who the Census places in the "Managers, officials, and proprietors, not elsewhere classified" category) have very high incomes. The average value is this group is close to $3,700, and 40 percent have topcoded income. This is not captured by the occupation wage score for this group, which is middling– about half of the score for doctors and two-thirds the score

[B.6] As with the rest of the analysis with labor market outcomes in the 1940 Census, we do not include entering cohorts after 1931 as their labor force participation rates have not convereged by 1940 and we are systematically less likely to observe the profession and income of future doctors and other professions that require years of graduate preparation.

for lawyers– despite the fact observed earnings for students in finance and top management are higher than for either of the two professional careers. Harvard graduates with successful careers in business and finance are, unsurprisingly, quite rich. The Census occupation categories do not capture this.

Table B.7 shows that the limits of Census occupation categories are quantitatively important. It reports income information about the Census occupations most common in our sample (panel A) and for all men from the same birth cohorts in the 1940 Census (panel B). Twenty percent of Harvard students with a Census occupation are in the broad category "Managers, officials, and proprietors (not elsewhere classified)" and over half of students with a Census occupation are in one of the five most frequent occupations. The fact that a handful of occupations covers a large portion of the sample suggests that Census occupation-level measures such as wage-score and occupational prestige may poorly reflect between student variation in our sample by discarding all within Census occupation variation.

The final point to make here is that limitations of the Census are asymmetric across the groups Harvard students we are particularly interested in here: private feeder students and final club members. For example, private feeder students are more likely to be in finance and/or in an upper management position which is likely to show up in Census as "professional, technical, and kindred workers (not elsewhere classified)". Over a third of our Harvard students are in that or a similarly broadly-defined Census occupation. Comparing panels A and B of Table B.6, we can examine how well-represented students in our sample are by their Census occupations. While students from our sample in all occupations report higher earnings on average than the general population, the discrepencies are larger for broadly-defined occupations. For example, Harvard students with a Census occupation of "managers, officials, and proprietors (n.e.c.)" are six times as likely to have top-coded wage income as similarly-aged men on average. In contrast, Harvard students categorized as lawyers and judges are only twice as likely to earn top wage income and Harvard physicians and surgeons are equally likely. From these comparisions, we would expect occupation-level measures to perform particularly poorly for the more than a third of students in broadly-defined Census occupations, which would introduce bias unevenly by a key characteristic of interest: high school background.

In section 5.6 we analyze room randomization using an alternate wage index that addresses most of these limitations by using Class Report records of social and occupational outcomes in concert with Census reports of wage earnings. This amounts to computing predicted wage earnings based on coarse occupation and adult social assocations. This has two main advantages over Census-only wage indices. The first is that it allows us to exploit the fine-grained variation within elite business careers that shows up in the Class Reports but not the Census. The second is that we can construct this index for more than twice as many students as we can for Census wage indices– match rates to Class Reports are higher, and Class Reports are available for more

cohorts.

See section B.5.1 below for a discussion of how the use of alternate measures of income and wealth affects the analysis of Census outcomes in section 4.

B.4 Computing private school indices

This section presents more detail on the construction of private school indices for activities, occupations, and adult social categories first described in section 3.4.3. To summarize how changes in first-year activities, occupation, and adult club participation relate to integration across social groups, we collapse activities (and occupations, and adult social categories) into standardized linear indices capturing how characteristic the set of outcomes is of private school students. We construct these indices by regressing an indicator for private feeder school background on a set of indicators for participation in different first-year activities (occupations, adult social clubs) and cohort fixed effects. We compute predicted values, and then standardize (mean zero, standard deviation one) to facilitate cross-outcome comparisons. To avoid using own schooling background as an input to prediction, we estimate this specification for each cohort c^* using cohorts $c \neq c^*$. We use a Lasso to select coefficients and use EBIC for model selection. Table B.8 reports coefficients estimated using data from all cohorts, and Figure B.13 displays the distributions of predicted values.

B.5 Additional descriptive analyses

B.5.1 Alternate income definitions

One limitation of earnings outcomes in the 1940 Census is that they do not include unearned income. To understand how this may affect the analysis in section 4, we consider proxies for income as well as exercises that impute income for occupations where unearned income may be particularly important.

Figure B.14 presents a version of Figure 3 with alternate outcome measures. The first outcome measure (Panels A and C) is an alternative measure of wage income where we adjust the wage incomes reported in the 1940 census by imputing top wages ($5,000) for all doctors and lawyers. We present this measure to address concerns that law and medicine may underreport earned income relative to unearned business income, and that these occupations may be disproportionately represented among non-club members, and non-private feeder students. The second is alternative measure (Panels B and D) is predicted earnings based on home values. We construct this measure for students who report being head of their household and owning their own home by predicting earnings on cohort indicators and a third-degree polynomial of home values. We present the predicted income measure as a means of stepping back from the limitations imposed by top coding of wage income and only an indicator for non-wage income.

Among home-owners, housing value provides an alternative measure of economic well-being. One drawback from this approach is that home values may also reflect inherited wealth. Using these alternate measures does not affect the conclusions one draws from Figure 3.

B.5.2 Sample description for within-family analysis

This section presents a more detailed description of the sample used for the within-family analysis in section 4.4. Families that send multiple sons to Harvard tend to be wealthier than other Harvard families. Tables B.9 and B.10 present descriptive statistics for 1) our full sample of students, 2) the sample of students for whom data on wages and academic class rank is available (i.e., the sample used in our non-family descriptive analyses), 3) the sample of students in multi-brother families who also have wage and rank data (i.e., the sample used in our family fixed effect analysis), and 4) the subset of group 3 for whom one there is within-family variation in selective final club membership. We see that 38% of students in the brothers sample went to private feeder schools, compared to 24% in wages and ranks sample overall. Cross-group differences in legacy status, and immigrant status all similarly indicate that the brothers' sample is higher-SES than the full sample.

There are relatively few families where the brothers have different final club outcomes. 8% of brothers fit this description, or 23 in total from 11 families. The rightmost columns of Tables B.9 and B.10 show that this group is from higher-status backgrounds than brothers over all. For example, 78% went to private feeder schools. The small number of brother pairs who differ in terms of final club membership helps explain the imprecision of the within-family estimates presented in Panel D of Table 4. Note that students for whom final club membership does not vary within family help identify the cohort effects and grade effects.

B.5.3 OLS specifications with third-year grades

This section discusses how the descriptive analysis in section 4.4 changes when we use other measures of academic performance. Our analysis of academic achievement in the main text focuses on first year class rank, because these data are available for all cohorts of Harvard students in our data. This appendix repeats the analysis for third-year class rank. These are the most advanced students for whom we observe class rank. We only observe these data through calendar year 1930.

We first show that first-year and third-year class rank are strongly correlated. The upper panel of Figure B.15 plots the distribution of third year rank group by each value of first-year rank group. Within each first year rank group except the sixth, the modal third-year rank group is the same as the first year group. The lower panel of Figure B.15 plots the mean and interquartile range of the third year rank group by the first year group.

Figure B.16 presents a version of Figure 3 with third year class rank. Table B.11 presents an alternate version of Table 4 that uses third-year class rank rather than first year class rank, and also includes some additional outcome measures. The first four columns repeat Table 4 but using third year class rank. Sample sizes are slightly smaller because class rank is missing more often in the third year than the first, even in cohorts for which both measures are available for some students. This change in specification does not affect the conclusions we draw about the size of the final club membership earnings premium, which remains large across all specifications. The main difference we see here is that the class rank premium rises somewhat and becomes more precisely estimated. For example, in Panel C of Table B.11, it is equal to $103 (SE=$20), compared to a $49 estimate (SE=$20) in Panel C of Table 4. The larger effects for third-year grades are still small relative to the final club membership premium. Continuing to focus on Panel C of Table B.11, for example, we see that the return to a one-group increase in rank is equal to one-eigth the return to membership in a selective final club when the outcome is earned income (column 2) and about 1/16th when the outcome is topcoding (column 3). There are only six rank groups, so for both of these outcomes the predicted effect of a move from the bottom rank group to the top rank group is substantially smaller than the effect of final club membership.

We continue our analysis of alternate income variables in the rightmost columns of Table B.11. In column 5, we display specifications that impute maximum income values for doctors and lawyers, as above. The second shows specifications where in addition to imputing outcome for doctors and lawyers, we follow Goldin and Katz (2008) and multiply all topcoded values by 1.4. These adjustments tend to raise the return to class rank. The first procedure tends to lower the return to final club membership relative to the column 2 baseline, while the second procedure tends to raise it. Regardless of which procedure we use, the return to final club membership remains very large relative to the return to improved academic standing across all specifications. In section 6.4 of the main text we use estimates from columns 5 and 6 of Panel C in this table to compare our findings to Goldin and Katz (2008). Note that the standard deviation of class rank in the sample used Column 2 of Panel C is 1.26.

B.6 Defining residential peer groups

We define residential peer groups using archival floor plans for dorms housing freshman students. Our aim is to identify peer groups that form as a result of residential proximity, i.e., the loci for casual interactions between residential peers. To define a peer group, we apply two rules. The first rule is that for two rooms to be part of the same residential peer group, a student must be able to get from one to the other without going outside. Most Harvard dorms are organized by entryways (a group of rooms surrounding a stairwell that can be accessed by a common exterior door), so this rule means that residential peer groups are composed of students within the same stairwell in the same dorm. The next rule is that floors (or hallways) are important

social groupings. This follows Marmaros and Sacerdote (2006), who show that being members of the same hall more than doubles interactions among college students. Not all Harvard dorms have hallways– some entryways are vertically oriented and consist of one or two rooms around a staircase entrance on each floor– but, for those that do, we split peer groups by floor. Applying these two rules results in almost all of our peer groups being defined as either the entire entryway of a dorm or the floor within an entryway of a dorm. (Exceptions to this rule are for two dorms with layouts that do not use the entryway model, described below.)

For clarity, we present examples of residential living arrangments that give rise to each main type of peer group. Figure B.9 presents the floor plans of two dorms, Gore and Grays. Gore is composed of entryways A through E. Within each entryway, multiple rooms are connected by shared hallway space. For example, entryway B in the northwest (upper left) corner of the building contains rooms B21 through B25 on the second floor. These rooms are connected by a hallway around the stairs and past B24 and B25. We consider these groupings of rooms on a single floor within an entryway analogous to the floor (hallway) peer groups in Marmaros and Sacerdote (2006), and define residential peer groups in Gore and similar dorms as floors within entryways. Grays, on the other hand, has only two rooms per floor within an entryway, and no common space. Due to the absence of floor-level common space, we define residential peer groups in Grays and similar dorms at the entryway level.

Looking across all dorms, we find that Hollis, Holworthy, Lionel, Mower, Stroughton, Straus, and Wigglesworth have similar layouts to Grays and define peer groups at the entryway level. Massachusetts, Matthews, McKinlock, Smith, Standish, Thayer, and Weld are similar to Gore, and we define peer groups at the entryway by floor level. Finally, Harvard Union and Shepherd are unusual in that they do not have multiple entryways. We define peer groups in those dorms as the entire dorm for Harvard Union and each floor within the dorm for Shepherd.

To understand how peer effects change with group definition and group size, we consider two alternate classifications based on residential proximity. The first is the *entryway*. This is a large grouping; it includes entire dorms when they have a single entrance, as is the case for Shepherd. The second is a smaller grouping that always defines peer groups as floor within entryway, regardless of the dorm's architectural features. We call this group *nearest neighbors*.

Figure B.17 displays histograms of group sizes for our main peer neighborhood definition (Panel A), for nearest neighborhors (Panel B), and for entryways (Panel C). Nearest neighbor groups are weakly smaller than our main peer neighborhood definition. Entryways are dramatically larger. The median size of a peer neighborhood is 8. For nearest neighbors it is 7, and for entryways it is 28, with the largest entryway groups containing 61 students.

Given the large differences in group sizes, direct comparisons of the effect of neighborhood rank across different spatial units are not sensible, especially for the larger entryway groups. Cross-group variation in mean group price is 44% as large for the entryway group as in our

main measure. This means that a given change in rank within large groups corresponds to a much smaller change in mean price. For example, moving from the 10th percentile to the 90th percentile in the distribution of peer group means is associated with an $188 increase in mean price, averaging over years. For the larger entryway groups, this value is $126 . To make comparing magnitudes of regression estimates more straightforward, we change the units for our nearest-neighbor and entryway group definitions so that they correspond to where a group would place in the peer neighborhood rank in the same year. That is, we compute the mean price within a designated spatial unit, and then assign it a rank based on where it would fall in the peer neighborhood distribution by our main definition.

We use these modified rank measures to re-estimate equation 2 for alternate spatial groupings. *We view these specifications not as robustness tests on our main results, but as ways to explore how residential peer effects operate in larger and smaller groups.* This is for three reasons. First, the effects are conceptually different: social multiplier effects may depend on group size, as may students' ability to sort *within* spatial groups (Glaeser et al., 2003; Carrell et al., 2013). Second, we are less confident in the internal validity of the alternate measures as reflections of the constraints the lived environment imposes on social interaction than we are for our main measure. We view both as being worse approximations of the component of the peer environment that varies with residential room assignment than our main definition; they are either too inclusive or not inclusive enough. Third, for the larger entryway definition, the variation in peer environment induced by randomization is more limited than in our main definition.

We report our findings in Tables B.12 through B.15. Tables B.12 and B.13 use the larger peer group definition, while Tables B.14 and B.15 use the smaller definition. Our central finding is that we see comparable estimates of peer effects at our alternative levels of aggregation. We leave the interpretation of this similarity for future work.

B.7 Long-run time series

B.7.1 Construction

This section describes the construction of the 1924-1990 long-run series used in section 6.1, and discusses the representativeness and consistency over time of the data sources we use. We make five main points. First, changes in the content and timing of Red Book publications mean that we cannot track first-year social activities over the long run. Second, we observe high school background over the full time series. Third, we are able to classify student gender on the basis of their first and middle names over the full time period, and these classifications closely track published statistics in the HEGIS/IPEDS in years when both are available. Fourth, we are able to observe students' Latin honors at graduation over the full period, and the rates of honors graduation that we observe track published aggregate statistics. Fifth, our coverage of occupation

declines somewhat over time, but we see no evidence of differential attrition on the basis of high school type. Our overall assessment is that the long-run series we assemble provides reliable evidence on trends in academic and occupational attainment, and differences in these outcomes by high school type.

We construct the long run series by extending the two main sources of Harvard records used in the main analysis: Freshman Red Books and 25th Reunion Class Reports. See section 3 for an extended discussion of these sources. We digitize Red Book records through the entering class of 2019. We have Class Report records for every graduating class through 1940 and then at five year intervals starting in 1945, and continuing through the graduating class of 1990. At the time of digitization, 1990 was the last 5-year interval for which 25-year Class Report outcomes were available. As in our main analysis, we define our sample universe using the Red Book records. Panel A of Figure B.18 shows how Red Book cohort sizes evolve across class years. Each point is the mean cohort size within a five year window around the central value. Red Book cohort sizes closely match cohort sizes reported in HEGIS/IPEDS data when these series become available in the late 1960s.

Ideally, our long run analysis would have included data on first-year social activities. However, changes in the information included in the Red Books makes doing so impossible. Prior to 1940, Red Books were published in the spring, at the end of students' first year. These spring books are the main source for social activity data in our analysis of data from the 1920s and 1930s. Beginning in 1940, Harvard published Red Books in both the fall and the spring. The fall Red Book lists students high schools and hometowns, but not college activities, while the spring Red Books continue to list college activities but with decreasing detail. In 1950, the spring Red Books cease publication, leaving only the fall Red Books. The fall Red Book series continues through the present. Our analysis uses spring Red Books through the end of their publication in 1949, then switches to the fall books for the remaining years.

Women were admitted to Harvard as undergraduates starting in the early 1970s. This change is visible in the cohort size data; class size rises to partially accommodate the addition of female students. We identify the gender of Harvard students using data on name frequency by gender and birth year from the Social Security Administration (2021). We first search for a perfect match of the first name in the cohorts close to the expected birth year of a student. If we find a unique perfect match, we assign the corresponding gender. If we find multiple perfect matchs, i.e. the name has been given to both male and female babies, we proceed by searching for perfect matches on the middle name if recorded in our data. If there is unique perfect match for the middle name, we assign the corresponding gender. Otherwise, we stop and assign the gender for which the first name frequency is higher. If no perfect match can be found, we calculate the Damerau-Levenshtein distances to all candidate names. We require the set of likely matches to have a distance of less than three. We then assign the gender for which the total frequency among

these likely matches is the highest. Panel B of Figure B.18 shows that this procedure works well. The share of students we identify as male is almost exactly one prior to 1971; it declines slowly towards 50% but does not reach this point by 1990. Gender shares in our data are almost exactly equal to those reported in HEGIS/IPEDS data in the years records of both types are available.

Match rates from Red Book data to Class Report data remain steady over time. The upper line in Panel C of Figure B.18 shows the share of Red Book records linked to Class Report data. We match roughly 80% of Red Book records to Class Report data over the full time series.

The detail of Class Report records falls somewhat over time, particularly after 1980. The lower two lines in Panel C of Figure B.18 show the share of Red Book records a) linked to a Class Report record that includes the student's occupation, and b) linked to a Class Report record that includes the student's educational history. The rates at which students report occupation conditional on Class Report match fall from roughly 90% in our main analysis period to roughly 70% from 1940 through 1970 to roughly 53% from 1980 on. The rates at which students report their educational history are steady from the 1920s through the 1970s at roughly 72% before falling to roughly 58% from 1980 on.

Two features of the data mitigate concerns about reduction in data quality in the later part of the series. The first is that the measures of academic achievement that we observe in our data closely track those published in aggregate statistics. Panel B of Figure 11 in the main text compares the rate of honors degree receipt reported in our data to those reported in public reports on academic honors receipt in Harvard (Healy, 2001). These series track closely in all years where both are available. The sample of students for whom educational histories are available in Class Report data is representative of the broader Harvard population.

The rates of graduate degree receipt we observe are also similar to those reported in previously published work. The key degrees of interest for us are MDs, JDs, and MBAs, as displayed in Figure 12. Goldin and Katz (2008) report the rates at which Harvard students receive these degrees. Their focus is on Harvard graduating classes within four-year windows around 1970, 1980, and 1990. There are many reasons that rates in our data might differ from rates in Goldin and Katz (2008), including differences in the graduating class years we analyze and differences in the timing of survey reporting in the life course.[B.7] Nevertheless, our findings are quite similar. For MD degrees, GK report rates of degree receipt among male students in classes near 1970, 1980, and 1990 as 18.5%, 18.0%, and 14.3%, respectively. For us, these values are 17.1%, 19.5%, and 13.6%. For JD degrees, GK report values of 24.8%, 24.1%, and 20.4%. We report values of 22.9%, 21.6%, and 20.7%. For MBA degrees, GK report values of 10.9%, 20%, and 19.4%, while we report values of 9.4%, 18.4%, and 15.1%.

The second feature is that there is little evidence of differential change in match rates to oc-

[B.7] Goldin and Katz (2008) conduct all surveys in 2006; for us, all degree reports are from 25 years following graduation.

cupation records. Panel C of Figure B.18 display match rates to occupational data for students in the Red Books, splitting by high school type. Match rates are similar for private feeder and public feeder students over the full period. These are the two groups for whom we draw cross-time comparisons in section 6.1. The observation that coverage declines similarly for these two groups makes us less worried that changes in selection into the data over time could affect the cross-group comparisons we draw.

Our overall conclusion is that data we have provide a credible picture of long-run changes in academic and occupational outcomes by group.

B.7.2 Job classification in the extended time series

Overview. Classifying occupations over the long run creates additional challenges relative to our main classification scheme because the way Harvard graduates describe their jobs may change over time. We address this issue in four ways.

1. We focus on a narrower set of careers than in our main analysis. Specifically, we limit our long-run analysis to finance, higher education, medicine, and law.

2. For medicine and law, we use descriptors of graduate education that map closely to career choices and are easy to code consistently over time, rather than the text descriptors of job title or occupation.

3. For all four career types, we expand the set of job descriptors we consider to cover new ways of describing these careers that emerge over time. In addition, we introduce regular expression representations of the existing identifiers, to account for spelling mistakes more flexibly.

4. For all four career types, we consider alternate classification approaches and reach substantively similar conclusions to those presented in the main analysis.

These supplementary analyses reinforce the argument in the main text that differences in career outcomes by high school type persist over the the 20th century, by showing that the conclusion holds across a variety of measures of career outcomes.

Extended coding of careers. Table B.16 shows the set of strings we add to those we code as finance, legal, higher education and medical careers for the extended analysis. These strings tend to pick up job title or detailed occupation descriptions, for example "fixed income." We also consider alternate specifications for finance careers that add names of large investment banks (current and historical) to list of strings we consider, to pick up people who do not list an occupation but do list an employer. We do not use the firm names in the analysis in the main text. Table B.16 also displays the list of banks we consider in the row referring to the subcategory "Firms (ext. definition)".

Alternate classification approaches. We consider alternate classification approaches for each of our four career categories. For finance careers, we consider an augmented classification approach that adds names of major investment banks to our baseline scheme. Figure B.19 compares results from this exercise with those from our main analysis. Including firm names raises our estimate of the rate at which students pursue finance careers slightly from the late 1970s and onward. Our central conclusion that there is a large gap by high school type that converges by the end of the period is unchanged.

For medical and legal careers, we compare our main measures based on MD and JD/LLB receipt with measures based on reported job outcomes. Figure B.20 shows how the share of students reporting MD and JD/LLB degrees relates to the share of students reporting medical and legal careers in the sample of students who report both career and degree information in the Class Report. We do not expect these measures to be identical; degree and career outcomes need not correspond precisely. For medical degrees, measures of medical careers based on job description are slightly higher than those based on degree receipt prior to 1950; after 1950 the measures are nearly identical. For law careers, measures are near identical prior to 1970. After 1970, measured rates of law careers in job descriptions fall off relative to JD receipt. Inspection of records suggests that this reflects a combination of people who receive law degrees but work outside of the law field, and the greater difficulty in classifying workers who list only their firm and not occupation descriptors in the more recent Class Reports.

Figure B.21 reports long-run trends by high school type for the alternate measure of career attainment. The "doctor" panel shows the rates at which students' job description fall into the medical doctor/health career category given in Table B.3. While public high school students are more likely to pursue medical careers throughout the period, we see more evidence of convergence by high school type for this measure than for MD degrees. This may reflect a higher propensity for public feeder students with MD degrees to work outside of the medical field, for example in research positions. For law, our results are essentially the same as for the JD measure, with little evidence of a gap by high school type. For PhD degrees, we see a similar overall pattern as we did for higher education careers in the main text, with similar rates for private and public feeder students in the 1920s and 1930s, but a large gap by the end of the period. In the 1980s and 1990s, the gap in PhD receipt by high school type is larger than the gap we observe in higher education careers.

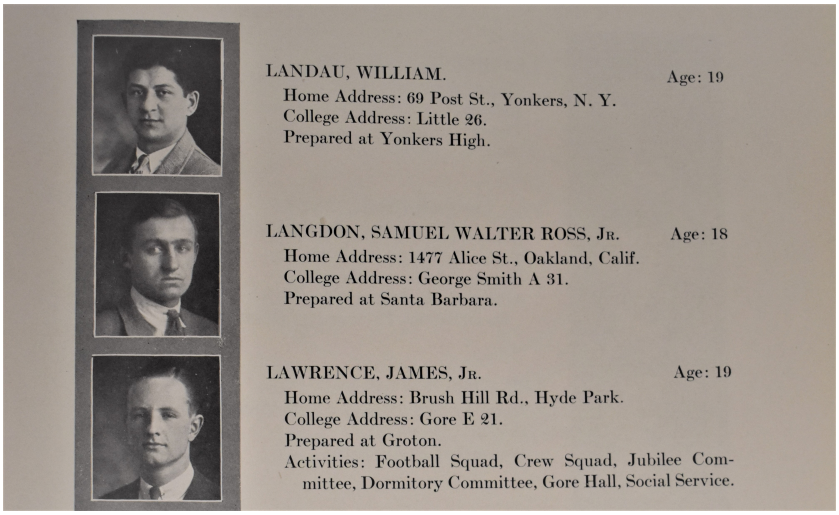
B.7.3 High school classification in the extended time series

Our long-run analysis extends the set of high schools in the "public feeder" and ''other private" categories. We add schools to these categories by examining the set of high schools listed at least ten times across in the cohorts 1925, 1930, 1935,..., 2010. We categorize each such school not already included in our main-sample categorization scheme. The only exceptions to this rule

are schools with common names, such as "Jesuit" or "Classical" that are difficult to consistently differentiate from one another. Through this process we add 18 public feeder schools and 72 private schools to the public feeder and other private categories. See Table B.17 for a list of these schools.

B.8 Figures

Figure B.1: Red book information on freshmen residence and activities



LANDAU, WILLIAM. Age: 19
Home Address: 69 Post St., Yonkers, N. Y.
College Address: Little 26.
Prepared at Yonkers High.

LANGDON, SAMUEL WALTER ROSS, JR. Age: 18
Home Address: 1477 Alice St., Oakland, Calif.
College Address: George Smith A 31.
Prepared at Santa Barbara.

LAWRENCE, JAMES, JR. Age: 19
Home Address: Brush Hill Rd., Hyde Park.
College Address: Gore E 21.
Prepared at Groton.
Activities: Football Squad, Crew Squad, Jubilee Committee, Dormitory Committee, Gore Hall, Social Service.

Missing dorm information for entering class of 1926. See Section B.1.1 for more details on the Red Books.

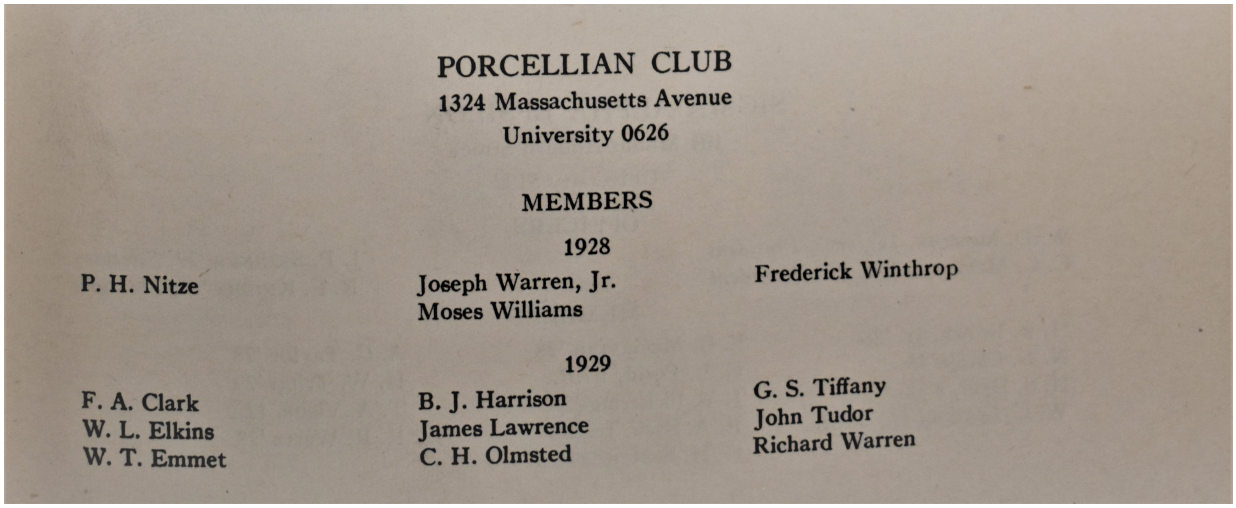
Figure B.2: Red book lists of freshmen activity participants

Crew O—The First Crew

Name	*Position*	*Age*	*Height*	*Weight*	*School*
J. Lawrence	B.	18	6′¾″	166	Groton
E. Hamlen	2	19	6′1″	173	Groton
W. T. Emmet	3	19	6′3¾″	176	St. Paul's
F. A. Clark	4	20	6′5″	198	St. Mark's
G. Murchie	5	19	6′4¾″	189	Kent
B. J. Harrison	6	19	6′4½″	186	St. Paul's
G. N. Saum	7	19	5′10½″	176	Watseka High
C. McK. Norton	S.	19	6′¼″	167	Groton
C. I. Neiman	Cox	17	5′	94	Boston Latin

See Section B.1.1 for more details on the Red Books.

Figure B.3: Student council register information including club membership



PORCELLIAN CLUB
1324 Massachusetts Avenue
University 0626

MEMBERS

1928

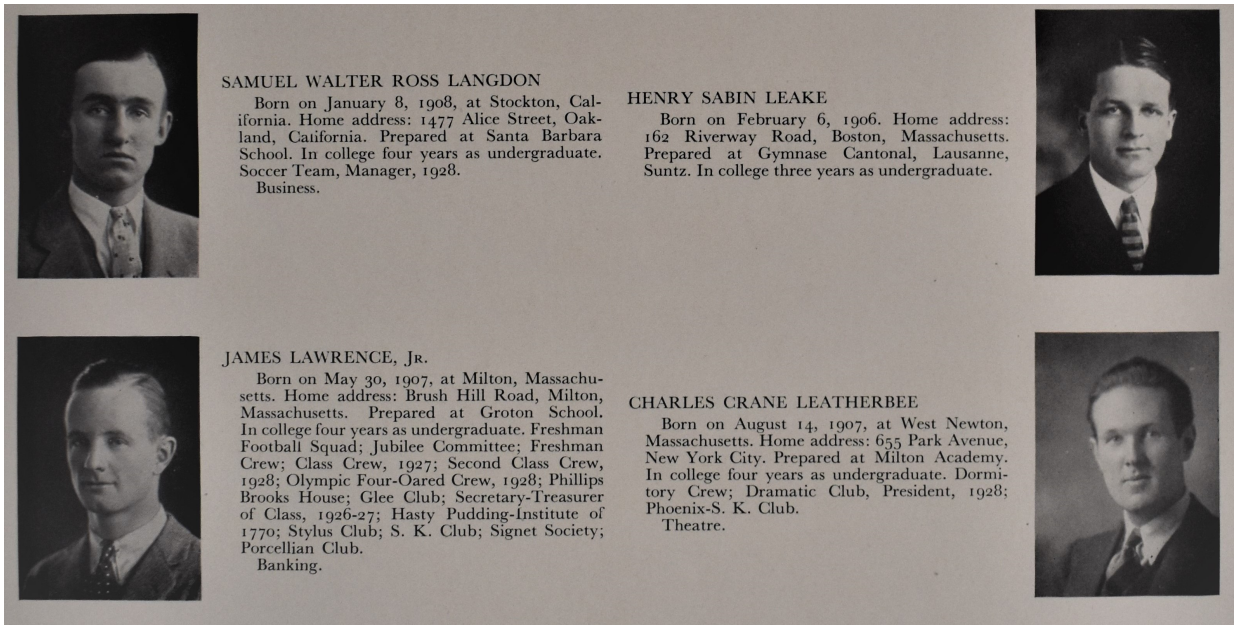
P. H. Nitze
Joseph Warren, Jr.
Moses Williams
Frederick Winthrop

1929

F. A. Clark
W. L. Elkins
W. T. Emmet
B. J. Harrison
James Lawrence
C. H. Olmsted
G. S. Tiffany
John Tudor
Richard Warren

Missing after school year 1927-28. See Section B.1.2 for more details on the Student Council Registers.

Figure B.4: Class albums, includes club membership



SAMUEL WALTER ROSS LANGDON

Born on January 8, 1908, at Stockton, California. Home address: 1477 Alice Street, Oakland, California. Prepared at Santa Barbara School. In college four years as undergraduate. Soccer Team, Manager, 1928.
Business.

HENRY SABIN LEAKE

Born on February 6, 1906. Home address: 162 Riverway Road, Boston, Massachusetts. Prepared at Gymnase Cantonal, Lausanne, Suntz. In college three years as undergraduate.

JAMES LAWRENCE, Jr.

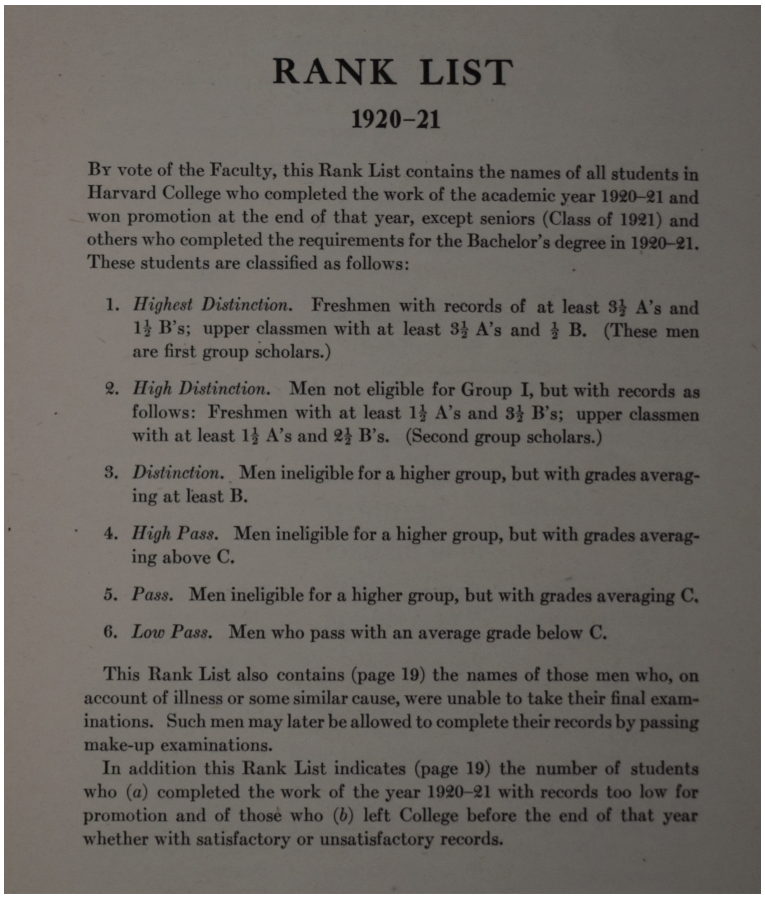
Born on May 30, 1907, at Milton, Massachusetts. Home address: Brush Hill Road, Milton, Massachusetts. Prepared at Groton School. In college four years as undergraduate. Freshman Football Squad; Jubilee Committee; Freshman Crew; Class Crew, 1927; Second Class Crew, 1928; Olympic Four-Oared Crew, 1928; Phillips Brooks House; Glee Club; Secretary-Treasurer of Class, 1926-27; Hasty Pudding-Institute of 1770; Stylus Club; S. K. Club; Signet Society; Porcellian Club.
Banking.

CHARLES CRANE LEATHERBEE

Born on August 14, 1907, at West Newton, Massachusetts. Home address: 655 Park Avenue, New York City. Prepared at Milton Academy. In college four years as undergraduate. Dormitory Crew; Dramatic Club, President, 1928; Phoenix-S. K. Club.
Theatre.

Missing for students who do not persist to graduation. Used to supplement Student Council Register information for years after the Register was discontinued. See Section B.1.3 for more details on the Class Albums.

Figure B.5: Rank group lists explanation of ranks and inclusion



RANK LIST

1920–21

By vote of the Faculty, this Rank List contains the names of all students in Harvard College who completed the work of the academic year 1920–21 and won promotion at the end of that year, except seniors (Class of 1921) and others who completed the requirements for the Bachelor's degree in 1920–21. These students are classified as follows:

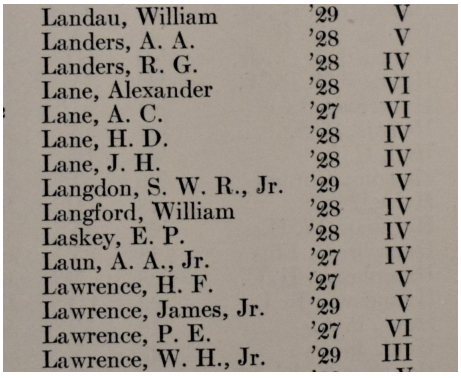
1. *Highest Distinction.* Freshmen with records of at least 3½ A's and 1½ B's; upper classmen with at least 3½ A's and ½ B. (These men are first group scholars.)
2. *High Distinction.* Men not eligible for Group I, but with records as follows: Freshmen with at least 1½ A's and 3½ B's; upper classmen with at least 1½ A's and 2½ B's. (Second group scholars.)
3. *Distinction.* Men ineligible for a higher group, but with grades averaging at least B.
4. *High Pass.* Men ineligible for a higher group, but with grades averaging above C.
5. *Pass.* Men ineligible for a higher group, but with grades averaging C.
6. *Low Pass.* Men who pass with an average grade below C.

This Rank List also contains (page 19) the names of those men who, on account of illness or some similar cause, were unable to take their final examinations. Such men may later be allowed to complete their records by passing make-up examinations.

In addition this Rank List indicates (page 19) the number of students who (*a*) completed the work of the year 1920–21 with records too low for promotion and of those who (*b*) left College before the end of that year whether with satisfactory or unsatisfactory records.

Available for all non-graduating students school years 1920-21 through 1930-31. Available for Freshmen only starting in school year 1931-32. See Section B.1.4 for more details on the rank group lists.

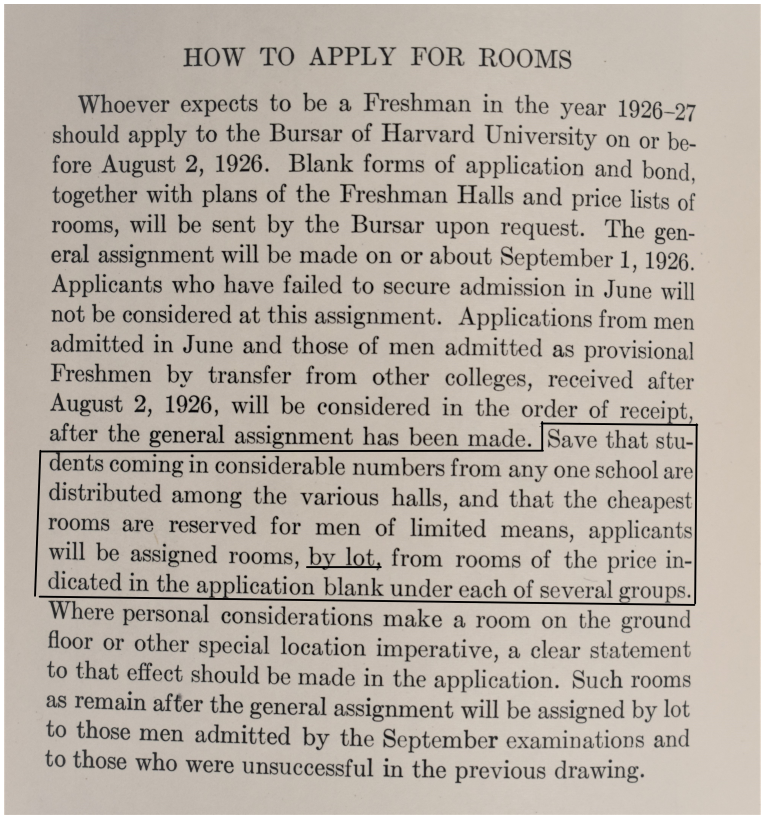
Figure B.6: Rank group lists



Landau, William	'29	V
Landers, A. A.	'28	V
Landers, R. G.	'28	IV
Lane, Alexander	'28	VI
Lane, A. C.	'27	VI
Lane, H. D.	'28	IV
Lane, J. H.	'28	IV
Langdon, S. W. R., Jr.	'29	V
Langford, William	'28	IV
Laskey, E. P.	'28	IV
Laun, A. A., Jr.	'27	IV
Lawrence, H. F.	'27	V
Lawrence, James, Jr.	'29	V
Lawrence, P. E.	'27	VI
Lawrence, W. H., Jr.	'29	III

Available for all non-graduating students school years 1920-21 through 1930-31. Available for Freshmen only starting in school year 1931-32. See Section B.1.4 for more details on the rank group lists.

Figure B.7: Room assignment procedure 1922-1941

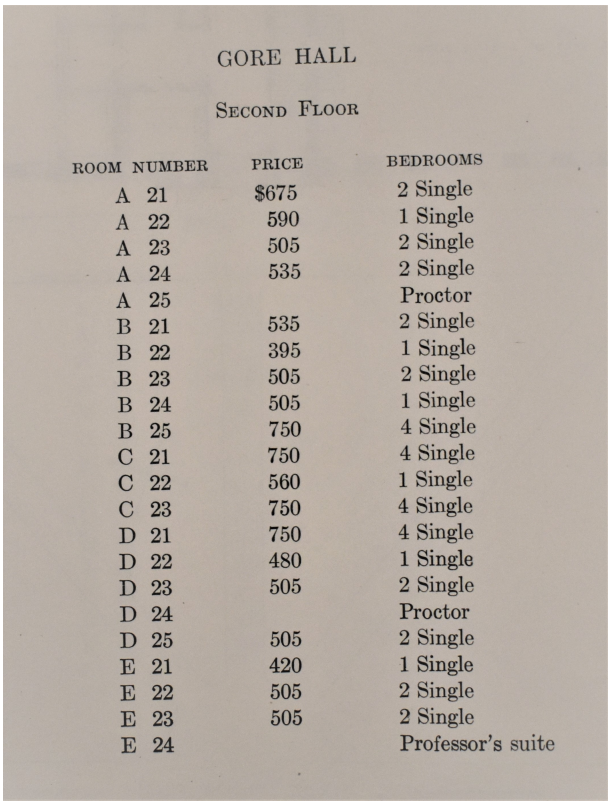


HOW TO APPLY FOR ROOMS

Whoever expects to be a Freshman in the year 1926–27 should apply to the Bursar of Harvard University on or before August 2, 1926. Blank forms of application and bond, together with plans of the Freshman Halls and price lists of rooms, will be sent by the Bursar upon request. The general assignment will be made on or about September 1, 1926. Applicants who have failed to secure admission in June will not be considered at this assignment. Applications from men admitted in June and those of men admitted as provisional Freshmen by transfer from other colleges, received after August 2, 1926, will be considered in the order of receipt, after the general assignment has been made. Save that students coming in considerable numbers from any one school are distributed among the various halls, and that the cheapest rooms are reserved for men of limited means, applicants will be assigned rooms, by lot, from rooms of the price indicated in the application blank under each of several groups. Where personal considerations make a room on the ground floor or other special location imperative, a clear statement to that effect should be made in the application. Such rooms as remain after the general assignment will be assigned by lot to those men admitted by the September examinations and to those who were unsuccessful in the previous drawing.

Emphasis added. Prior to 1922 and after 1944 students' ranked preferences were taken into account rooms assigned taking ranked preferences into account. Procedure from 1942-1944 unknown. See Section B.1.5 for more details on freshman housing.

Figure B.8: Price lists



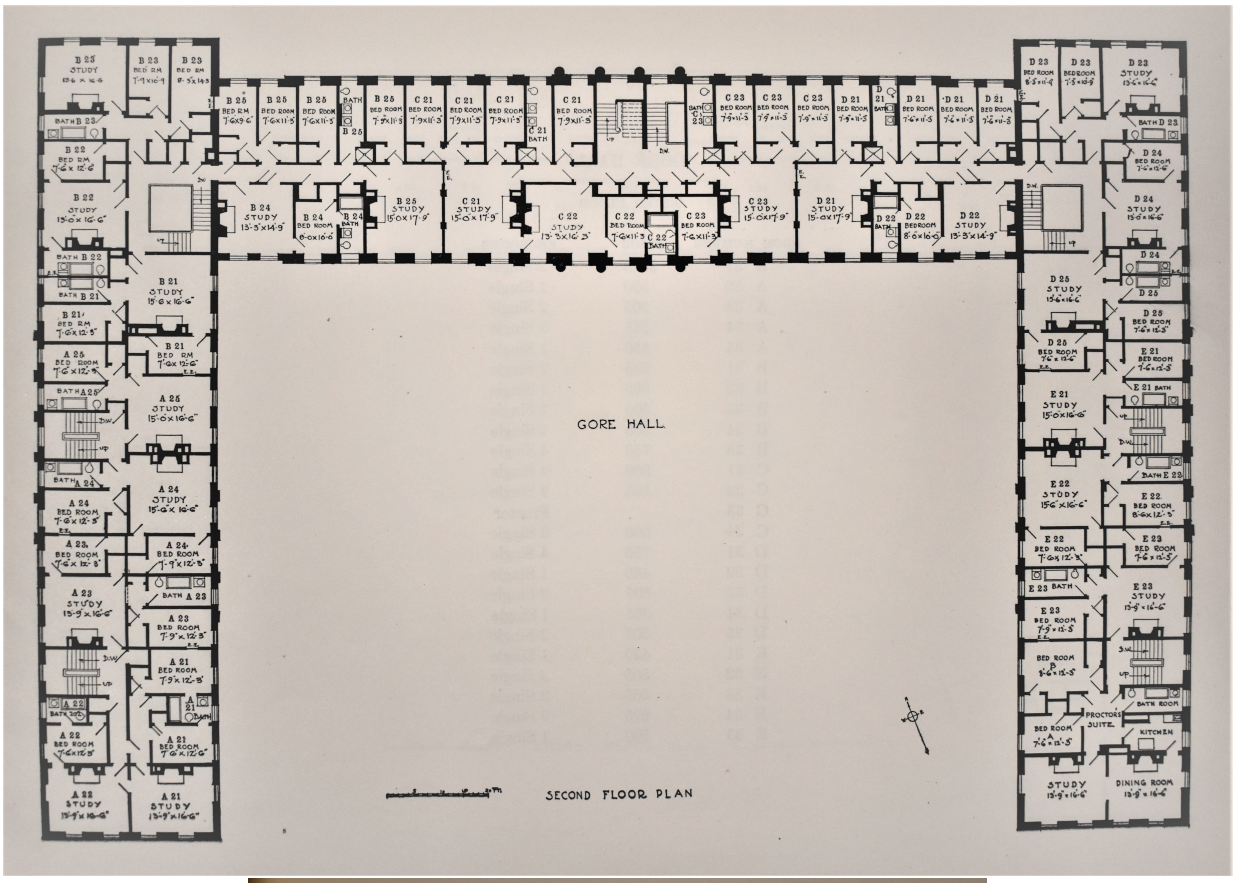
GORE HALL

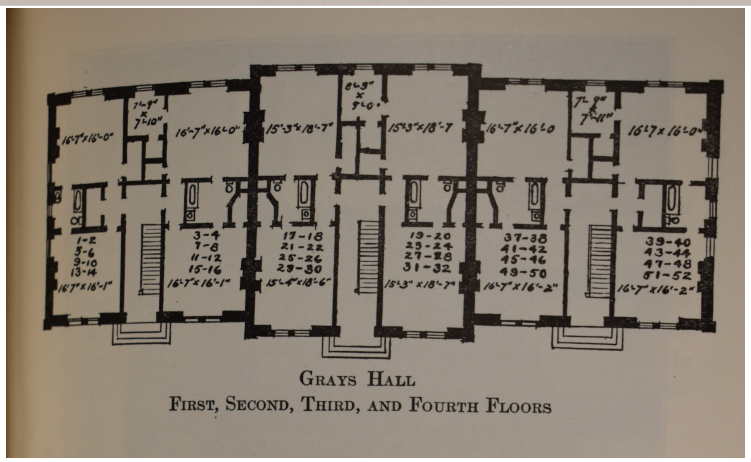
SECOND FLOOR

ROOM NUMBER	PRICE	BEDROOMS
A 21	$675	2 Single
A 22	590	1 Single
A 23	505	2 Single
A 24	535	2 Single
A 25		Proctor
B 21	535	2 Single
B 22	395	1 Single
B 23	505	2 Single
B 24	505	1 Single
B 25	750	4 Single
C 21	750	4 Single
C 22	560	1 Single
C 23	750	4 Single
D 21	750	4 Single
D 22	480	1 Single
D 23	505	2 Single
D 24		Proctor
D 25	505	2 Single
E 21	420	1 Single
E 22	505	2 Single
E 23	505	2 Single
E 24		Professor's suite

Available for 1920, 1932-1941. See Section B.1.5 for more details on freshman housing.

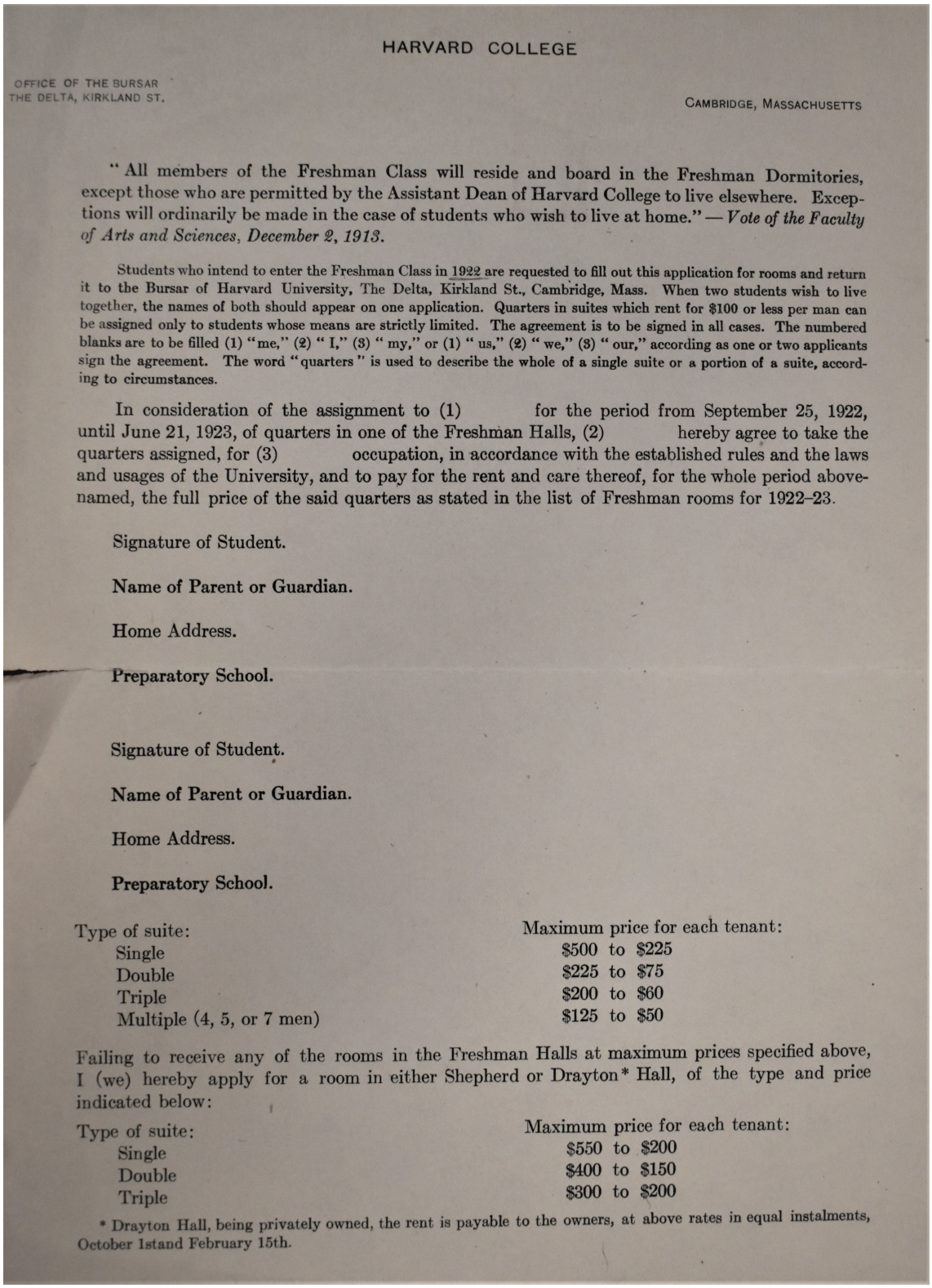
Figure B.9: Floor plan





Available for 1920, 1932-1941. The top image depicts Gore as an example of a dorm where the combination of floor and entryway (denoted by letter) is our preferred peer group definition. The bottom image depicts Grays as an example of a dorm where entryway is our preferred peer group definition. See Section B.1.5 for more details on freshman housing and Section B.6 for details on how we defined peer groups.

Figure B.10: Freshmen housing application blank



HARVARD COLLEGE

OFFICE OF THE BURSAR
THE DELTA, KIRKLAND ST.

CAMBRIDGE, MASSACHUSETTS

"All members of the Freshman Class will reside and board in the Freshman Dormitories, except those who are permitted by the Assistant Dean of Harvard College to live elsewhere. Exceptions will ordinarily be made in the case of students who wish to live at home." — *Vote of the Faculty of Arts and Sciences, December 2, 1913.*

Students who intend to enter the Freshman Class in 1922 are requested to fill out this application for rooms and return it to the Bursar of Harvard University, The Delta, Kirkland St., Cambridge, Mass. When two students wish to live together, the names of both should appear on one application. Quarters in suites which rent for $100 or less per man can be assigned only to students whose means are strictly limited. The agreement is to be signed in all cases. The numbered blanks are to be filled (1) "me," (2) "I," (3) "my," or (1) "us," (2) "we," (3) "our," according as one or two applicants sign the agreement. The word "quarters" is used to describe the whole of a single suite or a portion of a suite, according to circumstances.

In consideration of the assignment to (1) for the period from September 25, 1922, until June 21, 1923, of quarters in one of the Freshman Halls, (2) hereby agree to take the quarters assigned, for (3) occupation, in accordance with the established rules and the laws and usages of the University, and to pay for the rent and care thereof, for the whole period above-named, the full price of the said quarters as stated in the list of Freshman rooms for 1922–23.

Signature of Student.

Name of Parent or Guardian.

Home Address.

Preparatory School.

Signature of Student.

Name of Parent or Guardian.

Home Address.

Preparatory School.

Type of suite:	Maximum price for each tenant:
Single	$500 to $225
Double	$225 to $75
Triple	$200 to $60
Multiple (4, 5, or 7 men)	$125 to $50

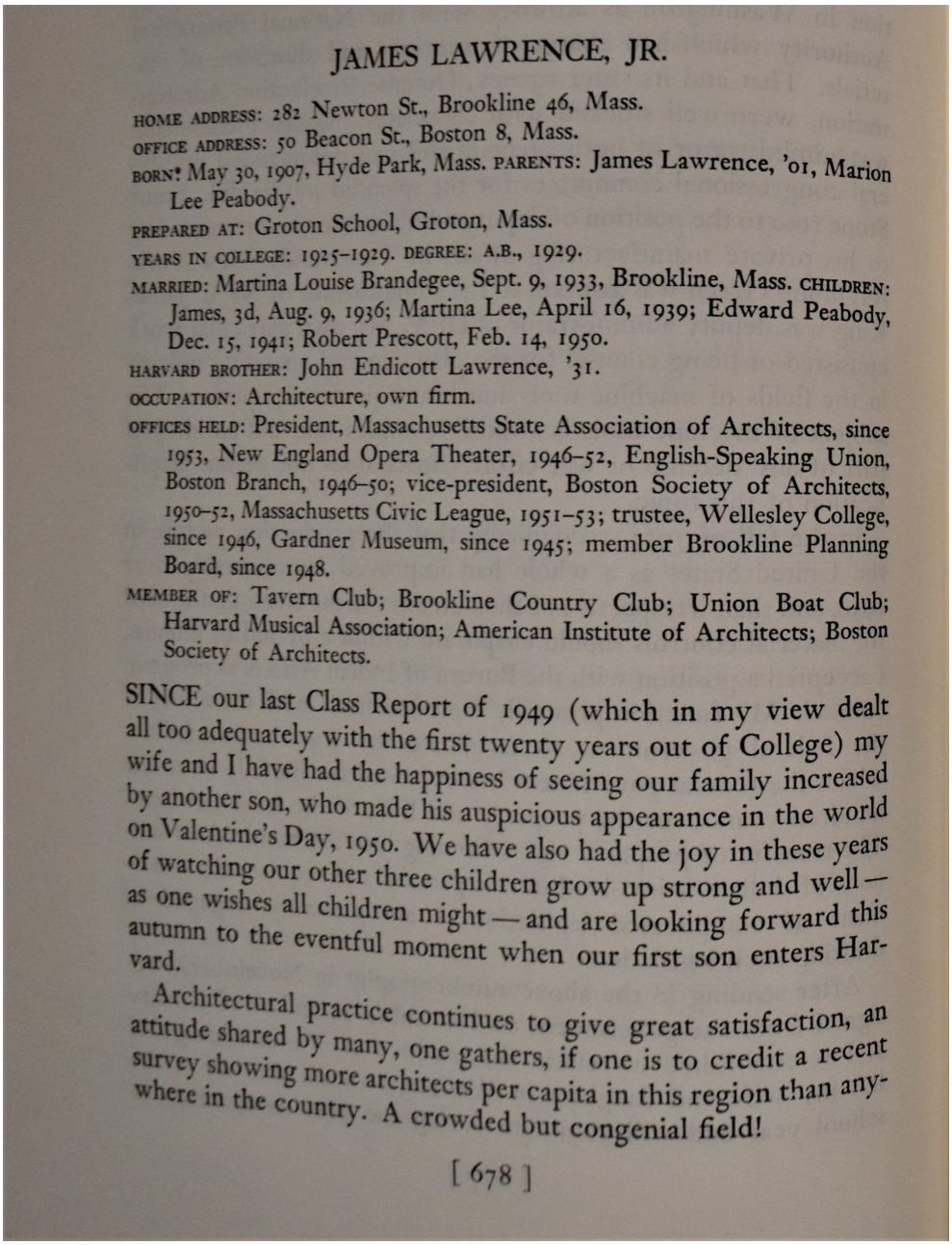
Failing to receive any of the rooms in the Freshman Halls at maximum prices specified above, I (we) hereby apply for a room in either Shepherd or Drayton* Hall, of the type and price indicated below:

Type of suite:	Maximum price for each tenant:
Single	$550 to $200
Double	$400 to $150
Triple	$300 to $200

* Drayton Hall, being privately owned, the rent is payable to the owners, at above rates in equal instalments, October 1stand February 15th.

See Sections B.1.5 and B.1.6 for more details on freshman housing and the housing application process.

Figure B.11: 25th Reunion Class Report



JAMES LAWRENCE, JR.

HOME ADDRESS: 282 Newton St., Brookline 46, Mass.
OFFICE ADDRESS: 50 Beacon St., Boston 8, Mass.
BORN: May 30, 1907, Hyde Park, Mass. PARENTS: James Lawrence, '01, Marion Lee Peabody.
PREPARED AT: Groton School, Groton, Mass.
YEARS IN COLLEGE: 1925–1929. DEGREE: A.B., 1929.
MARRIED: Martina Louise Brandegee, Sept. 9, 1933, Brookline, Mass. CHILDREN: James, 3d, Aug. 9, 1936; Martina Lee, April 16, 1939; Edward Peabody, Dec. 15, 1941; Robert Prescott, Feb. 14, 1950.
HARVARD BROTHER: John Endicott Lawrence, '31.
OCCUPATION: Architecture, own firm.
OFFICES HELD: President, Massachusetts State Association of Architects, since 1953, New England Opera Theater, 1946–52, English-Speaking Union, Boston Branch, 1946–50; vice-president, Boston Society of Architects, 1950–52, Massachusetts Civic League, 1951–53; trustee, Wellesley College, since 1946, Gardner Museum, since 1945; member Brookline Planning Board, since 1948.
MEMBER OF: Tavern Club; Brookline Country Club; Union Boat Club; Harvard Musical Association; American Institute of Architects; Boston Society of Architects.

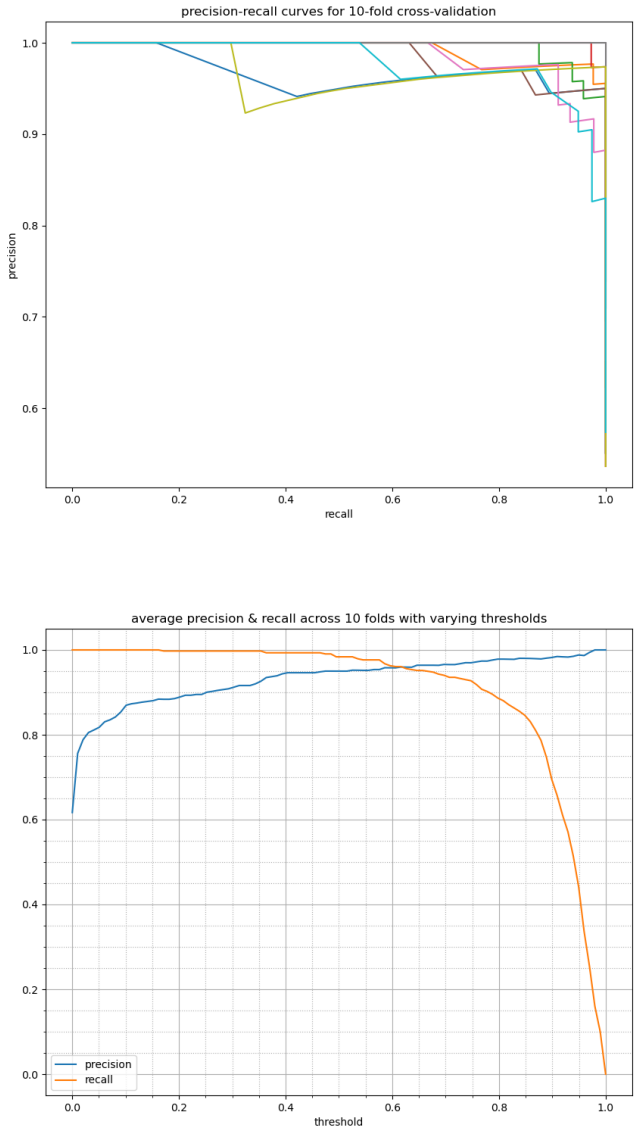
SINCE our last Class Report of 1949 (which in my view dealt all too adequately with the first twenty years out of College) my wife and I have had the happiness of seeing our family increased by another son, who made his auspicious appearance in the world on Valentine's Day, 1950. We have also had the joy in these years of watching our other three children grow up strong and well — as one wishes all children might — and are looking forward this autumn to the eventful moment when our first son enters Harvard.

Architectural practice continues to give great satisfaction, an attitude shared by many, one gathers, if one is to credit a recent survey showing more architects per capita in this region than anywhere in the country. A crowded but congenial field!

[678]

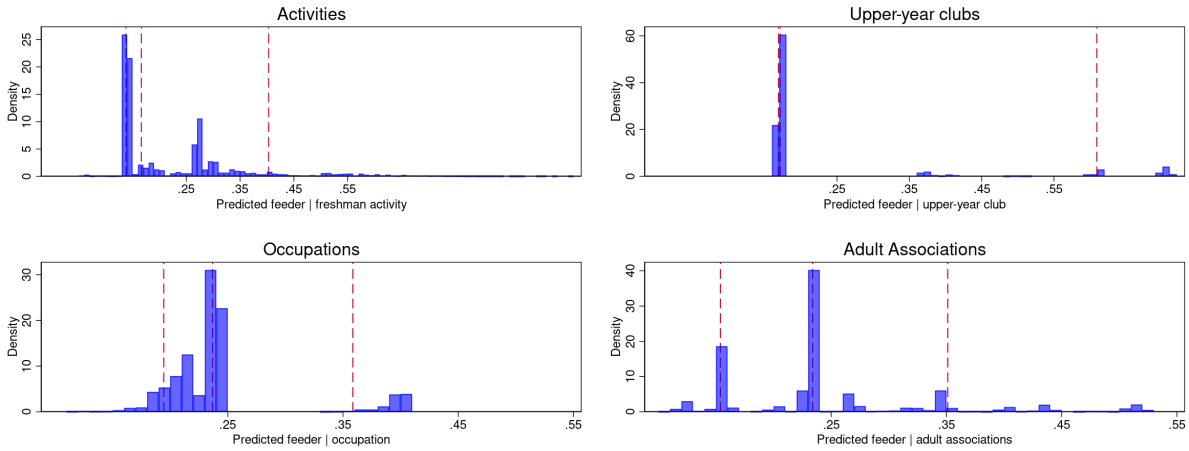
Classes 1923-1950 digitized. Includes birthplace through graduating class of 1937 (corresponding with Freshmen entering in fall 1933). See Section B.1.7 for more details on the Class Reports.

Figure B.12: Cross-validation for linking Red Books to Class Reports



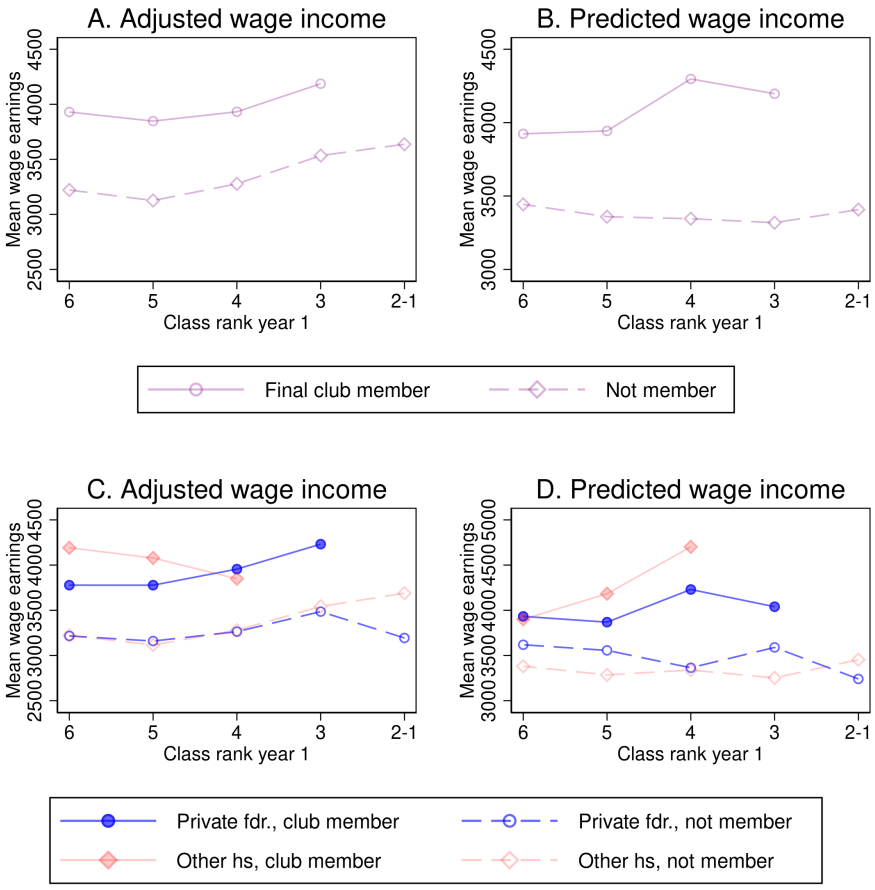
Precision and recall calculated for a 10-fold cross validation exercise on the manually labeled traing data set of 690 observations. The top figure presents the precision/recall curve for each of 10 folds. The bottom figure presents a precision curve and a recall curve by our choice of threshold (minimum similarity score at which we accept a candidate match), averaging over the results from the 10 folds. We use matchs above a threshold of 0.4 in our analysis. For more details on the matching of Red Book and Class Report records, see description in Section B.2.1.

Figure B.13: Occupation private school indices



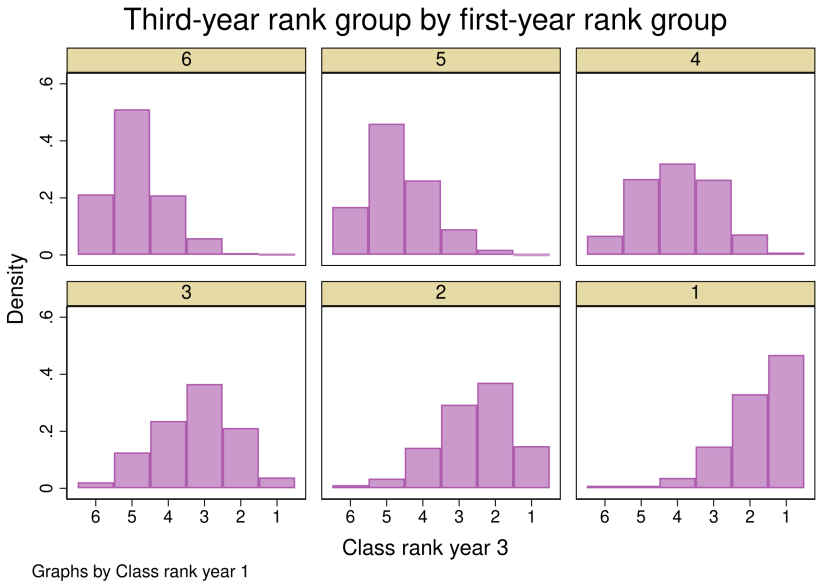
Histograms of activity private high school index (left panel) and occupation private high school index (right panel). Indices estimated using Lasso specifications as described in section 3.4.3. Predictions for each entering cohort are based on students in other entering cohorts. See Table B.8 for estimated coefficients from Lasso procedure in full sample. These distributions are of predicted values, prior to standardization.

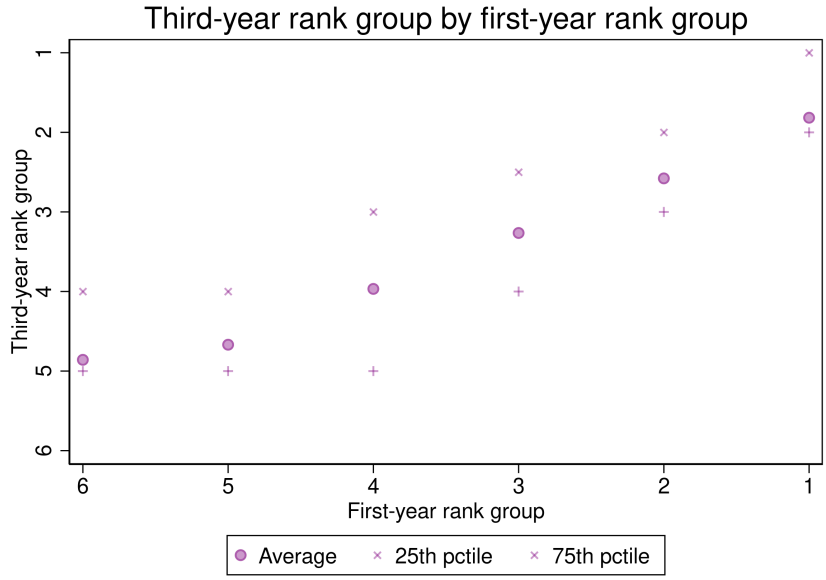
Figure B.14: Alternative earnings by academic performance and final club membership



Panels A-B depict earnings by freshman academic rank group and selective final club membership. Panels C-D present the same but also divide students by high school type. Rank groups 1 and 2 are collapsed and groups with fewer than 20 students not displayed. Includes students from cohorts 1920-1930 who matched to the 1940 census and reported wage income for Panels A and C (and owned own home for Panels B and D).

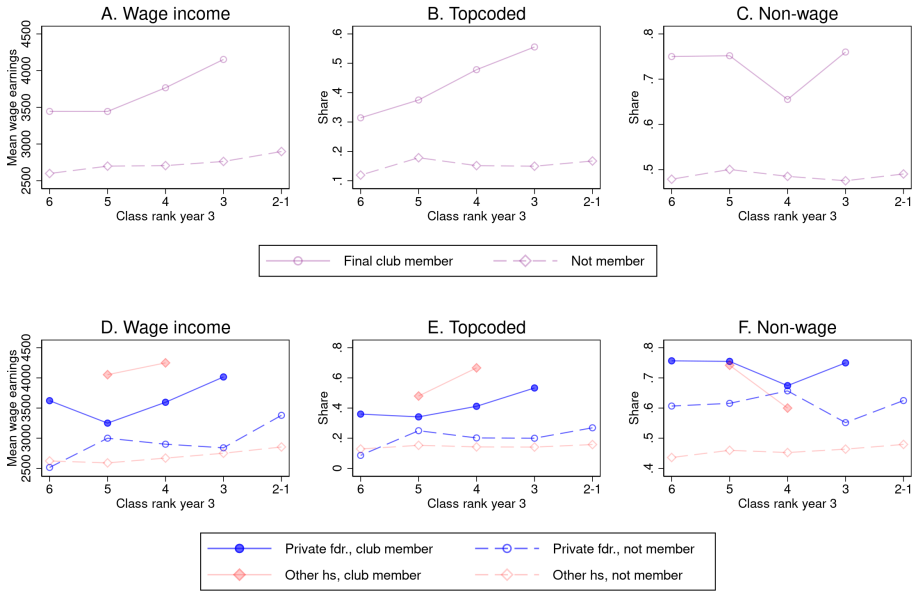
Figure B.15: Academic rank in years 1 and 3





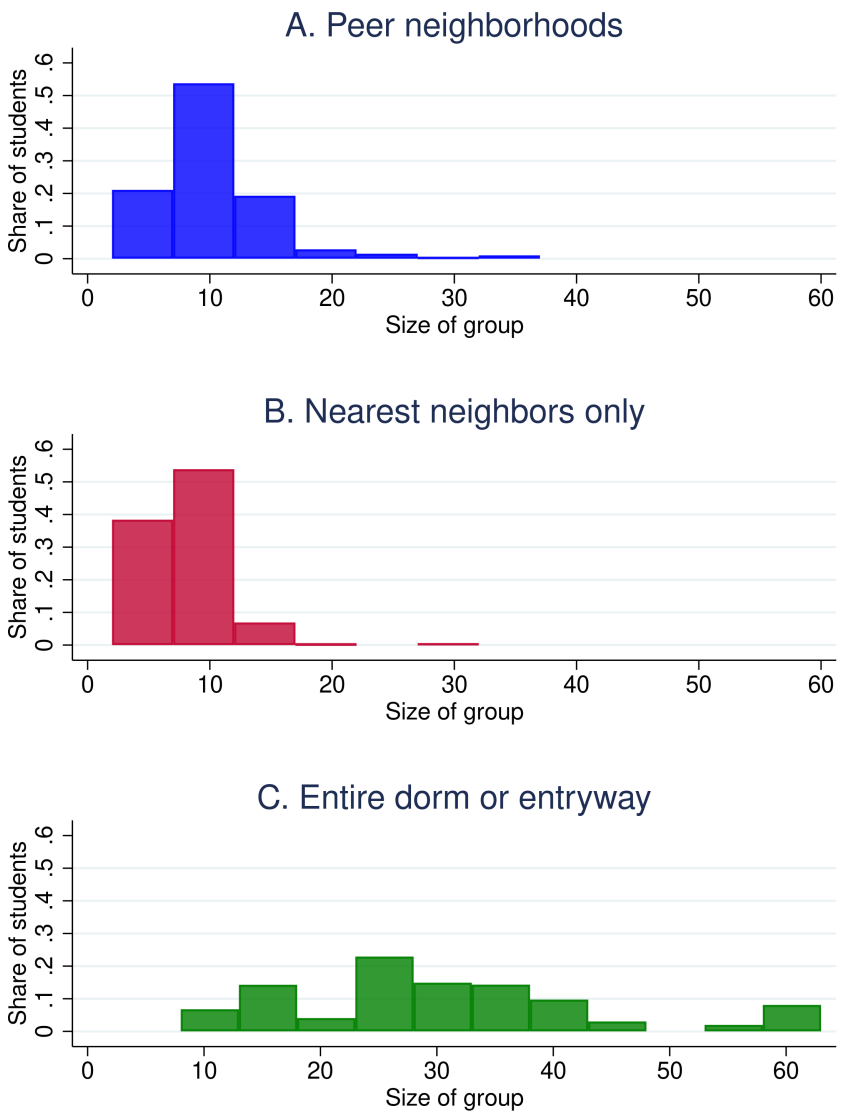
Figures present the within-student relationship between first- and third- year rank groups among students for whom both measures are available. See Section B.1.4 for more details on the rank group lists.

Figure B.16: Earnings by academic performance and final club membership



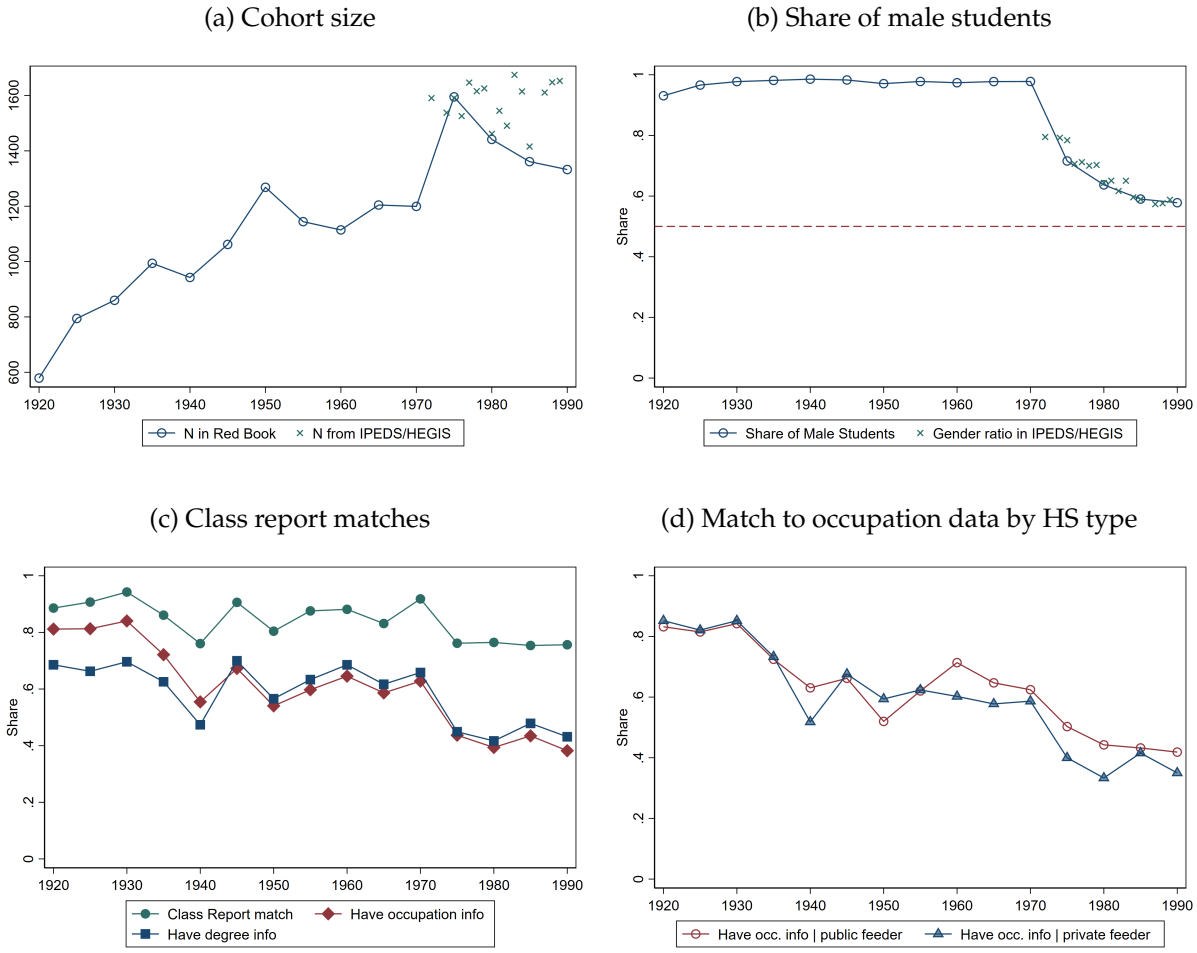
Panels A-C depict earnings by third year academic rank group and selective final club membership. Panels D-F present the same but also divide students by high school type. Rank groups 1 and 2 are collapsed and groups with fewer than 20 students not displayed. Includes students from cohorts 1920-1930 who matched to the 1940 census (and reported wage income for Panels A, B, D, and E).

Figure B.17: Histograms of group size for alternate spatial groupings



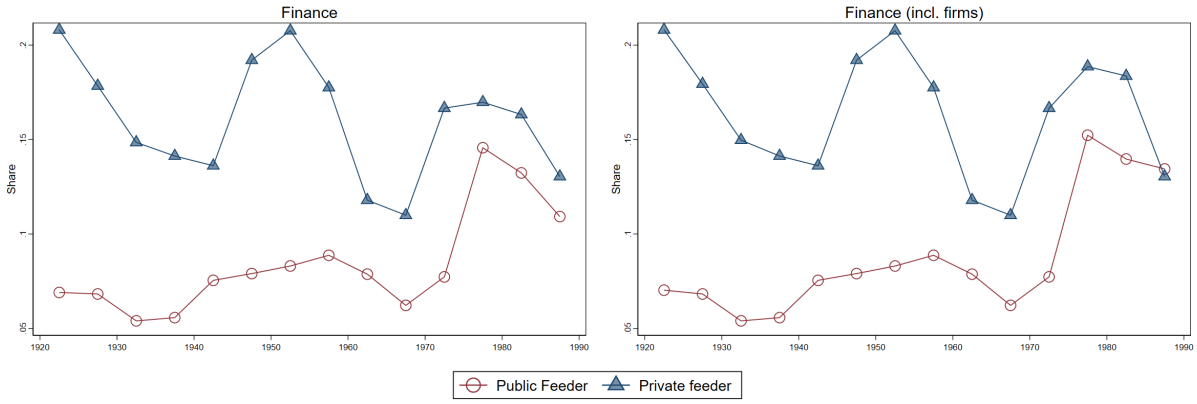
Histograms show the number of students who have different sized peer neighborhoods (Panel A; our main definition), nearest neighbor groups (Panel B), and larger groups of entire dorms (for small dorms) or entire entryway (Panel C). See B.6 for details.

Figure B.18: Long run data construction



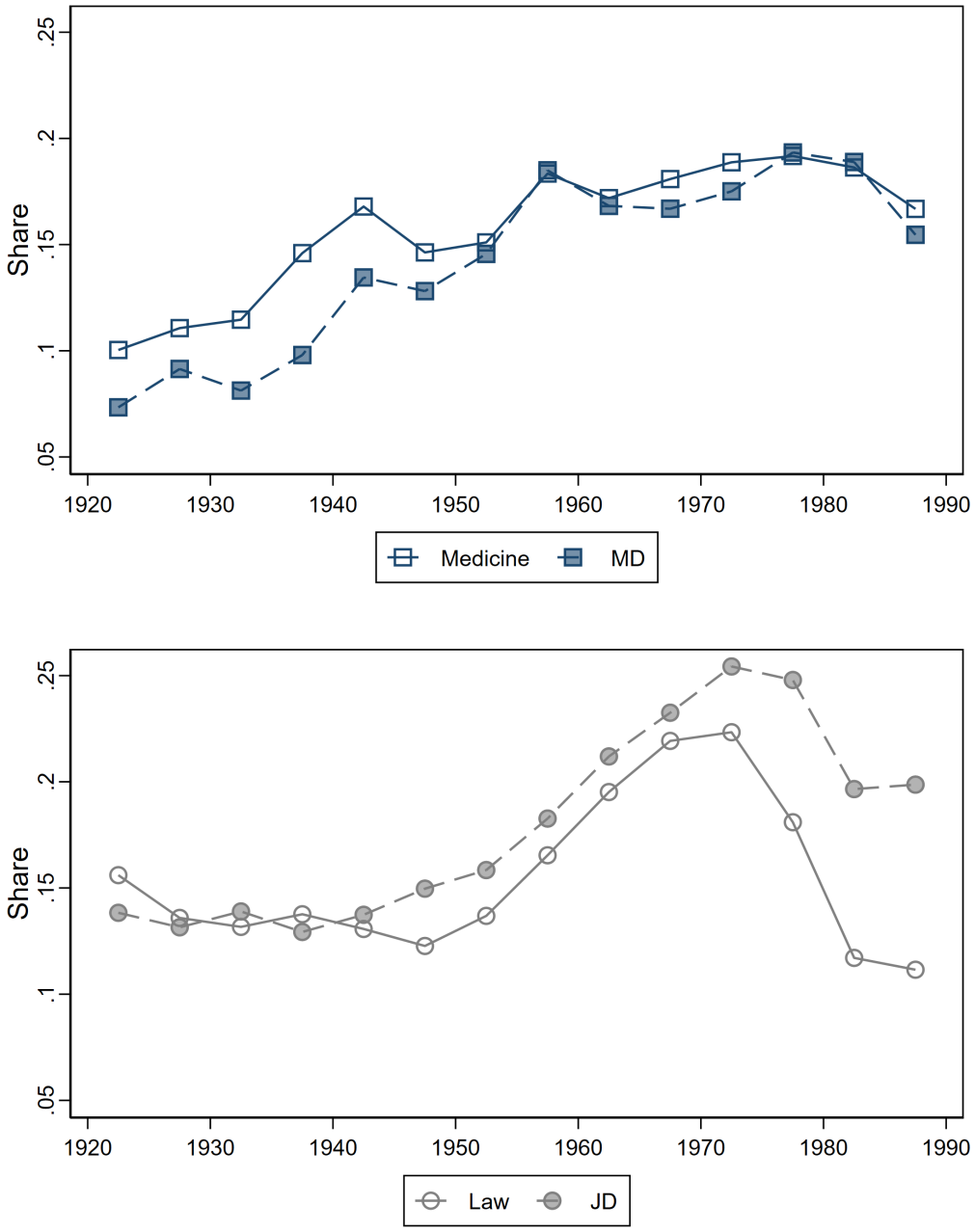
Horizontal axis in all panels is graduating class year. Panel A: Red Book cohort size. Points are averages within centered five-year bins. Panel B: share of male students in Red Books. Points are averages within centered five-year bins. Panel C: match rates from Red Book data to Class Report records, and to Class Report records that include either occupation data or educational history data. 1940 and earlier data are averages of centered 5-year bins. 1945 and later data are single-year means. Panel D: Merge rates to occupation data by high school type. 1940 and earlier data are averages of centered 5-year bins. 1945 and later data are single-year means.

Figure B.19: Main vs. alternative finance occupation measures



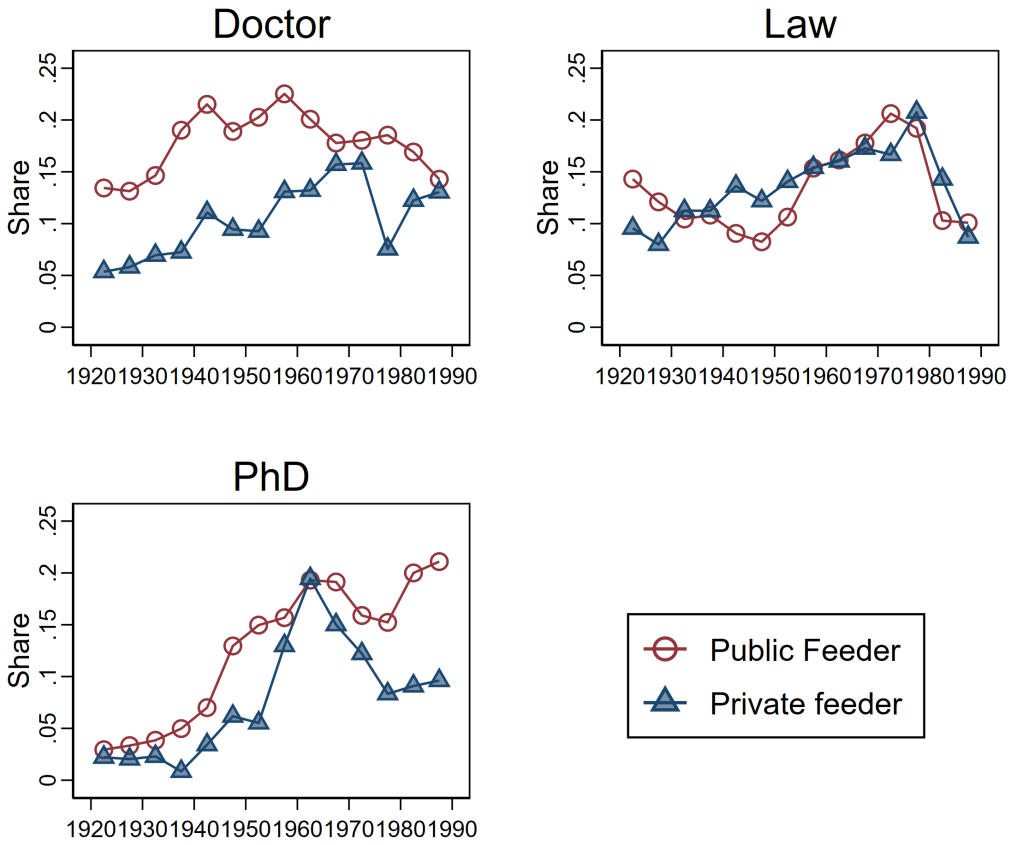
Share of students in finance careers by high school type, using main and supplemental definition. Supplemental definition uses names of major firms. Horizontal axis is graduating class year.. Sample is male students. Both panels use data from Class Reports that is available for each graduating class from 1924 through 1939 and then at five year intervals starting in 1940. Points display means over all years within 2.5 years on either side of the centered value. For example, the 1982.5 datapoint is an average of 1980 and 1985 class years, and the 1987.5 datapoint is an average of 1985 and 1990 class years.

Figure B.20: Trends in career classifications as measured by degree rececipt and job description



Share of students with listed degree/career outcome. Horizontal axis is graduating class year. Sample is students (male and female) with both degree and career information on Class Reports. Both panels use data from Class Reports that is available for each graduating class from 1924 through 1939 and then at five year intervals starting in 1940. Points display means over all years within 2.5 years on either side of the centered value. For example, the 1982.5 datapoint is an average of 1980 and 1985 class years, and the 1987.5 datapoint is an average of 1985 and 1990 class years.

Figure B.21: Long-run trends in alternate career outcomes by high school type



Share of students with listed degree/career outcome by high school type. Horizontal axis is graduating class year. Sample is male students. "Doctor" and "Law" measures are based on job text. "PhD" is students receiving PhD degree. Both panels use data from Class Reports that is available for each graduating class from 1924 through 1939 and then at five year intervals starting in 1940. Points display means over all years within 2.5 years on either side of the centered value. For example, the 1982.5 datapoint is an average of 1980 and 1985 class years, and the 1987.5 datapoint is an average of 1985 and 1990 class years.

B.9 Tables

Table B.1: Codes for common high schools

School Name	Private Feeder	Private	Public feeder
Albany Academy	0	0	1
Andover	1	1	0
Arlington High	0	0	1
Avon Old Farms	0	0	1
B. M. C. Durfee High	0	0	1
Belmont High	0	0	1
Belmont Hill	0	1	0
Berkshire School	0	1	0
Beverly High	0	0	1
Blake	0	0	1
Boston College High	0	0	1
Boston English High	0	0	1
Boston Latin	0	0	1
Brockton High	0	0	1
Bronx High School of Science	0	0	1
Bronxville High	0	0	1
Brookline High	0	0	1
Brooks	0	0	1
Browne and Nichols	0	1	0
Cambridge High and Latin	0	0	1
Chelsea High	0	0	1
Chicago Latin School	0	0	1
Choate	0	1	0
Culver Military Academy	0	0	1
Deerfield Academy	0	1	0
Dorchester High	0	0	1
East Boston High	0	0	1
Episcopal Academy	0	0	1
Erasmus Hall High	0	0	1
Evanston Township High	0	0	1
Everett High	0	0	1
Exeter	1	1	0
Fieldston/Ethical Culture	0	1	0
Fountain Valley	0	0	1
Governor Dummer Academy	0	0	1
Groton	1	1	0
Gunnery	0	1	0
Hackley	0	1	0
Haverford	0	0	1
Haverhill High	0	0	1
Hebron Academy	0	0	1
Hill School	0	1	0
Horace Mann	0	1	0
Hotchkiss	0	1	0
Huntington	0	0	1
James Madison High	0	0	1

Continued on next page

Table B.1 – continued from previous page

School Name	Private Feeder	Private	Public feeder
John Burroughs	0	0	1
Kent	0	1	0
Lawrence Academy	0	0	1
Lawrence High	0	0	1
Lawrenceville School	0	1	0
Lenox	0	0	1
Lexington High School	0	0	1
Loomis	0	1	0
Lowell High	0	0	1
Lynn Classical	0	0	1
Malden High	0	0	1
Medford High	0	0	1
Melrose High	0	0	1
Mercersburg Academy	0	0	1
Middlesex	1	1	0
Milton	1	1	0
Milwaukee Country Day	0	0	1
Montclair High	0	0	1
Morristown School	0	0	1
Moses Brown	0	1	0
Mount Hermon	0	0	1
New Bedford High	0	0	1
New Preparatory School	0	1	0
New Trier Township High	0	0	1
Newton Country Day School	0	1	0
Newton High	0	0	1
Noble and Greenough	0	1	0
North Quincy High	0	0	1
North Shore Country Day	0	1	0
Polytechnic Preparatory Country Day	0	1	0
Pomfret School	0	1	0
Portsmouth Priory	0	0	1
Quincy High	0	0	1
Reading High	0	0	1
Rindge Technical School	0	0	1
Riverdale Country School	0	0	1
Rivers School	0	1	0
Roosevelt High	0	0	1
Roxbury Latin	0	1	0
Salem High	0	0	1
Santa Barbara	0	1	0
Scarsdale High	0	0	1
Shady Side Academy	0	1	0
Somerville High	0	0	1
St. George's	1	1	0
St. Louis Country Day	0	1	0
St. Mark's	1	1	0
St. Paul Academy	0	0	1
St. Paul's, Concord, N.H.	1	1	0

Continued on next page

Table B.1 – continued from previous page

School Name	Private Feeder	Private	Public feeder
St. Paul's, Garden City, N. Y.	0	1	0
Stone School	0	0	1
Stuyvesant High	0	0	1
Tabor Academy	0	0	1
Taft	0	1	0
Thayer	0	1	0
Tome	0	0	1
University School	0	1	0
Volkmann	0	0	1
Walnut Hills High	0	0	1
Watertown High	0	0	1
Wellesley High School	0	0	1
Western Reserve Academy	0	0	1
Westminster School	0	0	1
William Penn Charter	0	0	1
Williston Academy	0	0	1
Winchester High	0	0	1
Winthrop High	0	0	1
Woodrow Wilson High	0	0	1
Worcester Academy	0	1	0

See Section B.1.1 for a discussion of categorizing high schools from the Red Books.

Table B.2: Codes for activities listed in red books

Category	Subcategory	Identifiers
Dormitory Committees		Smith Hall Dormitory Committee, Standish Dormitory Committee, Gore Hall Dormitory Committee, Dormitory Committee, McKinlock Hall Dormitory Committee, Dudley House Committee
Sports Teams	University Teams	baseball squad, soccer squad, track team, lacrosse squad, football squad, basketball squad, Interclass Basketball Team, cross country team, Golf Team, Hockey, Swimming Team, Wrestling Team, Tennis Team, Relay Team, Squash Raquets and Squash Tennis Champion, Freshmen Squash Tournament, Interclass Squash Team, Gym Team, University Fencing Squad, Freshman Golf Tournament, Polo Team, Freshman cheerleader, Handball, Fall Wherry Champtionship, rowing, Boxing, Freshman Football, Ski Team, Rugby, Association Football Team, 2d Football Team, Interclass Football, Sailing
	Dormitory Teams	Dormitory Basketball, Dormitory Football, Interdormitory Hockey, Dormitory Crew, Dormitory Hockey, Dormitory Squash Team, Dormitory Tennis, Dormitory Cross Country, Interhall Basketball, Interhall/Interdormitory athletics
	Intramural Teams	Intramural Football, Intramural Basketball, Intramural Boxing, Intramural Crew, Intramural Squash, Intramural Swimming
Musical Groups	University	University Orchestra, University Band, University Musical Clubs, University Instrumental Clubs, The College Choir, University Jazz Band, Appleton Chapel Choir, Gold Coast Orchestra, University Glee Club, Banjo Club, Mandolin Club, Pierian Sodality, Regimental Band, Concert Audience
	Freshman	freshman banjo club, freshman mandolin club, freshman glee club, Freshman Orchestra, Freshman Musical Clubs, Freshman Instrumental Clubs, Freshman Vocal Club
Red Book		Red Book
Foreign Clubs		International Council, Foreign Student Committee,
	Language Clubs	Cercle Francais, Circulo Espafiol, Circolo Italiano, Deutscher Verein, Russian Club
Outdoor Clubs		Gun Club, Harvard Flying Club, Harvard Mountaineering Club
Politics And Debate Clubs	Debate	Debating Team, Freshman Debating Council, Union Debating Society
	Politics	Harvard Chapter of League for industrial Democracy, Harvard Chapter of the Fellowship of Youth For Peace, Harvard Socialist Club, John Marshall Law club, Democratic Club, Liberal Club, Harvard Peace Society
	Other	Speakers committee, Discussion Club, Freshman Discussion Club
	Politics	Harvard Young Republican Club, Conservative League World Federalists
	Politics	Free Enterprise Society
Stem Clubs		Harvard Engineering Society, Boylston Chemical Club, Mathematics Club
Drama Clubs		Freshman Players, Dramatic Club, 47 Workshop, Freshman Vaudeville, Stage crew, Theatre Workshop

Continued on next page

Table B.2 – continued from previous page

Category	Subcategory	Identifiers
Newspaper And Literary Clubs		harvard crimson, The Advocate, Harvard Magazine, Literary Society, Journal, Harvard Lampoon, Harvard Monthly, Harvard Guardian, Service News, Yardling
Social Committees		finance committee, jubilee committee, Dinner Committee, entertainment committee, Smoker Committee, Regatta Committee, Visiting Teams Committee, Freshman library committee, Delegations Committee, Fall Tea Dance Committee, Confidential Guide Committee, Reviews Committee, Union Committee, Union Dance Committee, Winter Informal, Freshman Affairs Committee, Christmas Dance Committee, Dance Committee
Jewish Clubs		Harvard Menorah, Avukah Club
Other Organizations		Executive Board, Student Government, Chess Club, Endowment Fund, Hoover Fund, Social, Social Service, Register Board, Wireless club, Lowell House, H. U. B. Club, Phillips Brooks House Association, John Barnard Associates, Harvard Thomas-For-President Club, Harvard Memorial Society, Student Union, Stamp Club, Price Greenleaf Aid, Yacht Club, Photography Club, Outing Club, Aeronautical Society, Ornithological Club, WHRV - Harvard Radio Voice, Cosmopolitan Club, Wake, Undergraduate Faculty, Prospect Union, Railroading Association, American Civilization Group, Pre-Medical Society, Bridge Club, Film Society, [Major/Minor] Numerals, Social Relations Society
	High School Alumni Club	St. Paul's Club, Andover club, Harvard Choate Club, Harvard Classical Club
	Geographical	Western Club, Southern Club, Canadian Club
	Christian Groups	Christian Club, Catholic Club, Brotherhood of St. Andrew, Usher in Memorial Church, Usher of University Church
	War Effort	Student Defence League, Air Raid Precautions, Postwar Council
Military	Not deployed	R. O. T. C, U. S. Naval Reserve, Caisson Club, Harvard Naval Unit, Regiment
	Service	Infantry, Ambulance, Air Service, State Militia, Navy, Coast Artillery Corps, Royal Air Force, Marine Section, Artillery, British Ambulance, Machine Gun Battalion, Canadian Field Artillery, Signal Corps, U. S. Army, U.S. Coast Guard, other

See Section B.1.1 for a discussion of categorizing activities from the Red Books.

Table B.3: Codes for occupations listed in class reports

Category	Subcategory	Identifiers
Senior Management		director, president, executive, chairman
Low Management		manager, supervisor, administration
Finance	investment	securities, financial, investments, finance, trust, brokerage, broker, stock, investment, investors, trustee, stockbroker
	banking	banking, banker, bankers, bank
Accounting/Real Estate	treasury	treasurer, treasury
	accountancy	accounting, accountant, accountancy, bookkeeper, auditor, tax
	real estate	real, estate
	insurance	insurance, actuary
Lawyer	law	lawyer, attorney, law, judge, legal, counsel
Retail	business	business, owner
	sales	sales, salesman, selling, retail, merchandise, mercantile, merchant, store, stores, storekeeper, bookseller, retailer
	advertising	advertising, marketing
	hr	relations, personnel
	secretarial	secretary
	consulting	consultant, consulting
Doctor	doctor	physician, surgeon, doctor, medicine, anaesthesiologist, ophthalmologist, hospital, obstetrics, obstetrician, gynecologist, urologist, surgery, orthopedic, neurosurgeon, pediatrician
	psychiatry	psychiatrist, neuropsychiatrist, psychiatry, psychology, psychologist, psychoanalyst
	dentistry	dentist, orthodontist, dentistry
	pharmacy	druggist, pharmacist
Engineer	engineering	engineering, engineer, inventor
	design	designer, draftsman
Higher Education	professor	professor, university, college
Scientist	research	research, laboratory
	chemistry	chemist, chemistry, chemical
	other science	physical, botanist, biology, mineralogist, science, geologist, zoologist, metallurgist, physicist
	anthropology/history	anthropologist, historian
	economics	economics, economist, economic
	mathematics	mathematics, statistician
	library	library, librarian
Teacher	education	teacher, principal, school, admissions, lecturer, education, educational, teaching, tutor, educator, instructor, schoolmaster
Art/Publishing	publishing	publishing, copy, book, editor, proofreading, publisher
	writing	author, writer, literature, novelist
	journalism	journalist, newspaperman, newspaper, journalism, reporter, news
	art	painter, artist, arts, art, painting, curator
	music	composer, conductor, music, organist, musician
	entertainment	actor, theatre, theatres, theatrical, dramatist, entertainer
	photography	photography, photographer
Agriculture	agriculture	farmer, agriculture, ranching, farm, grower, dairy, farming, rancher
Government	government	government, state, bureau

Continued on next page

Table B.3 – continued from previous page

Category	Subcategory	Identifiers
	politics	political, senator, politics
Manufacturing	manufacturing	manufacturers, manufacturer, manufacturing, factory
Mining	mining	mining, coal, mine
	oil	petroleum, oil, gas
Construction/Architecture	construction	construction
	architecture	architect, architecture, architects
Religious	religious	rabbi, clergyman, priesthood, church, priest, ministry, bishop, episcopal, vicar, minister
Military	military	army, naval, navy, commander, military
Aviation	aviation	aviation, air, aircraft, airplane
Retired	retired	retired

See Section B.1.7 for a discussion of categorizing activities from the Class Reports.

Table B.4: Codes for common adult clubs and associations

Name	Social, any	Country Club	Gent. Club	Frat. Order	Honorary Politic.	Professional
Adventurers Club	1	0	1	0	0	0
Agawam Hunt Club	1	1	0	0	0	0
Algonquin club	1	0	1	0	0	0
Alpha Chi Sigma	0	0	0	0	1	1
Alpha Omega Alpha	0	0	0	0	1	0
American Academy of Arts and Sciences	0	0	0	0	1	1
American Academy of Pediatrics	0	0	0	0	0	1
American Anthropological Association	0	0	0	0	0	1
American Association for the Advancement of Science	0	0	0	0	0	1
American Association of University Professors	0	0	0	0	0	1
American Bar Association	0	0	0	0	0	1
American Board of Internal Medicine	0	0	0	0	0	1
American Board of Surgery	0	0	0	0	0	1
American Chemical Society	0	0	0	0	0	1
American Civil Liberties Union	0	0	0	0	1	1
American College of Physicians (fellow)	0	0	0	0	0	1
American College of Surgeons (fellow)	0	0	0	0	0	1
American Dental Association	0	0	0	0	0	1
American Economic Association	0	0	0	0	0	1
American Federation for clinical research	0	0	0	0	0	1
American Geophysical Union	0	0	0	0	0	1
American Heart Association	0	0	0	0	0	1
American Historical Association	0	0	0	0	0	1
American Institute of Accountant	0	0	0	0	0	1
American Institute of Architects	0	0	0	0	0	1
American Institute of Certified Public Accountants	0	0	0	0	0	1
American Institute of Electrical Engineers	0	0	0	0	0	1
American Institute of Mining and Metallurgical Engineers	0	0	0	0	0	1
American Jewish Committee	0	0	0	0	1	0
American Jewish Congress	0	0	0	0	1	0
American Judicature Society	0	0	0	0	0	1
American Law Institute	0	0	0	0	0	1
American Library Association	0	0	0	0	0	1
American Management Association	0	0	0	0	0	1
American Marketing Association	0	0	0	0	0	1
American Mathematical Society	0	0	0	0	0	1
American Medical Association	0	0	0	0	0	1
American Medical Society	0	0	0	0	0	1
American Meteorological Society	0	0	0	0	0	1
American Philosophical Association	0	0	0	0	0	1
American Physical Society	0	0	0	0	0	1
American Physiological Society	0	0	0	0	0	1
American Political Science Association	0	0	0	0	0	1
American Psychiatric Association	0	0	0	0	0	1
American Psychological Association	0	0	0	0	0	1
American Public Health Association	0	0	0	0	0	1

Continued on next page

Table B.4 – continued from previous page

Name	Social, any	Country Club	Gent. Club	Frat. Order	Honorary Politic.	Professional
American Society for Metals	0	0	0	0	0	1
American Society of Civil Engineers	0	0	0	0	0	1
American Society of Mechanical Engineers	0	0	0	0	0	1
American Statistical Association	0	0	0	0	0	1
American Trudeau Society	0	0	0	0	0	1
Anglers Club	1	0	1	0	0	0
Anteaters Club	1	0	1	0	0	0
Argyle Club	1	0	1	0	0	0
Arkwright Club	1	0	1	0	0	0
Association of Ex-Members of Squadron A	1	0	1	0	0	0
Association of Harvard Chemists	0	0	0	0	0	1
Association of the Bar of the City of New York	0	0	0	0	0	1
Athenaeum Club	1	0	1	0	0	0
B'nai B'rith	1	0	0	1	0	0
Belmont Country Club	1	1	0	0	0	0
Boston Bar Association	0	0	0	0	0	1
Boston Club of New Orleand	1	0	1	0	0	0
Boston Surgical Society	0	0	0	0	0	1
Boston Yacht Club	1	1	0	0	0	0
Brae Burn Country Club	1	1	0	0	0	0
Broad Street Club	1	0	1	0	0	0
Buffalo Club	1	0	1	0	0	0
Burlingame Country Club	1	1	0	0	0	0
Cape Fear Club	1	1	0	0	0	0
Capitol Hill club	1	0	1	0	1	0
Cavendish Club	1	0	1	0	0	0
Century Association	1	0	1	0	0	0
Chevy Chase Club	1	1	0	0	0	0
Chicago Club	1	0	1	0	0	0
Cincinnati Club	1	0	1	0	0	0
Cincinnati Country Club	1	1	0	0	0	0
City Club Coporation (The Lunch Club)	1	0	1	0	0	0
City Midday Club	1	0	1	0	0	0
City Tavern Association	1	0	1	0	0	0
Cliff Dwellers Club	1	0	1	0	0	0
Cloud Club	1	0	1	0	0	0
Clover Club	1	0	1	0	0	0
Club of Odd Volumes	1	0	1	0	0	0
Coffee House Club	1	0	1	0	0	0
Cohasset Golf Club	1	1	0	0	0	0
Cohasset Yacht Club	1	1	0	0	0	0
Cold Spring Harbor Beach Club	1	1	0	0	0	0
College Art Association	0	0	0	0	0	1
Colony Club	1	0	1	0	0	0
Columbia Country Club	1	1	0	0	0	0
Columbus Club	1	0	1	0	0	0
Concord Country Club	1	1	0	0	0	0
Cosmos Club	1	0	1	0	0	0

Continued on next page

Table B.4 – continued from previous page

Name	Social, any	Country Club	Gent. Club	Frat. Order	Honorary Politic.	Professional
Creek Club , Locust Valley	1	1	0	0	0	0
Crusing Club of America	1	1	0	0	0	0
Cumberland Club	1	0	1	0	0	0
Dedham Country and Polo Club	1	1	0	0	0	0
Denver Club	1	0	1	0	0	0
Des Moines Club	1	0	1	0	0	0
Detroit Club	1	0	1	0	0	0
Down Town Association (NYC)	1	0	1	0	0	0
Down Town Club (Boston)	1	0	1	0	0	0
Downtown athletic club (NYC)	1	0	1	0	0	0
Dublin Lake Club	1	1	0	0	0	0
Duquensne Club	1	0	1	0	0	0
Dutch Treat Club	1	0	1	0	0	0
Duxbury Yacht Club	1	1	0	0	0	0
Eastern Yacht Club	1	1	0	0	0	0
Elks	1	0	0	1	0	0
Essex Club, Newark	1	0	1	0	0	0
Essex County Club	1	1	0	0	0	0
Examiner Club	1	0	1	0	0	0
Explorers Club	0	0	0	0	0	1
Fay Club	1	0	1	0	0	0
Fort Orange Club	1	0	1	0	0	0
Fort Schuyler Club	1	0	1	0	0	0
Fraternal Order of Eagles	1	0	0	1	0	0
Genesee Valley Club	1	0	1	0	0	0
Graduates Club of New Haven	1	0	1	0	0	0
Grand Street Boys Association	1	0	1	0	0	0
Harmonie club	1	0	1	0	0	0
Hartford Club	1	0	1	0	0	0
Harvey Society	0	0	0	0	0	1
Home Club of Meriden	1	0	1	0	0	0
Hope Club	1	0	1	0	0	0
Houston Club	1	0	1	0	0	0
Idlewild Club	1	0	1	0	0	0
India House	1	0	1	0	0	0
Institute of Radio Engineers	0	0	0	0	0	1
Interdependent Order of Odd Fellows	1	0	0	1	0	0
Jonathan Club	1	0	1	0	0	0
Kittansett Club	1	1	0	0	0	0
Kiwanis [any]	1	0	0	1	0	0
Knickerbocker club	1	0	1	0	0	0
Knights of Columbus [any]	1	0	0	1	0	0
Knights of Pythias [any]	1	0	0	1	0	0
Links Club	1	0	1	0	0	0
Lions Club [any]	1	0	0	1	0	0
Loyal Order of Moose	1	0	0	1	0	0
Madison Square Garden Club	1	0	1	0	0	0
Manchester Yacht Club	1	1	0	0	0	0

Continued on next page

Table B.4 – continued from previous page

Name	Social, any	Country Club	Gent. Club	Frat. Order	Honorary Politic.	Professional
Manhattan Club	1	0	1	0	0	0
Maryland Club	1	0	1	0	0	0
Masons [any]	1	0	0	1	0	0
Massachusetts Bar Association	0	0	0	0	0	1
Massachusetts Medical Society	0	0	0	0	0	1
Mathematical Association of America	0	0	0	0	0	1
Merchants Club (NYC or Baltimore)	1	0	1	0	0	0
Merion Cricket Club	1	1	0	0	0	0
Metropolitan club	1	0	1	0	0	0
Milton-Hoosic Club	1	1	0	0	0	0
Minneapolis Club	1	0	1	0	0	0
Modern Language Association [any]	0	0	0	0	0	1
Myopia Hunt Club	1	1	0	0	0	0
Nation Council of Teachers of English	0	0	0	0	0	1
National Association of Cost Accountants	0	0	0	0	0	1
National Education Association	0	0	0	0	0	1
National Geographic Society	0	0	0	0	0	1
National Press Club	1	0	1	0	0	1
New England Pediatric Society	0	0	0	0	0	1
New York Academy of Sciences	0	0	0	0	0	1
New York County Lawyers Association	0	0	0	0	0	1
New York Society of Security Analysts	0	0	0	0	0	1
New York State Bar Association	0	0	0	0	0	1
New York Yacht Club	1	1	0	0	0	0
Newcomen Society [any]	0	0	0	0	0	1
Noonday Club of St Louis	1	0	1	0	0	0
Omicron Delta Kappa	0	0	0	0	1	0
PEN club	0	0	0	0	0	1
Pacific Club (either nantucket or honolulu)	1	0	1	0	0	0
Pacific Union Club	1	0	1	0	0	0
Pendennis Club	1	0	1	0	0	0
Phi Beta Kappa	0	0	0	0	1	0
Phi Delta Kappa	0	0	0	0	1	0
Phi Kappa Phi	0	0	0	0	1	0
Philadelphia Club	1	0	1	0	0	0
Pilgrims Society	1	0	1	0	0	0
Piping Rock Club	1	1	0	0	0	0
Pittsburgh Club	1	0	1	0	0	0
Public Relations Society of America	0	0	0	0	0	1
Quadrangle Club	1	0	1	0	0	0
Quequechan club	1	0	1	0	0	0
Quiet birdmen	1	0	1	0	0	0
Racquet and Tennis Club	1	0	1	0	0	0
Rainier Club	1	0	1	0	0	0
Rittenhouse Club	1	0	1	0	0	0
River club (NYC)	1	0	1	0	0	0
Rockefeller Center Luncheon Club	1	0	1	0	0	0
Rotary Club [any]	1	0	0	1	0	0

Continued on next page

Table B.4 – continued from previous page

Name	Social, any	Country Club	Gent. Club	Frat. Order	Honorary Politic.	Professional
Saturn Club	1	0	1	0	0	0
Savile Club	1	0	1	0	0	0
Sigma Xi	0	0	0	0	1	0
Society in Dedham for Apprehending Horse Theives	1	0	1	0	0	0
Somerset Club	1	0	1	0	0	0
Sons of the Revolution	1	0	0	1	0	0
St. Botolph Club	1	0	1	0	0	0
Standard Club	1	0	1	0	0	0
Tau Beta Pi (engineer honor society)	0	0	0	0	1	1
Tavern Club	1	0	1	0	0	0
Tennis and Racquet Club of Boston	1	0	1	0	0	0
The Brook Club	1	0	1	0	0	0
The Country Club, Brookline	1	1	0	0	0	0
The National Grange of the Order of the Patrons of Husbandry	1	0	0	1	0	0
The Recess	1	0	1	0	0	0
The Town Club of Scarsdale	1	0	1	0	0	0
The Travellers Club, Paris	1	0	1	0	0	0
Toledo Club	1	0	1	0	0	0
Tuxedo Club	1	1	0	0	0	0
Union Boat Club	1	1	0	0	0	0
Union Club	1	0	1	0	0	0
Wellesley Club	1	0	1	0	0	0
Wellesley County Club	1	1	0	0	0	0
Weston Golf Club	1	1	0	0	0	0
White Hall Club	1	0	1	0	0	0
White's Club London	1	0	1	0	0	0
Winchester Country Club	1	1	0	0	0	0
Worcester Club	1	0	1	0	0	0
Yorick Club	1	0	1	0	0	0
Zionist Organization of America	0	0	0	0	1	0
accounting association [any]	0	0	0	0	0	1
agriculture [any]	0	0	0	0	0	1
american academy of [any]	0	0	0	0	0	1
american club abroad [any]	1	0	1	0	0	0
arts professional association [any]	0	0	0	0	0	1
athletic club [any]	1	1	0	0	0	0
bar or legal assocaition [any]	0	0	0	0	0	1
beach club [any]	1	1	0	0	0	0
bond club [any]	0	0	0	0	0	1
business honor society	0	0	0	0	1	1
century club [any]	1	0	1	0	0	0
chamber of commerce [any]	0	0	0	0	0	1
city club [any]	1	0	1	0	0	0
commercial club [any]	0	0	0	0	0	1
cotillion clubs [any]	1	0	1	0	0	0
country club [any]	1	1	0	0	0	0

Continued on next page

Table B.4 – continued from previous page

Name	Social, any	Country Club	Gent. Club	Frat. Order	Honorary Politic.	Professional
day of the week club -cannot distinguish among cities [any]	1	0	1	0	0	0
education professional [any]	0	0	0	0	0	1
elite book club [any]	1	0	1	0	0	0
engineering professional association [any]	0	0	0	0	0	1
finance professional association [any]	0	0	0	0	0	1
insurance association [any]	0	0	0	0	0	1
legal honor society	0	0	0	0	1	1
leisure animals [any]	1	1	0	0	0	0
management association	0	0	0	0	0	1
medical [any]	0	0	0	0	0	1
misc industry and commerce assocation	0	0	0	0	0	1
officers club [any]	1	0	1	0	0	0
other professions	0	0	0	0	0	1
political club [any]	0	0	0	0	1	0
press club or jounalist professional association [any]	0	0	0	0	0	1
public administration [any]	0	0	0	0	0	1
science [any]	0	0	0	0	0	1
shriners [any]	1	0	0	1	0	0
social science [any]	0	0	0	0	0	1
technical associations [any]	0	0	0	0	0	1
union league club [any]	1	0	1	0	0	0
yacht club [any]	1	1	0	0	0	0
zionist organization [any]	0	0	0	0	1	0

See Section B.1.7 for a discussion of categorizing adult clubs and associations from the Class Reports.

Table B.5: Precision of matches from data sources with only name and year

Data source	N matched with Red Books	Red Book precision	N matched with Class Reports	Class Report precision
Rank Lists	176	100%	182	100%
Student Council Registers	192	98.4%	164	98.2%
Class Albums	189	98.9%	190	99.5%

Quality checks for matches relying only on names and years done on random samples of 200 records from the data source being matched to Red Book and Class Report records.

Table B.6: Most common Census occupation by Class Report occupation

Class report occ.	Has wage inc.	Wage inc.	Wage inc. 5k+	Has non-wage inc.	Modal census occ.	Occ. score	N	Share
Finance	0.75	3644	0.39	0.63	Managers, officials, and proprietors (n.e.c.)	4200	244	0.30
Senior management	0.79	3712	0.40	0.56	Managers, officials, and proprietors (n.e.c.)	4200	379	0.26
Law	0.48	3247	0.26	0.68	Lawyers and judges	6200	530	0.59
Doctor	0.35	1898	0.10	0.74	Physicians and surgeons	8000	346	0.54

Income measures of individuals in our main occupation groups (finance, upper-management, law, medicine) from the Class Reports who also report the modal Census occupation (using 1950 occupation classificiations) within Class Report occupation. Inlcudes cohorts 1920-1930. Occupation scores are the median total income reported by in the 1950 Census of all individuals in that occupation rounded to the nearest hundred dollars.

Table B.7: Most common Census occupations

Census occ.	N	Has wage inc.	Wage inc.	Wage inc. 5k+	Has non-wage inc.
A. Harvard cohorts 1920-1930					
Managers, officials, and proprietors (n.e.c.)	1085	0.75	3481	0.35	0.56
Lawyers and judges	671	0.50	3251	0.27	0.67
Physicians and surgeons	375	0.37	1939	0.12	0.73
Salesmen and sales clerks (n.e.c.)	363	0.91	2447	0.14	0.37
Teachers (n.e.c.)	342	0.95	2244	0.04	0.43
Professional, technical and kindred workers (n.e.c.)	270	0.87	2752	0.12	0.40
Clerical and kindred workers (n.e.c.)	247	0.90	2074	0.04	0.33
College professors	196	0.97	2587	0.08	0.40
All occupations	5607	0.73	2747	0.18	0.50
B. All men ages 27-37					
Managers, officials, and proprietors (n.e.c.)	722984	0.54	1997	0.06	0.49
Lawyers and judges	63532	0.39	2639	0.14	0.66
Physicians and surgeons	52091	0.37	2074	0.11	0.66
Salesmen and sales clerks (n.e.c.)	560366	0.90	1490	0.01	0.15
Teachers (n.e.c.)	123466	0.95	1487	0.00	0.21
Professional, technical and kindred workers (n.e.c.)	78307	0.89	1825	0.02	0.19
Clerical and kindred workers (n.e.c.)	424290	0.95	1406	0.00	0.09
College professors	17977	0.90	2185	0.03	0.29
All occupations	11161415	0.76	1167	0.01	0.25

Income measures for Harvard cohorts 1920-1930 (panel A) and similarly-aged men (panel B) with most common Census occupations (using 1950 occupation classificiations) reported by Harvard cohorts 1920-1930.

Table B.8: Coefficient estimates from Lasso estimation

Activities		UY Clubs		Occupations		Associations	
Have any activity	0.027	Selective final club	0.096	Low management	-0.011	Country club	0.109
Activity count	0.030	Hasty Pudding Inst. 1770	0.205	Finance	0.160	Gentleman's club	0.166
Activity leadership position	0.078	Any final club	0.236	Doctor	-0.049	Fraternal order	-0.030
Social	0.217			Law	-0.029	Prof. Association	-0.076
Sports	0.077			Higher ed.	-0.027		
Music	-0.028			Science	-0.033		
Redbook	0.047						
Dorm committee	0.054						
Language club	0.162						
Politics club	-0.131						
Other club	-0.023						
N	14383		13394		10767		12352

Estimated coefficients from Lasso regressions of private feeder indicators on freshman activity category indicators, upper-year social clubs, career type indicators, and adult associations. Model selection conducted based on EBIC. Sample: students in all entering cohorts. Indices estimated using Lasso specifications as described in section 3.4.3.

Table B.9: Brothers sample description: background and college outcomes

	All	Wages & rank	Brothers	Mixed membership
A. Demographics				
Have high school data	0.984	0.973	0.965	0.913
Any private high school	0.423	0.426	0.607	0.952
Private feeder	0.241	0.243	0.379	0.783
Any public feeder	0.266	0.264	0.193	0.043
From MA	0.509	0.517	0.491	0.522
From NY	0.172	0.162	0.214	0.217
Have Harvard father	0.071	0.105	0.228	0.217
Have Harvard brother	0.204	0.227	1.000	1.000
B. Census childhood household demographics				
Have Census pre-Harvard	0.609	0.705	0.839	0.826
N servants	0.937	0.877	1.617	2.100
S or E Eur. immg. gen. 1-2	0.098	0.078	0.021	0.000
Have father's occupation	0.689	0.680	0.704	0.667
Father's occupation: Doctor	0.101	0.088	0.000	0.000
Father's Occupation: Lawyer	0.109	0.117	0.211	1.000
C. First-year campus location				
Have address data	0.950	0.931	0.965	1.000
Live on campus	0.800	0.824	0.892	1.000
Have room attributes	0.712	0.709	0.792	0.955
Room price per occupant	209.2	198.6	214.4	224.4
Peer neighborhood price	213.9	208.0	217.4	213.4
25th pctile neighborhood rank	0.256	0.256	0.310	0.424
75th pctile neighborhood rank	0.711	0.685	0.740	0.852
D. Academic class rank groups				
Rank group 1	0.018	0.018	0.014	0.000
Rank group 2	0.069	0.058	0.039	0.000
Rank group 3	0.155	0.142	0.165	0.130
Rank group 4	0.242	0.237	0.267	0.217
Rank group 5	0.376	0.381	0.351	0.304
Rank group 6	0.141	0.164	0.165	0.348
Class rank year 1	4.311	4.398	4.396	4.870
Not ranked year 1	0.209	0.000	0.000	0.000
E. First-year activities				
Have any activity	0.526	0.569	0.695	0.913
N activities	1.046	1.250	1.726	3.043
Activity leadership position	0.064	0.079	0.105	0.304
Sports	0.367	0.387	0.484	0.652
Social	0.083	0.105	0.175	0.435
Music	0.133	0.178	0.214	0.261
Other activities	0.174	0.208	0.323	0.565
First-year activity index	0.000	0.050	0.377	1.235
F. Upper-year social clubs				
Hasty Pudding Inst. 1770	0.153	0.167	0.295	0.652
Selective final club	0.070	0.077	0.154	0.522
Any final club	0.136	0.156	0.274	0.652
Upper-year club index	-0.000	0.054	0.416	1.555
N	14383	3361	285	23

All: full Red Book sample universe. Wages and rank: All students in cohorts 1920-1930 who are matched to year 1 class rank and to the 1940 census with wage income reported. These are the students in the earnings specification of models A-C reported in 4. Brothers: Students in cohorts 1920-1930 matched to year 1 class rank and 1940 census wage income who have at least one brother meeting the same criteria. These are the students in the earnings specification of model E reported in 4. Mixed membership: Students with wages and rank group who have at least one brother who also has waged and rank group and has the opposite membership status in selective final clubs. These are the students whose within family variation identifies the final club membership returns parameter in the earnings specification of model E reported in 4. See text for detailed variable definitions.

Table B.10: Brothers sample description: adult outcomes

	All	Wages & rank	Brothers	Mixed membership
A. Adult associations				
Any social club	0.343	0.371	0.428	0.565
Country club	0.242	0.249	0.330	0.391
Gentleman's club	0.113	0.126	0.172	0.304
Fraternal order	0.103	0.114	0.081	0.043
Any honor/prof group	0.361	0.440	0.425	0.391
Prof. Association	0.337	0.409	0.393	0.391
Honor society	0.074	0.093	0.098	0.043
Adult association index	-0.000	0.005	0.215	0.564
B. Occupations				
Have occupation	0.872	0.939	0.979	0.957
Finance	0.102	0.121	0.122	0.091
Accounting	0.108	0.125	0.115	0.318
Doctor	0.091	0.054	0.047	0.000
Law	0.122	0.099	0.086	0.091
Higher ed.	0.070	0.075	0.072	0.000
Teach	0.081	0.109	0.118	0.000
Government	0.034	0.037	0.036	0.000
Art/pub	0.074	0.080	0.093	0.045
Senior management	0.213	0.229	0.215	0.318
Low management	0.118	0.124	0.129	0.091
Retail	0.136	0.137	0.151	0.273
Occupation index	-0.000	0.054	0.062	0.072
C. Adult census				
In school	0.030	0.026	0.014	0.000
In labor force	0.953	0.983	0.975	1.000
Wage income	2747	2771	2968	3051
Has wage income	0.726	1.000	1.000	1.000
Non-wage inc. 50+	0.503	0.398	0.513	0.619
Wage inc. 5000+	0.175	0.182	0.239	0.348
N	14383	3361	285	23

All: full Red Book sample universe. Wages and rank: All students in cohorts 1920-1930 who are matched to year 1 class rank and to the 1940 census with wage income reported. These are the students in the earnings specification of models A-C reported in 4. Brothers: Students in cohorts 1920-1930 matched to year 1 class rank and 1940 census wage income who have at least one brother meeting the same criteria. These are the students in the earnings specification of model E reported in 4. Mixed membership: Students with wages and rank group who have at least one brother who also has waged and rank group and has the opposite membership status in selective final clubs. These are the students whose within family variation identifies the final club membership returns parameter in the earnings specification of model E reported in 4. See text for detailed variable definitions.

Table B.11: Labor market outcomes by academic performance in third year and social success

	Has earnings	Earnings	Topcoded	Non-wage	Imputed wages	Topcodes x 1.4
A. Baseline						
Private feeder	-0.012	398	0.111	0.187	197	290
	(0.016)	(63)	(0.019)	(0.019)	(65)	(103)
Class rank year 3	0.003	86	0.009	0.004	153	241
	(0.005)	(20)	(0.006)	(0.007)	(21)	(34)
Sample mean	0.479	2795	0.178	0.507	3390	4184
N	5585	2674	2674	3617	3140	3140
B. Add most elite final clubs						
Private feeder	-0.016	225	0.063	0.149	63	91
	(0.017)	(66)	(0.020)	(0.021)	(69)	(109)
Class rank year 3	0.004	97	0.012	0.006	162	254
	(0.005)	(20)	(0.006)	(0.007)	(21)	(34)
Selective final club	0.016	735	0.205	0.165	586	872
	(0.027)	(101)	(0.035)	(0.030)	(98)	(162)
Sample mean	0.479	2795	0.178	0.507	3390	4184
N	5585	2674	2674	3617	3140	3140
C. Private feeders with HS FEs, legacy indicators						
Class rank year 3	0.008	96	0.022	0.008	110	182
	(0.013)	(51)	(0.016)	(0.015)	(49)	(78)
Selective final club	0.004	602	0.163	0.057	512	745
	(0.036)	(142)	(0.045)	(0.040)	(139)	(222)
Have Harvard father	-0.035	-2	0.031	0.075	210	387
	(0.035)	(134)	(0.042)	(0.039)	(136)	(220)
Sample mean	0.483	3054	0.256	0.650	3474	4302
N	1264	610	610	831	686	686
D. Within family						
Class rank year 3	-0.018	65	0.014	0.025	124	187
	(0.024)	(93)	(0.032)	(0.029)	(109)	(170)
Selective final club	-0.044	995	0.357	0.086	1057	1816
	(0.126)	(588)	(0.144)	(0.187)	(513)	(769)
Sample mean	0.512	2919	0.222	0.618	3478	4350
N	727	212	212	387	289	289
E. Within Hasty Pudding (approximate applicant pool)						
Private feeder	-0.046	127	0.025	0.118	61	107
	(0.036)	(138)	(0.044)	(0.042)	(139)	(221)
Class rank year 3	0.016	163	0.067	0.024	156	294
	(0.016)	(62)	(0.020)	(0.017)	(60)	(97)
Selective final club	-0.029	448	0.104	0.123	375	522
	(0.034)	(125)	(0.042)	(0.039)	(126)	(206)
Sample mean	0.497	3344	0.356	0.663	3691	4654
N	932	463	463	615	513	513

All models include cohort fixed effects. The sign on class rank groups is reversed so that high numbers correspond with higher academic performance (e.g., the lowest performing group is coded as -6 and the highest performing group is coded -1). Models A-B include students in cohorts 1920-1930 who are matched with third-year rank records (and, for columns 2-4, the 1940 Census). Model C includes high school fixed effects and additionally restricts to students from private feeder high schools. Model D includes family fixed effects and restricts to students who have at least one brother who also attended in cohorts 1920-1930 and is matched with first year class rank and 1940 census records. Model E includes only students who were members of Hasty Pudding which is described as a "proving ground" for final club membership. Robust standard errors in parentheses. Standard errors of model D are clustered at the family level.

Table B.12: Entryway rank effects on short-run outcomes

	All	Private	Non-private	Test
A. First-year activities				
Have any activity	0.052	0.193	-0.007	0.003
	(0.035)	(0.059)	(0.042)	
N activities	0.183	0.529	0.040	0.076
	(0.120)	(0.253)	(0.124)	
Activity leadership position	0.017	0.055	0.003	0.237
	(0.019)	(0.040)	(0.022)	
Social	0.054	0.199	-0.015	0.001
	(0.026)	(0.057)	(0.025)	
Sports	-0.011	0.101	-0.065	0.022
	(0.036)	(0.067)	(0.039)	
Music	0.015	0.058	-0.004	0.300
	(0.027)	(0.054)	(0.032)	
Other activities	0.027	0.025	0.041	0.816
	(0.032)	(0.059)	(0.036)	
First-year activity index	0.157	0.543	-0.027	0.004
	(0.086)	(0.178)	(0.092)	
N	9640	2854	6630	
B. Upper-year social clubs				
Selective final club	0.035	0.127	-0.009	0.022
	(0.024)	(0.057)	(0.019)	
Less selective final club	-0.020	-0.066	0.004	0.186
	(0.021)	(0.050)	(0.021)	
Hasty Pudding Inst. 1770	0.001	0.082	-0.023	0.148
	(0.033)	(0.066)	(0.035)	
Upper-year club index	0.055	0.251	-0.019	0.158
	(0.089)	(0.181)	(0.088)	
N	8886	2632	6108	
C. First-year academic rank				
Rank group 1	0.004	-0.006	0.010	0.395
	(0.011)	(0.009)	(0.016)	
Rank group 2	0.011	0.018	0.006	0.703
	(0.017)	(0.022)	(0.022)	
Rank group 3	-0.006	-0.034	0.008	0.359
	(0.023)	(0.035)	(0.029)	
Rank group 4	-0.007	0.072	-0.053	0.024
	(0.026)	(0.042)	(0.034)	
Rank group 5	0.075	0.073	0.077	0.953
	(0.033)	(0.055)	(0.043)	
Rank group 6	-0.022	-0.027	-0.024	0.954
	(0.023)	(0.043)	(0.028)	
Rank listed year 1	0.055	0.096	0.024	0.230
	(0.027)	(0.051)	(0.033)	
Class rank year 1	0.027	0.022	0.063	0.841
	(0.107)	(0.153)	(0.141)	
N	7253	2103	5000	

Coefficients on entryway price rank from regressions of form given in equation 2. Specifications are identical to those reported in Table 6 but with ranks based on much larger entryway groups, not peer neighborhood. See section B.6 for details. Standard errors clustered at entryway level.

Table B.13: Entryway rank effects on long-run outcomes

	All	Private	Non-private	Test
A. Adult associations				
Any social club	0.039	0.136	-0.005	0.098
	(0.036)	(0.069)	(0.047)	
Country club	0.034	0.119	0.000	0.121
	(0.034)	(0.066)	(0.041)	
Gentleman's club	0.059	0.099	0.034	0.313
	(0.026)	(0.059)	(0.027)	
Fraternal order	-0.015	-0.003	-0.019	0.718
	(0.022)	(0.033)	(0.029)	
Any honor/prof group	-0.026	0.034	-0.062	0.182
	(0.035)	(0.061)	(0.043)	
Prof. Association	-0.026	0.020	-0.062	0.255
	(0.033)	(0.059)	(0.042)	
Honor society	-0.001	0.038	-0.016	0.215
	(0.019)	(0.035)	(0.024)	
Adult association index	0.173	0.314	0.119	0.318
	(0.082)	(0.179)	(0.088)	
N	8315	2471	5690	
B. Occupation choice				
Finance	0.031	0.171	-0.045	0.001
	(0.025)	(0.059)	(0.027)	
Accounting	-0.034	-0.043	-0.035	0.886
	(0.024)	(0.050)	(0.030)	
Doctor	0.010	0.005	0.002	0.933
	(0.022)	(0.030)	(0.029)	
Law	-0.024	-0.057	-0.009	0.392
	(0.027)	(0.047)	(0.035)	
Higher ed.	-0.011	-0.033	-0.002	0.472
	(0.020)	(0.032)	(0.027)	
Teach	-0.022	-0.027	-0.014	0.764
	(0.020)	(0.034)	(0.027)	
Government	0.001	-0.001	0.011	0.732
	(0.014)	(0.030)	(0.018)	
Art/pub	0.018	0.018	0.023	0.899
	(0.021)	(0.036)	(0.028)	
Senior management	0.009	0.057	-0.015	0.332
	(0.034)	(0.063)	(0.041)	
Low management	0.053	0.058	0.031	0.606
	(0.027)	(0.044)	(0.032)	
Retail	-0.004	0.057	-0.025	0.123
	(0.027)	(0.043)	(0.034)	
Occupation index	0.089	0.530	-0.133	0.001
	(0.080)	(0.187)	(0.089)	
N	7169	2093	4928	
C. Adult income				
Wage income	-107.6	-148.0	-89.1	0.889
	(164.1)	(352.8)	(208.8)	
Wage inc. 5000+	0.021	0.038	0.007	0.787
	(0.046)	(0.098)	(0.057)	
Non-wage inc. 50+	0.046	0.082	-0.008	0.435
	(0.057)	(0.093)	(0.071)	
Class Report wage index	5.9	54.8	-19.9	0.259
	(31.1)	(58.9)	(35.7)	
N	2512	693	1726	

Coefficients on entryway price rank from regressions of form given in equation 2. Specifications are identical to those reported in Table 7 but with ranks based on much larger entryway groups, not peer neighborhood. See section B.6 for details. Standard errors clustered at entryway level.

Table B.14: Nearest neighbor rank effects on short-run outcomes

	All	Private	Non-private	Test
A. First-year activities				
Have any activity	0.014	-0.039	0.032	0.331
	(0.036)	(0.060)	(0.045)	
N activities	0.167	0.467	0.040	0.117
	(0.117)	(0.247)	(0.129)	
Activity leadership position	0.032	0.082	0.005	0.139
	(0.021)	(0.050)	(0.021)	
Social	0.025	0.155	-0.045	0.003
	(0.027)	(0.064)	(0.025)	
Sports	-0.017	0.006	-0.033	0.637
	(0.038)	(0.069)	(0.047)	
Music	0.054	0.095	0.040	0.414
	(0.029)	(0.057)	(0.036)	
Other activities	-0.002	-0.080	0.049	0.068
	(0.032)	(0.063)	(0.037)	
First-year activity index	0.060	0.238	-0.067	0.136
	(0.088)	(0.187)	(0.093)	
N	9117	2748	6222	
B. Upper-year social clubs				
Selective final club	0.037	0.114	-0.014	0.043
	(0.023)	(0.062)	(0.018)	
Less selective final club	-0.024	-0.045	-0.015	0.586
	(0.021)	(0.052)	(0.023)	
Hasty Pudding Inst. 1770	0.013	0.113	-0.031	0.041
	(0.030)	(0.065)	(0.031)	
Upper-year club index	0.061	0.299	-0.084	0.057
	(0.081)	(0.188)	(0.080)	
N	8409	2539	5733	
C. First-year academic rank				
Rank group 1	-0.001	-0.015	0.002	0.421
	(0.011)	(0.016)	(0.014)	
Rank group 2	0.019	0.009	0.020	0.763
	(0.017)	(0.026)	(0.022)	
Rank group 3	0.009	-0.005	0.009	0.789
	(0.025)	(0.044)	(0.031)	
Rank group 4	0.014	0.111	-0.026	0.021
	(0.031)	(0.049)	(0.038)	
Rank group 5	0.005	-0.131	0.082	0.006
	(0.034)	(0.065)	(0.041)	
Rank group 6	0.028	0.062	-0.005	0.251
	(0.025)	(0.050)	(0.030)	
Rank listed year 1	0.074	0.032	0.082	0.449
	(0.031)	(0.056)	(0.038)	
Class rank year 1	0.017	0.001	0.028	0.905
	(0.109)	(0.195)	(0.133)	
N	6856	2028	4686	

Coefficients on nearest-neighbor price rank from regressions of form given in equation 2. Specifications are identical to those reported in Table 6 but with ranks based on smaller nearest-neighborh groups, not peer neighborhood. See section B.6 for details. Standard errors clustered at nearest-neighbor level.

Table B.15: Nearest neighbor effects on long-run outcomes

	All	Private	Non-private	Test
A. Adult associations				
Any social club	0.046	0.218	-0.025	0.004
	(0.037)	(0.069)	(0.048)	
Country club	0.047	0.174	-0.013	0.019
	(0.034)	(0.068)	(0.044)	
Gentleman's club	0.006	0.068	-0.022	0.175
	(0.028)	(0.061)	(0.031)	
Fraternal order	-0.007	0.036	-0.021	0.245
	(0.024)	(0.039)	(0.031)	
Any honor/prof group	0.013	0.040	0.011	0.706
	(0.034)	(0.063)	(0.043)	
Prof. Association	0.004	0.028	-0.001	0.704
	(0.033)	(0.063)	(0.042)	
Honor society	-0.019	0.027	-0.027	0.216
	(0.020)	(0.035)	(0.025)	
Adult association index	0.066	0.291	-0.050	0.079
	(0.081)	(0.173)	(0.094)	
N	7857	2380	5331	
B. Occupation choice				
Finance	0.004	0.115	-0.057	0.014
	(0.027)	(0.064)	(0.031)	
Accounting	0.031	0.033	0.027	0.918
	(0.026)	(0.052)	(0.031)	
Doctor	-0.002	-0.064	0.024	0.074
	(0.024)	(0.038)	(0.031)	
Law	-0.023	0.031	-0.049	0.162
	(0.028)	(0.045)	(0.038)	
Higher ed.	-0.018	-0.021	-0.032	0.815
	(0.022)	(0.037)	(0.029)	
Teach	-0.010	-0.015	-0.008	0.887
	(0.023)	(0.042)	(0.029)	
Government	0.017	0.048	0.010	0.291
	(0.016)	(0.030)	(0.019)	
Art/pub	-0.020	-0.030	-0.007	0.618
	(0.022)	(0.036)	(0.029)	
Senior management	0.037	0.087	0.015	0.356
	(0.035)	(0.068)	(0.043)	
Low management	0.079	0.037	0.091	0.341
	(0.027)	(0.045)	(0.034)	
Retail	0.008	0.050	-0.016	0.299
	(0.029)	(0.052)	(0.039)	
Occupation index	0.009	0.377	-0.184	0.010
	(0.089)	(0.194)	(0.105)	
N	6776	2019	4616	
C. Adult income				
Wage income	6.7	-368.1	217.4	0.169
	(189.8)	(370.5)	(234.5)	
Wage inc. 5000+	0.022	-0.020	0.073	0.417
	(0.051)	(0.099)	(0.062)	
Non-wage inc. 50+	0.044	0.094	0.025	0.556
	(0.053)	(0.099)	(0.065)	
Class Report wage index	31.1	130.2	-17.6	0.011
	(27.2)	(49.2)	(34.1)	
N	2361	685	1587	

Coefficients on nearest-neighbor price rank from regressions of form given in equation 2. Specifications are identical to those reported in Table 7 but with ranks based on smaller nearest-neighbor groups, not peer neighborhood. See section B.6 for details. Standard errors clustered at nearest-neighbor level.

Table B.16: Additional codes for occupations used in long-run time series

Category	Subcategory	Identifiers
Financial	Investment	Capital Management, Financing, Fund Management, Fund Manager, Investing, Investor, Private Equity, Security Analyst, Trader, Trading, Venture Capital
	Banking	Lending
	Firms (ext. definition)	Barclays, Bear Stearns, Berkshire, BMO Capital Markets, BNP Paribas, Carl M Loeb, Citigroup, Clark Dodge, Credit Suisse, Deutsche Bank, E.F.Hutton, Evercore Partners, Goldman Sachs, HSBC, J.P. Morgan, Jefferies Group, John P. Chase, Kidder Peabody, Lazard, Lehman Brothers, Loeb Rhoades, Manhattan Chase, Merrill Lynch, Mizuho, Morgan Chase, Morgan Stanley, Nomura, Perice Fenner Beane, RBC Capital, S. F. Moseley, Scudder Stevens Clark, Solomon Brothers, UBS, Upham & Company, Wainwright & Company, Welch Forbes, Wells Fargo, White Weld
Law	Law	Justice, Litigation, Litigator, Solicitor, Supreme Court
Medical	Doctor	Cardiologist, Dermatologist, Gastroenterologist, Gynecology, M.D., Medical Director, Medical Officer, Neonatologist, Neurology, Nurse, Oncologist, Ophthalmolology, Opodiatrist, Optometrist, Otolaryngologist, Pathologist, Podiatrist, Pulmonologist, Radiologist, Rheumatologist
	Psychiatry	Mental Health Center, Psychological, Psychotherapist
Academics/Research	Professor	College Teacher, Faculty, Institute Of Technology, Lecturer, Senior Tutor, University Teacher

See Section B.7.2 for a discussion of using additional occupation identifiers in the long-run time series.

Table B.17: Additional codes of common high schools used in the long-run time series

School Name	Private Feeder	Private	Public feeder
Baltimore City College HS	0	0	1
Bellaire High School	0	0	1
Belmont High School	0	0	1
Belmont Hill School	0	1	0
Berkeley High School	0	0	1
Berkeley Prep	0	1	0
Bethesday-Chevy Chase	0	0	1
Beverly Hills High School	0	0	1
Braintree High School	0	0	1
Brighton High School	0	0	1
Brooklyn Technical High School	0	0	1
Cambridge School of Weston	0	1	0
Canterbury School	0	1	0
Collegiate School	0	1	0
Concord Academy	0	1	0
Concord High School	0	0	1
Concord-Carlisle	0	0	1
Cranbrook School	0	1	0
Delbarton School	0	1	0
Detroit Country Day	0	1	0
Edgemont High School	0	0	1
Eton College	0	1	0
George School, Pa.	0	1	0
Germantown Friends School	0	1	0
Gilman School	0	1	0
Great Neck North	0	0	1
Great Neck South	0	0	0
Greenwich High School	0	0	1
Harvard-Westlake	0	0	1
Henry M. Gunn High School	0	0	1
Highland Park High School	0	0	1
Hopkins School	0	1	0
Horace Greeley High School	0	0	1
Hunter College High School	0	0	0
Iolani School	0	1	0
La Jolla High School	0	0	1
Lakeside School	0	1	0
Lincoln-Sudbury	0	0	1
Lynbrook High School	0	0	1
Maimonides School	0	1	0
Mamaroneck High School	0	0	1
National Cathedral School	0	0	0
Needham High School	0	0	1
New Rochelle High School	0	0	1
Newton High School	0	0	1
Newton North	0	0	1
Newton South	0	0	1
Nichols School	0	1	0

Continued on next page

Table B.17 – continued from previous page

School Name	Private Feeder	Private	Public feeder
Palo Alto High School	0	0	1
Paul D. Schreiber	0	0	1
Pine Crest School	0	1	0
Polytechnic School	0	1	0
Princeton Day School	0	1	0
Princeton High School	0	0	1
Punahou School	0	1	0
Putney School	0	1	0
Regis High School	0	1	0
Ridgewood High School	0	0	1
Rye Country Day	0	1	0
San Francisco University High School	0	1	0
San Marino High School	0	0	1
Shaker Heights High School	0	0	1
Sidwell Friends School	0	1	0
St. Albans	0	1	0
St. Andrew's School	0	1	0
St. John's Prep	0	1	0
St. John's School	0	1	0
Staples High School	0	0	1
The Brearley School	0	1	0
The Chapin School	0	1	0
The College Preparatory School	0	1	0
The Dalton School	0	1	0
The Pingry School	0	1	0
The Winsor School	0	1	0
Thomas Jefferson	0	0	1
Torrey Pines High School	0	0	1
Trinity School	0	1	0
Troy High Schoo	0	0	1
Upper Canada College	0	1	0
Walt Whitman High School	0	0	1
Wayland High School	0	0	1
Westfield High School	0	0	1
Weston High School	0	0	1
Westwood High School	0	0	1
Weymouth High School	0	0	1
White Plains High School	0	0	1
Winston Churchill High School	0	0	1
Wyoming Seminary	0	1	0
Xaverian Brothers High School	0	1	0
Xaverian High School	0	1	0

"Public feeder" and "other private" high schools added to main classification scheme for long run series. See Section B.7.3 for a discussion of categorizing additional high schools for the long-run time series.