The G.I. Bill, Standardized Testing, and Socioeconomic Origins of the U.S. Educational Elite Over a Century

Ran Abramitzky, Jennifer Kowalski, Santiago Pérez, and Joseph Price*

Abstract

We compile, transcribe, and standardize historical records for 2.5 million students at 65 elite (private and public) U.S. colleges. By combining these data with more recent survey and administrative data, we assemble the largest dataset on the socioeconomic backgrounds of students at American colleges spanning the last 100 years. We document the following: First, despite a large increase in the share of lower-income students in the overall college-going population, the representation of these students at elite private or public colleges has remained at similarly low levels throughout the last century. Second, the representation of upper-income students at elite colleges decreased after World War II, but this group has regained its high representation since the 1980s. Third, while there has been no increase in the economic diversity of elite private and public colleges, these colleges have become more racially and geographically diverse. Fourth, two major policy changes in the history of American higher education, namely the G.I. Bill after World War II and the introduction of standardized tests for admissions, had little success in increasing the representation of lower- and middle-income students at elite colleges.

*Abramitzky: Stanford University and NBER (email: ranabr@stanford.edu); Kowalski: Stanford University (email: jennifer.kowalski@stanford); Pérez: University of California - Davis and NBER (email: seperez@ucdavis.edu), Price: Brigham Young University and NBER (email: joe_price@byu.edu). We thank Lukas Althoff, Raj Chetty, Gabriele Cristelli, Anna Carolina Dutra Saraiva, Liran Einav, Michela Giorcelli, Saumitra Jha, Jay Hamilton, Jon Levin, Jenny Martínez, and numerous seminar participants for helpful comments. In addition, we extend our deepest gratitude to a large team of research assistants who helped with data collection: Felix Ampofo, Brenda Angom, Kate Beal, Weston Blocker, Richmond Botchway, Tom Bryan, Annie Busath, Kelsey Carido, Jenny Charters, Cristine Chen, Kevin Chen, Michelle Chen, Ray Cheng, Carver Coleman, Khalel Corona, Eli Cox, Britton Davis, Mike Diaz, Anna Carolina Dutra Saraiva, Colby Eastmond, Laren Edwards, Peter Fortuna, Katelyn Gale, Hayley Gill, Myesha Gilliland, Jaxon Goble, Sam Goldrup, Ezra Gonzalez, Andrew Grisson, Jared Grooms, Stephanie Kalfayan, Stephen Kemboi, Shivali Korgaonkar, Jeffrey Lam, Acii Laourich, Tyan Lee, Kate Lindsay, Josh Nicholls, Mary Jane MacArthur, Nate McGhie, Tommy Morgan, Leah Morrell, Alexander Mudijire, Owen Murray, Barbara Nyagortey, Alexander Okyere, Doris Oppong, Francis Osilo, Allen Otterstrom, Peyton Petersen, Desmond Komla Pongo, Hector Gabriel Reyes Figueroa, Abby Sandoval, Nathan Schill, Ethan Simmons, Sarah Stanley, Nicholas Thatcher, Rebekah Todd, Lauren Tolley, Juan David Torres Alarcón, Gracia Totimeh, Thomas Vick, Ellie Wang, Hannah Weymouth, Johnny Weymouth, Daniel Wilkerson, Margaret Williams, Megan Wright, and 234 freelancers from the Upwork platform. Abramitzky, Pérez, and Price gratefully acknowledge funding from the National Science Foundation and the Russell Sage Foundation.

1 Introduction

Over the last 100 years, the United States has experienced a dramatic increase in college attendance rates, from less than 10% of the population at the beginning of the 20th century, to over 60% in the modern day (Goldin and Katz, 1999). Despite this "democratization" of access, colleges today are still heavily segregated on the basis of parental income, with most students at elite colleges coming from high-income families (Chetty et al., 2020). These two facts motivate our main research questions: Have students from low-income families always been similarly underrepresented at elite colleges? How did two of the most dramatic changes in American higher education, namely the G.I. Bill subsidies for returning World War II veterans and the introduction of standardized testing for admissions, affect the representation of lower-income students at elite colleges? And finally, have elite colleges become more diverse along *other* dimensions beyond parental income (such as race and students' geographic origins)?

We answer these questions by building the most comprehensive dataset of the socioeconomic backgrounds of students in U.S. colleges spanning the last 100 years. These data are based on a large-scale digitization of historical student records for nearly 2.5 million students at 65 elite (private and public) U.S. colleges.[1] Assembling these data involved compiling, scanning, transcribing, and standardizing more than 3,000 historical documents. We link the students from these college records to U.S. population censuses to observe their family backgrounds prior to attending college. Combining these data with existing surveys of college students, modern administrative data, as well as nationally representative samples of the general college-going population, we build a dataset on the socioeconomic backgrounds of students at American colleges spanning 1915-2013. Figure 1 illustrates the construction of one observation in our historical data – Edward Fasnacht, a University of Pennsylvania student from the early 1920s. Upon linking Edward to the 1920 U.S. Population Census, we observe him as a child living with his father, Jacob Fasnacht, who was working as a book keeper at the time. This process, done at scale, constructs the historical records.

In the first part of the paper, we establish *three new facts* about the evolution of the socioeconomic backgrounds of the U.S. college-going population over the past 100 years:

Fact #1: *Despite a large increase in the share of lower-income students in the overall college-going population, the representation of these students at elite private or public colleges has remained at similarly low levels throughout the last century.*

In the early 1920s, individuals growing up in the bottom 20% of the income distribution constituted about 8% of the *overall* college-going population. Nearly a century later, men in the bottom 20% comprise approximately 13% of the male college-going population. This increased representation of children from lower-income families is even more pronounced among women: In recent years women who grew in the bottom 20% of the income distribution represent close to 20% of female college goers.

These patterns for the overall college-going population starkly contrast with those at elite private colleges such as Harvard and Yale, where there has been virtually no change in the

[1] See Section A.1 for the full list of colleges.

representation of lower-income students. In particular, students with parents from the bottom 20% of the income distribution have consistently made up approximately 5% of the student bodies at these institutions throughout the last century. These patterns are remarkably similar across all elite private institutions, including Ivy League universities, other elite universities such as Chicago, Duke, MIT and Stanford, and elite small private liberal arts colleges. One exception are elite women-only colleges, where the representation of lower-income students was the lowest among all groups of elite private colleges in the past and is now the highest.

Elite public schools have, on average, also maintained a small and stagnant share of low-income students throughout the century. However, there is considerably more heterogeneity within elite public colleges than among private ones, with some public colleges experiencing notable gains in the representation of lower-income students. For instance, at UC Berkeley and UCLA the proportion of students that come from the bottom 20% rose from less than 3% in the 1920s to 10% by the early 2000s.

Fact #2: *The representation of upper-income students at elite colleges decreased after World War II but this group has regained its high representation since the 1980s.*

Prior to World War II, students from upper-income (top 20%) families represented 65-70% of students at elite private colleges and 50-55% of students at elite public colleges. This proportion fell in the post-WWII period, so that by the early 1980s, students from such families constituted only 50% of students at elite private colleges and 40% of students at elite public colleges. The representation of upper-income students rebounded in the 1980s, with students from the top 20% of the income distribution now comprising a similar proportion of the student body at elite colleges as in the pre-World War II period. Because the representation of lower-income students remained nearly constant throughout the period, the recent increase in the proportion of upper-income students has primarily occurred at the expense of middle-income students.

Fact #3: *While there has been, on average, no increase in the economic diversity of elite private and public colleges, these institutions have become more diverse in terms of students' race and geographic origins compared to the early 20th century.*

Prior to the 1960s, elite colleges (both private and public) had student bodies that were almost entirely white, yet today Black students represent around 5-7% of their enrollment (corresponding to approximately half of their representation in the general college-age population). The representation of Black students increased sharply in the mid-1960s (coinciding with the Civil Rights Movement), and then remained at similar levels throughout the early 2000s. Today, the relative representation of Black students in elite private colleges is at similar levels as in the early 2000s, whereas in elite public colleges it is actually *lower* than two decades ago.

Elite colleges (particularly the private ones) have also become more diverse with respect to students' geographic origins, integrating into a national market and in recent decades increasingly attracting students from all over the globe. Prior to the 1950s, international students accounted for less than 5% of enrollment in private colleges, and only 30% of the students in these colleges hailed from a town outside their college region. The proportion of out-of-region students increased sharply in private colleges in the late 1950s and early 1960s, and has continued to increase to nearly 60% recently. These colleges have also experienced a

sharp rise in the share of international students, who now account for about 15% of students. The proportion of students from hometowns outside of their college region has increased more modestly at elite public colleges (from 10% in 1960s to 20% today), implying that these schools still serve an overwhelmingly local student body.

In the second part of the paper, we use our student-level data to investigate the role of two major policies that had the potential to increase access of lower-income students to elite colleges. First, we investigate the effects of the G.I. Bill, which provided educational subsidies for returning World War II veterans and hence relaxed budget constraints for lower-income students. Second, we study the effects of the introduction of the Scholastic Aptitude Test (SAT), which might have enabled elite colleges to better identify high-achieving lower-income students. We focus on these two policies, as they have often been described as having helped "democratize" access to college in the United States (Karabel, 2005).

The G.I. Bill was enacted in 1944 and provided returning World War II veterans with educational benefits substantial enough to fully finance attendance at the most expensive colleges in the country. As such, it had the potential to help remove financial barriers preventing lower-income *men* from attending elite colleges.[2] Furthermore, although individuals from higher-SES families were more likely to participate in the War (Collins and Zimran, 2024), participation rates among men in the most affected birth cohorts were high (above 70%) across the parental income distribution (implying that large numbers of lower- and middle-income individuals could potentially benefit from the subsidies). On the other hand, the law did not explicitly address any pre-existing educational inequities (for instance, those by parental income or race), and so lower-SES students might have been less well-positioned to take advantage of the law's college benefits.

To assess whether the law led to changes in the socioeconomic origins of students at elite colleges, we exploit the fact that the likelihood of participating in WWII (and hence receiving G.I. Bill benefits) varied sharply based on a person's year and quarter of birth (a strategy similar to that in Bound and Turner (2002) and Fetter (2013)). For instance, close to 70% of male students in our data born in the third quarter of 1927 served in the War, whereas less than 20% of those born in the third quarter of 1928 did so. Intuitively, if the G.I. Bill increased the representation of lower-SES individuals at elite colleges, we should observe that students from cohorts more exposed to the program were, on average, of a lower socioeconomic background than those from the less exposed cohorts. Against this hypothesis, our results enable us to rule out even modest increases in the proportion of lower- and middle-income students at elite colleges. If anything, the representation of such students appears *lower* among likely veterans in elite public colleges. Similarly, we also see no evidence of an increase in the proportion of Black students, neither in private nor in public colleges (if anything, we observe a decline in the proportion of Black men in public colleges).

The fact that the *proportion* of lower- and middle-income students did not increase at elite colleges could be driven by: (i) lower- and middle-income students increasing their likelihood of attending an elite college but doing so to a similar degree as their higher-income

[2]Women constituted a small minority of WWII veterans (about 2%) so, in practice, the G.I. Bill benefits accrued mostly to men.

counterparts, or (ii) lower- and middle-income students simply not increasing their likelihood of attending an elite college. Our results are more consistent with the latter: we see little evidence that lower- and even middle-income students attended elite colleges at higher rates due to the law. These findings suggest that removing financial barriers to access was not enough to countervail pre-existing inequalities preventing lower-SES students from attending elite colleges.

We next investigate whether the introduction of standardized testing resulted in changes in students' socioeconomic origins at elite colleges. In the past, colleges lacked a standardized way to judge the merits of all applicants, and hence relied on "feeder" high schools that the college believed would produce suitable students (Hoxby, 2009). In principle, the introduction of standardized testing could have leveled the playing field by helping elite colleges identify high-achieving students across the income distribution. Indeed, this argument in favor of standardized exams has been made both historically (including when the exams were first introduced Lemann (2000)) and more recently.[3] On the other hand, there are substantial gaps in standardized test scores by race and parental income, which has prompted critics of these exams to question their potential equity consequences (see, for instance, Berger (2012)).

To assess the role of standardized testing in shaping students' socioeconomic backgrounds, we compile a new dataset with information on the date of introduction of the Scholastic Aptitude Test (SAT) across all the colleges in our data. The SAT was first administered in 1926, and by the end of our sample period, 80% of the colleges in our sample of elite institutions required the SAT for all applicants. We use this information in combination with our student-level data to estimate event-study models exploiting the staggered introduction of the SAT across colleges. This empirical strategy exploits the fact that comparable colleges (e.g., Princeton and Harvard) adopted the SAT at different points in time (1926 and 1935, respectively). We find that the introduction of standardized testing had no effects on the likelihood that lower- and middle-income students would enroll in elite private colleges, enabling us to rule out even modest changes in students' economic backgrounds at such colleges. We also find a relatively small and short-lived increase in the proportion of lower-income students at elite public colleges.

Overall, the findings in the second part of the paper suggest that two policies that have often been described as broadening access to college had more modest "democratizing" effects when focusing on the most highly selective colleges. A common feature of these policies is that they targeted individuals who were close to the age of applying to college. Hence, these interventions might have come "too late" to address pre-existing educational inequalities.

[3]For example, in its recent decision to reinstate the SAT, MIT argued that doing so helps the university to "*identify socioeconomically disadvantaged students who lack access to advanced coursework or other enrichment opportunities that would otherwise demonstrate their readiness for MIT*" (MIT Admissions, 2022). Similar statements were made by other elite universities. For instance, Dartmouth's announcement reinstating the SAT argued that the decision *"was guided by social science research that suggests we can improve our ability to identify students from a wide range of economic backgrounds who will succeed at Dartmouth."* (Dartmouth Office of the President, 2024).

1.1 Related Literature

Our paper contributes to four strands of literature. First, we contribute to the literature characterizing the socioeconomic backgrounds of students at different tiers of U.S. colleges (Astin and Oseguera, 2004; Pallais and Turner, 2006; Bailey and Dynarski, 2011; Hoxby and Avery, 2012; Chetty et al., 2020). Specifically, we provide the first long-run series on how socioeconomic composition has evolved since the 1910s across a nationally representative sample of elite colleges. Understanding who attends these colleges and how this has changed over time is important, as societal leaders in the U.S. are disproportionately drawn from graduates of these institutions (Chetty, Deming and Friedman, 2023).[4] Although data on the backgrounds of college students is available for the more recent period, there is no large-scale, high-frequency, individual-level data prior to 1966 (a period that spans key events in American history in general, and in the history of higher education in particular). Notably, our data encompasses the pre-World War II "formative years" of higher education (Goldin and Katz, 1999), the large-scale expansion of higher education after World War II and the G.I. Bill, and the onset of the Civil Rights Movement.[5]

Second, we contribute to the literature investigating the effects of the mid-century G.I. Bills on educational attainment (Angrist, 1993; Angrist and Krueger, 1994; Lemieux and Card, 2001; Bound and Turner, 2002; Turner and Bound, 2003; Stanley, 2003; Angrist and Chen, 2011; Larsen et al., 2015; Thomas, 2017; Collins and Zimran, 2024). This literature has mostly focused on the extensive margin of college attainment, asking whether those exposed to the program were more likely to attend college. By contrast, we study how the G.I. Bill shaped *who* goes to college and *which* college they attend. In investigating the heterogeneous impacts of the G.I. Bill on the basis of socioeconomic status, our results are related to work that looks at such heterogeneous impacts on the basis of race and ethnicity (Turner and Bound, 2003; Collins and Zimran, 2024). Our finding that, if anything, the G.I. Bill reduced the representation of lower-income and Black students in elite public colleges is consistent with earlier findings from Stanley (2003) (who focused on the overall college-going population) and Turner and Bound (2003) (who also focuses on the overall college-going population, but investigates inequities on the basis of race).

Third, our analysis of the consequences of introducing standardized tests in admissions contributes to the literature investigating the equity consequences of different college admission practices (see, for instance, Long (2004); Hurwitz (2011); Fletcher and Mayer (2014); Bleemer (2022); Black, Denning and Rothstein (2023)). Studies in this literature have fo-

[4]To the extent that we provide a long-term perspective on patterns of access to higher education, our results are related to those in Hendricks, Herrington and Schoellman (2021), which analyzes the evolving roles of a student's ability and family background in explaining whether a high school student attends college (the extensive margin). In contrast, we focus on the relationship between parental background and which *specific* colleges students attend. Understanding how students sort across specific colleges is important, as there is a wide disparity in adult outcomes even among those who attend college (Chetty et al., 2020). Astin and Oseguera (2004) do look at trends in access to selective institutions, but they focus only on the post-1980 period.

[5]There is a vast literature in sociology studying the social origins of the educational elite (and of elites more generally) across different countries and time periods, see for instance Bourdieu and Passeron (1990), Domhoff (2018), and Mills (2019).

cused on different policies used in the admission process, such as affirmative action and "top percent" policies. We contribute to this literature by providing the first evidence of the consequences of the *introduction* of standardized testing, a now common (and controversial) tool in the admission process.[6] More broadly, our analysis of the impacts of the G.I. Bill and the introduction of standardized testing contributes to the literature that attempts to explain the underrepresentation of lower-income students in higher education, particularly in selective colleges (see Page and Scott-Clayton (2016) for a summary).[7]

Finally, we contribute to the literature on the historical evolution of American higher education (Goldin and Katz, 1999; Hoxby, 2009; MacLeod and Urquiola, 2021). Our paper is the first to characterize the social backgrounds of students attending different tiers of colleges spanning the full 20th century (including the pre-WWII period and the crucial years immediately following the War), and to assess the consequences of two landmark changes in the history of American higher education (the G.I. Bill and the introduction of standardized testing in admissions) for student sorting across colleges.

In work concurrent to ours, Bleemer and Quincy (2024) investigate changes in the link between college attendance and upward mobility since 1900 in the U.S., and find that college has become more regressive since the 1960s. We complement Bleemer and Quincy (2024) by putting together new historical data on students from 65 elite colleges, by characterizing the socioeconomic backgrounds of students at elite colleges over the 1915-2013 period, by investigating changes in access by students' race and geographic origins in addition to parental income, and by studying the impact of two major policies, namely the G.I. Bill and the introduction of standardized testing.

2 Data

In this section, we outline our data on students in elite colleges spanning 1915-2013. We then describe our nationally representative data on the backgrounds of the overall college-going population. Table A1 provides an overview of all data sources used and the birth cohorts that each covers. Tables A2-A5 give additional documentation for our historical data collection. We emphasize that our data *do not* include information on who applies and gets admitted to college, we only observe those who effectively enroll. Section A in the Online Appendix provides further details on the data collection process.

[6] A substantial literature documents large differences in test scores based on socioeconomic status (Alvero et al., 2021). Autor and Scarborough (2008) and Moreira and Pérez (2022) investigate the equity consequences of introducing standardized testing for workers' recruitment in the private and public sector, respectively.

[7] Studies in this literature have considered explanations such as informational barriers (Hoxby and Avery, 2012; Goodman, 2016; Bettinger and Evans, 2019), behavioral biases (Dynarski et al., 2021), lack of mentoring (Oreopoulos, Brown and Lavecchia, 2017), differences in test-taking and application behavior between low- and high-income students (Smith, 2014; Bulman, 2015; Hurwitz et al., 2015, 2017), and financial barriers (see Dynarski, Page and Scott-Clayton (2023) for a summary of this literature).

2.1 Students in Elite Colleges Before 1966

College Register Data. We construct a dataset of undergraduate students who attended one of 65 elite U.S. institutions of higher education between 1915 and 1966.[8] These data constitute the largest and most comprehensive historical database of American college students at elite institutions. We stop the data collection in 1966, as this is the latest year for which we can link individuals of typical college-going age to their childhood households in the 1950 census (the most recent full-count census that is currently available). Our data on elite private colleges include the twelve "Ivy-Plus" institutions (the eight Ivy League schools plus Chicago, Duke, MIT and Stanford), eight elite historically women's colleges (the "Seven Sisters" plus any coordinating women's college to an Ivy-Plus institution), and fifteen elite small liberal arts colleges (the "Little Ivies"). In addition to these elite private colleges, we also include a set of 31 selective public schools. These public schools are a useful benchmark for the elite private colleges as they tend to enroll students of relatively comparable academic ability.[9] Tables A2-A5 provide a complete list of all colleges and the associated years included in our dataset. Data Appendix Section A.1 provides detailed information on the construction of each of these school groupings.

To obtain the list of all undergraduates at these 65 institutions, we locate publicly available documents (e.g., registers, directories, catalogs, bulletins, yearbooks) for each individual year for each school. For the 1915-1966 period, that amounts to individually sourcing 3,380 documents. Although these sources vary in format and the information they include, they all feature the names of students enrolled, and most of them also include information on the student's hometown, class standing, and college major. As an example, Panel (a) of Figure 1 shows a page from the 1921 University of Pennsylvania Catalogue. This source includes all students attending the University of Pennsylvania in 1921, and provides each student's name, hometown, residence during the academic year, and class standing.

Some of these documents are available in PDF format in online public repositories such as the HathiTrust Digital Library, the Internet Archive, or the digital collections of specific colleges. In cases where the records are not available online, we contact university librarians and request scanned copies of the documents. If both of these strategies prove unsuccessful, we hire local college students to visit the college library and make the necessary scans. Once the PDFs are acquired, we digitize them using a combination of automated optical character recognition (OCR), natural language processing (NLP), and hand transcription processes. In total, a team of 300 research assistants and freelancers have worked for over 40,000 hours to collect, digitize, review, and standardize the nearly 2.5 million unique student records at these 65 elite colleges.

One important strength of these documents is that, unlike student surveys, they do not

[8] We define a "degree-seeking undergraduate" to be any student enrolled full-time in their first degree program. This excludes, for example, special students, students with a previous degree, students only pursuing certificates, etc. See Data Appendix Section A.5 for more details.

[9] Unlike the elite private schools, these public institutions are defined to be "elite" based on relatively modern rankings. That is out of necessity, as we do not have a more historical concept of the prestige of public schools. In fact, prior to the 1980s, comprehensive comparative rankings of U.S. colleges were not widely available (Espeland and Sauder, 2009).

rely on students to voluntarily report their information. Although some document styles may not contain all classes (e.g., a yearbook may only include a list of seniors instead of all undergraduates), whenever possible we include all enrolled students in our dataset in order to improve comparability with the modern sources.[10,11] Doing so is particularly important earlier in the data, as college completion rates were relatively low (even when focusing on elite institutions).

Of the 65 schools in our sample, the average private elite college had 469 students enrolled in a given year in the 1910s and 727 in the 1960s. For elite public schools, those numbers were 527 and 1,761, respectively. In total, we have nearly 2.5 million unique students during the 1915-1966 period.

Linking College Registers to Population Censuses. The college registers include students' first and last names and often (82% of cases) their hometowns, but lack information on the socioeconomic backgrounds of the students.[12] To obtain this information, we link students from the registers to a U.S. census record prior to when they attended college so we can observe them as children living with their parents. The linking uses information on students' exact first and last names, hometowns when available, and an approximate birth year (assuming they would have been between the ages of 17 and 28 when they started college).[13] In this step, we take advantage of the fact that most female students were unmarried at the time at which they attended college and thus still had their maiden name, allowing us to use their last names as reported in the college register to link them back to an earlier census.[14] Therefore, unlike most studies using linked historical data, ours includes both men and women in the analysis (and has similar matching rates for both).

To obtain information on students' socioeconomic backgrounds, we link students to the *two closest* censuses prior to them first attending college (e.g., we link students who we first observe attending college in 1911-1920 back to the 1910 and 1900 censuses). We link to multiple census records because individuals who started college at older ages (say, at age 25) might be less likely to be observed living with their parents in the closest census (which we need so as to observe parental characteristics). Whenever we match someone to more than one census, we keep the observation corresponding to the census that is closest in time to the student's first year of college (e.g., if we observe an individual who started college in 1953 living with their parents at age 5 in the 1940 census and at age 15 in the 1950 census, we keep

[10]For those students who we observe in more than one year, we only keep one observation in our dataset.

[11]Since comparing students who survived until their senior year with those who attended for a shorter period could bias our comparison (e.g., if students who graduate from college are from wealthier backgrounds than those who simply attend), we also report separate analyses in which we include only schools that report 1) all enrolled students, 2) freshmen, and 3) seniors. In all samples, we find that our outcomes of interest are similar. Additionally, we note that contemporary dropout rates are extremely low at elite schools, and so a record of modern students who were ever enrolled is quite similar to, for example, a modern list of only freshmen or seniors.

[12]Appendix Figure A4 shows the proportion of successfully geocoded individuals by entering cohort.

[13]We use a relatively wide age band so as to be able to capture individuals who might have attended college at older ages (for instance, returning WWII veterans who benefited from educational benefits provided by the G.I. Bill).

[14]Among women aged 16 to 27 who were attending school as of the 1930 census, 97.3% were single.

the observation corresponding to the 1950 census). We make this decision so that parental background is measured as close as possible to when students' college enrollment decisions occurred.

Figure A1 shows the match rates, separately based on gender, year of attendance, and whether hometown information was used, that we achieved by linking the student records to different population censuses. Across all years, we match around 55% of individuals to at least one childhood census. These match rates are largely comparable across years, gender, and group of colleges (see Figure A2).[15] We are able to match a higher rate than is typical, as even though college registers do not have exact information on students' year or place of birth, we do have information on students' hometowns for a large proportion of the sample (and we additionally allow ourselves the possibility of matching to two census records, instead of just one).[16]

Because we are only able to match a subset of students, a concern with our approach is that the linked sample may not be representative of the population of students at a particular college. We deal with this issue in two main ways. First, we re-weight the data so as to adjust for differences in matching rates by college, gender, and entry cohort. By doing so, we make sure that our sample is balanced in terms of the relative size of each of the colleges. Second, we note that for the period in which our data overlap with alternative nationally representative sources capturing a similar population of students at selective colleges (such as the UCLA Freshmen Survey described in detail below), students' socioeconomic backgrounds in our matched sample are similar to those in such alternative data sources.

Once we have linked a student to a census record, we restrict the sample to students who we observe in the relevant census i) at the age of 18 or younger, and ii) living with their father. These restrictions enable us to observe the father's labor market outcomes.[17] After imposing these restrictions, we obtain information on parental background for close to 90% of the students who we match to at least one census. Combined with a matching rate of approximately 55%, this implies that we observe family background for close to 50% (90% of 55%) of the full sample of students from which we started (see Figure A3 for a breakdown of the different steps).

In our final linked sample, we have 1.2 million students across all elite schools in the 1915-1966 period. Of these, 375,000 are from elite private colleges and the remaining are from elite public. This sample is 69% male, which is similar to the male-to-female ratio in the register data prior to linking, as well as the gender ratio reported by Goldin, Katz and

[15] To deal with any remaining differences, throughout the paper we re-weight the data so as to account for differences in matching rates by gender, college, and entry cohort.

[16] A group that is missing from our analysis on family backgrounds are international students who moved to the U.S. on their own to attend college (in this case, we are unable to observe their parents in a U.S. Census). International students accounted for about 3-4% of students in our historical data (see Figure 4). We note that information on the economic backgrounds of international students is also absent from the data in Chetty et al. (2020) that we use to provide a past-present comparison of students' backgrounds at elite colleges.

[17] We use the father's, rather than the household's, income. As of the 1920 census, less than 7% of mothers in the population reported an occupation, approximately half of whom were engaged in unpaid family work (e.g., keeping house, family farm laborer, etc.). Of mothers with children who attended college, these rates of paid employment are even lower.

Kuziemko (2006) for the aggregate college-going population in this time period. In terms of demographics, our sample is 98% white, 35% were living in the northeast prior to attending college, and 89% have at least one American-born parent.

We provide further details on our linking strategy and sensitivity checks in Online Appendix Section A.8.

Constructing Proxies of Parental Income in the Historical Data. As the 1940 census is the first U.S. census to include information on individual earnings, we need to construct proxies ("income scores") for the incomes of fathers who we observe prior to 1940 in order to obtain measures of parental income that are consistent across all sample years.

Our preferred approach is to use a statistical model to predict income from a rich set of covariates for men aged 30–59 in non-agricultural occupations in the 1940 census.[18] We then use this model to predict income for fathers in all census years. This method for assigning income has recently been used in Abramitzky et al. (2021), and is similar to the machine-learning approach for calculating occupational scores proposed by Saavedra and Twinam (2020). Using a broader set of characteristics for income prediction, rather than directly using IPUMS "occscore" variable (which assigns each occupation its median earnings in the 1950 census), enables us to capture differences (such as racial and regional income gaps) within occupations.

For fathers who we observe in the 1940 or 1950 census, we can, in principle, also use actual individual-level income data. A disadvantage of this approach is that we do not observe a comparable measure in earlier years and hence cannot build a consistent long-run series. Furthermore, although the 1950 census includes a question on total income, it is challenging to use such a variable as it is measured with substantial error in the currently available version of the 1950 full-count census.[19] For the 1940 census, the income measure includes only wage and salary income and does not include capital income or income from self-employment (including farm income). We implement two adjustments so as to be able to assign a measure of income for all individuals (including the self-employed) in the 1940 census. In order to assign income to farmers we use a method developed by Collins and Wanamaker (2022), which makes use of the fact that the 1940 census records the incomes of farm laborers, and that later census years record how much farmers earn relative to farm laborers. We thus compute farmer income by multiplying the known income of farm laborers in 1940 with the ratio of earnings for farmers versus farm laborers in the 1960 census, by region, race, and immigration status.[20] Second, to assign income to other self-employed individuals, we

[18]Specifically, we regress (positive) log income in 1940 on the following fixed effects: detailed 3-digit occupation, age, race, current state of residence, and country of origin. We also include interaction terms between 3-digit occupation and census region fixed effects, country of origin and current census region fixed effects, as well as 3-digit occupation and country of origin fixed effects.

[19]IPUMS estimates that only 70-75% of observations in the current version (as of November of 2024) accurately reflect the actual value provided by respondents (`https://usa.ipums.org/usa/full_count.shtml#January2024`).

[20]Collins and Wanamaker (2022) use the ratio of earnings for farmers versus farm laborers by region and race. We add an additional adjustment by immigration status. We further follow Collins and Wanamaker (2022) by scaling up the earnings of farm laborers and farm managers to account for non-cash (in kind) earnings. This measure places the average farmer father in the 36th percentile of the income distribution in 1910. In 1910,

follow Collins and Wanamaker (2022) and use the 1960 census to calculate the ratio of mean income for self-employed non-wage workers in an occupation to the mean income of wage workers in the same occupation. We then use this ratio to scale predicted wage earnings of the self-employed in the 1940 census.

We use these income measures (either actual income for fathers we observe in 1940 or income scores) to rank fathers relative to all other fathers with children born in the same year in the national income distribution. In Section 3.4, we check the robustness of our results to alternative income scores. An important robustness check that we conduct is to show that for those samples in which it is possible to use either income scores or *actual* individual-level income, our results are similar regardless of which one we use. This exercise helps validate our use of predicted income to characterize the socioeconomic backgrounds of students who attended college prior to 1940.

2.2 Students in Elite Colleges After 1966

Our post-1966 data on the backgrounds of students in selective colleges are based on three sources: the publicly available version of Opportunity Insights, the CIRP Freshman Survey (TFS) collected by the UCLA Higher Education Research Institute (HERI), and institution-level data from the Integrated Postsecondary Education Data System (IPEDS). We describe these three sources in turn.

Opportunity Insights. To characterize the parental income of students at different U.S. colleges in the more recent time period, we use publicly available data from Chetty et al. (2020). These data were built by linking parents to children using U.S. tax records. The sample underlying these data comprises over 30 million children born between 1980 and 1991 whose parents filed a tax return in the U.S. at some point when the child was between the ages of 12 and 17. This sample covers more than 98% of all US-born children in these birth cohorts. Individuals in this sample typically started college in the 1998-2009 period. We use the publicly available version of the data, which reports college-by-college aggregate statistics (including the share of students by parental income quintile). These statistics, however, are not reported separately by gender on a college-by-college basis, and so in some figures we are forced to report statistics pooling all students together.

College attendance in these data is measured using two sources: federal tax records and Department of Education records. First, all higher education institutions qualifying for federal financial aid must file a 1098-T tax form in each calendar year for any student who pays tuition. From these records, it is possible to infer the college attended by those students who pay tuition. Second, since some students do not pay tuition (typically low-income students receiving financial aid), these data are complemented with Federal Pell Grant records from the Department of Education's National Student Loan Data System (NSLDS). In the baseline definition of college attendance, students who attended more than one college are assigned to the college they attended the most between the ages of 19 and 22. The data do not include information on whether a student actually graduated from college. Hence, just as in

farmers constituted 42% of all fathers and 36% of fathers of children with some college education.

our historical data, students who eventually drop out are included in the sample.

Data on parental income come from federal income tax records that span 1996 to 2014. Parental income is measured as the average total parental income when the child was between the ages of 15 and 19.[21] As with the historical cohorts, we use these measures to rank parents relative to all other parents with children born in the same year.

CIRP Freshman Survey (TFS). This nationally representative sample is based on an annual survey that has been given to full-time, first-enrolled freshmen students across different U.S. colleges since 1966. The full sample includes information on the demographics (including parental income, race, and hometown) of more than 14 million college students, as well as information on college characteristics (such as its selectivity – based on average SAT scores in 2022 – and private/public status). A key advantage of this survey is that it starts when our historical data collection effort ends. Hence, combining this survey with the college register data enables us to cover the full period from 1915-2008.

There are, however, two main challenges when using this survey in combination with the historical registers and Opportunity Insights data. The most important challenge is that we cannot identify specific institutions in TFS. (The data do, however, include a re-coded school identifier that we can use to track the same institution within the survey over time.) The second challenge is that schools do not necessarily participate in every year of the survey. Moreover, even among schools that do participate consistently, the available information varies from year to year.

To deal with these limitations, we identify institutions within the survey that fit two criteria: (1) are as comparable as possible to two target groups of institutions (elite private and elite public colleges) and (2) participate in the survey frequently enough so that we can follow a stable group of institutions over the full 1966-2008 period.

To create a set of comparable TFS institutions to that in our data, we use the following procedure. First, we create groups of institutions in our sample of 65 colleges based on: (1) whether they are private or public, (2) their geographic region, and (3) their median combined SAT score (rounded to the nearest hundred). This criteria divides our 65 universities into 26 different groups. Second, we create an equivalent grouping for the institutions we observe in TFS data.[22] Third, we only keep the observations in TFS that belong to groups of institutions that overlap with those of our 65 colleges. This procedure guarantees that the institutions that we follow using TFS data are similarly selective and are located in the same areas of the

[21] When we observe parental income in our historical cohorts, children are in some cases younger than 15. This could be problematic for our past-present comparison to the extent that parental income observed when children are below the age of 15 might be an imperfect proxy for parental income when children are closer in age to college attendance. In Section 3.4, we show that our results are similar if we restrict the sample to students for whom we observe parental income scores when students are closer to attending college (at ages 14-18).

[22] We define regions using the smallest level of geographic aggregation that is reported in TFS data (the "Office of Business Economics" region, or OBE region). States are categorized into different regions according to the following criteria: New England (CT, ME, MA, NH, RI, VT), Mid-East (DE, DC, MD, NJ, NY, PA), Great Lakes (IL, IN, MI, OH, WI), Plains (IA, KS, MN, MO, NE, ND, SD), Southeast (AL, AR, FL, GA, KY, LA, MS, NC, SC), Southwest (AZ, NM, OK, TX), Rocky Mountains (CO, ID, MT, UT, WY), and Far West (AK, CA, HI, NV, OR, WA).

country as those in our main sample.

Finally, to construct a stable group of colleges throughout the period, we focus on the set of colleges that appear in TFS at least once in *every decade* throughout the 1966-2008 period.[23] Following UCLA Higher Education Research Institute practice, we only include a school in our sample if it is included in "The American Freshman: National Norms" reports that are published every year, as inclusion in this report signals that the school had a satisfactory participation rate. We then aggregate the data over ten-year periods, re-weighting the data such that each college maintains a stable share of students within the ten-year period.[24] There are 15 private colleges and 15 public colleges in TFS that fit all of our criteria. The data include nearly 1.15 million students in these colleges, of which 250,000 attend a private institution and 900,000 attend a public one.

Information on parental income is self-reported by students, who are asked to provide their best estimate of the income of their parents (before taxes, and in intervals rather than exact values) in the previous year. To assign students to parental income quintiles, we first use yearly data from the Current Population Survey (CPS) to rank all parents of 17-year-old children on the basis of their combined income, by year, starting in 1966. We then assign students in TFS to parental income quintiles assuming that income is uniformly distributed within each of the self-reported intervals.[25] The top interval is always higher than the cutoff for the top quintile, and so we assign all students in the top income interval of the survey to the top quintile. This method for assigning parental income quintiles is similar to the one used in Jacome, Kuziemko and Naidu (2023).

As TFS data overlaps with Opportunity Insights data for the most recent period and with our college registers data for its earlier years, we can compare the share of students who are assigned to each parental income quintile based on these alternative sources.

Figure B2 shows that TFS appears to provide an overall close approximation to our target sample of institutions. For the late 1960s period, the survey estimates a similar fraction of students from bottom 20% (5-7% in private colleges and 8-9% in public colleges) and top 20% families (55-60% in private colleges and 35-40% in public colleges) compared to our register data. For the most recent period, the survey estimates a similar proportion of students from the bottom 20% in private colleges (4-5%) compared to the Opportunity Insights data, but a slightly higher proportion in public colleges (8-9% as opposed to around 5%). Finally, the fraction of students from the top 20% in private colleges appears underestimated by about 10 percentage points in TFS relative to the more recent Opportunity Insights data. This underestimation is smaller when focusing on public colleges.

Integrated Postsecondary Education Data System (IPEDS). We use data from IPEDS

[23]Note that even though we cannot identify specific colleges, we can still use the re-coded college identifiers in the sample to follow the same college over time. We require the college to show up in the survey at least once per decade (rather than every year) as there are no universities in our target set of institutions that participate in the survey every year.

[24]For instance, if a school shows up only once during a decade and another school appears ten times, observations from the first school will receive ten times the weight as those from the second school.

[25]For instance, if an interval in TFS spans $10,000 to $20,000 and the cutoff value between the first and the second parental income quintiles is $12,000, we randomly assign 20% of the students in the income interval to the first quintile and 80% to the second quintile.

spanning 1984 to 2022. These data are based on surveys conducted by the U.S. Department of Education and are reported at the level of the institution. An advantage of these data is that they enable us to identify specific institutions. A limitation is that while they include information on students' race and home state, there is no information on parental income. Figure B3 shows that the recent trends in the relative representation of Blacks in selective private and public colleges are similar regardless of whether we use the IPEDS data or TFS, again validating the notion that TFS provides a reasonably close approximation to our target group of schools.

2.3 Nationally Representative Data on the College-Going Population

We complement our college-by-college data with nationally representative data on the socioeconomic backgrounds of the general college-going population. Each of these samples enables us to characterize, for different college cohorts, the backgrounds of the "average" individual with some college education (regardless of the college that they attended). Table A1 shows the cohorts included in our analysis and the corresponding data source that we use for each.

1910-1940 and 1920-1950 Linked Samples. The data on the earliest birth cohorts (individuals born between 1894 and 1920) in our analysis are based on samples linking individuals from: (1) the 1910 to the 1940 full-count U.S. population censuses and (2) the 1920 to the 1950 full-count U.S. population censuses. The birth cohorts included in these samples are those that are: (1) old enough to have likely started college by the latter census year (that is, at least 30 years old in 1940 or 1950), and (2) young enough to be living with their parents in the earlier census year (that is, at most 16 years old in 1910 or 1920). These samples enable us to observe the family backgrounds of a large sample of individuals, some of whom reported at least some college education in the 1940 or 1950 censuses. For 1910-1940 links we use data from the Census Tree (Buckles et al., 2023), as it enables us to follow both men and women. Since the Census Tree does not link individuals to the 1950 census, we use the crosswalks built by IPUMS Multigenerational Longitudinal Project (Ruggles et al., 2024) for our 1920-1950 links.[26]

Survey of Occupational Change in a Generation (OCG). For *men* born between 1911 and 1943, we use data from the Survey of Occupational Change in a Generation (OCG). This survey was conducted as a supplement to the 1962 and 1973 CPS. It is a nationally representative survey of *men* and includes information on each respondent's educational and family background (including parental occupations). We restrict the data to individuals between the ages of 30 and 60 at the time of the survey. We impose this age restriction so as to maximize the chance of observing whether someone ever attended college while minimizing concerns of selective mortality at later ages. Unfortunately, this survey did not include women so women are not included in some of the figures focusing on these birth cohorts.

[26] Although the MLP 1920-1950 data include both men and women, the data on women only includes those who did not change their surnames in between census waves (e.g., women who remained single, women who did not change their surname upon marriage). Hence, we only use the data on men for the MLP 1920-1950 sample.

CIRP Freshman Survey (TFS). In addition to enabling us to characterize the backgrounds of students at selective colleges, as discussed above, TFS is a nationally representative sample that also allows us to characterize the general college-going population for those who first attended college no earlier than 1966.

U.S. Population Census and American Community Survey (ACS). Finally, we use cross-sectional data from the census and the ACS spanning 1940 to 2020 so as to characterize the educational attainment of the U.S.-born population by birth cohort, gender, and race. We restrict the data to individuals who are between the ages of 30 and 60 at the time of the survey.

3 Socioeconomic Origins of College Students

In this section, we provide evidence of how the socioeconomic origins of college students have changed since the 1910s. To do so, we first construct statistics on the parental income rank of students who attended any college over the course of the 20th century. We then compute similar statistics for students who attended elite private and elite public colleges. We finally characterize trends by race and geographic origin, both for the general college-going population and for those attending elite institutions.

3.1 Parental Income

All Colleges. Figure B1 shows that there has been a sharp increase in the proportion of individuals attending college. This figure, which focuses on the U.S.-born population, shows the proportion of individuals with at least one year of college, by typical college entry year (at age 18) and gender, using data from the U.S. census (1950-1990) and the ACS (2000-2020). Among individuals turning 18 prior to World War II, attendance rates were similar for men and women, hovering just over 15%. The proportion of individuals attending college increased sharply post World War II, a period that is often referred to as the "Golden Years" of American higher education Goldin and Katz (1999). Indeed, by the early 1970s close to 60% of individuals turning 18 eventually attended college - a nearly threefold increase in just two decades. The sharp rise in college attendance in the late 1960s, and subsequent decline in the mid-1970s, has been attributed to incentives generated by the Vietnam War draft (Card and Lemieux, 2001). The proportion of individuals attending college started rising again in the early 1980s, particularly among women, such that by 2010 nearly 70% of women and 60% of men attended college at some point in their adult lives.

Panels (a) and (d) in Figure 2 show that the large expansion in college enrollment that took place in the 1945-1970 period was accompanied by an increase in the economic diversity of the college-going population. In these two panels, we plot the proportion of college students from families in the bottom 20% and top 20% of the parental income distribution, by college entering class and gender. Panel (a) shows the results for men and Panel (d) for women. Prior to World War II, less than 10% of college students (male or female) came from families in the bottom 20% of the income distribution and approximately 40% came from families

in the top 20%. But by the end of the "Golden Years", the proportion of students from families in the bottom 20% increased by almost 50% (from less than 10% to nearly 15%) and the proportion of students from families in the top 20% declined by about a quarter (from 40% to 30%). Since the early 1970s, the proportion of students from families in the bottom 20% has continued to increase among women, but has remained fairly stagnant for men. Indeed, today women from bottom 20% families account for nearly 20% of the female college-going population, while bottom 20% men only account for around 13%. There has been a corresponding decline in the proportion of students from the top 20%, more so for women than for men.[27]

Elite Colleges. In contrast to the trends observed in the general college-going population, the proportion of lower-income students has not increased at most elite private institutions. In Panels (b) and (e) of Figure 2, we show analogous plots to those above, but now limited to students in elite private institutions. Panel (b) focuses on men in elite private colleges. This panel shows that the proportion of men students from families in the bottom 20% of the income distribution has remained close to 5% throughout the entire sample period. In Panel (e) of Figure 2 we pool together all women in elite private colleges (elite women's colleges plus women in coeducational private colleges) and also find that the proportion of women from lower-income families in these colleges remained around 5% throughout the sample period.

Although there have been limited changes in the representation of lower-income students over the 20th century, we find that the proportion of students from upper-income families in elite private colleges decreased after WWII but has bounced back since the 1980s. As a consequence, the proportion of students from the top 20% is now around 70%, similar to what it was in the earlier part of the 20th century. Because the proportion of lower-income students has remained mostly constant throughout the period, it follows that the recent gains in the representation of upper-income students have occurred mostly at the expense of those from *middle-income* (percentile 20-80) families (see Figure B5). These changes in the representation of the top 20% in elite colleges echo broader trends affecting the U.S. income distribution in the same time period, where inequality follows a similar U-shape pattern (declining after WWII to then revert (Piketty and Saez, 2003)). A possible interpretation of these concurrent trends is that when inequality is high, higher-income families have stronger incentives to invest in the "admissions race" as the stakes associated with attending a selective university increase.

We find relatively little heterogeneity in these trends across elite private colleges, in terms of both the levels and trends of enrollment of students by parental income. Panel (a) in Figure B4 recreates the same plots as Figure 2, but now for men in the Ivy Plus and Little Ivies

[27] Bailey and Dynarski (2011) investigate inequality in access to higher education in the U.S. and show that the income gap in access to college widened when comparing the 1961-1964 birth cohorts (who would typically start college between 1979 and 1982) to those born 1979-1982 (who would typically start college between 1997 and 2000). While the likelihood of attending college has increased more in *absolute* terms for children of higher-income families than for lower-income children, the *proportional* increase has been of a very similar magnitude. Hence, the *relative* probability that higher- and lower-income students would attend college (which is the focus of our figure) has remained mostly flat since the 1980s.

separately.[28] The figure shows similar trends, with the proportion of students from lower-income families consistently below 5% throughout the entire period. Furthermore, there is relatively little heterogeneity even when focusing on *specific* colleges (see Figure B6). This figure shows, for each private college in our data, the proportion of students from bottom 20% families in 1915-1925, 1955-1965, and 1998-2009. Most elite private colleges are clustered together, with the representation of bottom 20% students ranging between 2-7% in the earlier part of the sample and 3-5% today. For example, the share of lower-income students at Harvard, Yale, and Princeton has been less than 5% through the entirety of the last century. In these cases, there is at most a modest increase in the representation of poorer students. At Stanford, the historic period had an average share of lower-income students of less than 3%, compared to just over 4% in the more recent period. These examples are representative of the vast majority of elite private schools, although we observe some heterogeneity when focusing on women. Specifically, elite women-only colleges have seen larger increases in the representation of lower-income students compared to co-educational colleges, with the proportion reaching as high as 7% for the more recent cohorts (see Panel (b) of Figure B4 and Panel (c) of Figure B6).

At elite public institutions, we once again find a remarkably stable share (approximately 5%, slightly higher than at private colleges) of lower-income students (Panels (c) and (f) of Figure 2). Similar to elite private colleges, elite public colleges also experienced a relative decrease (and subsequent reversal) in the proportion of students from upper-income families after WWII. An important difference relative to private colleges, however, is that the representation of such students has always been lower at public colleges compared to their private counterparts. At elite public schools, both for men and women, the proportion of students from the top 20% has never exceeded 55% of the student body (compared to close to 65-70% in elite private colleges).

Nevertheless, we find considerable heterogeneity across elite public schools, with some schools experiencing a sharp increase in economic diversity over the course of the last century, and others experiencing a decline. Figure B7 shows the proportion of students from bottom 20% families in 1915-1925, 1955-1965, and 1998-2009 for the public colleges in our data. The figure shows that colleges such as UC Berkeley, UC Davis, UCLA, and SUNY, all experienced an increase in the proportion of students from the bottom 20%. Take UC Berkeley, for example, where the share of lower-income students increased from 4% in the 1960s to nearly 9% in the most recent period. In contrast, some schools (for instance, Georgia, Georgia Tech, and UNC) have experienced a decrease in the proportion of students from the bottom 20%.[29]

[28] Note that, when we split these figures by gender, the composition of universities changes over time as some colleges switched from men-only to coeducational over the 20th century.

[29] Part of these trends likely reflect regional convergence in income over the course of the 20th century (for instance, the convergence between the U.S. South and the rest of the country). Public colleges located in areas of the country that lagged behind in the early 20th century might have started attracting students ranked higher in the national income distribution as these areas converged to the rest of the country.

3.2 Race

We next investigate whether there have been changes in the racial mix of the college-going population, focusing on the relative representation of Black students.[30] We calculate, for each college entry year and gender,

$$\frac{\%\text{ of Black Students in College-Going Population}}{\%\text{ of Black Individuals in General Population}},$$

where the % of Black individuals in the general population is computed based on individuals age 16-20 at the time of college entry.[31]

All Colleges. We find a substantial increase in the relative representation of Black students in the general college-going population over the course of the 20th century. Panel (a) of Figure 3 shows the relative representation of Black students among those with some college education, by typical college entry year and gender. The relative representation of Black students among those with some college was only 0.3 among those starting college in the early 1920s. This representation increased rapidly in the post-WWII "Golden Years" of American higher education, reaching close to 0.8 among men and 0.9 among women by the early 1970s. This convergence has slowed down among more recent cohorts: the relative representation of Black students is similar among those starting college in the early 2010s compared to those starting 40 years prior. Moreover, this relative representation is lower today than it was for cohorts starting college in the late 1980s. Across almost the entire period (with the exception of the early 20th century), the relative representation of Black women is higher than that of Black men. Moreover, the gender gap in the relative representation of Black individuals has remained similar since the early 1970s.

Elite Colleges. Panel (b) of Figure 3 reports an analogous measure to the one documented above, but now limits the sample to students in elite private colleges. The figure illustrates two main facts. First, the relative representation of Black students in elite private colleges has been lower than that in the general college-going population throughout the entire sample period. However, relative to the early 20th century, there has been a notable increase in the relative representation of Black students in elite private colleges. The relative representation of Black students hovered around 0.1 until the 1950s. By the early 1960s, such representation had increased to around 0.2. It then increased sharply over the 1960s such that by 1970 it had reached 0.4 among men and 0.5 among women. After remaining nearly constant for close to four decades, the relative representation increased again in recent years and is now slightly above 0.5 for women and slightly below 0.5 for men. This pattern is consistent with the "episodic" nature of Black economic progress that has been documented in the broader labor market over the course of the 20th century, where Black individuals made substantial gains

[30] Other racial minorities are either not possible to track over time or constitute an even smaller fraction of the college-going population in the historical data. For instance, the Hispanic category was only introduced in the 1970 U.S. Population Census, preventing us from computing the share of Hispanic students, by college, historically. Asians constituted just 0.37% of students in our pre-1960 data.

[31] We focus on this measure rather than simply the proportion of Black students, as the proportions of Black individuals in the overall U.S. population is not constant over time (due to, for instance, changes in the proportion of immigrants in the U.S. population).

during the Civil Rights era but more modest gains since (Donohue and Heckman, 1991). By and large, these patterns are present across all school groups within the broader category of elite private colleges, as shown in Panels (a) and (b) of Figure B8. One notable trend is that, in recent years, the relative representation of Black students is larger at Ivy Plus colleges than at elite women-only colleges or the Little Ivies.

Panel (c) of Figure 3 shows that there has also been an increase in the relative representation of Black students at elite public colleges. Until the 2000s, Black students were similarly underrepresented at elite public colleges and elite private colleges. However, the relative representation of Black students is now slightly higher at elite private schools compared to their public counterparts. This higher representation at private colleges is driven by a slight increase in the representation of Blacks at such colleges, but also by a recent *decline* in their representation at public colleges.

3.3 Geographic Origin

We next investigate long-run trends in the proportion of students who are from outside the college's region.[32] We find that elite private institutions have seen sharper increases in the geographic diversity of their student bodies than have their public counterparts.

All Colleges. We do not have yearly data on the proportion of students attending a college outside of their region prior to 1966 (the first year of the TFS Survey). We can, however, use the 1950 U.S. Census to estimate such a proportion as of 1950. Specifically, we first identify college "freshmen" in the 1950 census by restricting the sample to individuals who: (1) report having attended school at some point since February 1st of 1950 (the census took place on April 1st) and (2) have some college but less than one full year of completed college (so as to capture individuals who would have started college in the Fall of 1949). We then use their location in the 1950 census to identify the location in which they attend college and their location in April of 1949 (based on the question on previous place of residence) to identify the location of their hometown.[33]

We combine this estimate with information from TFS to produce a series spanning 1950-2008.[34] Starting in its 1982 wave, TFS includes a student's home ZIP code, which we use in combination with the college's "Office of Business Economics" (OBE) region (that is, the smallest level of geography reported at the institution level) to calculate the proportion of students attending a college outside of the region of their hometown.

The gray bars in Panel (a) of Figure 4 show relatively modest changes in the proportion of students who attend college far from home, with a slight increase in recent cohorts. By 1950, less than 10% of students attended college outside of their region. This proportion is around 15% among more recent college cohorts.

[32] Regions are classified according to the OBE region rubric (discussed above).

[33] In this step, we take advantage of the fact that the 1950 census is the first to report college students where they attended school, rather than with their parents' household.

[34] Note that TFS survey only added a question on students' ZIP codes starting in 1982, so we lack information for students' geographic origins in the 1965-1980 period.

Elite Colleges. The relatively modest changes in the geographic origins of the "average" college student contrast with the substantial changes among those attending elite private colleges. Specifically, Panel (a) of Figure 4 shows that there has been an increase in the geographic diversity of students in such colleges. There was remarkable stability in the first half of the 20th century, with approximately 30% out-of-region students over the 1915-1950 period. This proportion increased gradually after 1950, such that by the mid-1960s more than 40% of students at these colleges came from out of region. Today, almost 60% of students at elite private colleges are from out of region - double the share at the beginning of the 20th century.[35] Panel (b) of Figure 4 shows that international students have also become an increasingly important group contributing to "geographic diversity" at elite private universities. In the first half of the 20th century, this group constituted less than 5% of students. This proportion started rising in the late 1970s, and international students now constitute close to 12% of students at elite private universities.

We document more modest changes in geographic diversity at elite public colleges (Panel (a) of Figure 4). The proportion of out-of-region students was approximately 10% over the 1915-1966 period. This proportion increased in recent decades such that today close to 20% of students in these colleges come from outside of the institution's region (about one third of the corresponding proportion at private colleges). Interestingly, the proportion of international students has increased substantially at elite public colleges. Indeed, a large fraction of the "out-of-region" students are in fact students from outside the country. The proportion of international students appears to increase sharply after the Great Recession, consistent with public colleges' attempts to substitute state funding with tuition revenue.

3.4 Robustness

Linking and Sample Selection. Figures B12 to B15 show that our main conclusions are robust to alternative samples and census linking approaches. First, because we do not observe hometown information for all students in our data, we reproduce Figures 2 and 3 using the sample of individuals that result from a linking strategy that does *not* use hometown for disambiguation (so as to link all individuals using a consistent set of variables) (Figures B12 and B14). Second, because hometown information may be less accurate for individuals who we observe in a census further in time from when we observe them in the college registers, we reproduce our main findings using a sample of students who we observe *at most* five years away from the time at which they started college (Figures B13 and B15). In all cases, we find similar trends in terms of access by parental income and race to our baseline results.

Alternative Income Measures. Figure B9 shows that our main conclusions with respect to access by parental income are robust to alternative methods for imputing income in the historical data. First, we use an alternative method for imputing farmers' income. This alternative approach uses county-level data from the 1900 Census of Agriculture to calculate farmers' net earnings (farm income minus expenses) by county. This measure enables us to obtain a

[35] Note that for this figure we can use the registers data directly (without matching to the census) as such data includes information on students' hometowns.

more granular income measure for farmers, a large and heterogeneous group. Second, we use data from the 1901 Cost of Living Survey to assign parental incomes to individuals who we observe in the 1900 or 1910 censuses so that we can have a more contemporaneous data source for such individuals.

Importantly, we also show that for the students who we observe in the 1940 census (for whom we can observe *actual*, rather than just predicted, income), we estimate similar trends in the proportion of students from different parts of the parental income distribution. This similarity is not surprising given that there is a high (and similar across college groups) correlation in parental income ranks based on our imputation procedure and those based on actual income (see Figure B11). Finally, we show that our results are similar if we only include individuals who we observe living with their parents at ages 14-18 (so that family resources are measured closer to the time at which the student started college) (Figure B10).

3.5 Summary of Findings

This section showed that the representation of lower-income students at elite (both private and public) colleges has remained remarkably stable over the course of the 20th century. These colleges have experienced changes in the relative representations of upper- and middle-income students, with upper-income students losing ground in the post-WWII period but re-gaining most of their representation since the 1980s. We also document that despite the relatively modest changes in students' economic origins, there have been important changes in the racial and geographic backgrounds of students at these institutions. Notably, there has been an increase in the representation of Black students as well as out-of-region and international students.

4 The Impacts of the World War II G.I. Bill on Students' Socioeconomic Backgrounds at Elite Colleges

The previous section provided a long-run characterization of the socioeconomic origins of students at elite colleges in the U.S. In this section, we investigate the extent to which such origins were impacted by one of the largest governmental interventions in the history of American higher education: the federal subsidies for college attendance for returning World War II veterans (commonly referred to as the "G.I. Bill"). Beyond its interest as one of the largest government programs of the 20th century, focusing on the impacts of the G.I. Bill will also enable us to shed light on the role that financial barriers might play in accounting for the low representation of lower-income individuals at elite colleges documented above.

4.1 Background and Eligibility Criteria

The G.I. Bill, formally known as the *Servicemen's Readjustment Act*, was enacted in 1944 to provide a range of benefits (including educational subsidies and mortgage assistance programs) to returning World War II veterans. Veterans were eligible for benefits if they either

had been in active duty for more than 90 days or were discharged early through service disability. Educational benefits were additionally restricted to veterans who had started schooling by July of 1951.

The G.I. Bill offered educational benefits for a minimum of one year and a maximum of four years, with the length of benefits depending on both length of service and age. Under these criteria, most returning WWII veterans would have been eligible for maximum benefits (Bound and Turner, 2002). The educational benefits covered tuition and other educational expenses, and were high enough so as to finance attendance at the most expensive private colleges in the country. In addition to covering the cost of tuition, benefits also included a monthly cash allowance paid directly to veterans. The G.I. Bill also provided benefits for "below college" education (for instance, to complete high school) and for job training. In fact, more veterans used the benefits to access such services rather than for college education (Thomas, 2017).

The consensus view is that the G.I. Bill increased the overall educational attainment of beneficiaries, although there is some controversy over the equity consequences of the program (Bound and Turner, 2002; Stanley, 2003; Collins and Zimran, 2024).[36] As such, we are interested in understanding the extent to which the G.I. Bill increased the representation of lower-SES individuals at elite colleges. We emphasize that, as in all existing studies on the G.I. Bill, we cannot separate the effects of the G.I. Bill benefits from the direct effects of participating in the War (nor we can separate the educational aspect of the bill from its other components).

4.2 Empirical Strategy

To assess whether the G.I. Bill increased the representation of lower-SES individuals in elite colleges, our main research strategy exploits the fact that the likelihood of serving in World War II varied sharply depending on an individual year and month of birth. Specifically, there was a sharp decline in the likelihood of WWII service when comparing individuals born before and after the last quarter of 1927. This sharp decline arises as those born after 1927 had not reached the conscription age by the time that World War II hostilities concluded in 1945. We exploit this variation to assess whether students from "highly exposed" cohorts (who likely participated in the War and hence would have been eligible to receive G.I. Bill benefits) had different socioeconomic backgrounds than students from less exposed cohorts attending the same college or group of colleges. Intuitively, if the G.I. Bill increased the representation of lower-SES individuals at elite colleges, we should see that the likely beneficiaries of the program at such colleges had lower SES than other students. We focus on a reduced form specification as we have information on veteran status only for a small fraction (around 10%)

[36] Bound and Turner (2002) show that there were important racial differences in the impacts of the program, with Blacks (particularly those from the U.S. South) benefiting less. Stanley (2003) shows that the impacts of the mid-century G.I. Bills disproportionately benefited the educational attainment of higher-SES individuals. By contrast, Collins and Zimran (2024) find relatively large positive educational impacts of the G.I. Bill on Black individuals.

of the sample.[37]

One challenge in implementing this strategy in our context is that the 1930-1950 censuses do not include information on quarter or year of birth (they just include a respondent's age). To deal with this limitation, we use the fact that for 45% of individuals in our sample we are able to obtain information on their *exact birth dates* by using data from profiles on the Family Tree at familysearch.org. We provide further details on how we assigned birth dates in Online Appendix Section A.10. Importantly, we find little correlation between the likelihood of observing such information and students' socioeconomic backgrounds.

We use this information to estimate the following reduced form specification (analogous to the one in Fetter (2013)):

$$y_{it} = \alpha + \beta \mathbb{1}(yqob_t \leq c) + \gamma(yqob_t - c)\mathbb{1}(yqob_t \leq c) + \delta(yqob_t - c)\mathbb{1}(yqob_t > c) + \varepsilon_{it}, \quad (1)$$

where y_{it} is a characteristic of student i (e.g., whether the student belongs to a family in the bottom 20% of the parental income distribution, whether the student is Black, etc.) born in birth cohort t who is observed attending college in our data. $yqob_t$ represents a student's year and quarter of birth, and c is the eligibility cutoff. Hence, $yqob_t - c$ represents the time in quarters between a student's quarter of birth and the eligibility cutoff. We follow Fetter (2013) and use the last quarter of 1927 as the cutoff. In our baseline specification, we also follow Fetter (2013) and focus on individuals born between 1925 and 1930 (that is, three years before and after the eligibility cutoff). These birth cohorts include particularly "at-risk" individuals who would have been young enough at the time of discharge (in their early twenties at most) to plausibly benefit from the educational benefits provided by the G.I. Bill (Collins and Zimran, 2024). Our coefficient of interest is β, which captures whether there is a jump in students' socioeconomic characteristics around the eligibility cutoff. In a robustness exercise, we drop individuals born in 1928 (who would have been partially exposed to the act given imperfect compliance with the cutoff) and also expand the set of birth cohorts that we include. We cluster standard errors at the birth quarter level.

Figure 5 shows a steep decline in the likelihood of participating in World War II when comparing individuals born before and after the last quarter of 1927. Panel (a) includes all students in our elite colleges data for whom we can observe information on both quarter of birth and veteran status, whereas Panels (b) and (c) focus on students in elite private and public colleges, respectively. The figure also includes two lines of best fit relating the likelihood of being a veteran and quarter of birth (each estimated using only data at either side of the discontinuity). The decline in the likelihood of serving in WWII when crossing the eligibility

[37] We have incomplete information on veteran status for the following reasons. First, among the full count censuses that we use, the 1950 census is the only one with information on WWII veteran status. Hence, to measure veteran status for students we observe living with their parents in a census prior to 1950, we need to match students to the 1950 census (which we do using the 1930-1950 and 1940-1950 census crosswalks provided by IPUMS). Further, the question on veteran status was a sample line question in 1950, and hence only available for 20% of the individuals that we match to the 1950 census. As a result, we end up observing veteran status for only 10% of the students in our sample.

cutoff is large and comparable in magnitude (about 40 percentage points) regardless of the group of colleges that we consider. The figure also shows that the well-known, very high rates of WWII participation for the 1925-1927 birth cohorts are also present when focusing on men who ultimately attended highly selective colleges.

4.3 Findings

Figure 6 shows the relationship between measures of students' socioeconomic backgrounds (y-axis) and their year and quarter of birth (x-axis). The figure focuses on two different measures of students' socioeconomic status: i) whether they grew up in families in the bottom 20% of the parental income distribution (Panels a, b, and c) and ii) whether they grew up in families in the top 20% of such distribution (Panels d, e, and f). If the G.I. BIll increased the representation of lower-SES students in elite colleges, we should see that students born in or prior to the last quarter of 1927 are of lower-SES than students born immediately after.

Against this hypothesis, the figure shows that students' socioeconomic backgrounds, both at private and public colleges, evolve relatively smoothly around the eligibility cutoff. If anything, there seems to be an *increase* in the proportion of students from top 20% families in public colleges (and a corresponding decline in the proportion of students from the bottom 20%). We obtain a similar conclusion (that is, no increase in the representation of lower-income students) if we focus on the proportion of students from the bottom 50% instead of the proportion of students from the bottom 20% (see Figure B16). There is also no evidence of an increase in the proportion of students from the "middle" (percentiles 20-80) of the parental income distribution (Figure B17).

To more formally assess the role of the G.I. Bill, Table 1 presents the results of estimating Equation 1. Panel (a) confirms that there is a large increase in the likelihood of WWII service (of around 40 percentage points relative to a control group mean of about 30% among the cohorts included) when crossing the eligibility cutoff. The increase is similar in magnitude regardless of whether we focus on private or public colleges. Note that, as explained above, this result is estimated on a smaller sample because it requires observing individuals' veteran status, which we only do for those individuals who we can match to the 1950 census and who were asked to answer the sample line questions in that census.

Panel (b) focuses on students' socioeconomic backgrounds. Columns 1-3 include all students in the data, whereas Columns 4-9 split the sample by private/public status. Consistent with the graphical evidence in Figure 6, Columns 1-2 shows that there are small changes in the likelihood that such students would belong to families in either the top or the bottom of the income distribution. When focusing on the full sample of students, the estimates are very small in magnitude and precisely estimated, enabling us to rule out even modest changes in students' socioeconomic backgrounds around the eligibility cutoff. For instance, the likelihood that students across all elite colleges in our data would belong to the bottom 20% of parental income is 0.48 percentage points lower among the "treated" cohorts (relative to a mean of 5.5%) and we can rule out (at the 95% level) increases of more than 1 percentage point. The point estimates are noisier when focusing on students at private colleges but they also show little evidence of changes in students' likelihood of belonging to the bottom or the

top of the parental income distribution. If anything, public colleges exhibit a 1.1 percentage point *decline* in the proportion of bottom 20% students and a corresponding 2 percentage point increase in the proportion of top 20% students, consistent with some adverse equity consequences of the program (although this result is not robust to excluding the 1928 birth cohort from the analysis, see Panel f in Figure B18).

We next assess whether the G.I. Bill changed the racial composition of students at elite colleges. Figure 7 shows the proportion of Black students at elite colleges (y-axis), by quarter of birth (x-axis). Here, we see that the proportion of Black students across all elite colleges appears lower among the treated cohorts, with the difference being driven by elite public colleges. Indeed, Table 1 shows some evidence of a decline in the representation of Black students at elite public colleges (consistent with the evidence in Turner and Bound (2003)).

4.4 Robustness

Main Specification. In our main specification, we include individuals in our sample even if they attended college after 1951 (that is, *after* the expiration of WWII G.I. Bill benefits). We do so in order to provide students of the latest birth cohorts in our sample (that is, our "control group") with enough time to start college. Otherwise, students from "treated" birth cohorts might have looked different than those from "control" cohorts simply because individuals who start college later in life are of a different socioeconomic background than those who start earlier. Furthermore, as discussed above, the G.I. Bill also provided funding for completing *non-college* education. So, in principle, the effects of the program on college enrollment could materialize after its actual expiration as students who completed high school thanks to the program go on to attend college.

However, one issue with this approach is that veterans from the Korean war (June 1950-July 1953) also received educational benefits. And so, individuals in our control group who attended college post-1950 might have done so through a government subsidy and thus not provide a clean comparison group for World War II veterans. To deal with this concern, Figures B18 and B19 show that our results are similar if we limit the analysis to individuals who started college *prior* to 1952 (the year in which the Korean War G.I. Bill law was enacted). The similarity is not surprising given that nearly all (94%) of men in our baseline sample (born 1925-1930) had already started college by 1951. The only result that is not robust to this restriction is the decline in the representation of Black students discussed above.

Finally, Figures B18 and B19 show that our results are also similar if we: (1) expand the range of birth cohorts to include 1923-1930 rather than just 1925-1930, (2) exclude the 1928 birth cohort from the analysis, and (3) include college fixed effects in the regression.

Exploiting Within-Cohort Variation in Exposure to WWII. The empirical strategy above exploits variation *between* cohorts in exposure to WWII. In Figure B20, we instead exploit variation *within* birth cohorts. To do so, we estimate:

$$y_{itc} = \alpha_t + \alpha_c + \beta Veteran_{itc} + \varepsilon_{itc}, \tag{2}$$

where $Veteran_{itc}$ is an indicator that takes on a value of one if person i born in year t who attended college c is listed as a World War II veteran in the 1950 census. One advantage of

this approach is that it does not require that we observe birth quarter information. However, a key disadvantage is that there is likely selection in WWII participation rates complicating the interpretation of the results. Note that, similar to Figure 5, for students we observe in the 1930 or 1940 censuses we obtain information on their veteran status by linking them to the 1950 census (where veteran status was included for only one fifth of male respondents).

Yet, consistent with the findings from the model exploiting between-cohorts variation in exposure to World War II, Figure B20 shows limited evidence that enrolled veterans were of lower socioeconomic status than non-veterans at the same college. Each row in this figure corresponds to the estimate of β in Equation 2 using a different socioeconomic characteristic. If anything, veterans in elite public colleges, compared to non-veteran students at the same college, appear less likely to belong to families in the bottom 20% of the parental income distribution, more likely to belong to families in the top 20%, and less likely to be Black.

4.5 Discussion

Our findings indicate that the G.I. Bill did not appear to substantially alter the socioeconomic mix at the country's most selective colleges and, if anything, possibly lowered the representation of lower-income students in elite public colleges. Below, we provide additional discussion and empirical results to help interpret these findings.[38]

Unlike other programs aimed at increasing enrollment of lower-income students (Burland et al., 2022), G.I. Bill benefits were not limited to such students but rather were available to *anyone* who had served in the War. If students across the parental income distribution benefited similarly from the program, then we might observe no changes in the *proportions* of students from different segments of the parental income distribution in elite colleges (even if the program increased the *absolute* probability of attending an elite college for lower-income students). Furthermore, students higher in the parental income distribution might have been better prepared academically to benefit from the program, in which case the law could potentially even *reduce* the proportion of lower-income students at elite colleges.

Relatedly, G.I. Bill benefits were not targeted on the basis of previous educational attainment. Hence, we might find limited effects on the socioeconomic composition of students at the more selective colleges because academic (rather than financial) barriers were the binding constraint for most of the lower-income individuals "treated" by the G.I. Bill. Indeed, Figure B21 shows that, by 1950, only 35% of men born 1920-1930 at the bottom half of the parental income distribution had completed *high school* (whereas this proportion was 65% for those in the top 20%). We focus on high school completion as of 1950 (after WWII) because individuals in these birth cohorts would have been too young by 1940 to have completed high school. However, note that the G.I. Bill also provided benefits for *high school* completion, so 35% is likely an upper bound of the proportion of bottom 50% students who could have plausibly used the G.I. Bill college benefits.

To assess whether the proportion of lower-income students did not increase in elite colleges because these students did not benefit at all from the program or rather because they

[38] We emphasize that our findings *do not* imply that G.I. Bill beneficiaries did not attend elite colleges. Indeed, the evidence in Figure 5 shows that students in elite colleges exhibited high rates of WWII participation.

benefited to a similar degree as their higher-income counterparts, we estimate a version of Equation 1 using as outcome variable the *probability* of attending an elite college for students at different parts of the parental income distribution. To do so, we extend the sample to *all* US-born men born 1925-1930 for whom we observe exact birth quarter information, rather than just those at elite colleges.

Figure B22 shows the relationship between quarter of birth (x-axis) and the likelihood of attending an elite college (y-axis) for individuals at different parts of the parental income distribution (bottom 20%, middle 20-80%, and top 20%).[39] The outcome in the top row is an indicator that takes on a value of one if an individual is observed attending any elite college, whereas in the middle and bottom rows is an indicator that takes on a value one if the individual attends an elite private or an elite public college, respectively.

Regardless of the group of colleges that we consider, the figure shows little evidence of an increase in the probability of attending an elite college for students at either the bottom or the middle of the parental income distribution. The only discontinuity that we observe is that students from the *top 20%* of the parental income distribution become more likely to attend an elite public college. This finding is consistent with the results above, which showed that the G.I. Bill increased the proportion of top 20% students in elite public colleges.

Table B1 confirms that there is little evidence of the G.I. Bill increasing the likelihood of attending an elite college (private or public) for students at either the bottom or the middle of the parental income distribution. The only statistically significant change is an increase in the probability of attending an elite public college for students at the top 20% of the parental income distribution (a 0.13 percentage point increase in the probability of attendance, relative to a control group mean of 2%).[40]

Overall, these findings suggest that lower- and middle-income students did not become more likely to attend an elite college as a consequence of the G.I. Bill. A potential explanation for this result is that, as discussed above, only a minority of individuals in this group had even achieved the minimum requirement for attending college (that is, completing a high school degree). There are, however, two main additional channels which could have contributed to this muted response: First, WWII veterans might have faced a more "congested" higher education landscape upon returning, which might have made access to elite colleges more competitive (particularly for lower-income students). We note, however, that our data enable us to track veterans' enrollment decisions not just during the "peak" G.I. Bill years but also for several years after (where this congestion effects were likely less relevant). Yet, we still

[39] We calculate such likelihood by dividing the number of students in our sample of elite college students born in a given quarter by the number of all men born in that same quarter. Because we only observe whether someone attends an elite college if we successfully match them to the census, we normalize this ratio using the inverse matching probability by college and entry cohort.

[40] Note that this estimate is based on a reduced form regression relating the likelihood of attendance to being born in or before the fourth quarter of 1927. The first stage (which we can only estimate in a smaller sample) indicates an increase of roughly 40 percentage points in the likelihood of being a veteran for individuals born before before the cutoff. Hence, a 0.13 percentage points increase in the reduced form estimate would correspond to roughly 0.3 ($0.13 \times 0.4^{-}1$) percentage points increase (a 15% increase) in the likelihood of attending an elite public college for individuals at the top 20% of the parental income distribution).

find no increases in the attendance rates of lower-income students at elite colleges.[41]

Second, as emphasized above, we cannot separately identify the effects of the educational subsidies that were part of the G.I. Bill from the effects of the *War itself*. The War experience might have had a disruptive effect on students' educational trajectories, particularly those of lower-income backgrounds. It is possible that this disruption might have reduced the representation of lower-income students relative to a counterfactual in which these students obtained the college subsidies but did not participate in a war. In any case, our findings suggest a more modest role of the G.I. Bill in "changing the face of American colleges" than has previously been argued, at least with respect to elite colleges.

5 The Impacts of Standardized Testing on Students' Socioeconomic Backgrounds at Elite Colleges

The previous section shows that one of the largest government interventions in the history of American higher education had limited impacts in shaping students' socioeconomic backgrounds at elite colleges. We next consider the impacts of one of the major changes in colleges' admission practices over the course of the 20th century: the introduction of standardized testing in admissions.

5.1 Background

In the past, colleges predominantly relied on "feeder" high schools - that is, local high schools that a college believed produced graduates who were both academically and socially prepared to succeed at their institution (Hoxby, 2009). Today, elite colleges select from a much larger pool of potential applicants. Hoxby (2009) argues that a defining factor in this change was the emergence of national standardized tests (such as the SAT and ACT), which reduced the screening costs that colleges face when selecting students.

It is unclear, however, whether the introduction of such tests translated into a more socioeconomically diverse student body at elite colleges. In principle, standardized testing might have given lower-income students the chance to signal their academic ability and colleges the ability to identify the "talented poor". Indeed, this type of argument was often made at the time of the introduction of standardized testing (Lemann, 2000) (and also more recently as elite colleges throughout the country reinstated such exams). On the other hand, individuals from lower-SES backgrounds have, on average, lower scores on standardized tests (see, for instance, Rothstein (2020)) and concerns regarding the equity consequences of standardized exams have often been used as an argument against their use for college admissions (Nettles, 2019).

[41] A related issue is that the G.I. Bill likely had general equilibrium effects, whereby individuals in the "control group" might have been indirectly affected by facing a more competitive admissions environment. Note that such general equilibrium effects would tend to depress the attendance rates of the control group and hence *overestimate* any impacts of the program on attendance. Yet, we find limited evidence of an increase in the attendance rates of individuals in the treatment group even with this potential overestimation.

Below, we investigate whether the *first introduction* of standardized tests, in particular the *Scholastic Aptitude Test* (SAT), translated into increased enrollment of lower-SES students at elite institutions. The SAT, whose design was highly influenced by military IQ tests from World War I (Lemann, 2000), was first administered in 1926 as a general aptitude test. We focus on the SAT as the other main standardized exam that is used today (the ACT) was only introduced in 1959, 33 years after the SAT and close to the end of our sample period. Focusing on the SAT first introduction has the advantage that we can not only investigate the impacts of the exam in an era prior to the emergence of a test-preparation industry, but also see how the impacts change as this industry emerges.

5.2 Empirical Strategy

Our research design exploits the staggered adoption of the SAT across colleges. Table A6 shows the dates at which each of the colleges in our data adopted the SAT together with the data source that we use to obtain each of these dates. We classify a college as having adopted the SAT by year t if the school *required* the SAT for all of its applicants in that year.[42]

Figure 8 shows the proportion of colleges in our data with an SAT requirement, separately for private and public colleges. At the beginning of our sample period the SAT had yet to be created. The exam was first administered in 1926, with only a handful of selective private colleges in the Northeast requiring it to their applicants in that year. By the end of our sample period (the mid-1960s), close to 80% of our colleges had such a requirement.

Figure 8 shows that private colleges adopted standardized testing considerably earlier than their public counterparts. The first public college to adopt the SAT did so in 1955; by that year, more than 80% of the private colleges in our data were requiring the SAT for admissions. However, there is substantial heterogeneity in dates of adoption within private universities (and even within more narrowly defined groupings of elite private schools). For instance, within the Ivy League Princeton was the first adopter of the exam in 1926, whereas Harvard only started requiring it for all applicants in 1935. Among elite women's colleges Wellesley adopted the SAT in 1926, but Barnard only did so six years later in 1932.

We exploit this staggered adoption in an event-study framework. We estimate:

$$y_{ict} = \alpha_c + \alpha_t + \sum_{j=-5}^{j=5} \beta_j SAT_{c,t-j} + \varepsilon_{ict}, \tag{3}$$

where y_{ict} is a characteristic (e.g., whether the student belongs to a family in the bottom 20% of the parental income distribution, whether the student is Black, etc.) of student i who attended college c in year t , α_c are college fixed effects, and α_t are entry cohort fixed effects. Our main variables of interest are the $SAT_{c,t-j}$ dummies, which indicate that a given university was j number of years away from the adoption of the SAT. We omit the year before the

[42]This distinction is important as some colleges first required the SAT for applicants who wanted to be considered for scholarships, rather than for all applicants.

adoption of the SAT, so that each β_j measures the difference in outcomes relative to $j = -1$.[43] In some specifications, we include college-grouping dummies (Ivy Plus, Top Public, Little Ivies, Elite Women) interacted with entry cohort dummies, effectively exploiting variation in the timing of adoption within these more narrowly defined groups of colleges. Alternatively, we also estimate separate models for private and public colleges. We cluster standard errors at the level of the college.

In addition to this dynamic model, we summarize our findings by estimating a standard two-way fixed effects model, namely:

$$y_{ict} = \alpha_c + \alpha_t + \beta SAT_{ct} + \varepsilon_{ict}, \tag{4}$$

where SAT_{ct} is an indicator that takes on a value of one if school c had adopted the SAT by year t. We also estimate versions in which we use the event-study estimators suggested by Sun and Abraham (2021), Borusyak, Jaravel and Spiess (2024), and Callaway and Sant'Anna (2021).

The identification assumption in this research design is that in the absence of SAT adoption, the socioeconomic backgrounds of students in colleges that adopted earlier would have continued trending similarly to the socioeconomic backgrounds of students in colleges that adopted later.[44] This assumption becomes less credible when we pool all colleges (private and public) together, because as shown in Section 3 there are important differences in students' socioeconomic backgrounds (both in levels and trends) across these two groups of colleges. Moreover, Section 3 also shows that there is much more heterogeneity in students' backgrounds within the public rather than the private sector, suggesting that the research design may also be less credible in the former group of colleges compared to the latter. With this in mind, we interpret the evidence of the SAT's impacts on public colleges in a more suggestive manner.

To assess the plausibility of this design, in Figures 9, 10, and B23 we plot the β_j coefficients from Equation 3 using as outcome variables several students' characteristics. Overall, there is little evidence of a pre-trend regardless of the students' socioeconomic characteristic or the group of colleges that we consider. The only exception is that the proportion of Black students in elite private colleges appears to be trending upwards before the adoption of the SAT.

[43] We note that within our sample period, adopting the SAT is an absorbing state (that is, once a school adopts, it keeps using it).

[44] A potential violation of this assumption is that colleges in the "control group" might have been indirectly affected by another school's adoption of the SAT. In this case, the estimated effects might stem from changes in the control group rather than by changes in the college that adopted the exam. For instance, if the exam actually helped a college to recruit high-achieving lower-income students (and there was a relatively fixed pool of such students), adopting the exam in one college might have reduced the representation of lower-income students in other competing colleges, thus overestimating the impacts of the exam. We note, however, that we find limited increases in the proportion of lower-income students even with this potential overestimation.

5.3 Findings

Table 2 presents our baseline two-way fixed effects estimate of the impacts of adopting the SAT. Each of the outcome variables in this table is an indicator that takes on a value of one if a student has a certain characteristic (multiplied by 100 so that can we interpret β in Equation 4 as percentage point differences). Columns 1-3 include students from all colleges in our data, whereas Columns 4-6 and 7-9 restrict the analysis to students at private and public colleges, respectively. Across all colleges, there is a 0.25 percentage point increase in the proportion of students from the bottom 20% of the parental income distribution following the introduction of standardized testing (Column 1). This increase, however, is entirely driven by public colleges, which exhibit a statistically significant 1.2 percentage point increase (relative to an average proportion of 5%). By contrast, private colleges exhibit very limited changes in such a proportion: we can rule out an increase of more than 1 percentage point in the proportion of students from bottom 20% in elite private colleges.

Panels (a) to (c) in Figure 9 show the results of estimating Equation 3 using the likelihood that students would belong to the bottom 20% as the dependent variable. The event-study post-treatment coefficients are all small (less than a 1 percentage point increase) and statistically insignificant when focusing on private colleges. The coefficients are larger in magnitude and become statistically significant two years after the introduction of standardized testing when focusing on public colleges. Note, however, that the increase in the representation of bottom 20% students elite public colleges appears to shrink starting four years after the exam's introduction (suggesting a short-lived impact).

Table 2 shows limited evidence of a change in the proportion of students from top 20% families in private colleges. The point estimate is very small in size (corresponding to 0.04 percentage point increase relative to a mean of 65%) and we can rule out a 2 percentage point decrease in the proportion of such students. Table 2 also shows that the increase in the proportion of students from bottom 20% families in public colleges does not come at the expense of families at the top 20% of the distribution, as these colleges experience a simultaneous *increase* in the proportion of such students (of 1.6 percentage points). Rather, it seems to be driven by a decline in the proportion of individuals from families in the remainder (i.e., the middle) of the income distribution.

Panels (d)-(f) in Figure 9 show the event-study version of these findings. The post-event coefficients are relatively small in size and statistically insignificant when focusing on private colleges, so that the proportion of students from top 20% families remains flat in these colleges following the adoption of the SAT. The coefficients turn significant in public colleges five years after the treatment.

Table B2 and Figure B23 show that we reach a similar conclusion if we focus on alternative definitions of low income (bottom 50%) or on the proportion of "middle-income" students (percentiles 20-80). In private colleges, the introduction of standardized testing is not associated with any detectable changes in students' economic origins regardless of which segment of the parental income distribution we focus on. In public colleges, there is an increase in the proportion of students from the bottom of the parental income distribution which seems to be mostly driven by a decline in the representation of students at the middle,

although this effect is short lived.

In Columns 3, 6, and 9 of Table 2, we next assess whether the introduction of standardized testing shaped students' racial mix, focusing on the likelihood that students would be Black. Across all colleges, there is a 0.008 percentage point non-statistically significant decrease in the proportion of Black students, and we can rule out even small changes in such a proportion (Column 3). In private colleges, there is a 0.17 percentage point decrease (roughly a 20% percent decrease) in the representation of Black students, which is significant at the 1% level (Column 6). Public colleges exhibit a small increase (0.07 percentage points), but this is not statistically significant at conventional levels (Column 9).

Figure 10 shows event-study figures focusing on the proportion of Black students as the outcome variable. Panel (a), which includes all colleges in the data, shows very limited changes in such a proportion following the introduction of standardized tests, regardless of the number of lags that we consider. Panel (b), which focuses on private colleges, shows a decrease (of 0.3 percentage points) in such a proportion - and this remains statistically significant five years after the introduction of exams. However, as discussed above, there is some evidence of pre-trends when focusing on this outcome at private colleges.

5.4 Robustness

The previous subsection shows that the introduction of standardized testing had mostly limited impacts on students' socioeconomic mix at elite private colleges but led to a (short-lived) increase in the proportion of lower-income students in elite public colleges. Figures B28 and B29 show that our main conclusions are similar if we use alternative estimators to measure the effects of introducing standardized testing. These figures show event-study plots based on the Sun and Abraham (2021), Borusyak, Jaravel and Spiess (2024), and Callaway and Sant'Anna (2021) estimators. Figures B26 and B27 show the average value of post-treatment coefficients when using the same set of estimators. Again, we find limited evidence that private colleges became more socioeconomically diverse after the introduction of standardized testing, but some evidence of a (short-lived) increase in the proportion of lower-income students in elite public colleges.

Finally, Figures B24 and B25 and Table B3 show that we reach similar conclusions if we include interactions between school groupings and entry cohort fixed effects in our baseline specification.

5.5 Discussion

Overall, the results in this section suggest that the introduction of standardized testing (in particular the SAT exam) was not associated with persistent changes in the socioeconomic mix of students at elite colleges in general, with even more limited effects when focusing on private colleges.

There are two main potential reasons why the introduction of standardized exams did not persistently increase the representation of lower- and middle-income students in elite colleges. First, the introduction of such exams might have not dramatically altered the *applicant*

pool at such colleges. After all, even today (several decades after the first introduction of these exams), high-achieving low-income students sometimes not apply to selective colleges (Hoxby and Avery, 2012).[45] Because our data do not include information on the socioeconomic backgrounds of *applicants*, we cannot assess whether standardized exams changed colleges' applicant pool. Second, unlike in other countries (where college admissions are based solely on exam scores), colleges in the U.S. retained substantial discretion regarding who to admit even after the introduction of standardized testing. If colleges decided to give a relatively small weight to exam scores in their admission decisions (or if the availability of exam scores did not alter their ranking of applicants), then such exams would have had a modest effect on the composition of a college incoming class. But overall, and regardless of whether they are driven by the application or the admissions margin, our results are less consistent with the idea that the introduction of standardized exams dramatically reshaped elite colleges' socioeconomic mix.

6 Conclusions

We combine newly digitized historical student records with survey and administrative data to construct a new dataset on the socioeconomic origins of students at elite U.S. colleges over the course of the 20th century. The first part of our paper shows that there has been virtually no change in the representation of lower-income students at elite colleges, both private and public. Upper-income students decreased their representation at such colleges in the post-WWII period, but have since regained ground. The relative stability of the economic backgrounds of students at elite colleges contrasts with the sharp increase in students' racial and geographic diversity at these institutions. These patterns suggest that changes in racial and geographic diversity have all taken place within the middle and upper portions of the parental income distribution.

In the second part of the paper we investigate the effects of two policy changes that have often been described as increasing the representation of lower- and middle-income students at elite colleges: the G.I. Bill after World War II, which provided educational subsidies to returning veterans and hence might have helped ease financial constraints, and the introduction of standardized testing that might have helped elite colleges to better identify high-achieving lower-income students. We find no evidence that the G.I. Bill increased the representation of lower- and middle-income students at elite colleges. We also find no evidence that the introduction of standardized testing persistently increased the representation of lower- and middle-income at elite colleges. The limited effects of these two policies suggest that increasing the representation of lower-income students in elite colleges may require earlier or more targeted interventions.

This persistence of the underrepresentation of lower income students at elite institutions

[45] Moreover, a difference relative to the modern day is that until 1958 the College Board disclosed information on applicants' test scores only to colleges, but not to applicants themselves (Karabel, 2005). As a result, high-achieving lower-SES students might have been unaware of their own academic standing and hence refrained from applying to the more selective colleges.

is particularly striking given the large changes that have occurred in higher education over the last century. Our findings reveal that there is substantial continuity in the economic origins of students at elite colleges and that such persistence has in general been difficult to alter. The one important exception to this continuity is the relative decline in the proportion of students from the top 20% following WWII and its subsequent reversal since the 1980s.

Our data on students' parental income ends in 2013. In recent years, some elite colleges have introduced initiatives aimed at increasing their socioeconomic diversity. For instance, at Princeton and Stanford admitted students from low-income families are exempt from paying tuition, room, and board.[46] Although our data do not enable us to conclude whether these efforts have been effective in achieving the goal of increasing socioeconomic diversity, there is some evidence suggesting that this might be starting to happen. For instance, the proportion of first-generation college students at Princeton and Stanford is now around 20% compared to 10-15% in the early 2010s (Appendix Figure B30).[47]

[46]In 2023 Princeton and Stanford announced that students from families making less than $100,000 would not need to pay tuition, room, and board, see Stanford Report, 2023 and Princeton University, 2023

[47]Although first-generation students are not necessarily low income, the proportion of first-generation students is likely a useful proxy for characterizing short-run *trends* in the proportion of low-income students.

References

Abramitzky, Ran, Leah Boustan, Elisa Jacome, and Santiago Perez. 2021. "Intergenerational Mobility of Immigrants in the United States over Two Centuries." *American Economic Review*, 111(2): 580–608.

Abramitzky, Ran, Leah Boustan, Katherine Eriksson, Santiago Pérez, and Myera Rashid. 2020. "Census Linking Project: Version 2.0 [dataset]."

Abramitzky, Ran, Leah Platt Boustan, and Katherine Eriksson. 2012. "Europe's Tired, Poor, Huddled Masses: Self-Selection and Economic Outcomes in the Age of Mass Migration." *American Economic Review*, 102(5): 1832–56.

Alvero, AJ, Sonia Giebel, Ben Gebre-Medhin, Anthony L. Antonio, Mitchell L. Stevens, and Benjamin W. Domingue. 2021. "Essay content and style are strongly related to household income and SAT scores: Evidence from 60,000 undergraduate applications." *Science Advances*, 7(42).

Angrist, Joshua, and Alan B Krueger. 1994. "Why do World War II veterans earn more than nonveterans?" *Journal of labor economics*, 12(1): 74–97.

Angrist, Joshua D. 1993. "The Effect of Veterans Benefits on Education and Earnings." *ILR Review*, 46(4): 637–652.

Angrist, Joshua D., and Stacey H. Chen. 2011. "Schooling and the Vietnam-Era GI Bill: Evidence from the Draft Lottery." *American Economic Journal: Applied Economics*, 3(2): 96–118.

Astin, Alexander, and Leticia Oseguera. 2004. "The Declining "Equity" of American Higher Education." *The Review of Higher Education*, 27: 321–341.

Autor, David H, and David Scarborough. 2008. "Does job testing harm minority workers? Evidence from retail establishments." *The Quarterly Journal of Economics*, 123(1): 219–277.

Avery, Christopher, Caroline Hoxby, Clement Jackson, Kaitlin Burek, Glenn Pope, and Mridula Raman. 2006. "Cost Should Be No Barrier: An Evaluation of the First Year of Harvard's Financial Aid Initiative." National Bureau of Economic Research, Inc NBER Working Papers 12029.

Bailey, Martha J, and Susan M Dynarski. 2011. "Gains and gaps: Changing inequality in US college entry and completion." National Bureau of Economic Research.

Beale, Andrew V. 1970. "The Evolution of College Admission Requirements." *The National ACAC Journal*, 15(3): 20–22.

Berger, SJ. 2012. "The rise and demise of the SAT." *American Educational History Journal*, 39(1): 165–180.

Bettinger, Eric P., and Brent J. Evans. 2019. "College Guidance for All: A Randomized Experiment in Pre-College Advising." *Journal of Policy Analysis and Management*, 38(3): 579–599.

Bettinger, Eric P., Bridget Terry Long, Philip Oreopoulos, and Lisa Sanbonmatsu. 2012. "The Role of Application Assistance and Information in College Decisions: Results from the H&R Block Fafsa Experiment." *The Quarterly Journal of Economics*, 127(3): 1205–1242.

Black, Sandra E, Jeffrey T Denning, and Jesse Rothstein. 2023. "Winners and losers? the effect of gaining and losing access to selective colleges on education and labor market outcomes." *American Economic Journal: Applied Economics*, 15(1): 26–67.

Bleemer, Zachary. 2018. "The Uc Cliometric History Project And Formatted Optical Character Recognition." Center for Studies in Higher Education, UC Berkeley University of California at Berkeley, Center for Studies in Higher Education.

Bleemer, Zachary. 2022. "Affirmative action, mismatch, and economic mobility after California's Proposition 209." *The Quarterly Journal of Economics*, 137(1): 115–160.

Bleemer, Zachary, and Sarah Quincy. 2024. "Changes in the College Mobility Pipeline Since 1900."

Borusyak, Kirill, Xavier Jaravel, and Jann Spiess. 2024. "Revisiting event-study designs: robust and efficient estimation." *Review of Economic Studies*, rdae007.

Bound, John, and Sarah Turner. 2002. "Going to War and Going to College: Did World War II and the G.I. Bill Increase Educational Attainment for Returning Veterans?" *Journal of Labor Economics*, 20(4): 784–815.

Bourdieu, Pierre, and Jean-Claude Passeron. 1990. *Reproduction in education, society and culture.* Vol. 4, Sage.

Buckles, Kasey, Adrian Haws, Joseph Price, and Haley E.B. Wilbert. 2023. "Breakthroughs in Historical Record Linking Using Genealogy Data: The Census Tree Project." National Bureau of Economic Research, Inc NBER Working Papers 31671.

Bulman, George. 2015. "The Effect of Access to College Assessments on Enrollment and Attainment." *American Economic Journal: Applied Economics*, 7(4): 1–36.

Burland, Elizabeth, Susan Dynarski, Katherine Michelmore, Stephanie Owen, and Shwetha Raghuraman. 2022. "The power of certainty: Experimental evidence on the effective design of free tuition programs." National Bureau of Economic Research.

Callaway, Brantly, and Pedro HC Sant'Anna. 2021. "Difference-in-differences with multiple time periods." *Journal of econometrics*, 225(2): 200–230.

Card, David, and Thomas Lemieux. 2001. "Going to College to Avoid the Draft: The Unintended Legacy of the Vietnam War." *American Economic Review*, 91(2): 97–102.

Chetty, Raj, David Deming, and John Friedman. 2023. "Diversifying Society's Leaders? The Determinants and Causal Effects of Admission to Highly Selective Private Colleges." National Bureau of Economic Research, Inc NBER Working Papers 31492.

Chetty, Raj, John N Friedman, Emmanuel Saez, Nicholas Turner, and Danny Yagan. 2020. "Income Segregation and Intergenerational Mobility Across Colleges in the United States." *The Quarterly Journal of Economics*, 135(3): 1567–1633.

Collins, William J, and Ariell Zimran. 2024. "Who Benefited from World War II Service and the GI Bill? New Evidence on Heterogeneous Effects for US Veterans." National Bureau of Economic Research.

Collins, William J., and Marianne H. Wanamaker. 2022. "African American Intergenerational Economic Mobility since 1880." *American Economic Journal: Applied Economics*, 14(3): 84–117.

Dillon, Eleanor Wiske, and Jeffrey Andrew Smith. 2017. "Determinants of the match between student ability and college quality." *Journal of Labor Economics*, 35(1): 45–66.

Domhoff, G William. 2018. "Who Rules America?" In *Social Stratification*. 297–302. Routledge.

Donohue, John J., and James Heckman. 1991. "Continuous Versus Episodic Change: The Impact of Civil Rights Policy on the Economic Status of Blacks." *Journal of Economic Literature*, 29(4): 1603–1643.

Dynarski, Susan, CJ Libassi, Katherine Michelmore, and Stephanie Owen. 2021. "Closing the gap: The effect of reducing complexity and uncertainty in college pricing on the choices of low-income students." *American Economic Review*, 111(6): 1721–1756.

Dynarski, Susan, Lindsay Page, and Judith Scott-Clayton. 2023. "College costs, financial aid, and student decisions." In *Handbook of the Economics of Education*. Vol. 7, 227–285. Elsevier.

Espeland, Wendy, and Michael Sauder. 2009. "Rating the Rankings." *Contexts*, 8(2): 16–21.

Fetter, Daniel K. 2013. "How Do Mortgage Subsidies Affect Home Ownership? Evidence from the Mid-Century GI Bills." *American Economic Journal: Economic Policy*, 5(2): 111–47.

Fletcher, Jason, and Adalbert Mayer. 2014. "Tracing the Effects of Guaranteed Admission Through the College Processs: Evidence from a Policy Discontinuity in the Texas 10% Plan." *Contemporary Economic Policy*, 32(1): 169–186.

Goldin, Claudia. 1999. "A Brief History of Education in the United States." National Bureau of Economic Research, Inc NBER Historical Working Papers 0119.

Goldin, Claudia, and Lawrence F. Katz. 1998. "The Origins of State-Level Differences in the Public Provision of Higher Education: 1890-1940." *The American Economic Review*, 88(2): 303–308.

Goldin, Claudia, and Lawrence F. Katz. 1999. "The Shaping of Higher Education: The Formative Years in the United States, 1890 to 1940." *Journal of Economic Perspectives*, 13(1): 37–62.

Goldin, Claudia, Lawrence F. Katz, and Ilyana Kuziemko. 2006. "The Homecoming of American College Women: The Reversal of the College Gender Gap." *Journal of Economic Perspectives*, 20(4): 133–156.

Goodman, Joshua, Michael Hurwitz, Jonathan Smith, and Julia Fox. 2015. "The relationship between siblings' college choices: Evidence from one million SAT-taking families." *Economics of Education Review*, 48: 75–85.

Goodman, Sarena. 2016. "LEARNING FROM THE TEST: RAISING SELECTIVE COLLEGE ENROLLMENT BY PROVIDING INFORMATION." *The Review of Economics and Statistics*, 98(4): 671–684.

Greene, H., and M.W. Greene. 2001. *Greenes' Guide to Educational Planning:The Public Ivies. Greenes' Guides to Educational Planning*, HarperCollins.

Hendricks, Lutz, Christopher Herrington, and Todd Schoellman. 2021. "College Quality and Attendance Patterns: A Long-Run View." *American Economic Journal: Macroeconomics*, 13(1): 184–215.

Herzog, Thomas, Fritz Scheuren, and William Winkler. 2007. *Data Quality and Record Linkage.*

Hoxby, Caroline, and Sarah Turner. 2019. "The Right Way to Capture 'College Opportunity' – Popular Measures Can Paint the Wrong Picture of Low-Income Student Enrollment." *Education Next*, 19(2): 66–74.

Hoxby, Caroline M. 2009. "The Changing Selectivity of American Colleges." *Journal of Economic Perspectives*, 23(4): 95–118.

Hoxby, Caroline M., and Christopher Avery. 2012. "The Missing 'One-Offs': The Hidden Supply of High-Achieving, Low Income Students." National Bureau of Economic Research, Inc NBER Working Papers 18586.

Hoxby, Caroline M., and Sarah Turner. 2015. "What High-Achieving Low-Income Students Know about College." *American Economic Review*, 105(5): 514–17.

Hurwitz, Michael. 2011. "The impact of legacy status on undergraduate admissions at elite colleges and universities." *Economics of Education Review*, 30(3): 480–492.

Hurwitz, Michael, Jonathan Smith, Sunny Niu, and Jessica Howell. 2015. "The Maine Question: How Is 4-Year College Enrollment Affected by Mandatory College Entrance Exams?" *Educational Evaluation and Policy Analysis*, 37(1): 138–159.

Hurwitz, Michael, Preeya P. Mbekeani, Margaret M. Nipson, and Lindsay C. Page. 2017. "Surprising Ripple Effects: How Changing the SAT Score-Sending Policy for Low-Income Students Impacts College Access and Success." *Educational Evaluation and Policy Analysis*, 39(1): 77–103.

Jacome, Elisa, Ilyana Kuziemko, and Suresh Naidu. 2023. "Mobility for All: Representative Intergenerational Mobility Estimates Over the 20th Century." National Bureau of Economic Research, Inc NBER Working Papers 29289.

Karabel, Jerome. 2005. *The chosen: The hidden history of admission and exclusion at Harvard, Yale, and Princeton.* Mariner Books.

Kingsley, Clarence D. 1913. "The Reorganization of Secondary Education." *Journal of Education*, 77(20): 543–544.

Larsen, Matthew F, Theodore J McCarthy, Jeremy G Moulton, Marianne E Page, and Ankur J Patel. 2015. "War and marriage: assortative mating and the World War II GI Bill." *Demography*, 52(5): 1431–1461.

Lemann, Nicholas. 2000. *The big test: The secret history of the American meritocracy.* Macmillan.

Lemieux, Thomas, and David Card. 2001. "Education, Earnings, and the "Canadian G.I. Bill"." *Canadian Journal of Economics*, 34: 313–344.

Long, Mark. 2004. "Race and College Admissions: An Alternative to Affirmative Action?" *The Review of Economics and Statistics*, 86(4): 1020–1033.

MacLeod, W. Bentley, and Miguel Urquiola. 2021. "Why Does the United States Have the Best Research Universities? Incentives, Resources, and Virtuous Circles." *Journal of Economic Perspectives*, 35(1): 185–206.

Markovits, Daniel. 2019. *The Meritocracy Trap.* New York: Penguin Press.

Mazumder, Bhashkar. 2005. "Fortunate Sons: New Estimates of Intergenerational Mobility in the United States Using Social Security Earnings Data." *The Review of Economics and Statistics*, 87(2): 235–255.

Mills, C Wright. 2019. "The power elite." In *Social Stratification, Class, Race, and Gender in Sociological Perspective, Second Edition*. 202–211. Routledge.

Moll, R. 1985. *The Public Ivys: A Guide to America's Best Public Undergraduate Colleges and Universities.* Viking.

Moreira, Diana, and Santiago Pérez. 2022. "Who Benefits from Meritocracy?" National Bureau of Economic Research.

Nettles, Michael T. 2019. "History of testing in the United States: Higher education." *The ANNALS of the American Academy of Political and Social Science*, 683(1): 38–55.

Olson, Keith W. 1973. "The G. I. Bill and Higher Education: Success and Surprise." *American Quarterly*, 25(5): 596–610.

Oreopoulos, Philip, Robert S. Brown, and Adam M. Lavecchia. 2017. "Pathways to Education: An Integrated Approach to Helping At-Risk High School Students." *Journal of Political Economy*, 125(4): 947–984.

Page, Lindsay C., and Judith Scott-Clayton. 2016. "Improving college access in the United States: Barriers and policy responses." *Economics of Education Review*, 51: 4–22. Access to Higher Education.

Pallais, Amanda. 2015. "Small Differences That Matter: Mistakes in Applying to College." *Journal of Labor Economics*, 33(2): 493–520.

Pallais, Amanda, and Sarah Turner. 2006. "Opportunities for Low Income Students at Top Colleges and Universities: Policy Initiatives and the Distribution of Students." *National Tax Journal*, 59(2): 357–386.

Piketty, Thomas, and Emmanuel Saez. 2003. "Income inequality in the United States, 1913–1998." *The Quarterly journal of economics*, 118(1): 1–41.

Piketty, Thomas, and Emmanuel Saez. 2014. "Inequality in the long run." *Science*, 344(6186): 838–843.

Price, Joseph, Kasey Buckles, and Mark Clement. 2021. "Census Tree Data: Version 1.0 [dataset]."

Robinson, Michael, and James Monks. 2005. "Making SAT scores optional in selective college admissions: a case study." *Economics of Education Review*, 24(4): 393–405.

Rothstein, Jesse. 2020. "Standardized College Entrance Exams: Current Uses and Academic Perspectives on Use." *Meetings on the Regents of University of California, Testimony*.

Ruggles, Steven, Julia A Rivera Drew, Catherine A Fitch, J David Hacker, Jonas Helgertz, Matt A Nelson, Nesile Ozder, Matthew Sobek, and John Robert Warren. 2024. "The IPUMS Multigenerational Longitudinal Panel: Progress and Prospects."

Saavedra, Martin, and Tate Twinam. 2020. "A machine learning approach to improving occupational income scores." *Explorations in Economic History*, 75: 101304.

Smith, Jonathan. 2014. "The Effect of College Applications on Enrollment." *The B.E. Journal of Economic Analysis & Policy*, 14(1): 151–188.

Snyder, Thomas D. 1993. "120 Years of American Education: A Statistical Portrait."

Stanley, Marcus. 2003. "College Education and the Midcentury GI Bills." *The Quarterly Journal of Economics*, 118(2): 671–708.

Sun, Liyang, and Sarah Abraham. 2021. "Estimating dynamic treatment effects in event studies with heterogeneous treatment effects." *Journal of econometrics*, 225(2): 175–199.

Taft, R. L. 1970. "Name Search Techniques." *New York State Identification and Intelligence System, Albany, New York*.

Thomas, Megan D. 2017. "WWII GI Bill and its Effect on Low Education Levels: Did the World War II GI Bill have an Effect on High School Completion, Poverty, and Employment?" *Education Finance and Policy*, 12(4): 492–515.

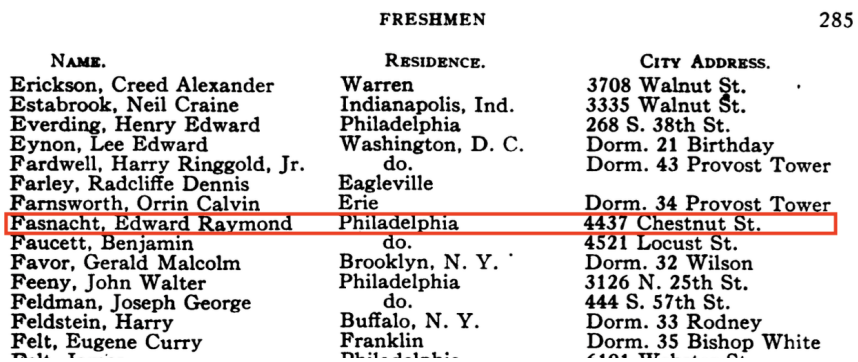
Turner, Sarah, and John Bound. 2003. "Closing the Gap or Widening the Divide: The Effects of the G.I. Bill and World War II on the Educational Outcomes of Black Americans." *The Journal of Economic History*, 63(1): 145–177.

Ward, Zachary. 2020. "Internal Migration, Education, and Intergenerational Mobility: Evidence from American History." *Journal of Human Resources*.

Zimmerman, Seth D. 2019. "Elite Colleges and Upward Mobility to Top Jobs and Top Incomes." *American Economic Review*, 109(1): 1–47.

Figure 1: Illustration of Data Digitization and Linking Strategy

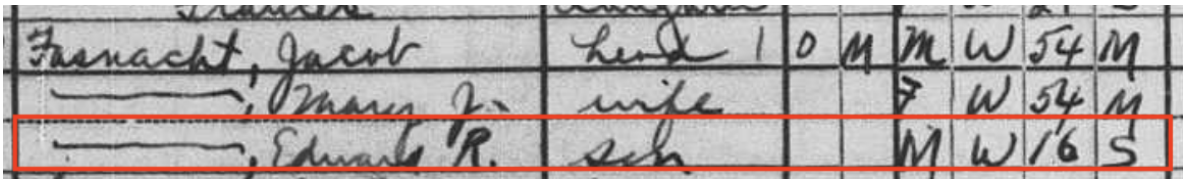
(a) University of Pennsylvania Catalogue, 1921



FRESHMEN 285

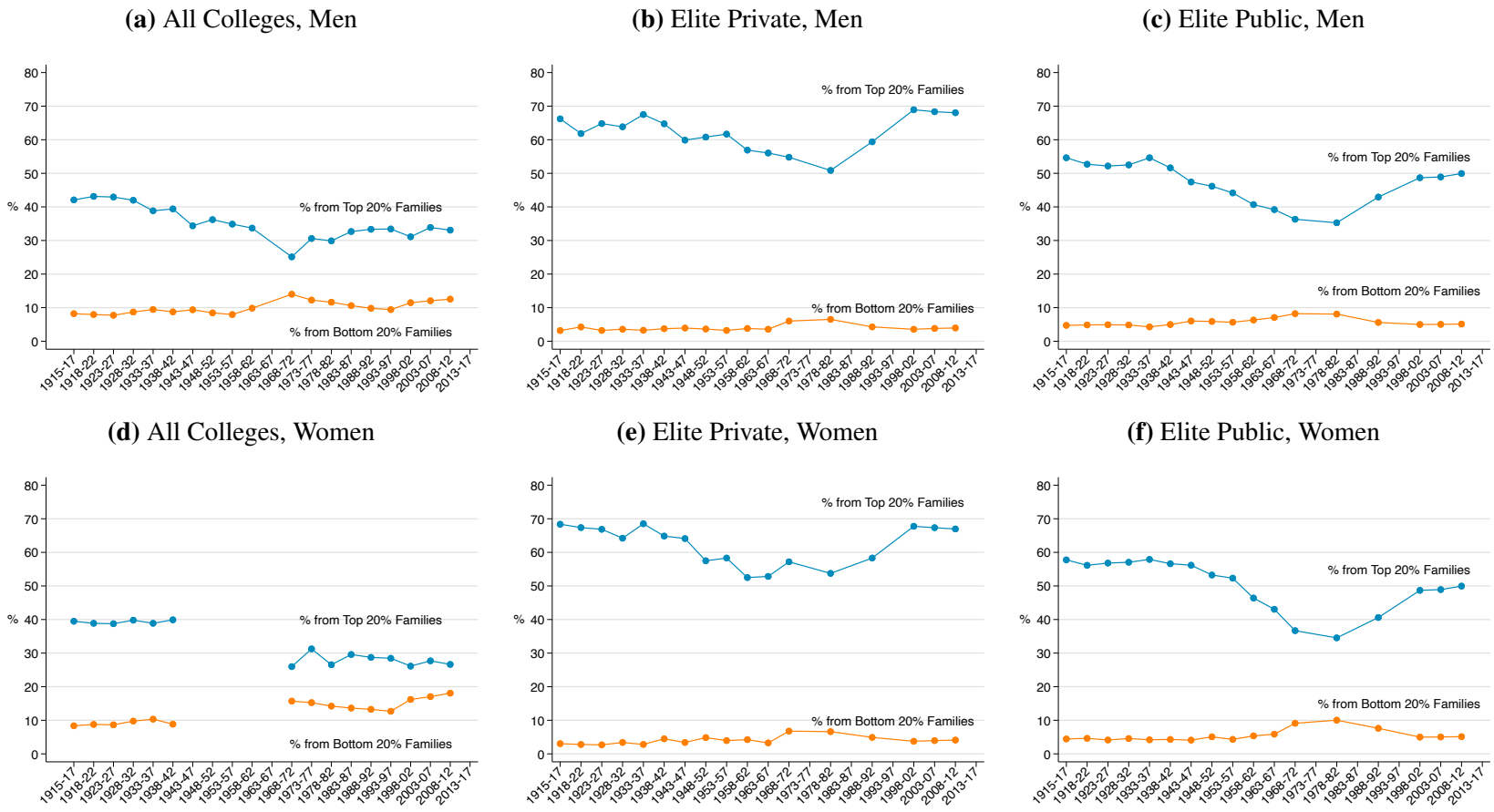
NAME.	RESIDENCE.	CITY ADDRESS.
Erickson, Creed Alexander	Warren	3708 Walnut St.
Estabrook, Neil Craine	Indianapolis, Ind.	3335 Walnut St.
Everding, Henry Edward	Philadelphia	268 S. 38th St.
Eynon, Lee Edward	Washington, D. C.	Dorm. 21 Birthday
Fardwell, Harry Ringgold, Jr.	do.	Dorm. 43 Provost Tower
Farley, Radcliffe Dennis	Eagleville	
Farnsworth, Orrin Calvin	Erie	Dorm. 34 Provost Tower
Fasnacht, Edward Raymond	Philadelphia	4437 Chestnut St.
Faucett, Benjamin	do.	4521 Locust St.
Favor, Gerald Malcolm	Brooklyn, N. Y.	Dorm. 32 Wilson
Feeny, John Walter	Philadelphia	3126 N. 25th St.
Feldman, Joseph George	do.	444 S. 57th St.
Feldstein, Harry	Buffalo, N. Y.	Dorm. 33 Rodney
Felt, Eugene Curry	Franklin	Dorm. 35 Bishop White

(b) Edward Fasnacht, 1920 Census



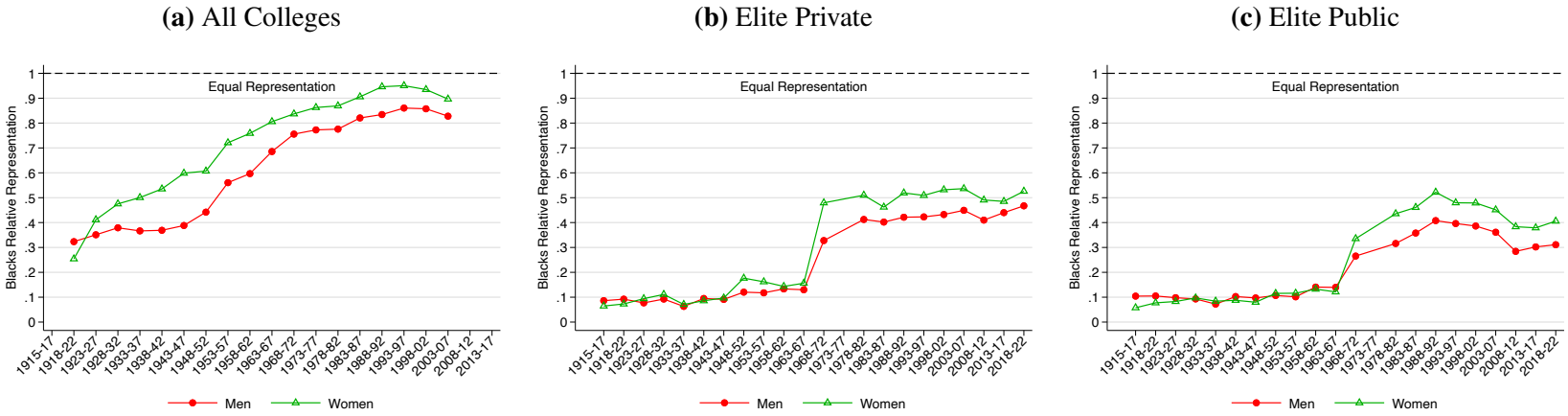
Notes: Panel (a) shows an excerpt from the 1921 University of Pennsylvania Catalogue. It highlights one student, Edward Fasnacht, who was enrolled at the University of Pennsylvania at the time. Panel (b) shows this student at age 16 living with his family in the 1920 US Population Census.

Figure 2: Economic Origins of College-Going Population, by College Group and Entering Cohort



Notes: This figure plots the proportion of students in the college-going population from families in the bottom and top quintiles of the parental income distribution, by entering cohort. Panels (a) and (d) consider the overall college-going population, with Panel (a) restricting to men and Panel (d) restricting to women. Panels (b) and (e) consider elite private colleges, with Panel (b) restricting to men and Panel (e) restricting to women. Panels (c) and (f) consider elite public colleges, with Panel (c) restricting to men and Panel (f) restricting to women. Data comes from a combination of our 1910-1940 and 1920-1950 linked US population census samples, our linked college registers, the Survey of Occupational Change in a Generation (for men only, conducted as a supplement to the 1962 and 1973 CPS), the UCLA Freshman Survey, and Opportunity Insights.

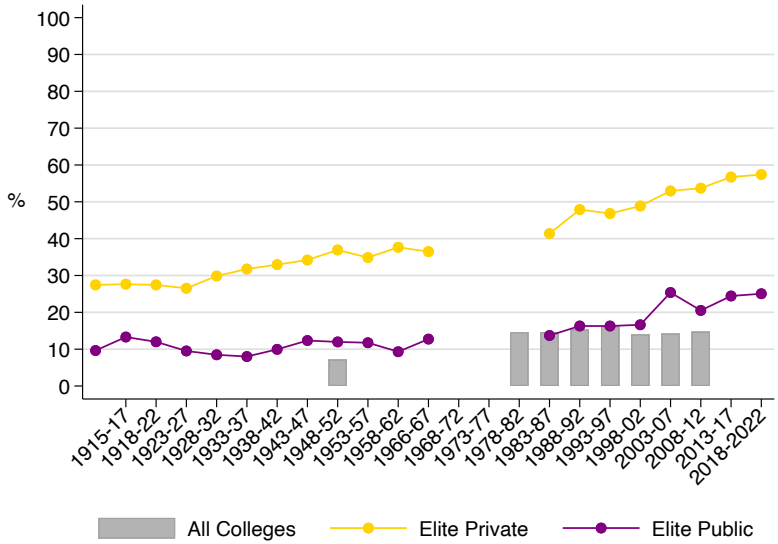
Figure 3: Relative Representation of Black Students in College-Going Population, by College Group and Entering Cohort



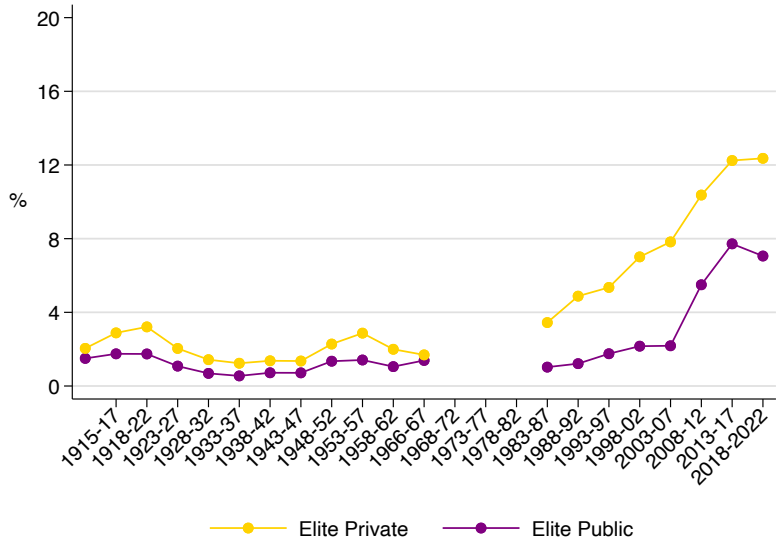
Notes: This figure plots the relative representation of Black students, by gender and cohort, in the aggregate college-going population (Panel (a)), at elite private institutions (Panel (b)), and at elite public institutions (Panel (c)). For Panel (a), data comes from the US population census and the American Community Survey. For Panels (b) and (c), historical data come from our linked college registers, reweighed by school-year-gender match rates, and modern data comes from the Integrated Postsecondary Education Data System (IPEDS).

Figure 4: Geographic Origins of College-Going Population, by College Group and Entering Cohort

(a) Share of Out-of-Region Students

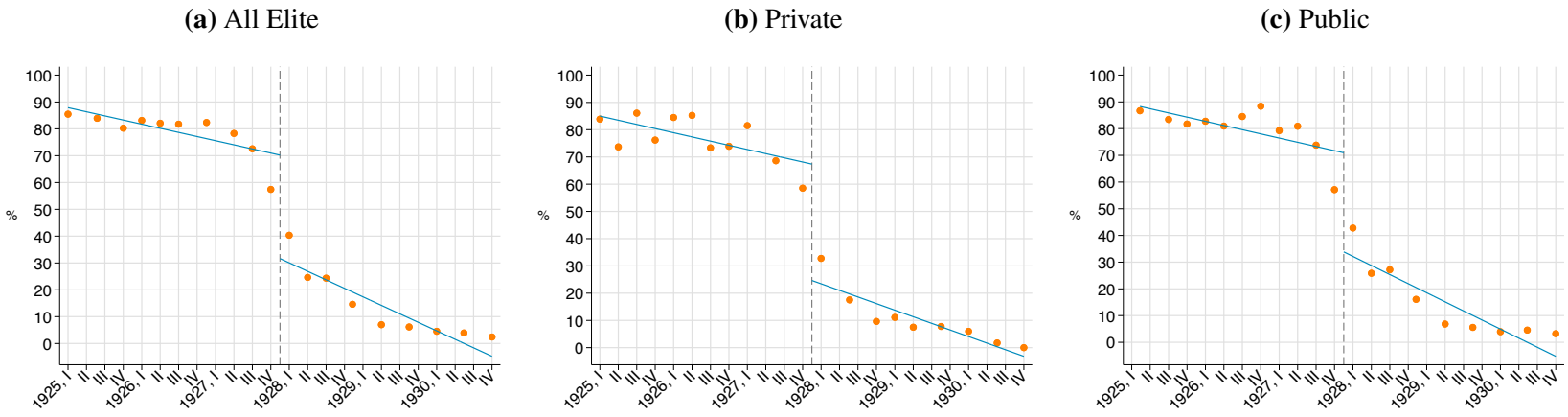


(b) Share of International Students



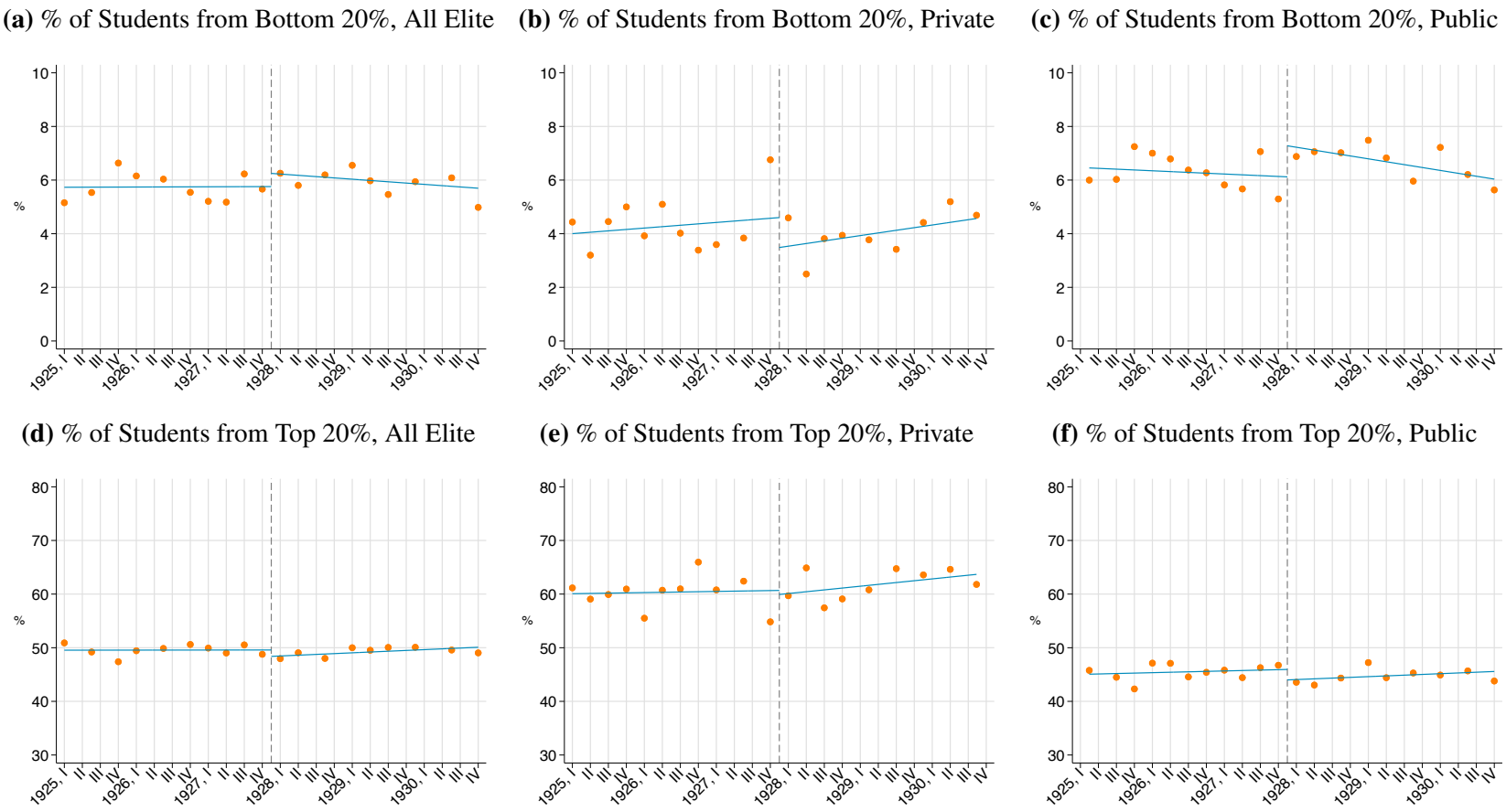
Notes: This figure plots the share of students in a college that hail from outside the college's region (Panel (a)) or from outside the country (Panel (b)), by entering class and college group. Regions are based on the smallest level of geographic aggregation that is reported for colleges in the UCLA Freshman Survery (the OBE region). States are categorized into different regions according to the following criteria: New England (CT, ME, MA, NH, RI, VT), Mid East (DE, DC, MD, NJ, NY, PA), Great Lakes (IL, IN, MI, OH, WI), Plains (IA, KS, MN, MO, NE, ND, SD), Southeast (AL, AR, FL, GA, KY, LA, MS, NC, SC), Southwest (AZ, NM, OK, TX), Rocky Mountains (CO, ID, MT, UT, WY), and Far West (AK, CA, HI, NV, OR, WA). The data on the general college-going population are from the 1950 US Population Census and the UCLA Freshman Survey. Note that this survey only added a question on students' ZIP codes starting in 1982, and so we lack information on students' geographic origins in the 1965-1980 period. The historical data on private and public colleges are from our digitized college registers, whereas the modern data are from the Integrated Postsecondary Education Data System.

Figure 5: Share of World War II Veterans at Elite Colleges, by Year and Quarter of Birth



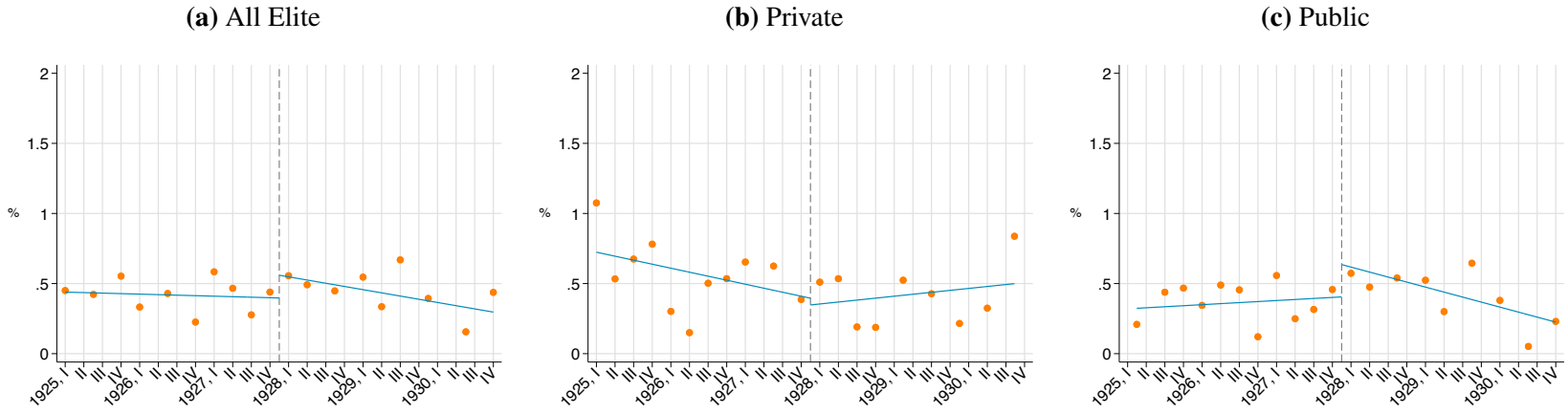
Notes: This figure shows the proportion of WWII veterans among men born between 1925 and 1930 in our sample of students in elite colleges. Panel (a) pools all students in elite colleges, whereas Panels (b) and (c) split the data based on whether the student attended a private or a public college, respectively. The data are from our linked college registers data matched to birth quarter information from FamilySearch.org.

Figure 6: Economic Origins of Students at Elite Colleges, by Year and Quarter of Birth



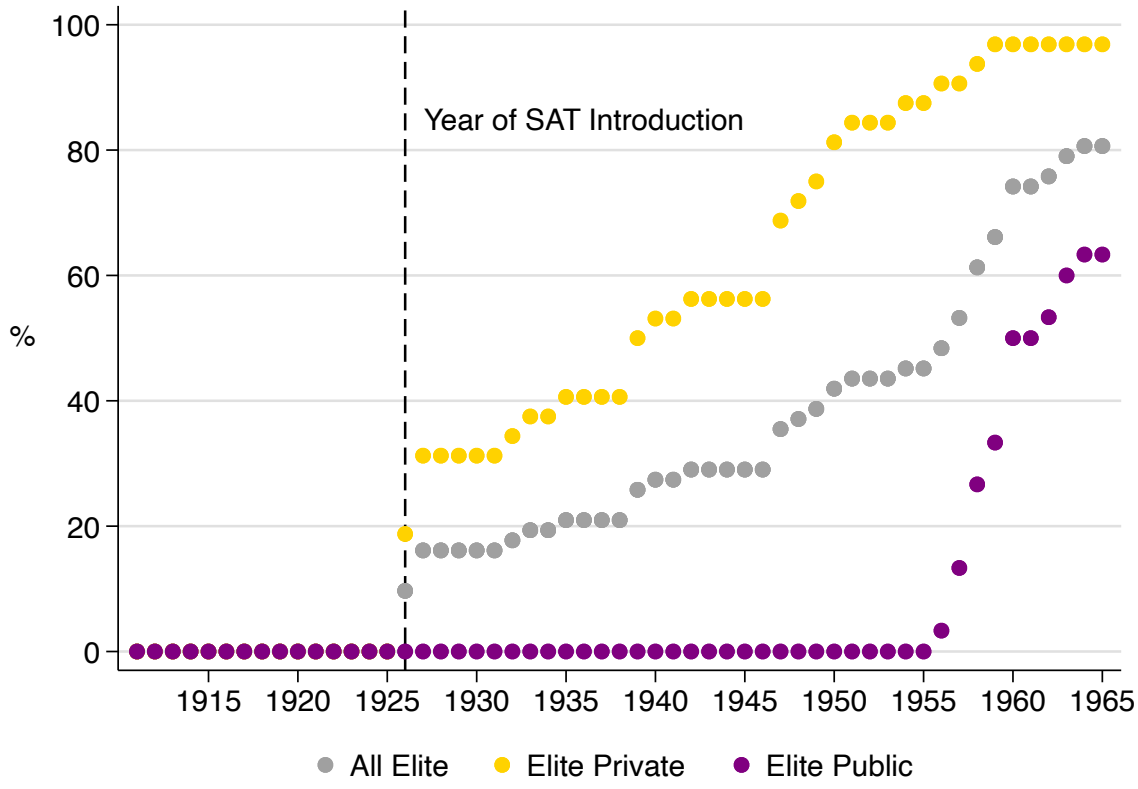
Notes: This figure shows the economic origins among men born between 1925-1930 in our sample of students in elite colleges, by year and quarter of birth. Panels (a) and (d) pool together all students in elite colleges, whereas Panels (b) and (e) and (c) and (f) split the data based on whether the student attended a private or a public college, respectively. The outcome in Panels (a)-(c) is an indicator if the student grew up in a family in the bottom 20%. The outcome in Panels (d)-(f) is an indicator if the student grew up in a family in the top 20% of the parental income distribution. All outcome variables are multiplied by 100 so that we can directly interpret the coefficient estimates as percentage point differences. The data are from our linked college registers data matched to birth quarter information from FamilySearch.org.

Figure 7: Proportion of Black Students at Elite Colleges, by Year and Quarter of Birth



Notes: This figure shows the proportion of Black individuals among men born between 1925-1930 in our sample of students in elite colleges, by year and quarter of birth.. Panel (a) pools together all students in elite colleges, whereas Panels (b) and (c) split the data based on whether the student attended a private or a public college, respectively. All outcome variables are multiplied by 100 so that we can directly interpret the coefficient estimates as percentage point differences. The data are from our linked college registers data matched to birth quarter information from FamilySearch.org.

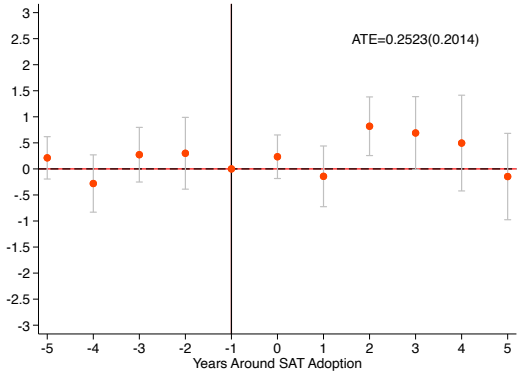
Figure 8: Proportion of Elite Colleges with an SAT Requirement



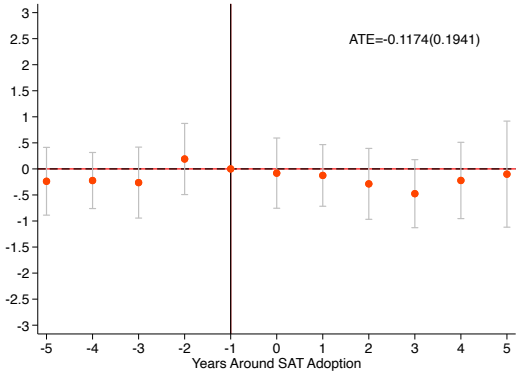
Notes: This figure shows the proportion of colleges in our sample of 65 selective colleges that required the Scholastic Aptitude Test (SAT) for all applicants. Similar statistics are then shown separately based on whether the college is private or public.

Figure 9: Impacts of Introducing the SAT on Students' Economic Origins at Elite Colleges, Event Study Regressions

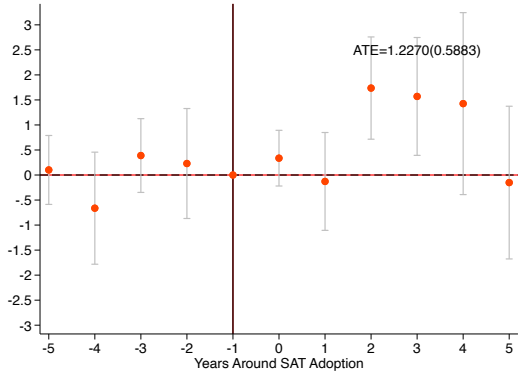
(a) % of Students from Bottom 20%, All Elite



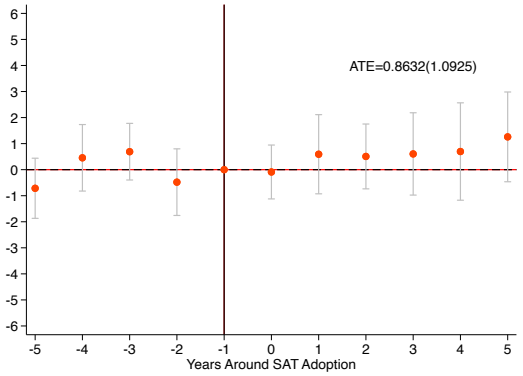
(b) % of Students from Bottom 20%, Private



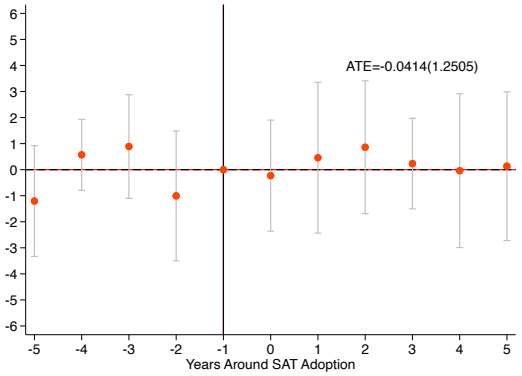
(c) % of Students from Bottom 20%, Public



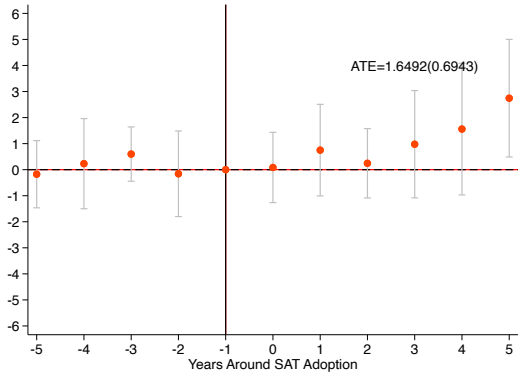
(d) % of Students from Top 20%, All Elite



(e) % of Students from Top 20%, Private

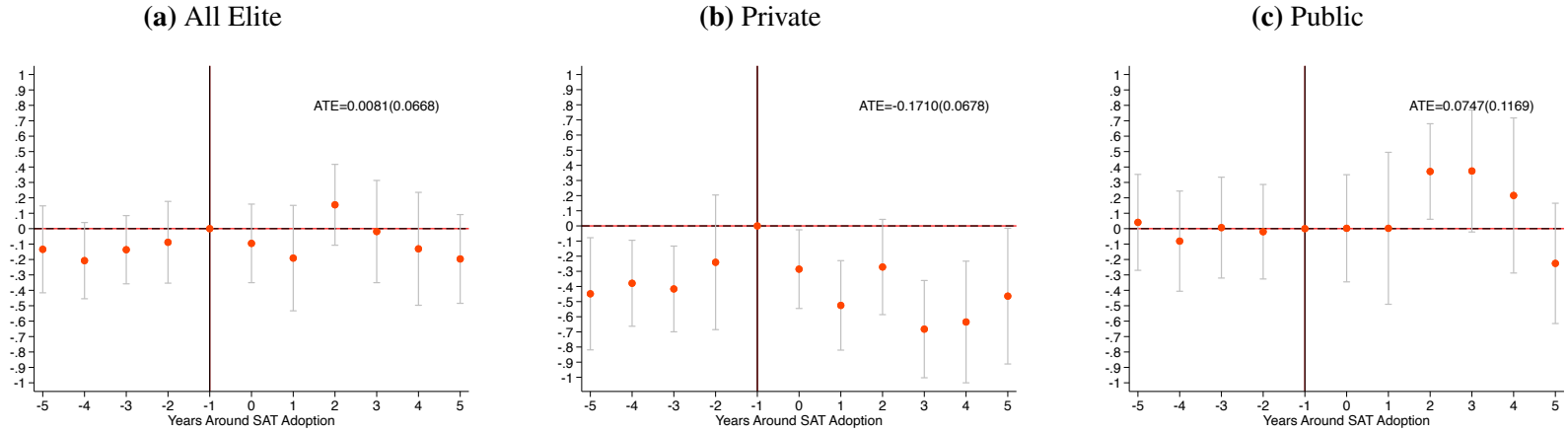


(f) % of Students from Top 20%, Public



Notes: This figure shows estimated coefficients from event study two-way fixed effects regressions. Each panel corresponds to a different outcome (the top row is whether the student grew up in a family in the bottom 20% of the parental income distribution, the bottom row is whether the student grew up in a family in the top 20% of the parental income distribution), and sample of colleges (all elite colleges in Panels (a) and (d), only private colleges in Panels (b) and (e), only public colleges in Panels (c) and (f)). Each of the panels also includes the coefficient and standard error associated with an indicator variable that is one if the school had adopted an SAT requirement. All outcome variables are multiplied by 100 so that we can directly interpret the coefficient estimates as percentage point differences.

Figure 10: Impacts of Introducing the SAT on the Proportion of Black Students at Elite Colleges, Event Study Regressions



Notes: This figure shows estimated coefficients from event-study two-way fixed effects regressions. The outcome variable is an indicator that takes a value of one if the student is Black. Each panel corresponds to a different estimation sample (all elite colleges in Panel (a), only private colleges in Panel (b), only public colleges in Panel (c)). Each of the panels also includes the coefficient and standard error associated with an indicator variable that is one if the school had adopted an SAT requirement. All outcome variables are multiplied by 100 so that we can directly interpret the coefficient estimates as percentage point differences.

Table 1: The G.I. Bill and the Socioeconomic Backgrounds of Students at Elite Colleges

(a) First Stage

	WWII Veteran		
	(1) All Elite Colleges	(2) Elite Private	(3) Elite Public
Born <= Q4 1927	38.56***	42.72***	37.12***
	(6.249)	(5.432)	(6.839)
Observations	4790	1188	3602
Control Group Mean	30.28	29.63	30.52

(b) Reduced Form

	All Elite Colleges			Elite Private			Elite Public		
	(1) Bottom 20%	(2) Top 20%	(3) Black	(4) Bottom 20%	(5) Top 20%	(6) Black	(7) Bottom 20%	(8) Top 20%	(9) Black
Born <= Q4 1927	-0.480	1.157*	-0.162**	1.134	0.706	0.0499	-1.149**	1.936**	-0.231**
	(0.315)	(0.652)	(0.0718)	(0.927)	(2.720)	(0.217)	(0.473)	(0.829)	(0.0893)
Observations	48434	48434	48434	13044	13044	13044	35390	35390	35390
Control Group Mean	5.506	50.29	0.581	3.605	62.81	0.562	6.277	45.20	0.588

Notes: *** p<0.01, ** p<0.05, and * p<0.1. The variable *Born <= Q*41927 is an indicator that takes a value of one if the individual was born in or before the fourth quarter of 1927. Panel (a) presents the results of estimating Equation 1 from the text using an indicator that takes a value of one if the individual is a World War II veteran as the outcome variable. Column 1 includes all students in our sample of elite colleges, whereas Columns 2 and 3 restrict the analysis to students at private and public colleges, respectively. Note that we estimate the first stage in a smaller sub-sample of individuals who we can link to the 1950 US Population Census and who responded to the sample line questions in this census. Panel (b) estimates the same reduced-form equation but focuses instead on students' socioeconomic characteristics: whether they belong to families in the bottom 20% of the parental income distribution (Columns 1, 4, and 7), whether they belong to families in the top 20% (Columns 2, 5, and 8), and whether they are Black (Columns 3, 6, and 9). All outcome variables are multiplied by 100 so that we can directly interpret the coefficient estimates as percentage point differences. Each group of columns corresponds to a different sample of colleges, as indicated by the column headings. The sample in both panels is restricted to men born between 1925 and 1930. Standard errors are clustered at the birth quarter level.

Table 2: Impacts of Introducing the SAT on Students' Socioeconomic Backgrounds at Elite Colleges

	All Elite			Elite Private			Elite Public		
	(1) Bottom 20%	(2) Top 20%	(3) Black	(4) Bottom 20%	(5) Top 20%	(6) Black	(7) Bottom 20%	(8) Top 20%	(9) Black
SAT	0.252	0.863	0.00809	-0.117	-0.0414	-0.171**	1.227**	1.649**	0.0747
	(0.201)	(1.092)	(0.0668)	(0.194)	(1.251)	(0.0678)	(0.588)	(0.694)	(0.117)
College FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1116267	1116267	1116267	364985	364985	364985	751282	751282	751282
Mean of Dep. Var.	4.704	53.68	0.953	3.366	63.10	0.894	5.354	49.11	0.981

Notes: $***$ p<0.01, $**$ p<0.05, and $*$ p<0.1. The variable *SAT* is an indicator that takes on a value of one if a college had implemented the SAT by year *t*. The outcome variable in Columns 1, 4, and 7 is an indicator that takes on a value of one if a student grew up in a family in the bottom 20% of the parental income distribution. The outcome variable in Columns 2, 5, and 8 is an indicator that takes on a value of one if the student grew up in a family in the top 20% of such distribution. The outcome variable in Columns 3, 6, and 9 is an indicator that takes a value of one if the student is Black. All outcome variables are multiplied by 100 so that we can directly interpret the coefficient estimates as percentage point differences. All regressions include college and year fixed effects. Standard errors are clustered at the college level.

Online Appendix - Not for Publication

A Data Appendix

A.1 Sample Selection - Schools

Our dataset consists of students who attended one of 65 selective institutions of higher education between 1915 and 1966. Our data on selective private colleges come from historic groupings of institutions that are still meaningful today. They include the following 34 institutions:

Ivy-Plus. This group of twelve highly selective private institutions is often used in the literature (Chetty et al., 2020) and includes the Ivy League (Brown University, Columbia University, Cornell University, Dartmouth College, Harvard University, University of Pennsylvania, Princeton University, and Yale University), plus the University of Chicago, Duke University, Massachusetts Institute of Technology, and Stanford University.

Highly Selective Women's Colleges. This group of eight schools includes the "Seven Sisters", a group of women's colleges that provided women with an educational experience similar to that of the Ivy League (that largely did not enroll women in the first half of the 20th century), as well schools directly affiliated with any Ivy-Plus institution. The former includes Barnard College, Bryn Mawr College, Mount Holyoke College, Radcliffe College, Smith College, Vassar College, and Wellesley College. The latter includes Pembroke College (the coordinate institution to Brown University) and Radcliffe College (a "Seven Sisters" college that is also the coordinate institution to Harvard University).

"Little Ivies." This is an informal grouping of 15 elite small liberal arts colleges that include Amherst College, Bates College, Bowdoin College, Colby College, Connecticut College, Hamilton College, Haverford College, Lafayette College, Middlebury College, Swarthmore College, Trinity College, Tufts University, Vassar College, Wesleyan University, and Williams College.

Note that since Vassar College is both an elite historically women's college and a Little Ivy, there are a total of 34 *unique* institutions across these three groups.

Our data on selective public colleges include the following 31 institutions:

1985 Public Ivies. The term "public ivy" was first coined in 1985 by Richard Moll in the book *"The Public Ivys: a Guide to America's Best Public Undergraduate Colleges and Universities"* (Moll, 1985). From this book, he provides a list of 15 "public ivies" and 9 "worthy runner-ups". The former includes the University of California (which applies to all eight campuses as of 1985 - Berkeley, Davis, Irvine, Los Angeles, Riverside, San Diego, Santa Barbara, and Santa Cruz), Miami University, University of Michigan, University of North Carolina at Chapel Hill, University of Texas - Austin, University of Vermont, University of Virginia, and the College of William & Mary. The latter includes the University of Colorado - Boulder, New College of Florida, Georgia Institute of Technology, University of Illinois at Urbana–Champaign, the State University of New York at Binghamton, Pennsylvania State University, University of Pittsburgh, University of Washington, and University of Wisconsin–Madison. Within the University of California system, the Irvine, Riverside, San Diego,

and Santa Cruz campuses were not founded until the 1950s and 1960s and so we exclude them from our sample. We also exclude the New College of Florida as it no longer exists as a standalone institution.

2001 "Public Ivies". A notable update to the 1985 list is from the 2001 book *"Greenes' Guides to Educational Planning"* (Greene and Greene, 2001). This book includes 31 colleges and universities, 19 of which overlap with the 1985 list. The institutions that overlap include the University of California (which applies to the Berkeley, Davis, Irvine, Los Angeles, San Diego, and Santa Barbara campuses), University of Colorado - Boulder, University of Illinois at Urbana–Champaign, Miami University, University of Michigan, State University of New York at Binghamton, University of North Carolina at Chapel Hill, Pennsylvania State University, University of Pittsburgh, University of Texas - Austin, University of Vermont, University of Virginia, University of Washington - Seattle, the College of William & Mary, and University of Wisconsin - Madison. The new additions include the University of Arizona, University of Connecticut, University of Delaware, University of Florida, University of Georgia, Indiana University Bloomington, University of Iowa, University of Maryland, Michigan State University, University of Minnesota - Twin Cities, Ohio State University, and Rutgers University.

A.2 PDF Acquisition and Compilation

To obtain the list of all undergraduates at these 65 institutions, we locate publicly available, primary source documents (e.g., registers, directories, catalogs, bulletins, yearbooks, etc.) for each individual year for each individual school. For the 1915-1966 period, that amounts to individually sourcing 3,380 primary source documents. The common feature of all these document types is that they include the universe of names of students enrolled and importantly do not rely on students to voluntarily report their information. Otherwise, these document types do vary in terms of the exact information that they include (e.g., class standing, hometown, major, etc.), both across schools as well as within schools across years. For cases in which we are able to locate multiple document styles for a single school, we prioritize selecting documents that are the most comprehensive in terms of students included (e.g., all undergraduates as opposed to just graduating seniors). In Tables A2-A5, Column (b) reports, for each school, the exact years for which we have successfully located documents. Column (c) reports the document type, and Column (d) reports the students included. In additional online documentation, we also report a description of the school type (e.g., coeducational vs. single sex) and an outline of what information is available within each document.

Some of these documents are available in PDF format in online public repositories such as the HathiTrust Digital Library, the Internet Archive, or the digital collections of specific colleges. This represents the best-case scenario, as acquiring the PDFs from such sources is relatively straightforward. In cases where the documents are unavailable online, we next contact a university's library system and request scanned copies of the physical documents housed within the university's archives. If such a service is unavailable, we request that the university librarian simply locate the documents for us, and then we send a research assistant into the university's archives to make the necessary scans. There are instances in which all

of these strategies prove unsuccessful, either because the documents do not exist (e.g., were not published during wartime), or because we were unable to successfully get in contact with a university librarian for assistance. Overall, our data cover 95% of all possible school-year combinations. For those that we are able to acquire, we then compile a ready-for-transcription PDF for each school-year pair.

A.3 Transcription

Once the PDFs are acquired, we digitize the student lists. Since the page formats of these documents vary drastically not only by school, but also within school by year, standard OCR techniques generally prove unsuccessful. Instead, we either entirely hand transcribe each document, or rely on a combination of OCR and hand transcription. In either case, a human eye sees each and every student entry.

This transcription work is carried out via the Upwork freelancing platform. We utilize the "Enterprise Suite", which gives us access to the top 1% of talent on the platform. Exclusively to this group of freelancers, we posted a job titled "PDF-to-Excel transcription of college registers", with the following description:

> "The task is to (quickly and accurately) transcribe PDFs into an Excel spreadsheet. (The PDF contains information on the names of students attending a specific college in the US, as well as their hometowns and class standing.) For your reference, an example and explicit instruction sheet are attached. This is part of a large-scale project, and so if the task goes well there is the potential for many more tasks in the immediate future!"

See Figure A6 for the explicit instruction sheet that we included in the job posting. Interested freelancers submit a proposal, and we use this as the main selection criteria. Each freelancer with a satisfactory proposal is then given a short 10-page trial document, from which we assess their ability. If the submitted transcription is 100% accurate, we then offer them a contract, and continue to do so as long as their work remains of high quality. It is a rare exception that a freelancer makes it past the trial assignment but subsequently does not perform well. As such, the team of freelancers is largely consistent across the entire data collection effort.

For each set of PDFs, we write explicit instructions to ensure that the freelancers understand the new document format and how it needs to be transcribed. The process of writing these instructions is necessarily carried out in a detail-oriented fashion, as we need to ensure that the freelancers correctly understand each piece of information that is provided in the PDF. For example, if a location is provided, we need to understand whether it is the student's home address or their college residence. If the document is a yearbook, we need to know which class (e.g., junior or senior) publishes the yearbook at this particular university and whether or not that has changed over time. Throughout the transcription process, any questions or uncertainties not addressed in the instruction sheet are handled via message within the Upwork platform on a daily basis. The final submitted spreadsheet contains a column for each piece of data included in the PDF, and each row represents an individual student.

In total, a team of 300 freelancers have worked for over 40,000 hours over the course of four years to digitize approximately 3,000 documents, which corresponds to nearly 2.5 million unique students in 65 colleges.

A.4 Transcription Quality Control

Once the PDFs are digitized using the transcription process just described, we then put each submitted transcription through a regimented quality control process. We have a team of RAs who first and foremost ensure that the PDF was transcribed according to the instructions with which we provided the transcribers. First, they check to make sure that all required information is recorded (e.g., that every piece of information outlined in the instructions was in fact transcribed). Second, they check to make sure that this information is recorded in the correct format (e.g., that if a name is in the "Last Name, First Name" format, that there is in fact a comma after the last name). And lastly, they check to make sure that this information is recorded in the correct column (e.g., that the degree the student is pursuing is recorded under "Degree" as opposed to "Previous Degree"). These are all "high-level" checks, as these errors are not likely to idiosyncratically occur among individual rows of the data, but rather are errors that would apply to the entire dataset. If it is found that any information was missed in the transcription, we go back to the transcribers and request a corrected file be submitted. This is in practice, however, not an error that we see from the transcribers that have made it past the trial-stage and into our team. A formatting error is rare, but if we do find one then this is normally corrected on our end via coding in the standardization process. The transcription of information into the incorrect column is also exceedingly rare, but if we do find this issue then this is fixed via a simple renaming of columns. These high-level checks are done on each and every submitted transcription, but errors of this sort virtually never appear.

Once these high-level checks are complete, the reviewers then proceed to check for idiosyncratic individual-level errors. At this stage, the reviewers check one random entry for every ten pages transcribed. Here, a large portion of the review is looking to see if any idiosyncratic transcription errors (e.g., misspellings) exist. In addition, the reviewers specifically check any data that comes from a heading (e.g., there is a heading for "Seniors" and then every student under that heading is a senior) to ensure that the correct heading was applied. Lastly, for schools for which we have information on the student's place of residence, we need to ensure that this location is correctly recorded. Within the transcription process, the transcriber needs to correctly identify whether the location is inside/outside of the U.S. and then record accordingly. If the location is inside the US, the transcriber has to identify the town versus state and then record each in their respective column. If the location is outside the US, the transcriber has to identify the country name and record this in a third location-related column. The reviewer ensures that all parts of the place of residence were correctly identified and then transcribed accordingly.

Lastly, within the transcription process transcribers highlight rows for which text is particularly blurry or obstructed enough to cause uncertainty in the transcription. At this stage, we have the reviewers put a second set of eyes on each of these highlighted entries to either

confirm what was already transcribed, or to provide a correction. This is done in addition to the above-described random checks.

If there are, at most, exceedingly rare random errors, then the file passes our quality control process. If any issues of note are identified, then the transcriber is informed and a corrected file is requested. Once received, the corrected version once again goes through the exact same quality control process. Note that it is exceedingly rare for us to need to request a corrected version.

A.5 Sample Selection - Students

Once the transcription successfully goes through our quality control process, we then need to determine which students are to be included in our sample. Our goal is to include all degree-seeking undergraduates. We define a "degree-seeking undergraduate" to be any student enrolled full-time in their first degree program. "Enrolled" implies that the student must be formally pursuing a degree (e.g., not auditing), but does not imply that the student must eventually graduate with said degree. "Full-time" implies that part-time students, special students, etc. will be excluded. "First" implies that the student does not have a previous degree. This will include the average modern undergraduate, as well as students in the past who may have entered a professional school (e.g., law, medicine, etc.) directly out of high school. For our purposes, "degree program" implies that the student is enrolled in a program that culminates in a bachelor's or bachelor's-equivalent degree. This means that students pursuing an associate degree, a diploma, a certificate, etc. will be excluded.

To identify these students, we have a team of RAs go through the following process for each school:

First, we identify the class standing values that are present in each year. In the best case, this is recorded as a column within the transcription. If not, we go back to the original PDF (that is, the raw PDF from which the ready-for-transcription PDF was created) and look to the omitted pages for any clues (e.g., if there is a header that contains the class standing information on a page not included in the ready-for-transcription PDF). If not, we look for clues from the document type itself (e.g., if it is a commencement program then we know everyone is graduating, if it is a yearbook then we look to the dedication page and see if it was signed by the "Class of YYYY", etc.). Once we have this information, we exclude all class standings that are not likely associated with a "full" degree program: special, unclassified, extension, evening, phantom roll, part-time, etc. Additionally, we exclude any student who is recorded as "graduate", as these students are not likely to be pursuing their first degree. (A "graduate" student is not to be confused with a "graduating" student, and care is taken to make sure that these are correctly distinguished.) No standard undergraduate class standings (freshmen, sophomore, junior, senior, graduating, undergraduate) are excluded at this stage. If there are non-standard class standing values (e.g., middle, lower, etc.) then we do research to better understand what that means for this particular school. This research involves first going back to the raw PDFs in order to see if any description or reference exists. If no helpful information is found, we conduct an online search.

Second, we see which years, if any, have previous degrees reported. If at least one pre-

vious degree is reported in a year, then we automatically know that the students who have a previous degree should be excluded (as they are clearly not currently pursuing their first degree). Additionally, we automatically know that students in that year who do not have a previous degree reported are necessarily pursuing their first degree. (Note that this alone does not guarantee inclusion into our sample, as it is still possible that they are not currently pursuing a full degree and thus ultimately need to be excluded.)

Third, we see which years, if any, report the degree currently being pursued. Here, we automatically exclude any degree that is less than a bachelor's degree (e.g., diploma, certificate, associate's, etc.). If the year does not have previous degrees reported and the degree listed is advanced (e.g., master's, doctorate, M.D., etc.), then we exclude it. All standard bachelor's degrees are included. If a Bachelor of Laws (also known as LLB or LL.B.) exists for the school, then we do outside research to determine whether or not this degree program required a previous degree for admission. Note that the Bachelor of Law degree was phased out of the U.S. system of higher education in the 1960s in favor of the Juris Doctor that we see today. Prior to its phase-out, there was a general (non-standardized) trend of increasing admission requirements. Thus, we need to determine the inclusion/exclusion status of the Bachelor of Law degree on a school-year basis. The outside research to determine this will come from admission requirements present in the original PDF documents themselves, or from an online search for the history of the law program at this particular school. Another degree that requires additional care is the Bachelor of Theology or Bachelor of Divinity. While it is a bachelor's degree, which we generally associate with a first degree, in this case, it is actually most often a postgraduate degree. And so, we conduct research just as above with the case of the Bachelor of Law degree. From this point forward, we no longer proceed with additional exclusion checks for students for which a degree is reported.

Fourth, we see which years, if any, report the school in which the student is enrolled. Here, we carefully think about each and every value. Any professional school (e.g., law, business, medicine, nursing, pharmacy, dental, education, etc.) requires outside research to determine whether or not there is a blanket inclusion/exclusion rule that should be applied at the school level. For example, perhaps we are able to locate a timeline for the history of the Law School, and it shows that starting in 1949 a previous degree was required for admission. Then, we know that law students who appear prior to 1949 should be included, and those who appear starting in 1949 and in all subsequent years should be excluded. As another example, consider the case in which we look to the original PDFs and see that the School of Education did not start awarding bachelor's degrees until 1936, and that prior to that graduates were awarded certificates. In that case, we know that education students who appear prior to 1936 should be excluded, and that students who appear in 1936 and subsequent years should be included. From this point forward, we no longer proceed with additional exclusion checks for students for which a school is reported and that school's inclusion/exclusion is certain.

Fifth, we see which years, if any, report the program (i.e., major) in which the student is enrolled. To begin, we see if any program values reasonably map to a school value, and then apply the decision for that school to all of the corresponding programs. For those that do not reasonably map, we conduct additional outside research. Here we can also exploit the existence of programs to clarify the inclusion/exclusion of schools that did not have a certain

blanket inclusion/exclusion rule above. For example, perhaps from our school research we know that dental hygienists at the Dental School only received a certificate, but that the pre-dental program awarded bachelor's degrees. In that case, we can explicitly exclude those students who were enrolled in the dental hygiene program within the Dental School, and include those students who were enrolled in the pre-dental program within the same school. The program variable may also give us further help in identifying students who are pursuing less than a bachelor's degree, as we may see such words as "Certificate", "Vocational", or "Training" included in the program name. These are all clues that the student is likely not pursuing a bachelor's or bachelor's-equivalent degree and should thus be excluded.

It is important to note that these documents for a particular school are not necessarily formatted in a consistent way across years. Hence, within each of the above steps, we may find that we have the desired information for some, but not all, years. In this case, we can use the years for which we do have information to help us make decisions about the years for which we do not. For example, assume that in 1921 and 1923 we see that every student within the School of Education is enrolled in a certificate program. However, in 1922 we only have the school, but not the degree, recorded. From the information from 1921 and 1923, we can reasonably conclude that all students in 1922 were also enrolled in a certificate program, and thus should also be excluded from our sample. But, what if we do not have any desired information for any of the years? Fortunately, that is never the case. At a minimum, we will know the student's class standing and can use that to make inclusion/exclusion decisions.

A.6 Sample Selection Quality Control

Acknowledging the subjectivity of some of the decisions that are made in the above process, we have a single reviewer who goes over each and every decision made to ensure that they are all made correctly and consistently. Once the sample selection process is completed and reviewed, we then implement these decisions in code and plot raw total enrollment trends for each school. This allows us to visualize the implications of our sample selection. Here, we are looking for any sudden dips or peaks, or overall trends that fundamentally go against what we would expect given the enrollment trends of the overall college-going population (Snyder, 1993). Note that we expect to see a general rise in enrollment over time, with a sharp decline during WWII and a sharp rise after the end of the war.

If an abnormality is present, we first check the code to make sure all exclusions were made correctly. If so, we conduct an online search in an attempt to find a historic enrollment plot/table that the particular university may have published. If we are able to find such a resource, this allows us to confirm that our enrollment trend aligns. If we are unable to find such a resource, then we attempt to understand what is driving the abnormality. Data permitting, this involves seeing which school(s) or program(s) are driving the sudden change. If we can pinpoint a plausible driver, then we conduct an online search to see if any notable shift occurred in that school or program in that particular year. For each abnormality, we note our findings.

A.7 Standardization

Once the base sample is defined, we then standardize the data as follows:

First, we standardize the Excel files that are the output of the transcription process. If multiple years of a school were transcribed in a single file, then we split that spreadsheet to create a separate file for each individual year. If a single year was transcribed across multiple files, then we append those spreadsheets together to create a single file for that individual year. We then rename all files in a systematic way ("School_Year.xlsx").

Second, we begin the main portion of the standardization process. First, we create consistent variables that we use to identify each school-year pair. We do so with three variables: Year, School, and OPEID. The latter is a unique identification number used by the U.S. Department of Education's Office of Postsecondary Education (OPE), and it allows us to easily identify schools across our datasets without having to rely on the school's name.

Third, we rename variables, as necessary, to ensure consistency across files. The complete set of final variables are as follows: Name, Married, Gender, Previous Degree, Degree, School, Program, Class Standing, Town, State, Country (if outside US), and Page Number. Note that not all variables are available for all school-year pairs.

Fourth, we use this set of uniform variables to implement the school-year-specific sample restrictions as determined by our sample selection process.

Fifth, with this undergraduate sample in hand, we then proceed with cleaning. Given that names are recorded in various formats (e.g., "First Name Last Name", "Last Name, First Name", etc.), we extract the first and last name separately and create a new variable for each. We then clean the names for non-standard symbols (e.g., accents), prefixes (e.g., "Miss"), and suffixes (e.g., "Jr.").

Sixth, we assign a gender to each individual. For schools that were always all-women, we know that every student across all years is a woman. For schools that were all-women (all-men) and then turned coeducational, we know that every student in each year until the school went coeducational was a woman (man). In some school-year pairs we have "Mrs.", "Ms.", or "Miss." recorded, which enables us to infer that everyone who has one of these prefixes is a woman, and that everyone else is a man. In other cases, the school lists students by gender (e.g., a list of senior men and a separate list of senior women), and so gender was already recorded during the transcription process. These three cases are considered assignments with certainty. For all students who do not yet have a gender reported, we are forced to create a gender likelihood variable. To do so, we calculate the probability that the first name is male using data from the U.S. population census. If the probability is above 80%, we record male. If the probability is below 20%, we record female. In both cases, we flag these as "uncertain" assignments. If the probability is between 20% and 80%, we do not assign either gender.

Seventh, we turn to our geographic variables. They include: "Town", "State", and "Country (if outside US)". We perform string cleaning on all three variables, and standardize state and country names. In the rare case that a town is reported without a state, we manually match it to a state if this can be done with certainty. (If that town exists in multiple states, for example, we would not match it.) Through geocoding (using OpenStreetMaps data via Nominatim), we then assign latitude and longitude information for each hometown using the

centroid of the town (if in the US) or country (if outside the US). If any location is unable to be geo-coded, we manually go through each and every one to see if a simple transcription error caused the problem (e.g., "Stanfodr" instead of "Stanford"). In this case, we make the correction and re-run the original geo-coding process. However, for the cases in which the transcription error is too large, or the location is too uncertain (e.g., "Venice" was recorded as the location without any other information, but this could be referring to Los Angeles, California or Italy), we omit the location from the geo-coding process and proceed.

Eighth, we only want to include an individual in our final, ready-to-link sample the first time that they appear in our data. Hence, we drop any duplicate name that appears within a given school within a rolling eight-year period.

Ninth, we construct a standardized entry cohort variable for each student. This first involves mapping all possible class standing values for a particular school into a standardized class standing variable. Our standardized class standing variables takes on values of zero to five. Zero corresponds to an undergraduate with an unknown specific class standing, one corresponds to freshmen, two corresponds to sophomores, three corresponds to juniors, four corresponds to seniors, and five corresponds to "super seniors". As an example, "Freshmen", "Freshman", "1", and "First Year" would all map to a value of one in our standardized class standing variable. For a school that reported semesters, "1" and "2" would map to a standardized value of "1", "3" and "4" would map to a standardized value of "2", etc.

Tenth, we use our standardized class standing variables, coupled with the document year, to estimate the year in which each student entered the institution. For example, if a student has a standardized class standing value of "1", then we estimate their entering cohort to be the first year in which we observe them in our data. This creates our "Entry Cohort" variable that is used as a basis for much of our analysis.

Eleventh, we append all years and schools together into a master file that is ready to be linked.

A.8 Linking Strategy

We link students from the college registers to a childhood census using the following linking strategy:

First, we clean names in the college register dataset to remove any non-alphabetic characters and to account for common misspellings and nicknames (e.g., Ben and Benjamin should be treated as the same name).

Second, we restrict the college register sample to individuals who are unique based on their exact names within an eight-year period for a given school. We implement this step because individuals might show up multiple times in the student-level data (for instance, as a Freshman and a Sophomore).

Third, we match individuals in the college register sample to the childhood census using their exact names and gender.[48] We match individuals in the registers data to the two closest

[48]For individuals with no assigned gender, we match them simply based on their exact names. Students are assigned a gender based on the university they attended (whether men only or women only) or their name if they attended a coed institution. See Section A.7.

pre-college censuses available (e.g., we link students who we observe starting college in 1941-1950 to both the 1940 and 1930 censuses). We restrict the set of potential matches in the earlier census to individuals who: (1) would have been between the age of 17 and 28 at the time at which we observe them attending college, and (2) if age 6-18 at the time at which we observe them in the childhood census, attended school in the previous year.[49] At this step, there are three possibilities:

1. If there is a unique match, then this pair of observations is considered a match.

2. If there are no matches, then the observation is discarded.

3. If there are multiple matches to the same census, we use a student's hometown information (when available) for disambiguation. Specifically, we count the number of potential matches that resided within 100 kilometers of a student's reported hometown. If there is a single potential match within that radius, the observation is considered a match.

4. If there are unique matches to more than one census, we keep the most recent observation in which the student is still observed living with their parents (e.g., if we match a student to both the 1920 and 1930 censuses at ages 5 and 15, respectively, we keep the census observation at which the student was 15).

A.9 Collecting New Data on Dates of SAT Adoption

We collected data on the dates at which each of the colleges in our data adopted the Scholastic Aptitude Test (SAT) as a formal admission requirement. To do so, we gathered information from annual college catalogues detailing each college's admission requirements in each year. Table A6 provides a comprehensive list of the date at which each college adopted the SAT and the data source that we use to establish such date.

A.10 Assigning Exact Birth Dates

One innovation in our analysis of the G.I. Bill is to use more precise information on an individual's birth date than is available in U.S. census records. U.S. census records provide the age of each person, and so even mapping this information to a birth year can be problematic. This is additionally subject to any measurement errors associated with miscalculating age or age heaping. Instesad, we use data from profiles on the Family Tree at familysearch.org in order to obtain the *exact* date of birth for individuals in our sample.

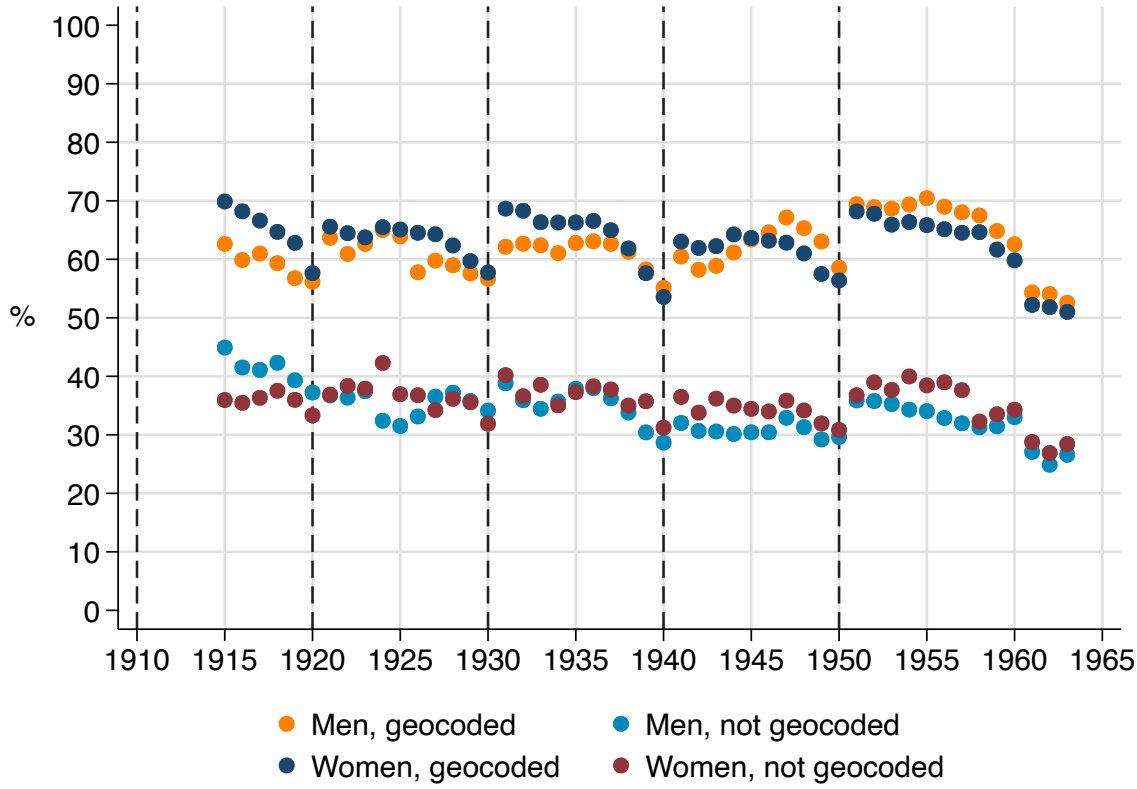
Profiles on FamilySearch are edited and curated by individuals doing family history research, usually for people to whom they are related. The profile has place holders where birth, marriage, and death information can be updated (and this is done in as detailed of a

[49] We do so to avoid including in the sample individuals who would have been very unlikely to attend college later in life and hence are likely to be false positives. All our results are nevertheless similar if we included these individuals in the sample.

fashion as the individual updating the profile decides to include). Sometimes, a profile can include just a birth year or something as imprecise as "about 1920". We focus on individuals who have both a month and year of birth (most of whom have a full precise birth date).

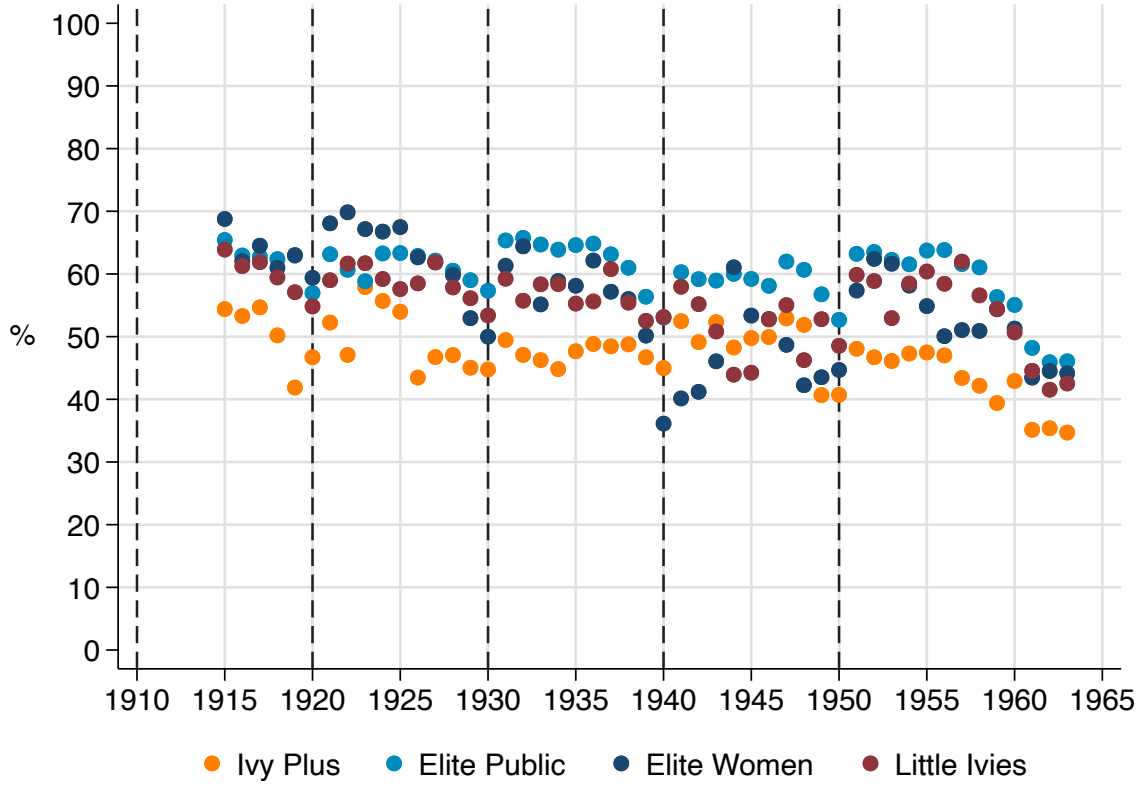
As part of work that the Record Linking Lab has done with FamilySearch, we have access to a crosswalk that allows us to observe if a person in the census has a match to a profile on the Family Tree. Thus, we start with individuals in our sample who are in the census, find their profile on the Family Tree (if they have one), and then extract their birth date from the profile. This provides us with a final sample that includes the birth month and year for approximately 45% of our sample. Importantly, there is little correlation between observing such information and students' socioeconomic backgrounds (see Figure A5).

Figure A1: Matching Rates of College Registers to Full-Count U.S. Population Censuses, by Entering Cohort, Gender, and Whether Hometown Information Available



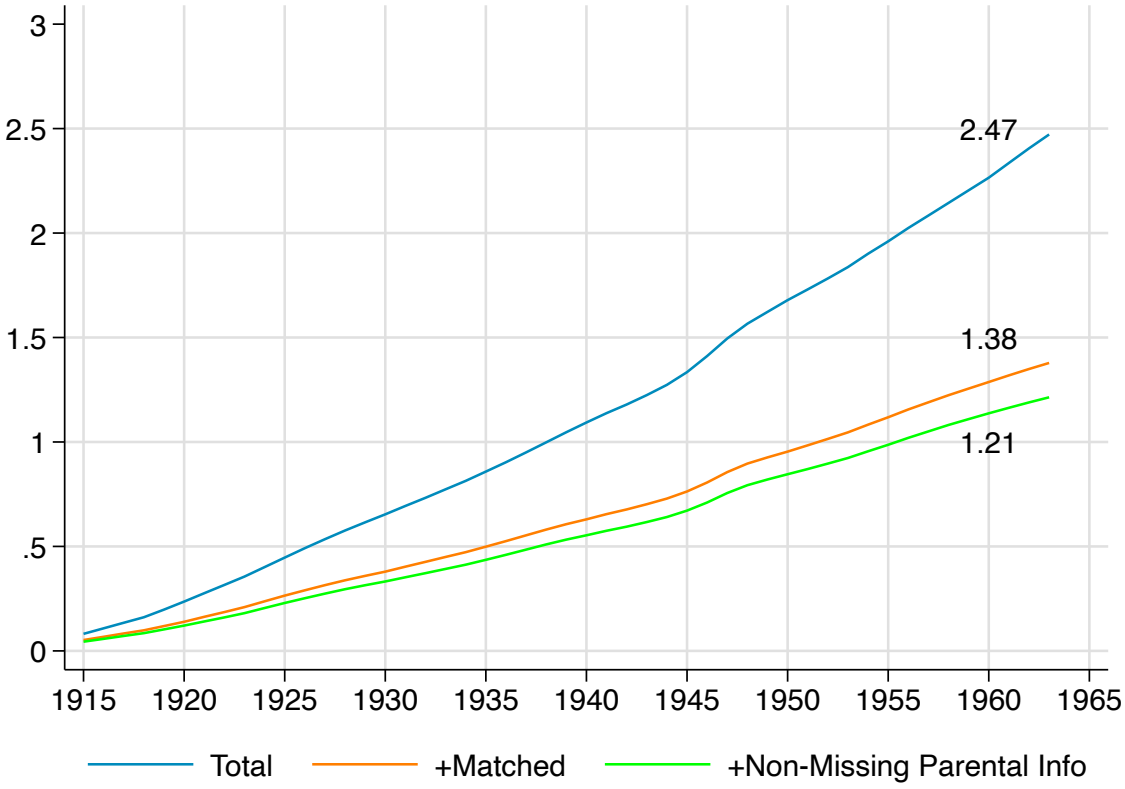
Notes: This figure presents the match rates, by entering cohort, gender, and whether the observation included hometown information, that we achieved by linking college registers to full-count U.S. population censuses. We link students to the two closest censuses prior to them attending college (e.g., we link students who we first observe attending college in 1911-1920 back to the 1910 and 1900 censuses). The linking uses information on students' exact first and last names, hometowns when available, and approximate birth years (assuming they would have been between the age of 17 and 28 when we observe them attending college).

Figure A2: Matching Rates of College Registers to Full-Count U.S. Population Censuses, by Entering Cohort and Group of Colleges



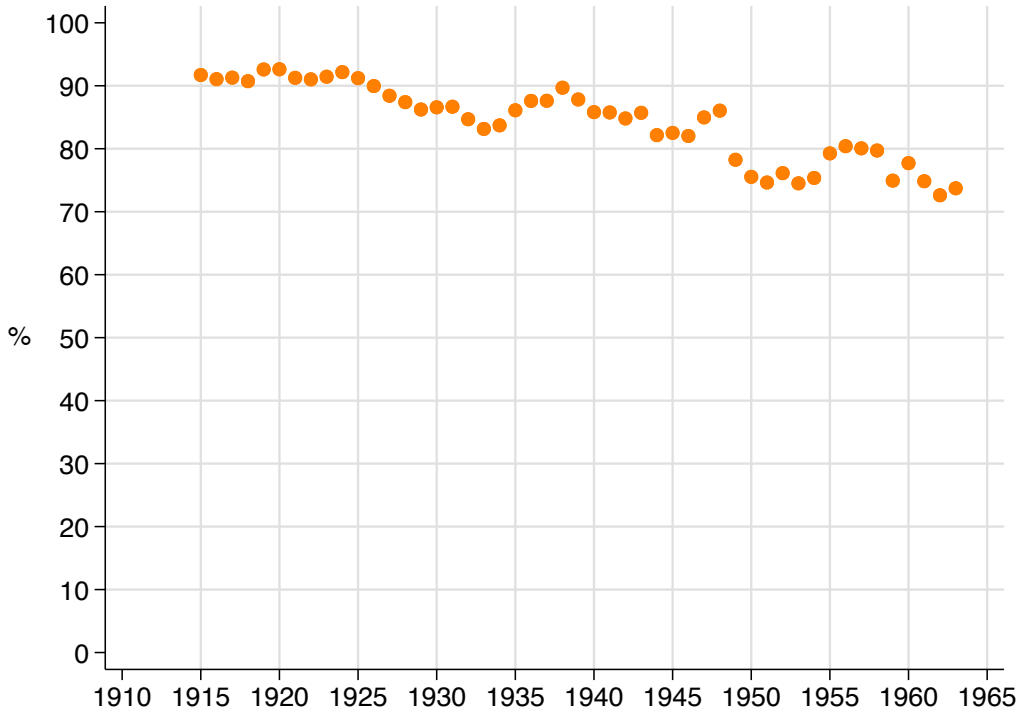
Notes: This figure presents the match rates, by entering cohort and gender, that we achieved by linking college registers to full-count U.S. population censuses, separately for each group of colleges in our data.

Figure A3: Cumulative Sample Size, Before and After Census Linking (in Millions)



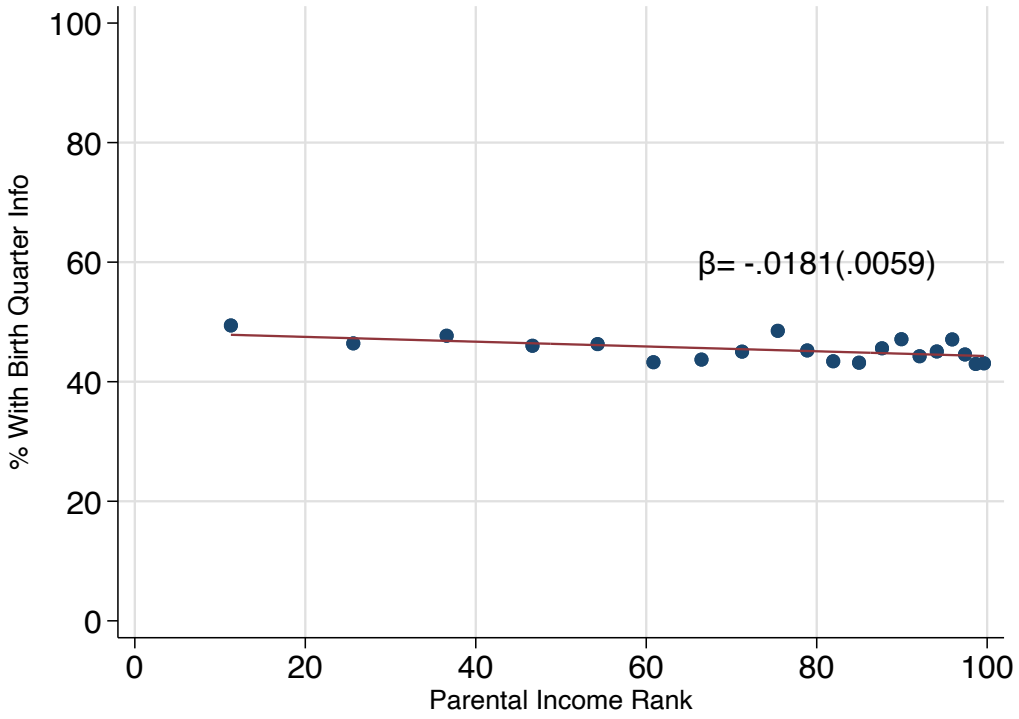
Notes: This figure shows the cumulative number of observations (1) before linking students to the census, (2) after linking students to the census, and (3) after restricting the sample to students for whom we observe parental background information in the census.

Figure A4: Proportion of Geocoded Observations by Entering Cohort



Notes: This figure shows the proportion of individuals in our data who report hometown information and are successfully geocoded, by entering cohort.

Figure A5: Relationship Between Parental Income Rank and Likelihood of Observing Birth Quarter Information



Notes: This figure shows the relationship between students' parental income rank (x-axis) and the likelihood of observing exact birth quarter information.

Figure A6: Illustration of Upwork Job Posting

Instructions:

Please refer to the PDF (ToTranscribe.pdf) that you were given. You will see that each page of this PDF consists of a list of students. Your job is to transcribe information from the PDF into the Excel spreadsheet (Transcription.xlsx) that you were given. The Excel spreadsheet already has the columns filled out, with column 1 labeled "Name", column 2 labeled “College”, column 3 labeled “Degree”, column 4 labeled “Program”, column 5 labeled “Town”, column 6 labeled “State”, column 7 labeled “Country (if outside US)”, and column 8 labeled “Page Number”.

1. Please transcribe the name exactly as it appears in the PDF into the “Name” column of the Excel spreadsheet.

2. Students are listed in groups, with each group corresponding to a specific college (e.g. “College of Agriculture”) and a specific degree (e.g. “Bachelor of Science”). Please record the college under the “College” column of the Excel spreadsheet, and the degree under the “Degree” column of the Excel spreadsheet.

3. Following each student’s name, you will find the program that the student is enrolled in (e.g. “Animal Science”, “Agricultural Education”, “Plant Science”, etc.) Record this under the “Program” column of the Excel spreadsheet.

4. Following the program, you will find the student’s place of residence. Record this information as follows:
    a. If the location is inside the US: Record the town under the “Town” column and the state under the “State” column. Leave the “Country (if outside US)” column blank. If a town is listed without a state, then the state is *always* “CA” by default.
    b. If the location is outside the US: Leave the “Town” and “State” columns blank. Record the country in the “Country” column.

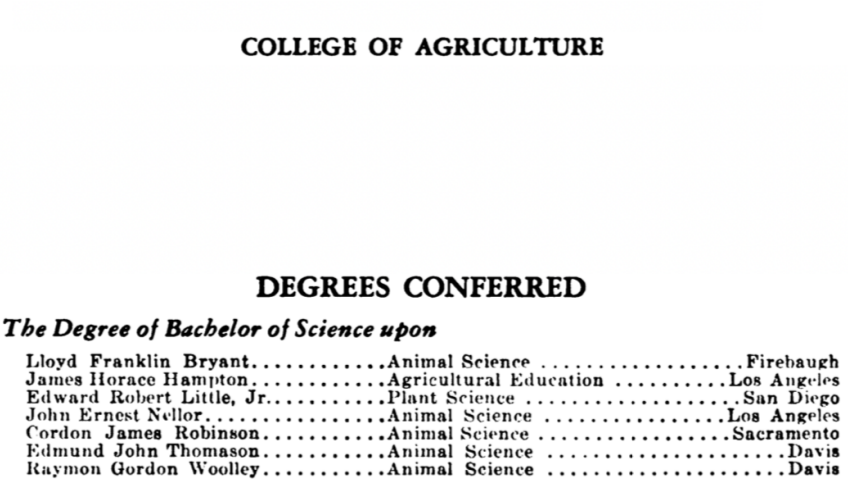
5. Record the page in which each individual appears under the “Page Number” column of the Excel spreadsheet. This should be the page number *inside* the PDF, not of the PDF itself.

Tips:

- Transcribe names without accents or other characters.
- To save time, we recommend filling out the “Page Number” column once per page. In other words: First transcribe all relevant information from a single page, and then go back and easily add the (same) page number to all of these observations.

Example:

Help us take this:



COLLEGE OF AGRICULTURE

DEGREES CONFERRED

The Degree of Bachelor of Science upon

Lloyd Franklin Bryant.....Animal Science .....Firebaugh
James Horace Hampton.....Agricultural Education .....Los Angeles
Edward Robert Little, Jr.....Plant Science .....San Diego
John Ernest Nellor.....Animal Science .....Los Angeles
Cordon James Robinson.....Animal Science .....Sacramento
Edmund John Thomason.....Animal Science .....Davis
Raymon Gordon Woolley.....Animal Science .....Davis

And turn it into this:

Name	School	Degree	Program	Town	State	Country (if outside US)	Page Number
Lloyd Franklin Bryant	College of Agriculture	Bachelor of Science	Animal Science	Firebaugh	CA		6
James Horace Hampton	College of Agriculture	Bachelor of Science	Agricultural Education	Los Angeles	CA		6
Edward Robert Little, Jr.	College of Agriculture	Bachelor of Science	Plant Science	San Diego	CA		6
John Ernest Nellor	College of Agriculture	Bachelor of Science	Animal Science	Los Angeles	CA		6
Cordon James Robinson	College of Agriculture	Bachelor of Science	Animal Science	Sacramento	CA		6
Edmund John Thomason	College of Agriculture	Bachelor of Science	Animal Science	Davis	CA		6
Raymon Gordon Woolley	College of Agriculture	Bachelor of Science	Animal Science	Davis	CA		6

Notes: This figure provides an example of an instruction sheet given to freelancers on the Upwork platform as part of the data digitization process.

Table A1: Data Compilation Overview

Entry Cohort	Elite Colleges	Any College
1915-1940	College Registers	1910-1940, 1920-1950 Linked US Population Censuses
1940-1965	College Registers	Survey of Occupational Change in a Generation (OCG)
1966-2000	CIRP Freshman Survey (TFS)	CIRP Freshman Survey (TFS)
2000-2010	Opportunity Insights	CIRP Freshman Survey (TFS)

Notes: This table provides an overview of all birth cohorts included in our dataset and the accompanying data source employed for each. For each birth cohort, we present the data source for both the elite colleges and "any college" (i.e., aggregate college-going population) samples.

Table A2: Data Collection Documentation, Ivy Plus

School	Years	Document	Students
Brown	1915-1943,1946-1966	**Catalogues** (1915-1926); **Yearbooks** (1927-1943,1946-1966)	**All undergraduates** (1915-1926); **Seniors** (1927-1943,1946-1966)
Chicago	1915-1966	**Comm. records** (1915-1966)	**Seniors** (1915-1966)
Columbia	1915-1966	**Catalogues** (1915-1944); **Comm. records** (1945-1966)	**All undergraduates** (1915-1944); **Seniors** (1945-1966)
Cornell	1915-1932,1935-1966	**Catalogues** (1915-1931); **Yearbooks** (1932,1937,1959-1966); **Directories** (1935-1936,1938-1945,1948-1958); **Telephone directories** (1946-1947)	**All undergraduates** (1915-1931,1935-1936,1938-1958); **Seniors** (1932,1937,1959-1966)
Dartmouth	1915-1966	**Catalogues** (1915-1940); **Bulletins** (1941-1945); **Comm. records** (1946-1950); **Directories** (1951-1966)	**Seniors** (1915-1940,1946-1950); **All undergraduates** (1941-1945,1951-1966)
Duke	1915-1966	**Catalogues** (1915-1928); **Bulletins** (1929-1966)	**All undergraduates** (1915-1953); **Seniors** (1954-1966)
Harvard	1915-1948,1951-1966	**Catalogues** (1915-1948); **Directories** (1951-1966)	**All undergraduates** (1915-1948, 1951-1966)
MIT	1915-1922,1924-1966	**Yearbooks** (1915-1922,1924-1966)	**All undergraduates** (1915-1922,1924-1944); **Seniors** (1945-1966)
UPenn	1915-1966	**Catalogues** (1915-1924); **Directories** (1925-1929,1931-1949,1951-1966); **Yearbooks** (1930,1950)	**All undergraduates** (1915-1929,1931-1949,1951-1966); **Seniors** (1930,1950)
Princeton	1915-1966	**Catalogues** (1915-1926); **Alumni lists** (1927-1966)	**All undergraduates** (1915-1926); **Seniors** (1927-1966)
Stanford	1915-1966	**Registers** (1915-1946); **Yearbooks** (1947-1966)	**All undergraduates** (1915-1946); **Seniors** (1947-1966)
Yale	1915-1966	**Catalogues** (1915-1950); **Bulletins** (1951-1966)	**All undergraduates** (1915-1918,1920-1926,1951-1966); **Seniors** (1919,1927-1950)

Notes: This table provides, for each Ivy Plus college in our sample, a thorough overview of the following: years for which we have historical documents, the general type of historical documents (e.g., registers, catalogues, etc.) for each year, and the students included (e.g., all undergraduates, only seniors, etc.) in each year. "Comm. records" indicate commencement records.

Table A3: Data Collection Documentation, Little Ivies

School	Years	Document	Students
Amherst	1915-1966	**Catalogues** (1915-1933); **Yearbooks** (1934-1940); **Bulletins** (1941-1947,1949-1966); **Comm. records** (1948)	**All undergraduates** (1915-1947,1949-1966); **Seniors** (1948)
Bates	1915-1966	**Yearbooks** (1915-1966)	**All undergraduates** (1915-1916,1918-1933,1936,1942-1948,1950) **Seniors** (1917,1934-1935,1937-1941,1949,1951-1966)
Bowdoin	1915-1966	**Catalogues** (1915-1966)	**All undergraduates** (1915-1966)
Colby	1915-1966	**Bulletins** (1915-1921,1930,1939-1966); **Catalogues** (1922-1929,1931-1938)	**All undergraduates** (1915-1960); **Seniors** (1961-1966)
Conn	1916-1917, 1919-1966	**Bulletins** (1916-1917); **Yearbooks** (1919-1966)	**Freshmen** (1916); **Freshmen and sophomores** (1917); **All undergraduates** (1919-1948); **Seniors** (1949-1966)
Hamilton	1915-1966	**Catalogues** (1915-1918,1920-1922); **Bulletins** (1919,1923-1936); **Comm. records** (1937-1948,1950-1963,1965-1966); **Yearbooks** (1949,1964)	**All undergraduates** (1915-1918,1920-1922,1949); **Seniors** (1919,1923-1948,1950-1966)
Haverford	1915-1943, 1946-1966	**Yearbooks** (1915-1943,1946-1966)	**All undergraduates** (1915-1935); **Seniors** (1936-1943,1946-1966)
Lafayette	1915-1933	**Catalogues** (1915-1933)	**All undergraduates** (1915-1933)
Middlebury	1915-1966	**Bulletins** (1915-1930,1939-1959,1961,1964-1966); **Yearbooks** (1931-1938,1960,1962-1963)	**All undergraduates** (1915-1937,1939-1951); **Sophomores, juniors, and seniors** (1938); **Seniors** (1952-1966)
Swarthmore	1915-1966	**Yearbooks** (1915-1953,1955-1964,1966); **Catalogues** (1954,1965)	**All undergraduates** (1915-1933,1935,1938-1940,1943,1947); **Juniors and seniors** (1934,1936-1937,1944); **Juniors** (1941-1942,1949,1951-1953,1959); **Seniors** (1945-1946,1948,1950,1954-1958,1960-1966)
Trinity	1915-1966	**Bulletins** (1915-1921,1923-1942,1944-1966); **Comm. records** (1922,1943)	**All undergraduates** (1915-1921,1923-1942,1944-1966); **Seniors** (1922,1943)
Tufts	1915-1966	**Comm. records** (1915-1966)	**Seniors** (1915-1966)
Wesleyan	1915-1965	**Yearbooks** (1915-1965)	**All undergraduates** (1915-1944,1946-1947); **Seniors** (1945, 1948-1965)
Williams	1915-1944,1946, 1948-1966	**Yearbooks** (1915-1944,1946,1948-1966)	**All undergraduates** (1915-1944,1948-1956); **Seniors** (1946,1958-1966); **Freshmen and seniors** (1957)

Notes: This table provides, for each Little Ivy college in our sample, a thorough overview of the following: years for which we have historical documents, the general type of historical documents (e.g., registers, catalogues, etc.) for each year, and the students included (e.g., all undergraduates, seniors only, etc.) in each year. "Comm. records" indicate commencement records.

Table A4: Data Collection Documentation, Elite Women's

School	Years	Document	Students
Barnard	1915-1966	**Yearbooks** (1915-1918,1921-1953,1955-1966); **Alumni registers** (1919-1920); **Newspapers** (1954)	**All undergraduates** (1915-1918,1921-1953,1955-1964); **Juniors and Seniors** (1919-1920); **Seniors** (1954, 1965-1966)
Bryn Mawr	1915-1966	**Catalogues** (1915-1942,1945-1947); **Yearbooks** (1943-1944,1948-1949); **Newspapers** (1950,1957-1958,1962-1964); **Alumni lists** (1951-1956,1959-1961,1965-1966)	**All undergraduates** (1915-1942,1945-1947); **Seniors** (1943-1944,1948-1966)
Mount Holyoke	1915-1966	**Catalogues** (1915-1966)	**All undergraduates** (1915-1966)
Pembroke	1915-1945,1949-1966	**Catalogues** (1915-1922,1924-1927); **Yearbooks** (1923,1928-1945,1949-1966)	**All undergraduates** (1915-1945,1949-1966)
Radcliffe	1915-1966	**Registers** (1915); **Directories** (1916-1966)	**Seniors** (1915); **All undergraduates** (1916-1966)
Smith	1915-1966	**Catalogues** (1915-1929,1935,1940,1943,1947,1955); **Yearbooks** (1930-1933,1953); **Comm. records** (1934,1936-1939,1941-1942,1944-1946, 1948-1952,1954,1956-1966)	**All undergraduates** (1915-1929,1935,1940,1943,1947); **Seniors** (1930-1934,1936-1939,1941-1942,1944-1946, 1948-1966)
Vassar	1915-1957	**Catalogues** (1915-1928); **Alumni registers** (1929-1957)	**All undergraduates** (1915-1928); **Seniors** (1929-1957)
Wellesley	1915-1966	**Bulletins** (1915-1966)	**Seniors** (1915-1966)

Notes: This table provides, for each elite women's college in our sample, a thorough overview of the following: years for which we have historical documents, the general type of historical documents (e.g., registers, catalogues, etc.) for each year, and the students included (e.g., all undergraduates, seniors only, etc.) in each year. "Comm. records" indicate commencement records.

Table A5: Data Collection Documentation, Elite Public

School	Years	Document	Students
Arizona	1915-1966	**Yearbooks** (1915-1966)	**Juniors and seniors** (1915-1916,1925,1928-1929,1931-1935); **All undergraduates** (1917-1924,1926); **Freshmen, sophomores, and seniors** (1927); **Seniors** (1930,1936-1966)
UC Berkeley	1915-1945,1947-1966	**Registers** (1915-1945,1947); **Comm. records** (1948-1966)	**All undergraduates** (1915-1945, 1947); **Seniors** (1948-1966)
UC Davis	1923-1934,1936-1945,1947-1959, 1961-1964	**Registers** (1923-1934,1936-1945,1947); **Comm. records** (1948-1959,1961-1964)	**All undergraduates** (1923-1934,1936-1945,1947); **Seniors** (1948-1959,1961-1964)
UCLA	1922-1945,1947-1966	**Registers** (1922-1945,1947); **Comm. records** (1948-1959); **Yearbooks** (1960-1966)	**All undergraduates** (1922-1945, 1947); **Seniors** (1948-1966)
UCSB	1946-1966	**Yearbooks** (1946-1948,1960-1966); **Comm. records** (1949-1959)	**Seniors** (1946-1966)
Colorado	1915-1949,1951-1960,1962-1966	**Catalogues** (1915-1949,1951-1960,1962-1966)	**All undergraduates** (1915-1949,1951-1960,1962-1966)
Connecticut	1915-1966	**Comm. records** (1915-1966)	**Seniors** (1915-1966)
Delaware	1915-1966	**Yearbooks** (1915-1966)	**Seniors** (1915,1920,1922,1924,1928-1930,1932-1937,1939, 1941,1943,1946); **All undergraduates** (1916-1919, 1921,1923,1925-1927,1931, 1938,1940,1942,1944-1945,1947-1966)
Florida	1915-1943,1945-1954,1957-1966	**Yearbooks** (1915-1916,1919-1920,1927-1943, 1946-1949,1953,1957-1965); **Directories** (1917-1918,1921-1926,1945, 1950-1952,1954, 1966)	**Seniors** (1915-1916,1919,1927-1932, 1939,1957-1966); **All undergraduates** (1917-1918,1921-1926,1933,1936-1937, 1940-1942,1945-1952,1954); **Mixed** (1920,1934-1935,1938, 1943,1953)
Georgia	1915-1966	**Yearbooks** (1915-1966)	**All undergraduates** (1915-1916,1923-1924,1926,1928-1938, 1940-1963,1965-1966); **Seniors** (1917,1919,1921-1922,1927,1964) **Juniors and seniors** (1918,1920); **Sophomores, juniors, and seniors** (1925); **Freshmen, sophomores, and seniors** (1939);
Georgia IT	1915-1966	**Yearbooks** (1915-1966)	**Seniors** (1915-1926,1931,1934-1940,1942-1943, 1951-1956,1965-1966) **Juniors and seniors** (1927-1928,1930,1932,1944,1946); **All undergraduates** (1929,1933,1941,1945,1947-1950, 1957-1958,1960-1964); **Sophomores, juniors, and seniors** (1959);
Illinois	1915-1966	**Yearbooks** (1915-1966)	**Seniors** (1915-1919,1925-1966); **Juniors** (1920-1924)

School	Years	Document	Students
Indiana	1915,1917-1922,1924-1938, 1941-1947,1949-1955, 1957-1966	**Registers** (1915,1917-1922,1925-1938); **Comm. records** (1924); **Yearbooks** (1941-1947,1949-1955,1957-1966)	**All undergraduates** (1915,1920-1922,1925-1938); **Seniors** (1917-1919,1924,1941-1947,1949-1955, 1957-1966)
Iowa	1915-1945,1947-1966	**Yearbooks** (1915-1945,1947-1966)	**Juniors** (1915,1918-1919,1921-1926,1928-1940,1943,1945); **Seniors** (1916-1917,1920,1927,1942,1947-1966); **Juniors and seniors** (1941,1944)
Maryland	1915-1922,1925-1966	**Yearbooks** (1915-1922,1925-1966)	**Seniors** (1915-1917,1921-1922,1925-1966); **Juniors and seniors** (1918-1920)
Miami (OH)	1915-1966	**Yearbooks** (1915-1966)	**All undergraduates** (1915-1920,1922-1927,1934,1937-1948, 1950-1963); **Freshmen, sophomores, and seniors** (1921); **Juniors and seniors** (1928-1933,1935-1936,1949); **Seniors** (1964-1966)
Michigan	1915-1966	**Registers** (1915-1926); **Yearbooks** (1927-1966)	**All undergraduates** (1915-1926); **Seniors** (1927-1966)
Michigan State	1915-1920,1922,1924-1966	**Comm. records** (1915-1920,1922,1924-1966)	**Seniors** (1915-1920,1922,1924-1966)
Minnesota	1915-1966	**Yearbooks** (1915-1966)	**Juniors** (1915-1926) **Seniors** (1927-1966)
SUNY - Binghamton	1949-1966	**Yearbooks** (1949-1966)	**All undergraduates** (1949-1966)
North Carolina	1915-1966	**Registers** (1915-1939); **Comm. records** (1940-1966)	**All undergraduates** (1915-1923,1926-1933,1935-1939); **Seniors** (1924-1925,1934,1940-1966)
Ohio State	1915-1966	**Yearbooks** (1915-1966)	**Seniors** (1915-1939,1961-1966); **All undergraduates** (1940-1960)
Penn State	1915-1966	**Yearbooks** (1915-1966)	**Juniors and seniors** (1915-1930); **Seniors** (1931-1966)
Pitt	1915-1966	**Yearbooks** (1915-1966)	**All undergraduates** (1915-1921,1923-1925,1947-1951); **Juniors** (1922,1927-1928,1930); **Juniors and seniors** (1926, 1929); **Seniors** (1931-1932,1934-1946,1952-1966) **Seniors & medical undergrads** (1933);
Rutgers			

School	Years	Document	Students
Texas	1915-1966	**Yearbooks** (1915-1966)	**Seniors** (1915-1926); **Juniors and seniors** (1927-1931); **All undergraduates** (1932-1966)
Vermont	1915-1954,1956-1966	**Yearbooks** (1915-1954,1956-1966)	**All undergraduates** (1915-1917,1919-1920,1923-1931,1933,1936-1948); **Seniors** (1918,1949,1956-1966); **Juniors and seniors** (1921,1934,1950); **Sophomores, juniors, and seniors** (1922); **Freshmen, sophomores, and seniors** (1932); **Juniors** (1935,1951-1954)
Virginia	1915-1944,1946-1966	**Registers** (1915-1936); **Yearbooks** (1937-1944, 1946-1966)	**All undergraduates** (1915-1943,1946-1948,1958-1965); **Seniors** (1944,1949-1957,1966)
Washington	1915-1966	**Yearbooks** (1915-1966)	**Seniors** (1915-1966)
William & Mary	1915-1966	**Registers** (1915-1946); **Comm. records** (1947-1966)	**All undergraduates** (1915-1946); **Seniors** (1947-1966)
Wisconsin	1915-1966	**Yearbooks** (1915-1966)	**Seniors** (1915-1966)

Notes: This table provides, for each elite public institution in our sample, a thorough overview of the following: years for which we have historical documents, the general type of historical documents (e.g., registers, catalogues, etc.) for each year, and the students included (e.g., all undergraduates, seniors only, etc.) in each year. "Comm. records" indicate commencement records.

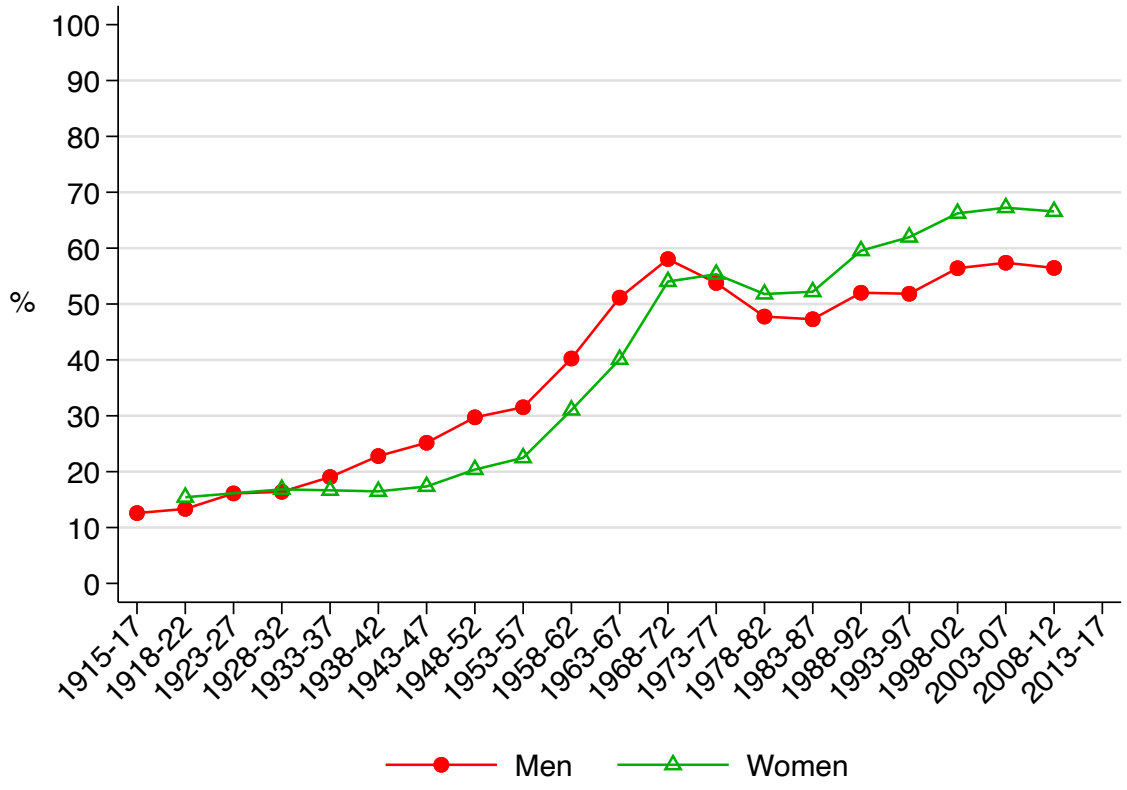
Table A6: Dates of SAT Adoption

School	Year of Adoption	Source
University of Pennsylvania	1926	University Newspaper
Princeton University	1926	University Librarian
Wellesley College	1926	College Bulletin (1926-1927)
Smith College	1926	College Librarian
Vassar College	1926	College Newspaper
Trinity College	1926	College Newspaper
Yale University	1927	University Librarian
Mount Holyoke College	1927	College Bulletin
Bryn Mawr College	1927	College Bulletin
Swarthmore College	1927	College Bulletin
Barnard College	1932	College Bulletin
Haverford College	1933	College Bulletin
Harvard University	1935	University Bulletin
Radcliffe College	1935	College Bulletin
Brown University	1939	University Bulletin
Pembroke College	1939	College Bulletin
Amherst College	1939	College Bulletin
Connecticut College	1939	College Bulletin
Williams College	1940	College Bulletin
Columbia University	1942	University Bulletin
Massachusetts Institute of Technology	1947	Institute Bulletin
Bowdoin College	1947	College Catalogue (1946-1947 and 1947-1948)
Colby College	1947	College Catalogues (1945-1946 and 1946-1947)
Hamilton College	1947	College Librarian
Rutgers University - New Brunswick	1947	University Librarian
Bates College	1948	College Librarian (1946 and 1948 catalogues)
Middlebury College	1949	College Bulletins (1948-1949 and 1949-1950)
Stanford University	1950	University Librarian (article)
Tufts University	1950	University Librarian
Dartmouth College	1951	College Librarian
Lafayette College	1951	College Librarian (Lafayette College Catalogue)
University of Chicago	1951	University Librarian
University of California - Berkeley	1955	The Atlantic
University of Georgia	1957	University Catalogue (1957-1958)
University of North Carolina	1957	University Record (1956-1957 and 1957-1958)
Duke University	1958	University Bulletin (1957-1958 and 1958-1959)
University of Virginia	1958	University Librarian (Newspaper)
William and Mary	1958	College Bulletin (1957-1958 and 1958-1959)
University of Connecticut	1958	University Bulletin (1957-1958 and 1958-1959)
University of Pittsburgh	1958	University Librarian
Cornell University	1959	University Catalogues (1958-1959 and 1959-1960)
University of Iowa	1960	University Librarian
University of Vermont	1960	University Bulletin (1957-1958)
University of California - Davis	1961	University Catalogue (1959-1960 and 1960-1961)
University of California - Los Angeles	1961	University Catalogue (1960-1961 and 1961-1962)
University of California - Santa Barbara	1962	University Catalogue (1960-1961 and 1961-1962)
University of Colorado	1962	University Catalogue (1958-1960 and 1960-1961)
University of Minnesota	1962	University Bulletin (1960-1961)
University of Maryland	1963	University Catalogues (1960-1961 and 1961-1962)
University of Arizona	1963	University Catalogues (1961-1963 and 1963-1964)
Indiana University	1963	University Librarian (IU Board of Trustees Meeting)
University of Michigan	1964	University Librarian
Case Western Reserve University	1965	A Report of the Committee on Admissions and Entrance Examinations
University of Washington	1965	University Librarian
University of Florida	1978	University Record (1978-1979 and 1979-1980)

Notes: This table provides, for each college in our sample, an overview of the year the SAT was first required for all applicants at each institution, as well as the source for the information. ”Librarian” indicates instances where the information was acquired by contacting the relevant institution’s librarian.

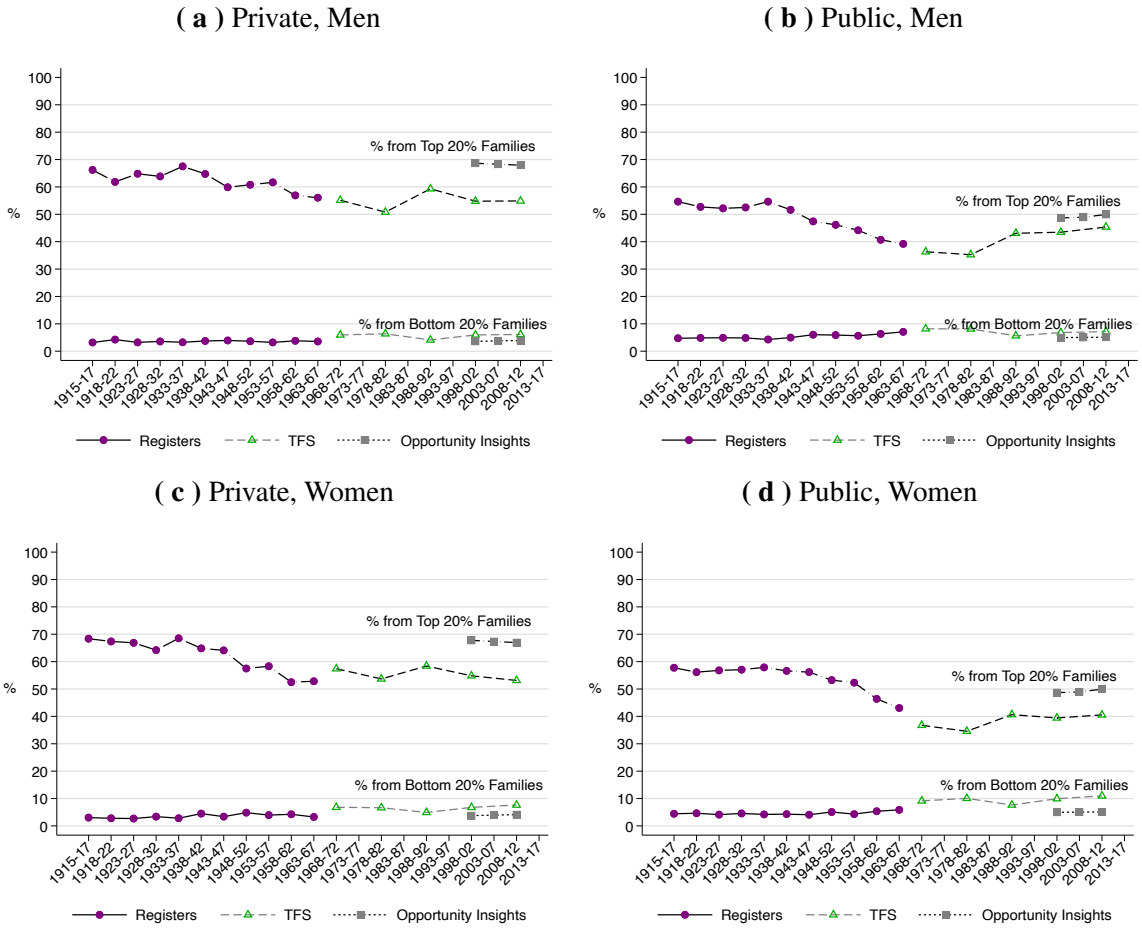
B Additional Results

Figure B1: Share of Individuals With At Least Some College, by Entering Cohort and Gender



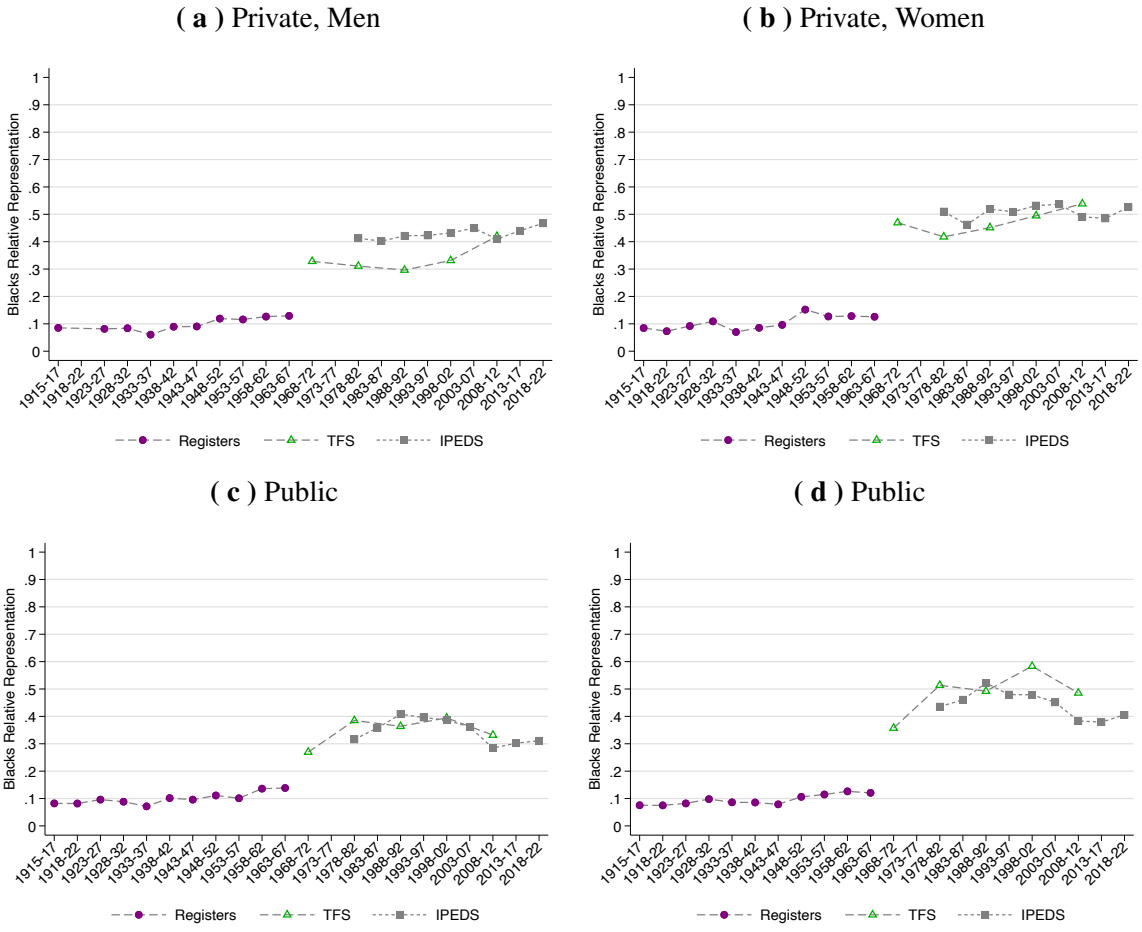
Notes: This figure shows the proportion of individuals with at least one year of college, by typical college entry year (at age 18) and gender. Data are from the U.S. population census (1950-1990) and the ACS (2000-2020). The sample is restricted to U.S.-born individuals.

Figure B2: Comparison Between the UCLA Freshman Survey and Alternative Data Sources, Students' Economic Origins at Elite Colleges



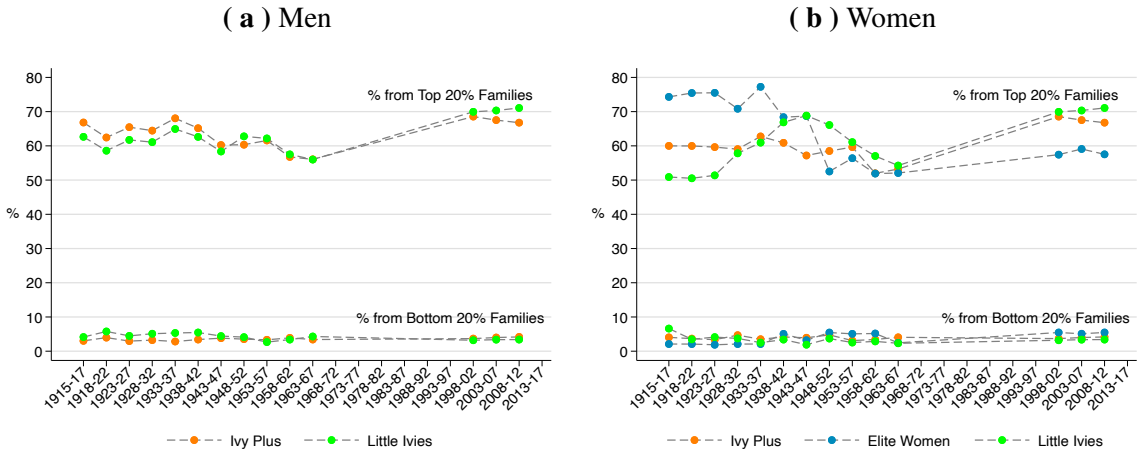
Notes: This figure plots the proportion of students from bottom 20% and top 20% families in elite private (Panels (a) and (c)) and public (Panels (b) and (d)) colleges, by gender (men in Panels (a)-(b) and women in Panels (c)-(d)) and entering cohort. The figure separates the results depending on whether they are based on our linked college registers (1915-1965), the UCLA Freshmen Survey (1970-2010), or Opportunity Insights (2000-2010).

Figure B3: Comparison Between the UCLA Freshman Survey and Alternative Data Sources, Relative Representation of Black Students at Elite Colleges



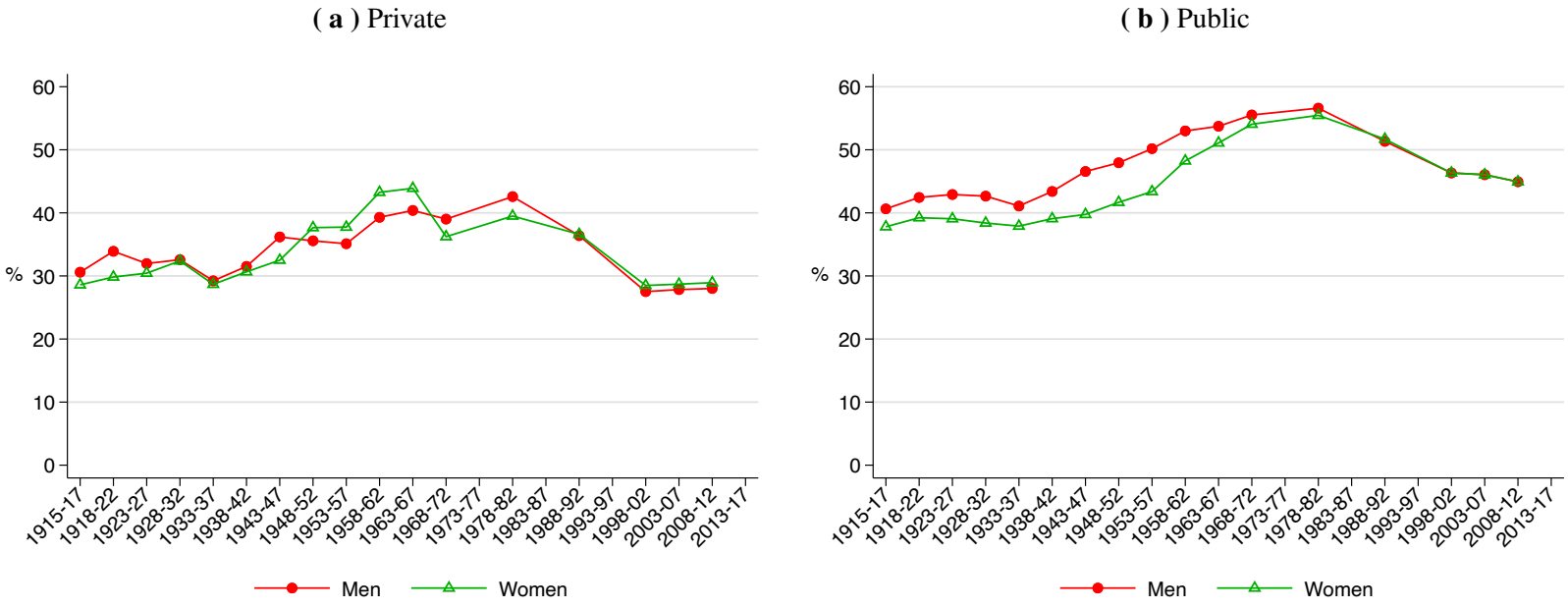
Notes: This figure plots the relative representation of Black students, by gender and cohort, at elite private colleges (Panels (a) and (b)), and at elite public colleges (Panels (c) and (d)). The figure separates the results depending on whether they are based on our linked college registers (1915-1965), the UCLA Freshmen Survey (1970-2010), or data from the Integrated Postsecondary Education Data System (IPEDS) (1980-2022).

Figure B4: Economic Origins of Students at Elite Private Colleges, by School Group and Gender



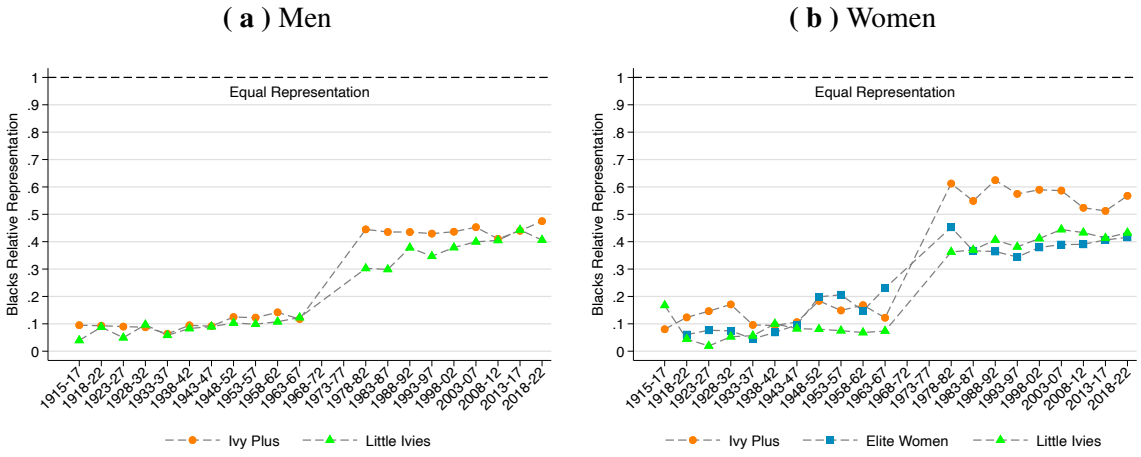
Notes: This figure plots the fraction of students from bottom 20% and top 20% of the parental income distribution, by gender (men in Panel (a) and women in Panel (b)) and cohort, in the Ivy Plus, Little Ivies, and elite women's institutions. Historical data comes from our linked college registers (1915-1965) and modern data come from Opportunity Insights (2000-2010). Note that we cannot produce a full series spanning the 1970-2000 period as TFS survey does not enable us to identify specific groups of colleges.

Figure B5: Proportion of Students from Middle (Percentiles 20% - 80%) of Parental Income Distribution at Elite Colleges



Notes: This figure shows the proportion of students from families in the middle (percentiles 20-80) of the parental income distribution, by entering cohort. Results are presented for elite private schools in Panel (a), and elite public schools in Panel (b). Data comes from a combination of our linked college registers (1915-1966), the UCLA Freshman Survey (1966-1998), and Opportunity Insights (1998-2012).

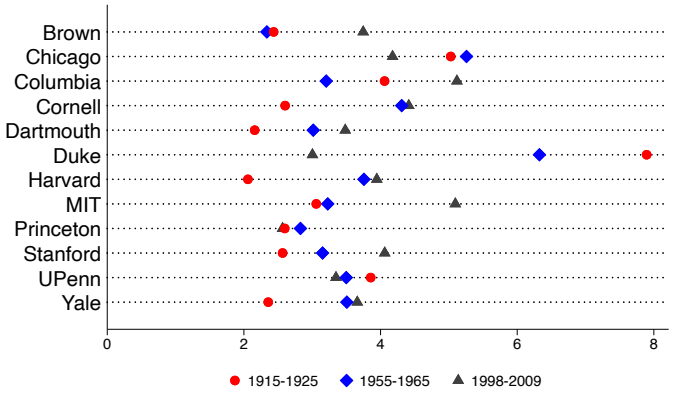
Figure B8: Relative Representation of Black Students at Elite Private Colleges, by College Group and Gender



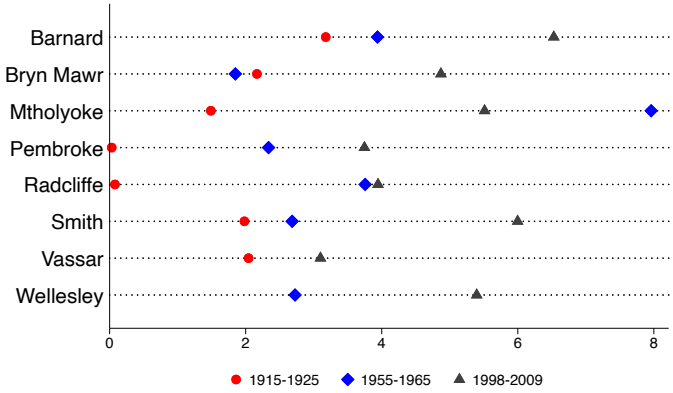
Notes: This figure plots the relative representation of Black students, by gender and cohort, in the Ivy Plus, Little Ivies, and elite women's institutions. Men are shown in Panel (a) and women are shown in Panel (b). Historical data are from our linked college registers and modern data are from the Integrated Postsecondary Education Data System (IPEDS).

Figure B6: Proportion of Students from Bottom 20% at Elite Private Colleges

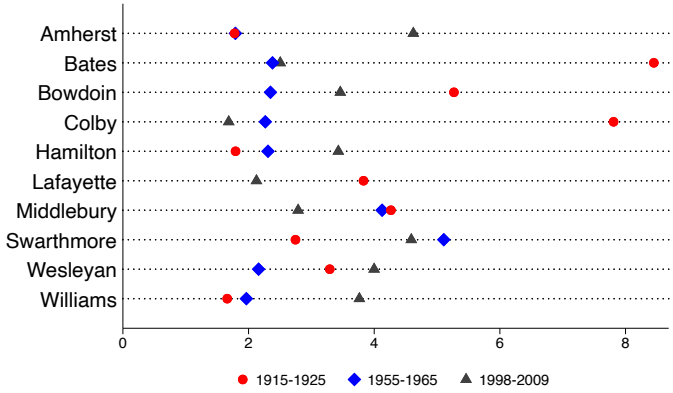
(a) Ivy Plus



(b) Elite Women

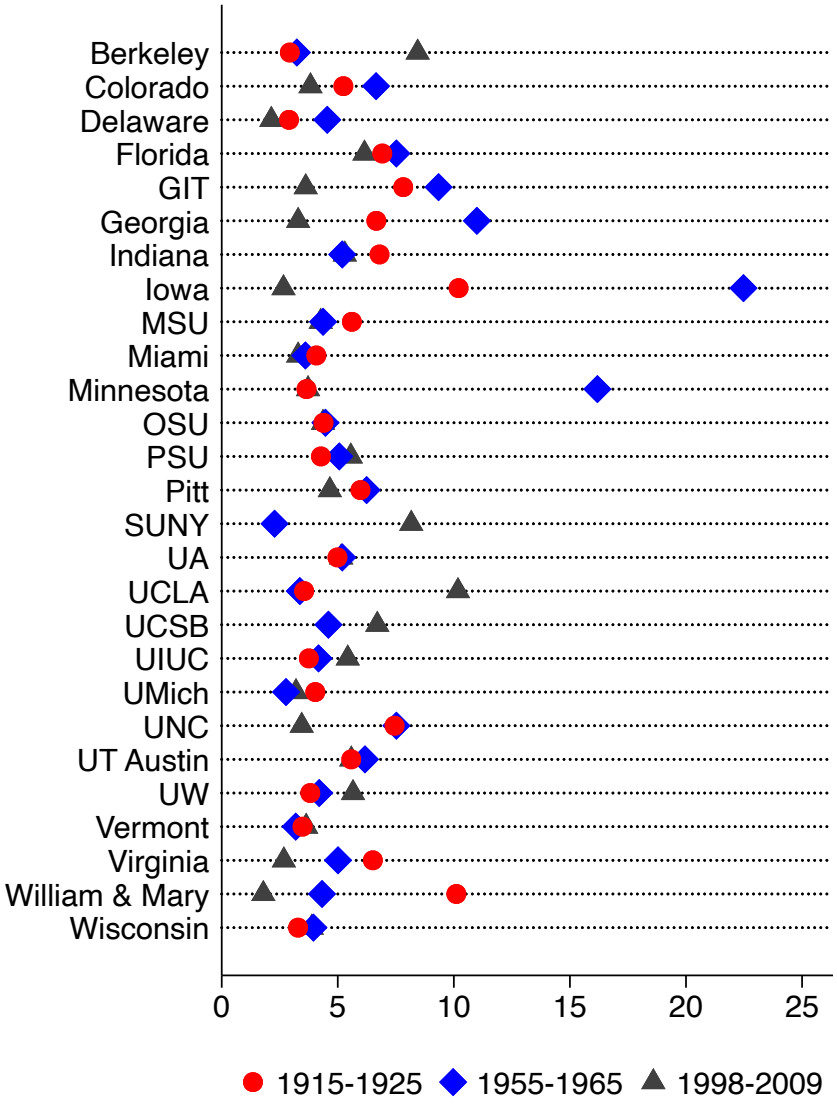


(c) Little Ivies



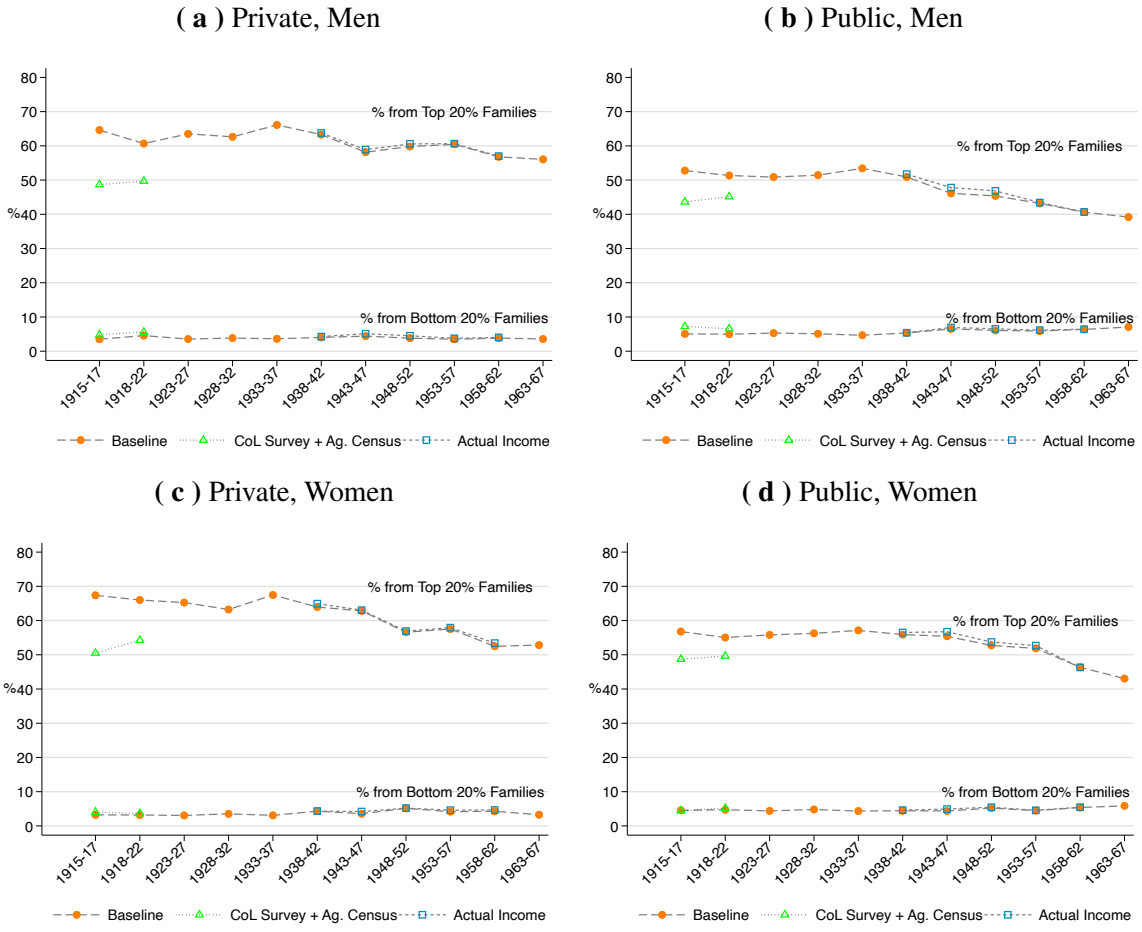
Notes: This figure shows the proportion of students from bottom 20% families at the private colleges in our data in 1915-1925, 1955-1965 and 1998-2009. Panel (a) includes the Ivy Plus, Panel (b) includes Elite Women's colleges, and Panel (c) includes the Little Ivies. 1915-1925 and 1955-1965 data are from our linked college registers and 1998-2009 data are from Opportunity Insights.

Figure B7: Proportion of Students from Bottom 20% at Elite Public Colleges



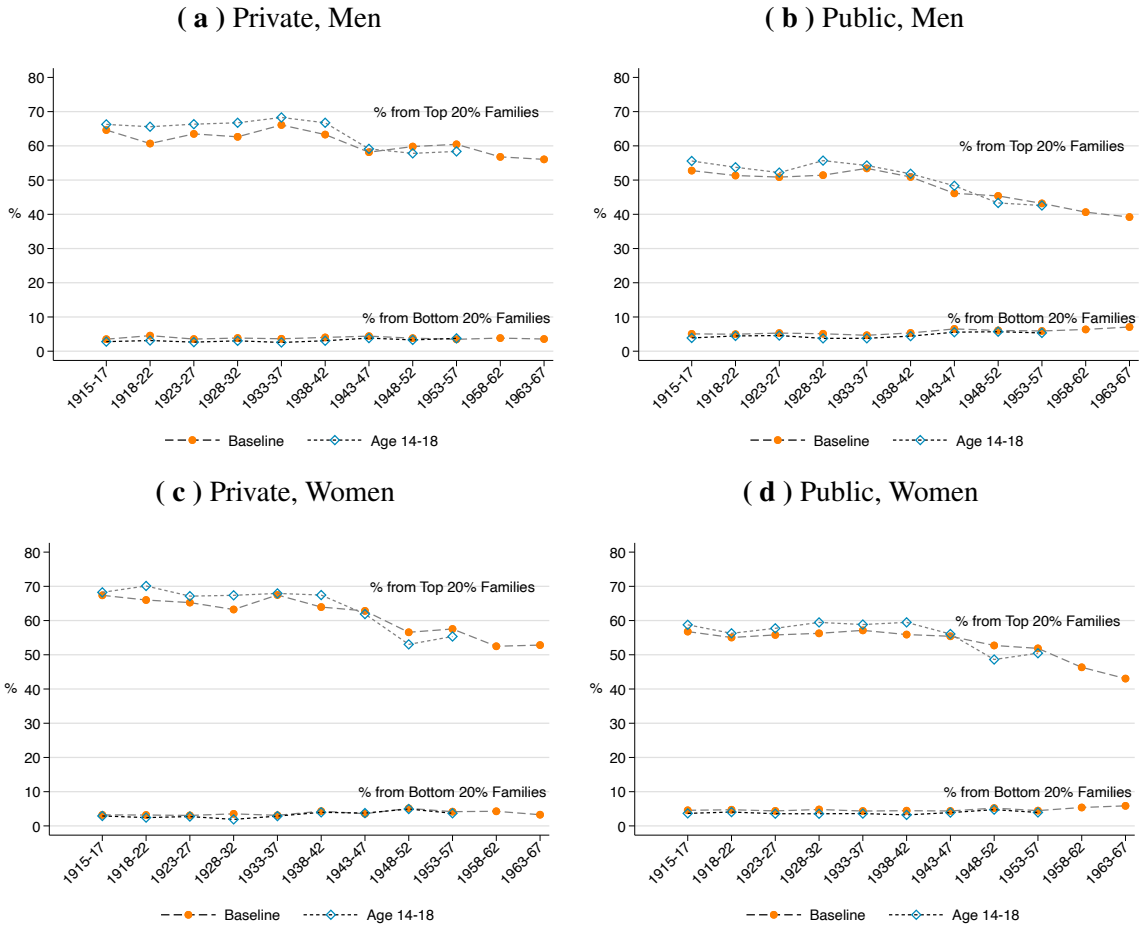
Notes: This figure shows the proportion of students from bottom 20% families at the public colleges in our data in 1915-1925, 1955-1965 and 1999-2008. 1915-1925 and 1955-1965 data are from our linked college registers and 1998-2009 data are from Opportunity Insights.

Figure B9: Economic Origins of Students at Elite Colleges, Robustness to Alternative Income Scores



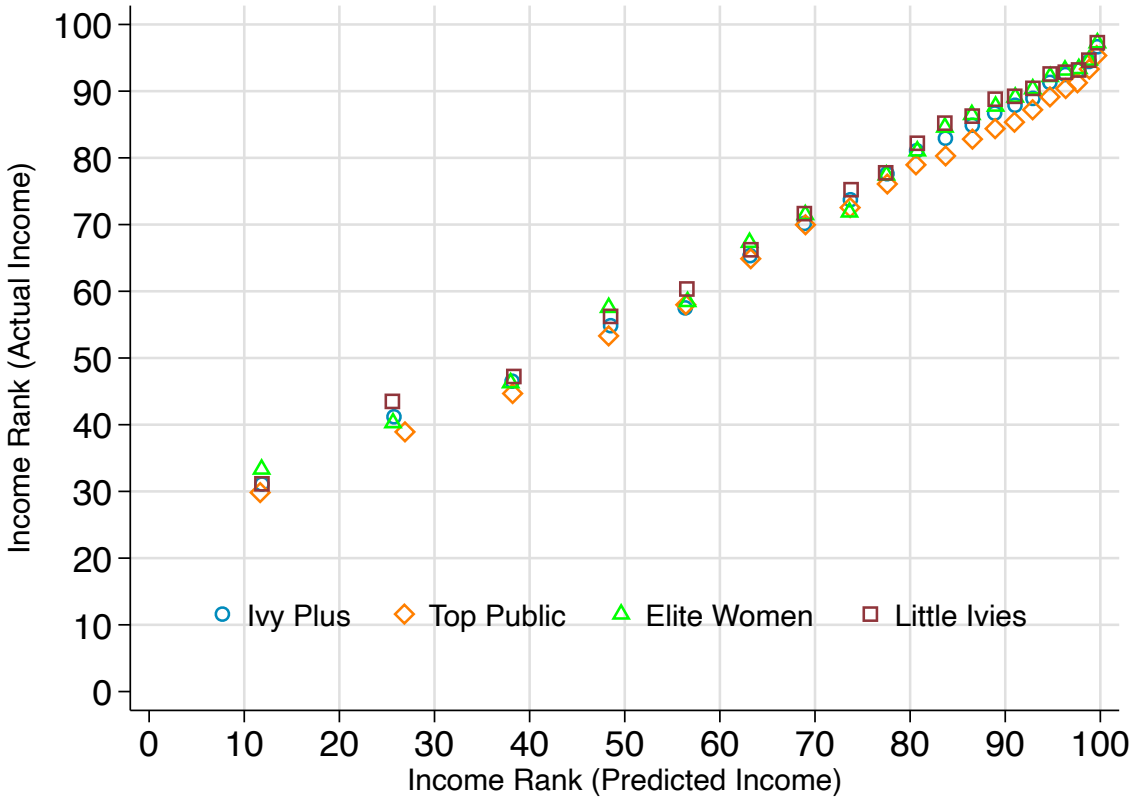
Notes: This figure shows the robustness of our results on the economic origins of students at elite colleges to using alternative methods to impute parental income in the historical census. For students we observe in the 1940 census (starting college up to 1955), we can use actual income rather than predicted income. For students we observe in the 1900 or 1910 censuses, we replace our baseline measure with income scores based on data from the 1900 Census of Agriculture and the 1901 Cost of Living Survey. Results are presented for both private (Panels (a) and (c)) and public (Panels (b) and (d)) colleges, by gender (men in Panels (a)-(b) and women in Panels (c)-(d)).

Figure B10: Economic Origins of Students at Elite Colleges, Robustness to Only Including Students Age 14-18 at the Time of the Census



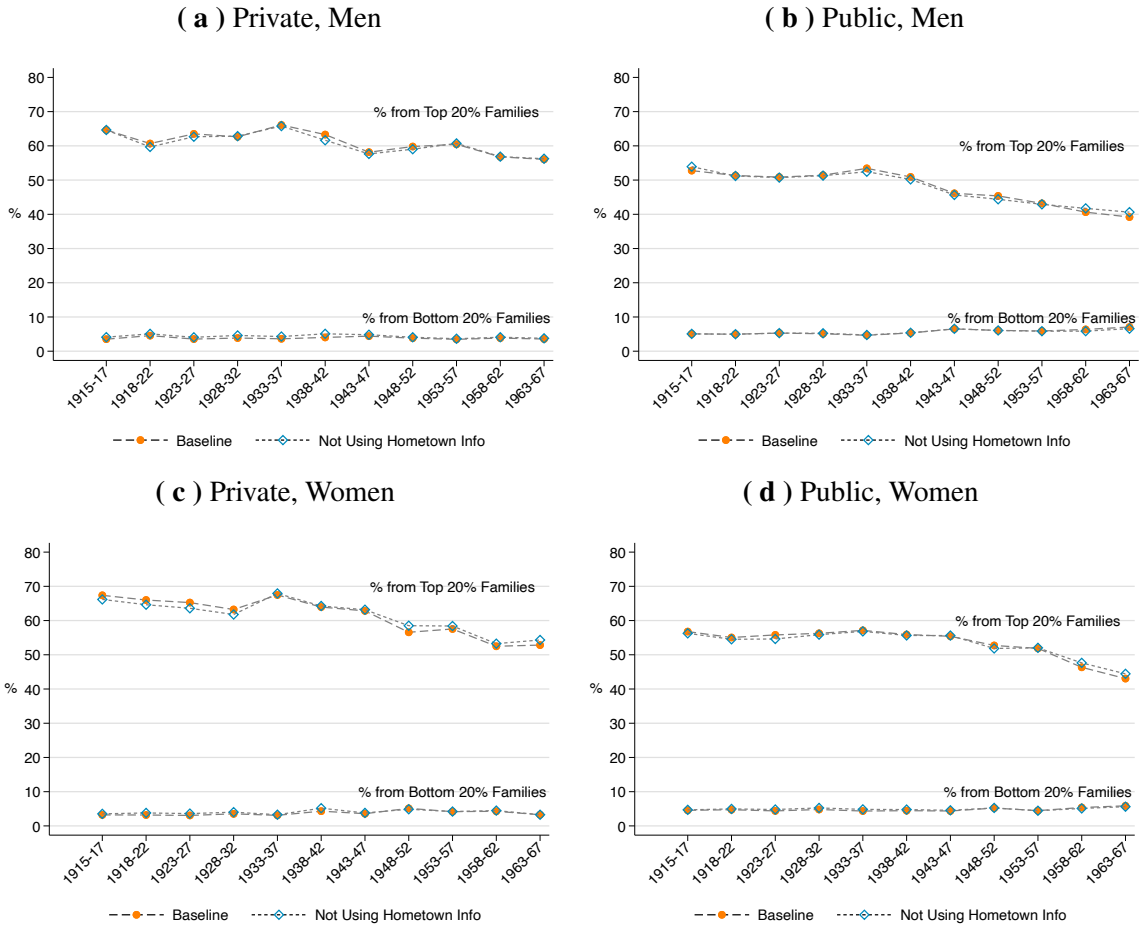
Notes: This figure shows the robustness of our main results on students' economic origins at elite colleges to restricting the sample to students who we observe in the census at ages 14-18 (that is, closer to them enrolling in college). Results are presented for both private (Panels (a) and (c)) and public (Panels (b) and (d)) colleges, by gender (men in Panels (a)-(b) and women in Panels (c)-(d)).

Figure B11: Actual versus Predicted Income, by College Group



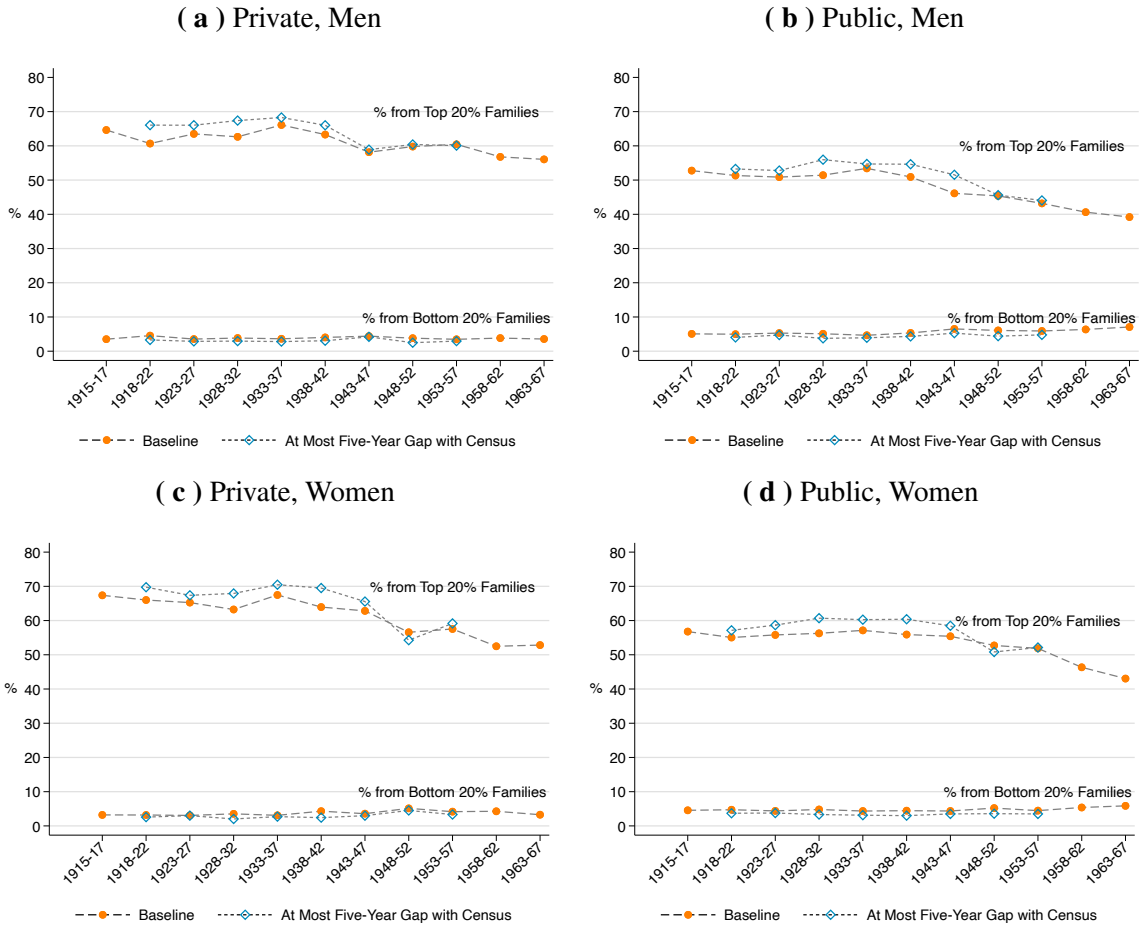
Notes: This figure shows a binned scatter plot of parental income ranks based on our baseline measure of predicted income (x-axis) and parental income ranks based on actual income (y-axis), separately for each group of colleges in our data. The sample is restricted to students for whom we observe parental information in the 1940 census.

Figure B12: Economic Origins of Students at Elite Colleges, Robustness to Not Using Hometown for Matching



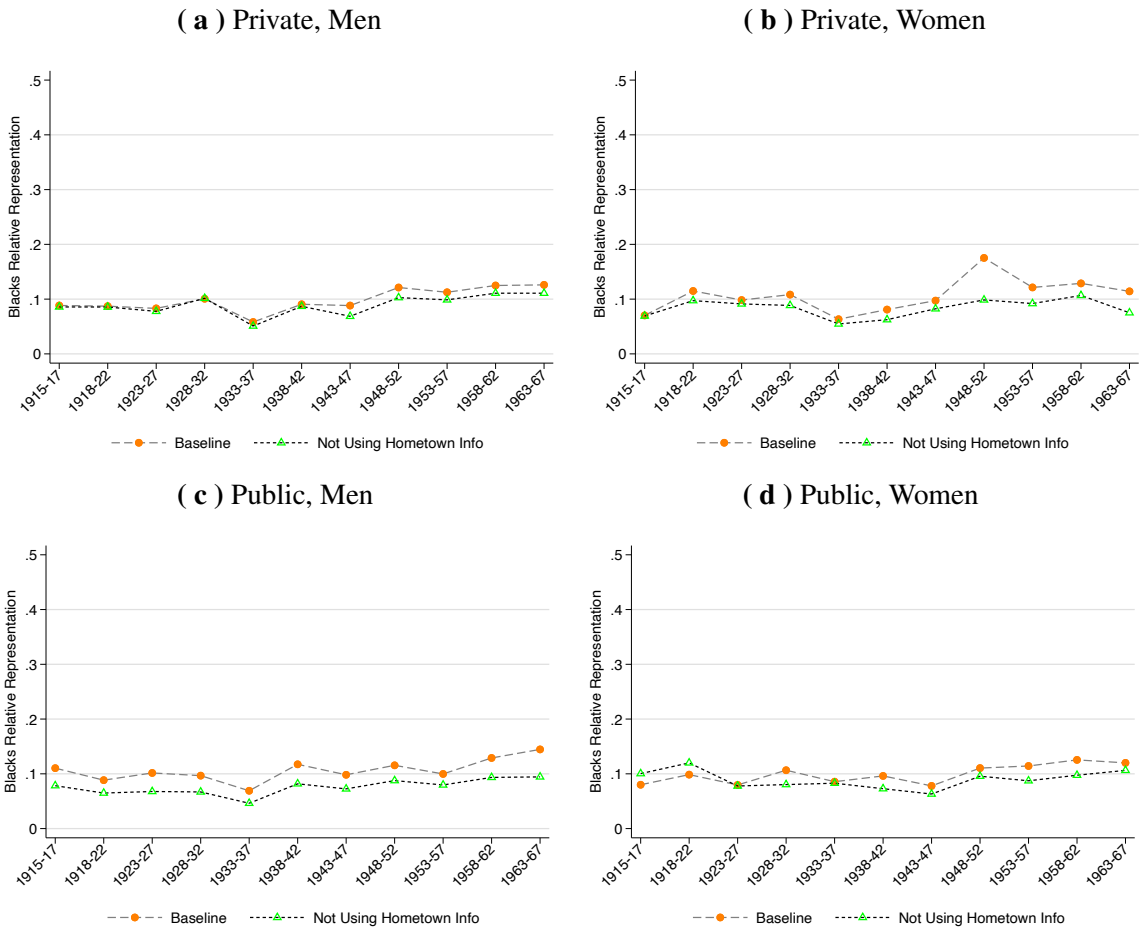
Notes: This figure shows the robustness of our results on students' economic origins at elite colleges to not using the hometown information for disambiguation (so as to use a consistent set of matching variables across the entire sample). Results are presented for both private (Panels (a) and (c)) and public (Panels (b) and (d)) colleges, by gender (men in Panels (a)-(b) and women in Panels (c)-(d)).

Figure B13: Economic Origins of Students at Elite Colleges, Robustness to Only Using Matches With Less than a Five-Year Gap Between Census and Registers



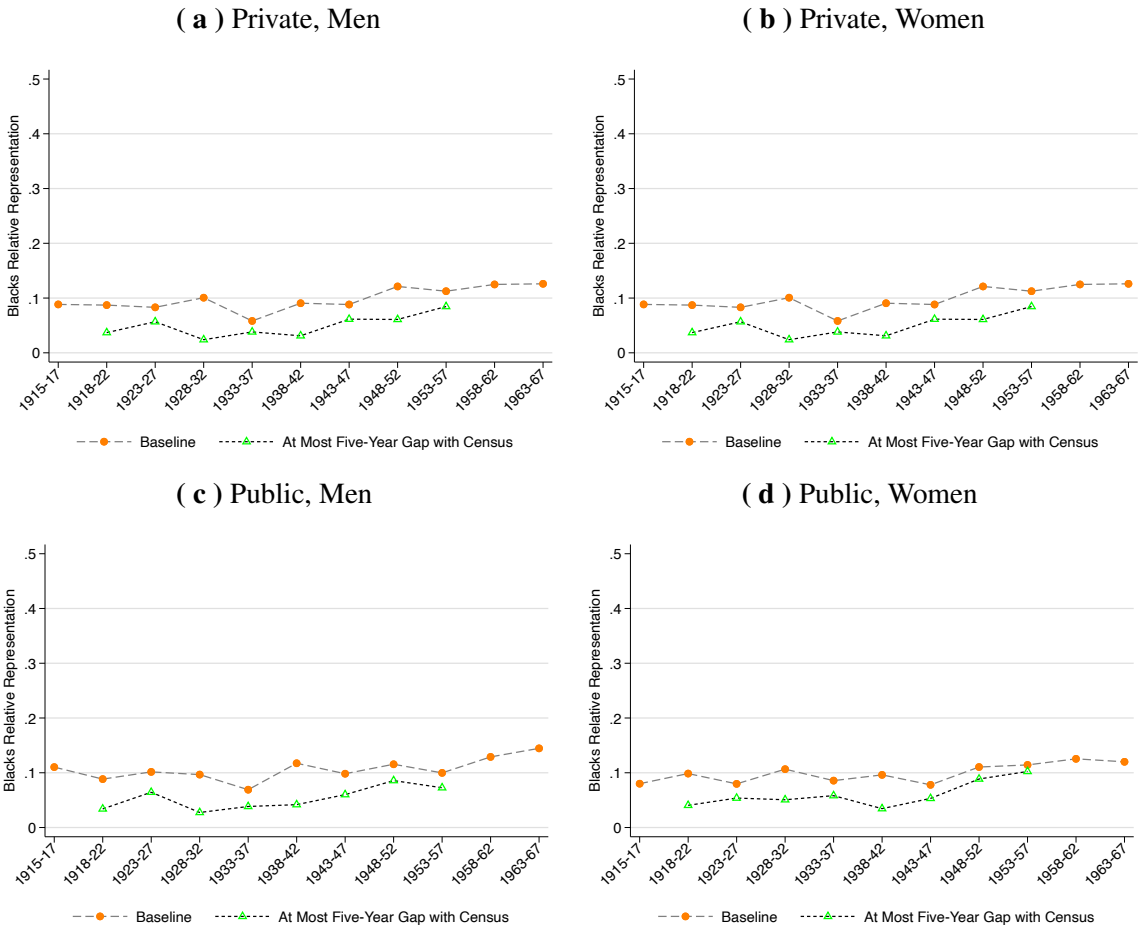
Notes: This figure shows the robustness of our main results on students' economic origins to restricting the sample to students who we observe in the census at most five years prior to observing them in the college registers. Results are presented for both private (Panels (a) and (c)) and public (Panels (b) and (d)) colleges, by gender (men in Panels (a)-(b) and women in Panels (c)-(d)).

Figure B14: Relative Representation of Black Students at Elite Colleges, Robustness to Not Using Hometown Information for Matching



Notes: This figure shows the robustness of our main results on the relative representation of Black students at elite colleges to not using the hometown information for disambiguation (so as to use a consistent set of matching variables across the entire sample). This figure plots the relative representation of Black students, by gender and cohort, at elite private institutions (Panels (a) and (b), and at elite public institutions (Panels (c) and (d)).

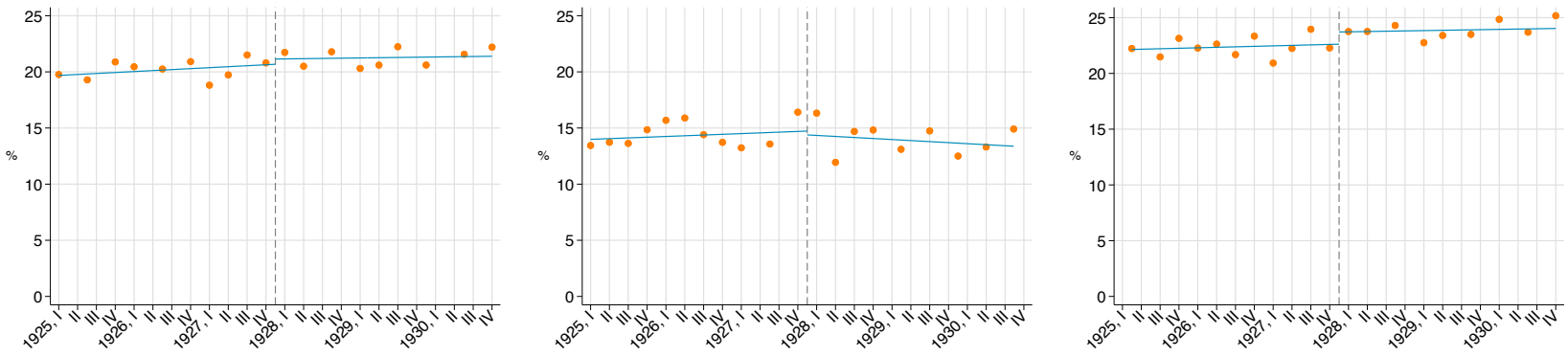
Figure B15: Relative Representation of Black Students at Elite Colleges, Robustness to Only Using Matches With Less than a Five-Year Gap Between Census and Registers



Notes: This figure shows the robustness of our main results on the relative representation of Black students to restricting the sample to students who we observe in the census at most five years prior to observing them in the college registers. This figure plots the relative representation of Black students, by gender and cohort, at elite private institutions (Panels (a) and (b)), and at elite public institutions (Panels (c) and (d)).

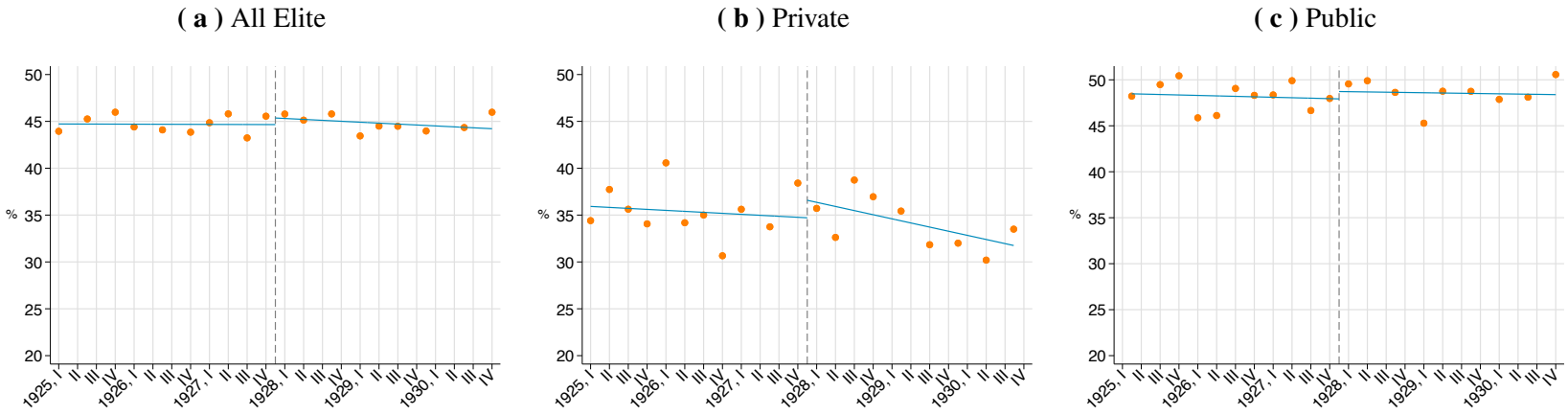
Figure B16: Impacts of the G.I. Bill on Representation of Students from Bottom 50% of Parental Income Distribution

(a) % of Students from Bottom 50%, All Elite **(b)** % of Students from Bottom 50%, Private **(c)** % of Students from Bottom 50%, Public



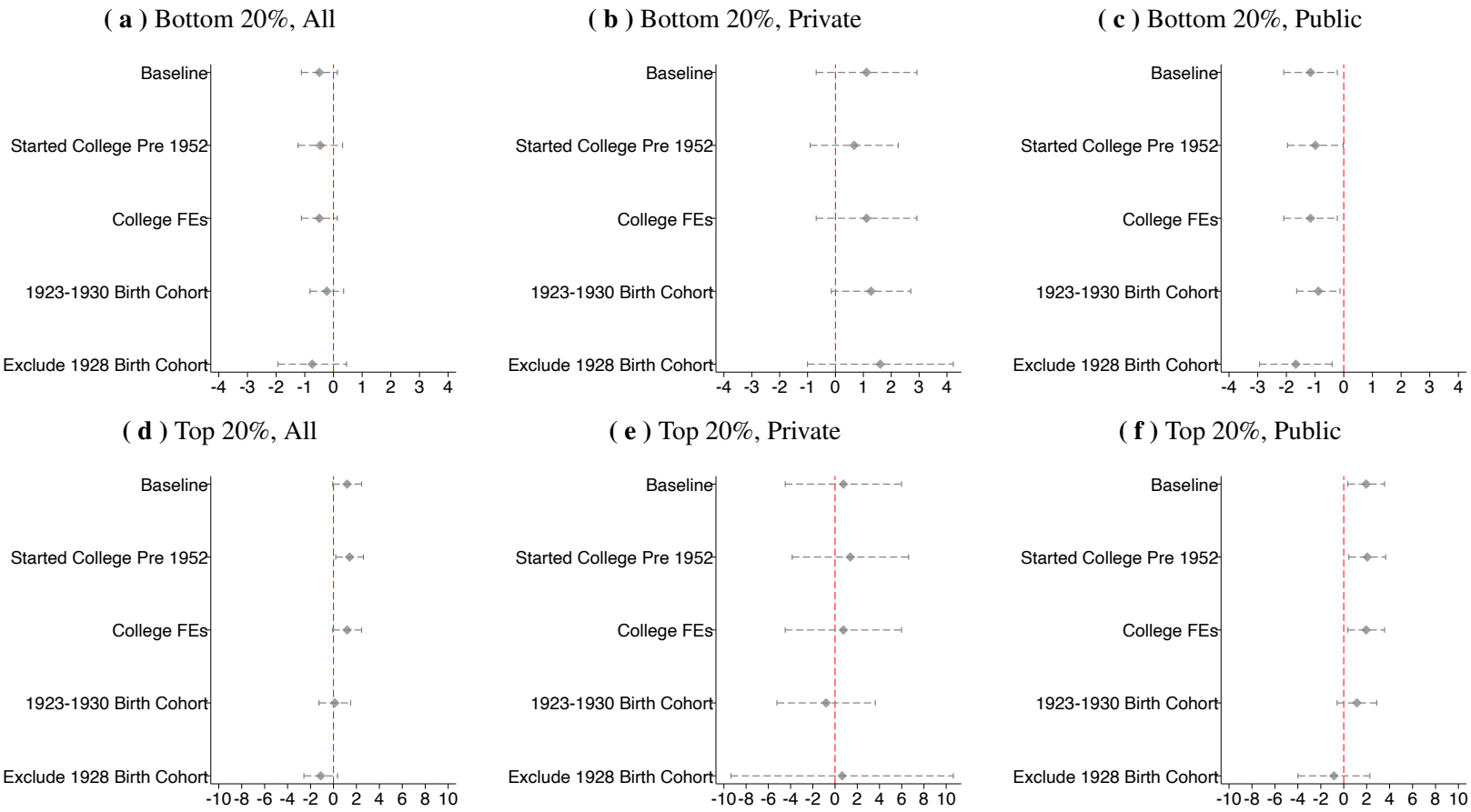
Notes: This figure shows the proportion of students from the bottom 50% of the parental income distribution among men born 1925-1930 in our sample of students at selective colleges, by year and quarter of birth. Results are presented for all elite schools in Panel (a), elite private schools in Panel (b), and elite public schools in Panel (c).

Figure B17: Impacts of the G.I. Bill on Representation of Students from the Middle (Percentiles 20-80) of Parental Income Distribution



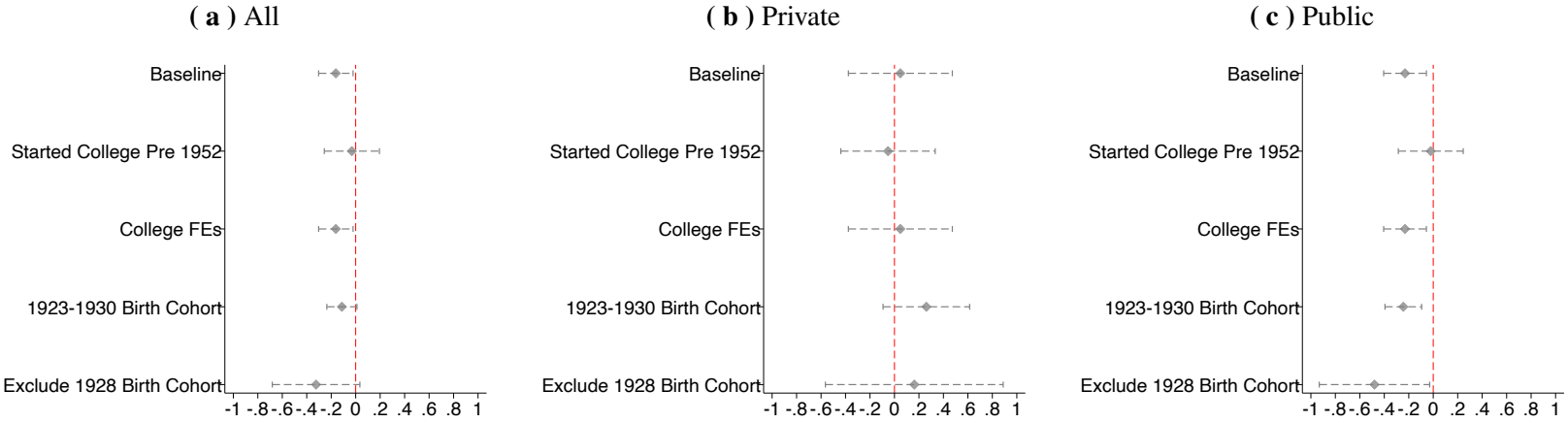
Notes: This figure shows the proportion of students from the middle (percentiles 20-80) of the parental income distribution among men born 1925-1930 in our sample of students at elite colleges, by year and quarter of birth. Results are presented for all elite schools in Panel (a), elite private schools in Panel (b), and elite public schools in Panel (c). All outcome variables are multiplied by 100 so that we can directly interpret the coefficient estimates as percentage point differences.

Figure B18: Impacts of the G.I. Bill on Students' Economic Origins at Elite Colleges, Robustness



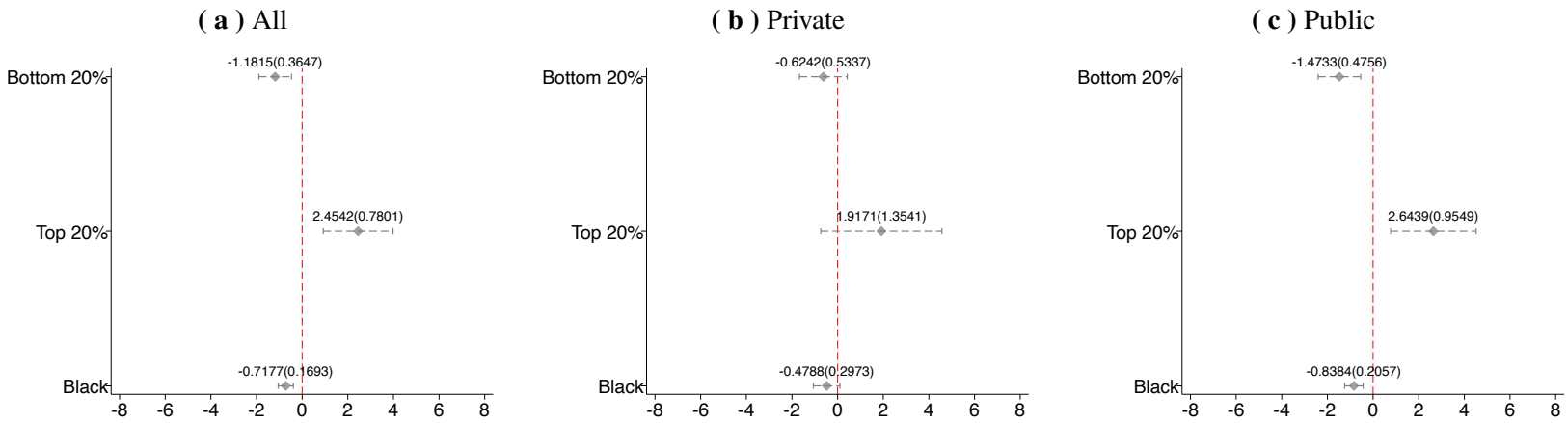
Notes: This figure shows the robustness of the main results presented in Table 1. The "Baseline" row corresponds to our baseline specification. The "Started College Pre 1952" row restricts the sample to students we observe starting college before 1952. The "College FEs" row adds college fixed effects to our main specification. The "1923-1930 Birth Cohort" row expands the set of birth cohorts included in the sample (from 1925-1930 to 1923-1930). The "Exclude 1928 Birth Cohort" row drops from the sample individuals born in 1928. Panels correspond to different characteristics: whether the student grew up in a family in the bottom 20% of the parental income distribution in Panels (a)-(c) and whether they grew in a family in the top 20% in Panels (d)-(f). Panels (a) and (d) includes all elite institutions, whereas Panels (b) and (e) focus on elite private and Panels (c) and (f) focus on elite public.

Figure B19: Impacts of the G.I. Bill on the Proportion of Black Students at Elite Colleges, Robustness



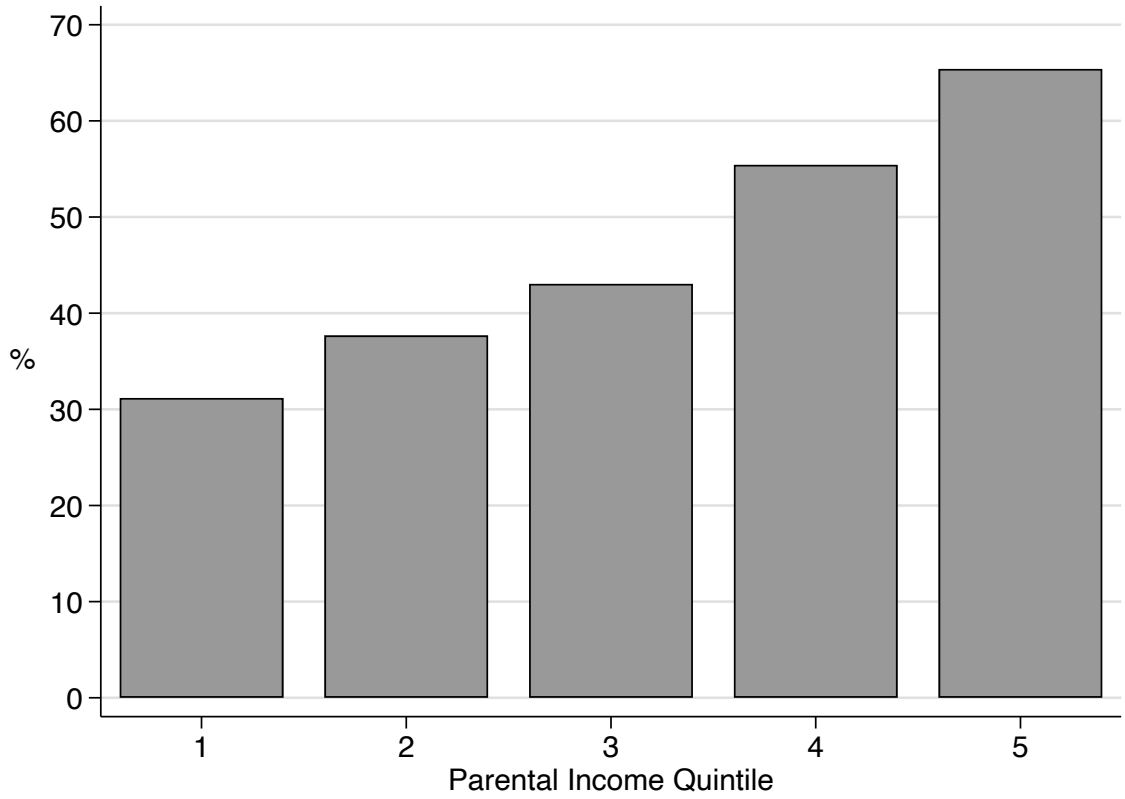
Notes: This figure shows the robustness of the main results presented in Table 1. The "Baseline" row corresponds to our baseline specification. The "Started College Pre 1952" row restricts the sample to students we observe starting college before 1952. The "College FEs" row adds college fixed effects to our main specification. The "1923-1930 Birth Cohort" row expands the set of birth cohorts included in the sample (from 1925-1930 to 1923-1930). The "Exclude 1928 Birth Cohort" row drops from the sample individuals born in 1928. Panel (a) includes all elite institutions, whereas Panels (b) and (c) focus on elite private and elite public, respectively.

Figure B20: Impacts of the G.I. Bill on Students Socioeconomic Backgrounds at Elite Colleges, Within-Cohort Comparison



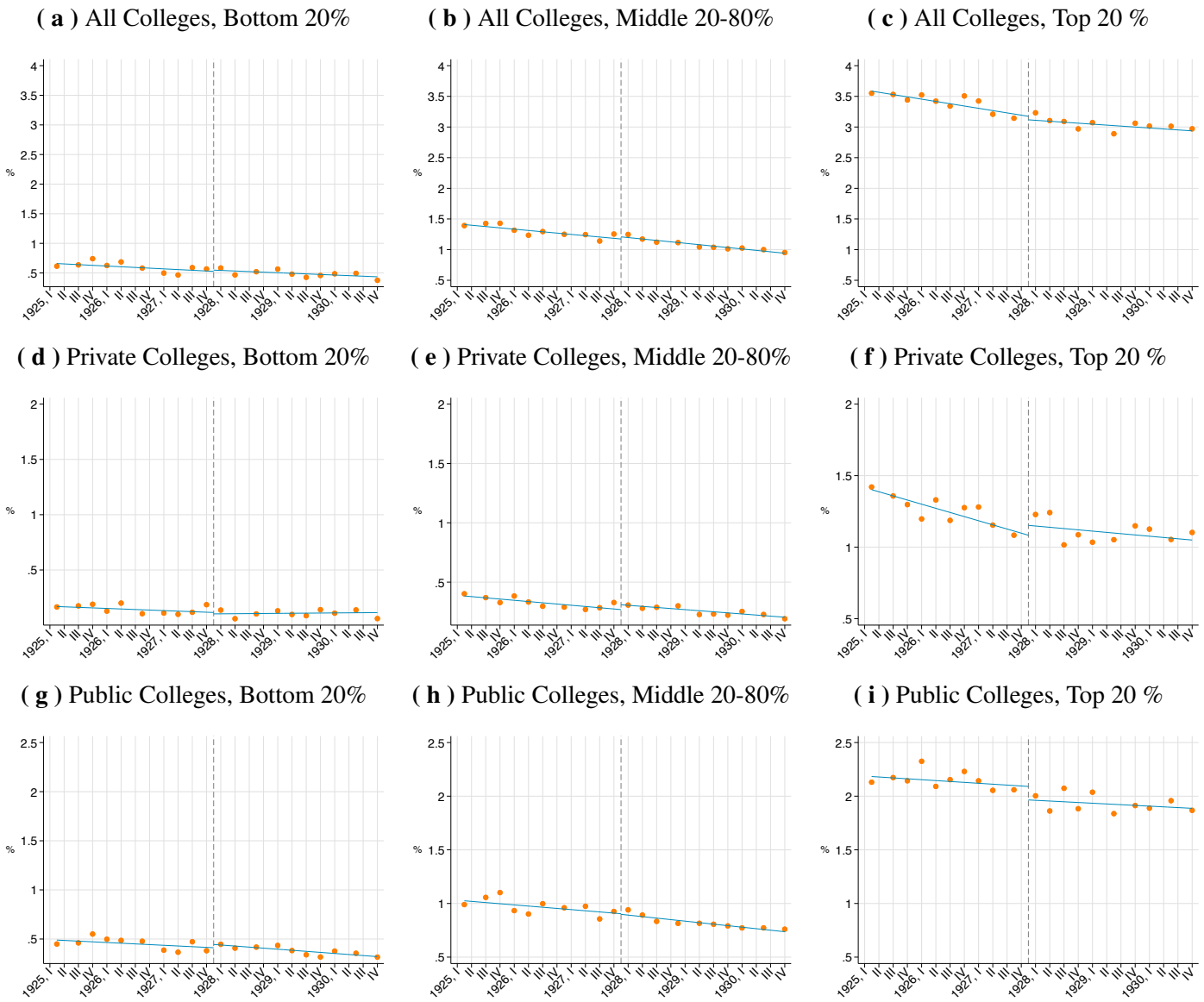
Notes: This figure reports the results of estimating the coefficient associated to a World War II veteran dummy in Equation 2 in the text. Each row corresponds to a different characteristic: whether the student grew up in a family in the bottom 20% of the parental income distribution, whether they grew in a family in the top 20%, or whether they are Black. All outcome variables are multiplied by 100 so that we can directly interpret the coefficient estimates as percentage point differences. Panel (a) includes all male students in our sample of elite colleges, whereas Panels (b) and (c) focus on male students at private or public colleges, respectively.

Figure B21: Share of Men With a High-School Degree, by Parental Income



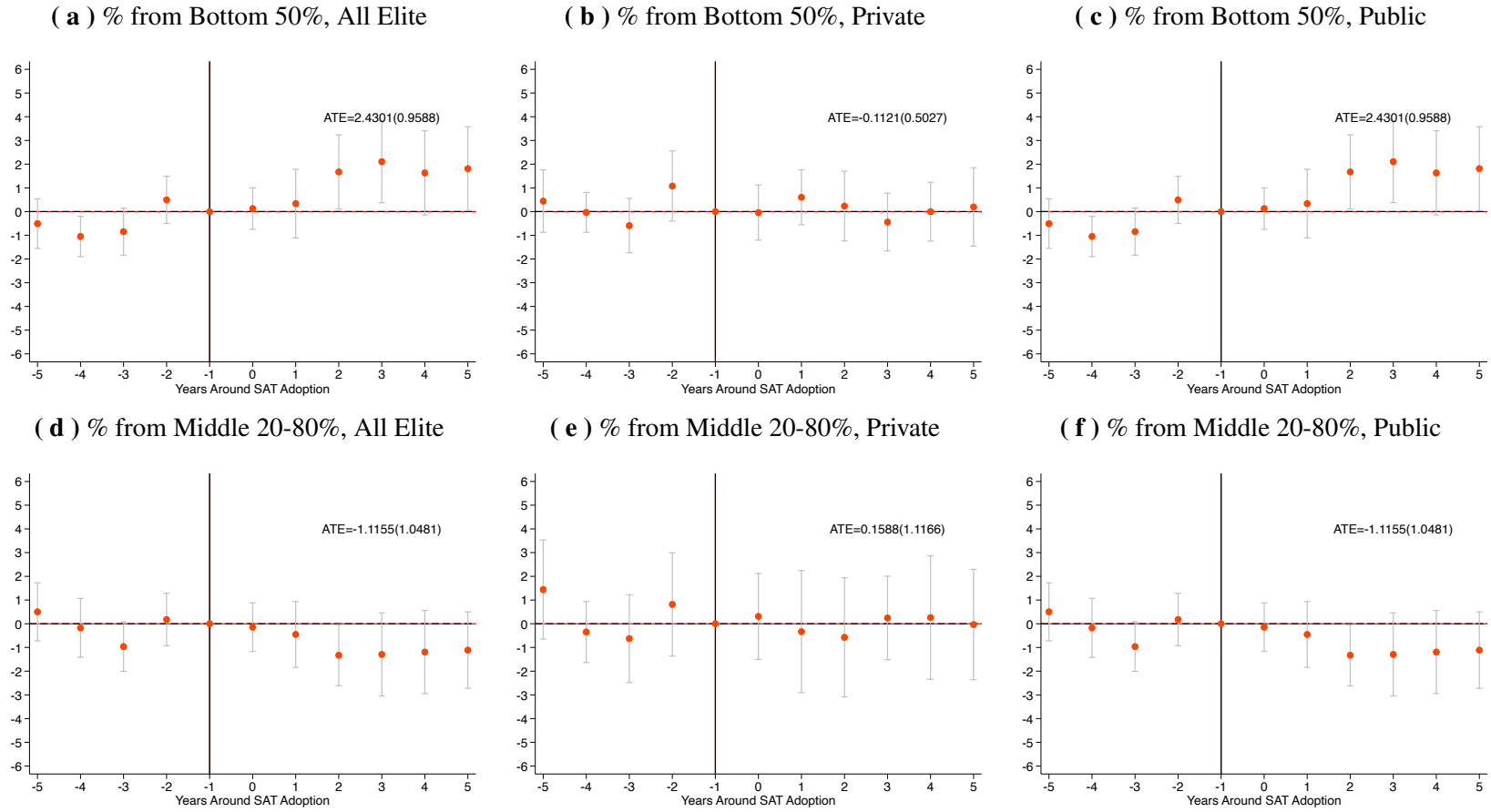
Notes: This figure shows the proportion of individuals with completed high school by parental income, using data linking the 1930 to 1950 censuses (Ruggles et al., 2024). The sample is restricted to U.S.-born men born 1920-1930.

Figure B22: Impacts of the G.I. Bill on the Likelihood of Attending an Elite College



Notes: This figure shows the proportion of individuals who attended an elite college, by year and quarter of birth. The top row includes all elite colleges, whereas the middle and bottom row focus on elite private and elite public colleges, respectively. All outcome variables are multiplied by 100 so that we can directly interpret the coefficient estimates as percentage point differences. Each column corresponds to a different estimation sample: individuals born in bottom 20% families, individuals born in middle 60-80% families, and individuals born in top 20% families. The sample includes all men born 1925-1930

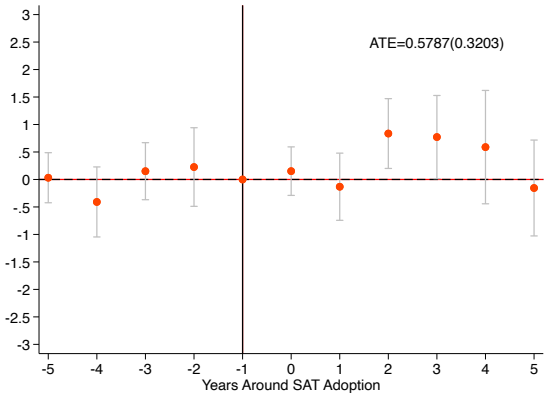
Figure B23: Impacts of Introducing the SAT on Students' Economic Origins at Elite Colleges, Other Segments of the Parental Income Distribution



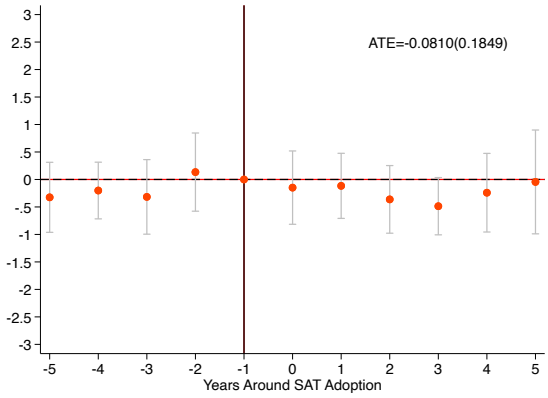
Notes: This figure shows estimated coefficients from event study two-way fixed effects models. Each panel corresponds to a different outcome (whether the student grew up in a family in the bottom 50% of the parental income distribution in Panels (a)-(c) or whether the student grew up in a family in percentiles 20-80 of the parental income distribution in Panels (d)-(f)), and sample of colleges (all elite in Panels (a) and (d), elite private in Panels (b) and (e), and elite public in Panels (c) and (f). Each of the panels also include the coefficient and standard error associated with an indicator variable that is one if the school had adopted an SAT requirement. All outcome variables are multiplied by 100 so that we can directly interpret the coefficient estimates as percentage point differences.

Figure B24: Impacts of Introducing the SAT on Students' Socioeconomic Backgrounds at Elite Colleges, Event Study Regressions Adding Interactions Between School Group and Entry Cohort

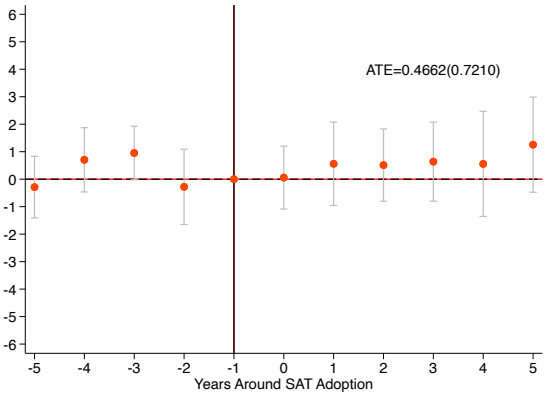
(a) % of Students from Bottom 20%, All Elite



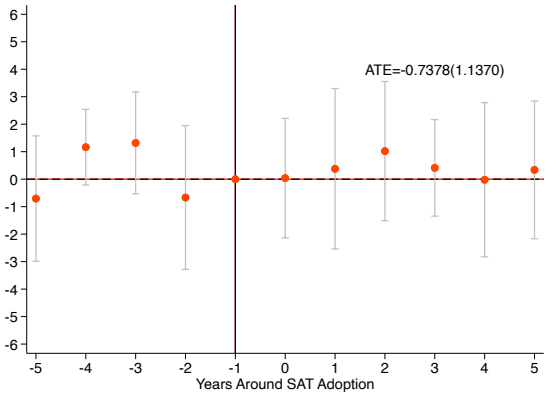
(b) % of Students from Bottom 20%, Private



(c) % of Students from Top 20%, All Elite

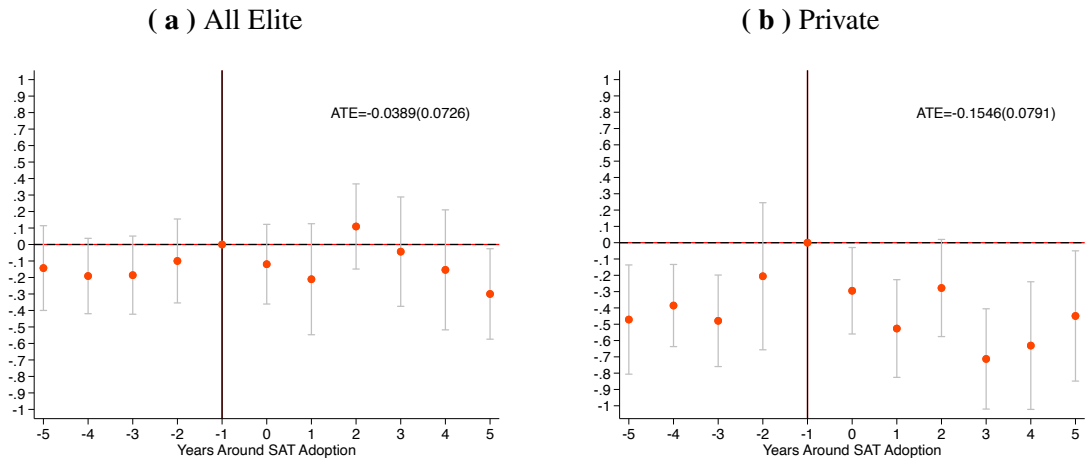


(d) % of Students from Top 20%, Private



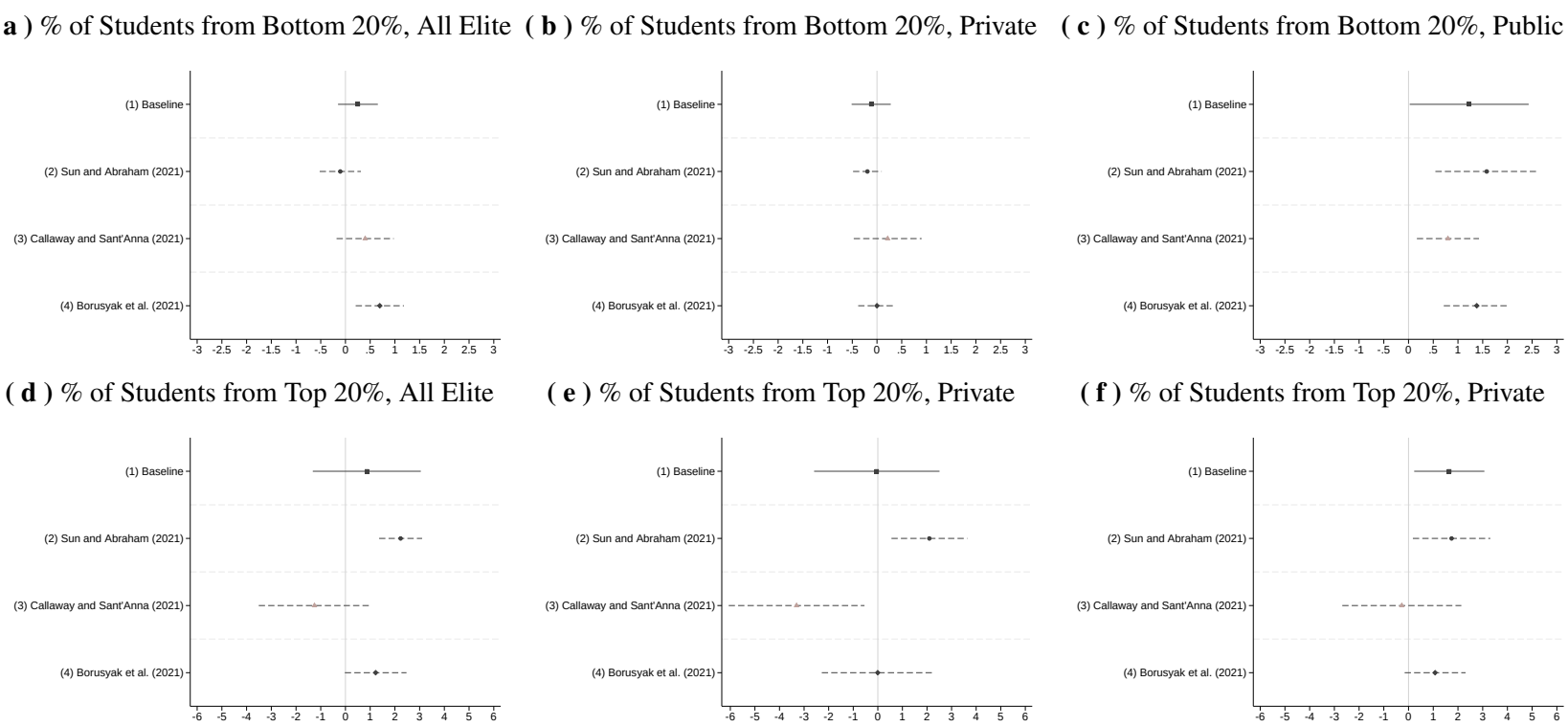
Notes: This figure shows estimated coefficients from event study two-way fixed effects regressions in which we add interaction terms between entry cohort and school group (e.g., Ivy Plus, Little Ivies, Elite Women, Top Public). Note that this model cannot be estimated when restricting the sample to just top public universities. Each panel corresponds to a different outcome (whether the student grew up in a family in the bottom 20% of the parental income distribution in Panels (a)-(b) or whether the student grew up in a family in the top 20% of the parental income distribution in Panels (c)-(d)), and sample of colleges (all private in Panels (a) and (c) and elite private in Panels (b) and (d)). Each of the panels also include the coefficient and standard error associated with an indicator variable that is one if the school had adopted an SAT requirement. All outcome variables are multiplied by 100 so that we can directly interpret the coefficient estimates as percentage point differences.

Figure B25: Impacts of Introducing the SAT on % of Black Students at Elite Colleges, Event Study Regressions Adding Interactions Between School Group and Entry Cohort



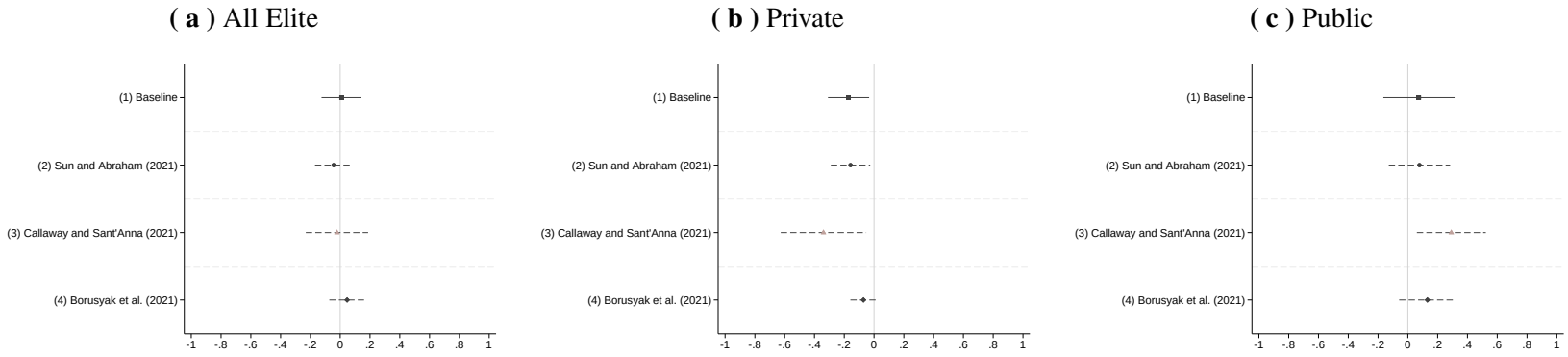
Notes: This figure shows estimated coefficients from event study two-way fixed effects regressions in which we add interaction terms between entry cohort and school group (e.g., Ivy Plus, Little Ivies, Elite Women, Top Public). Note that this model cannot be estimated when restricting the sample to just top public universities. The outcome variable is an indicator that takes on a value of one is the student is Black. Panel (a) shows results for all elite colleges and Panel (b) restricts to elite private. Each of the panels also include the coefficient and standard error associated with an indicator variable that is one if the school had adopted an SAT requirement. All outcome variables are multiplied by 100 so that we can directly interpret the coefficient estimates as percentage point differences.

Figure B26: Impacts of Introducing the SAT on Students' Economic Origins at Elite Colleges, Average Post-Treatment Coefficient



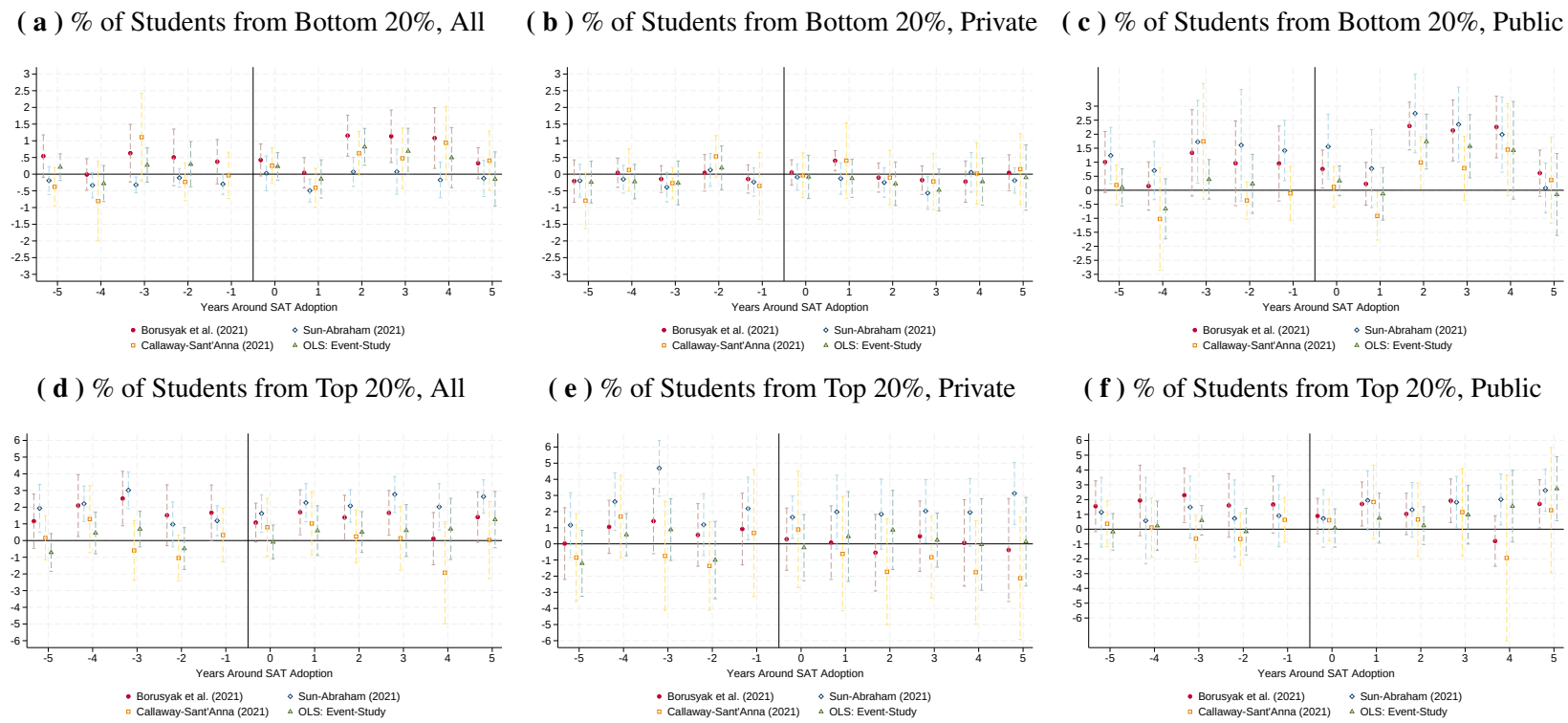
This figure shows the average post-treatment coefficients that we obtain when estimating the impacts of the SAT introduction using alternative estimators. Each panel corresponds to a different outcome (the percent of students who grew up in the bottom 20% of the parental income distribution in Panels (a)-(c) and the percent of students who grew up in the top 20% in Panels (d)-(f)) and sample of colleges (all elite colleges in Panels (a) and (d), only elite private in Panels (b) and (e), and only elite public in Panels (c) and (f)). The baseline row corresponds to the baseline TWFE estimator. The Sun-Abraham row corresponds to the estimator proposed in Sun and Abraham (2021). The Callaway and Sant'Anna (2021) row corresponds to the estimator proposed in Callaway and Sant'Anna (2021) The BJS row corresponds to the estimator proposed in Borusyak, Jaravel and Spiess (2024).

Figure B27: Impacts of Introducing the SAT on % of Black Students at Elite Colleges, Average Post-Treatment Coefficient



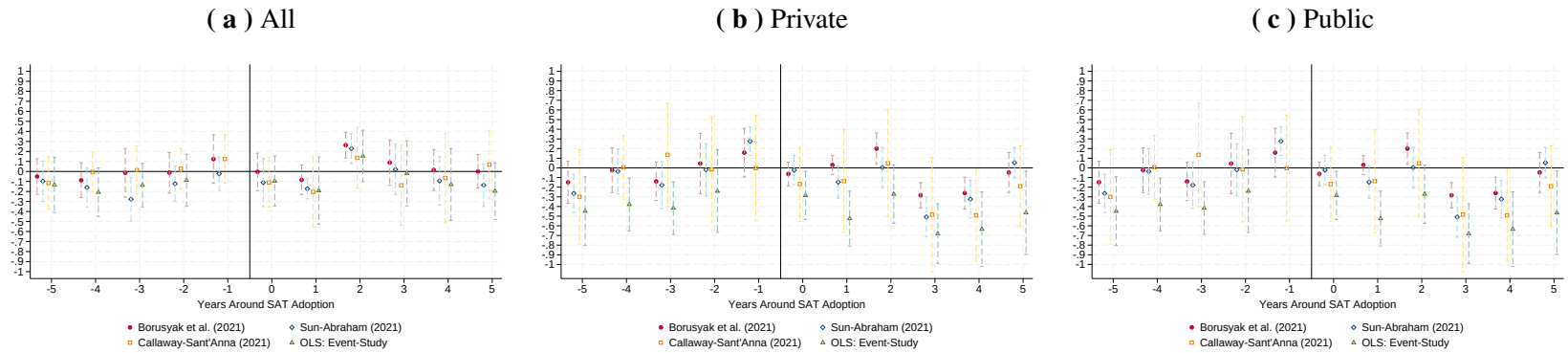
This figure shows the average post-treatment coefficients that we obtain when estimating the impacts of the SAT introduction using alternative event-study estimators. Panel (a) includes all elite colleges, Panel (b) restricts to elite private, and Panel (c) restricts to elite public. The baseline row corresponds to the baseline TWFE estimator. The Sun-Abraham row corresponds to the estimator proposed in Sun and Abraham (2021). The Callaway and Sant'Anna (2021) row corresponds to the estimator proposed in Callaway and Sant'Anna (2021) The BJS row correspond to the estimator proposed in Borusyak, Jaravel and Spiess (2024).

Figure B28: Impacts of Introducing the SAT on Students' Economic Origins at Elite Colleges, Event Study Regressions, Robustness to Alternative Estimators



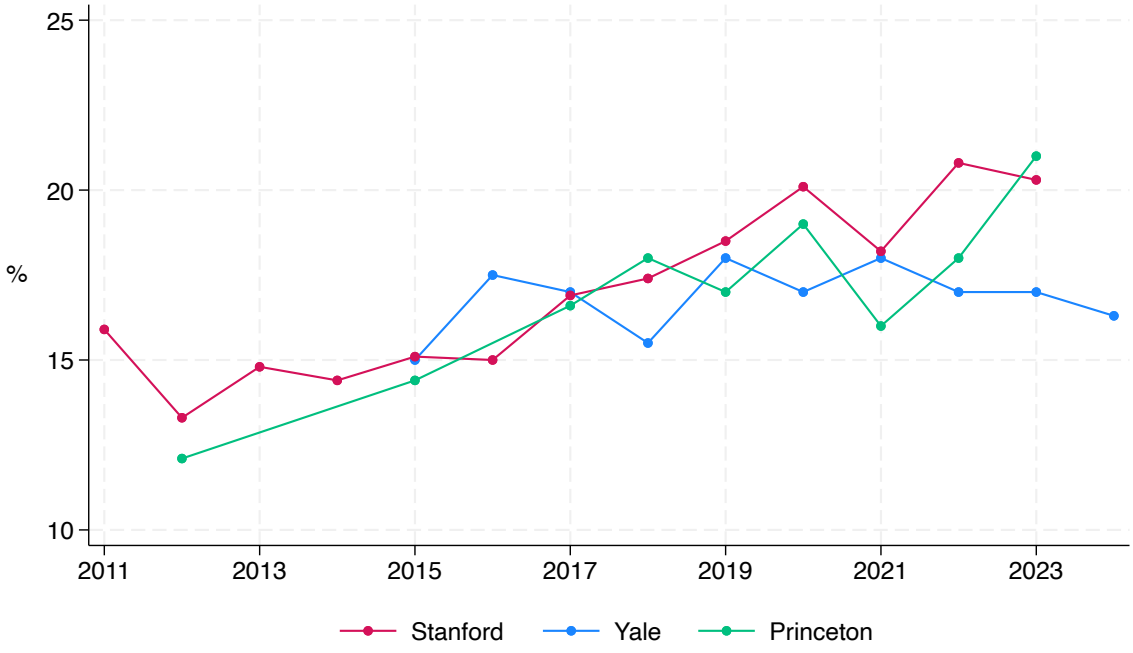
Notes: This figure replicates Figure 9 using alternative event-study estimators. Each panel corresponds to a different outcome (the percent of students who grew up in the bottom 20% of the parental income distribution in Panels (a)-(c) and the percent of students who grew up in the top 20% in Panels (d)-(f)) and sample of colleges (all elite colleges in Panels (a) and (d), only elite private in Panels (b) and (e), and only elite public in Panels (c) and (f)). OLS: Event Study corresponds to the standard two-way fixed effects event-study estimator. Sun and Abraham (2021) corresponds to the estimator proposed in Sun and Abraham (2021). Borusyak, Jaravel and Spiess (2024) corresponds to the estimator proposed in Borusyak, Jaravel and Spiess (2024). Callaway and Sant'Anna (2021) corresponds to the estimator proposed in Callaway and Sant'Anna (2021).

Figure B29: Impacts of Introducing the SAT on % of Black Students at Elite Colleges, Event Study Regressions, Robustness to Alternative Estimators



Notes: This figure replicates Figure 10 using alternative event-study estimators. Panel (a) includes all elite colleges, Panel (b) restricts to elite private, and Panel (c) restricts to elite public. OLS: Event Study corresponds to the standard two-way fixed effects event-study estimator. Sun and Abraham (2021) corresponds to the estimator proposed in Sun and Abraham (2021). Borusyak, Jaravel and Spiess (2024) corresponds to the estimator proposed in Borusyak, Jaravel and Spiess (2024). Callaway and Sant'Anna (2021) corresponds to the estimator proposed in Callaway and Sant'Anna (2021).

Figure B30: Share of First-Generation Students at Selected Universities



Notes: This figure shows the proportion of individuals who are the first in their family to attend college in each entry cohort, using each institution's self-reported data. Data for Stanford comes from the Office of Institutional Research & Decision Support. Data from Princeton comes from the Office of Communications' press releases on the composition of each year's incoming class and data from Yale comes from the First-Year Class Profile fact sheet.

Table B1: The G.I. Bill and the Likelihood of Attending an Elite College

	All Elite			Elite Private			Elite Public		
	(1) Bottom 20%	(2) Middle	(3) Top 20%	(4) Bottom 20%	(5) Middle	(6) Top 20%	(7) Bottom 20%	(8) Middle	(9) Top 20%
Born <= Q4 1927	-0.0264	-0.0548	0.0755	0.0107	-0.0333	-0.0569	-0.0371	-0.0215	0.132**
	(0.0413)	(0.0349)	(0.0586)	(0.0251)	(0.0218)	(0.0634)	(0.0297)	(0.0364)	(0.0527)
Observations	46638	161428	66165	46638	161428	66165	46638	161428	66165
Control Group Mean	0.556	1.216	3.252	0.128	0.305	1.198	0.428	0.911	2.054

Notes: *** p<0.01, ** p<0.05, and * p<0.1. The variable *Born <= Q*41927 is an indicator that takes on a value of one if the individual was born in or before the fourth quarter of 1927. The outcome in columns 1 to 3 is an indicator that takes on a value of one if the individual attended any of the elite colleges in our data. The outcome in columns 4 to 6 is an indicator that takes on a value of one if the individual attended an elite private college. The outcome in columns 7 to 9 is an indicator that takes on a value of one if the individual attended an elite public college. All outcome variables are multiplied by 100 so that we can directly interpret the coefficient estimates as percentage point differences. Each column title indicates the sample that we use to estimate the result (children from bottom 20% families, children from families at percentiles 20-80, children from top 20% families). The sample is restricted to men born between 1925 and 1930. Standard errors are clustered at the birth quarter level.

Table B2: Impacts of Introducing the SAT on Students' Economic Origins at Elite Colleges, Other Segments of the Parental Income Distribution

	All Elite		Elite Private		Elite Public	
	(1) Bottom 50%	(2) Middle 20-80	(3) Bottom 50%	(4) Middle 20-80	(5) Bottom 50%	(6) Middle 20-80
SAT	2.430**	-1.115	-0.112	0.159	2.926**	-2.876***
	(0.959)	(1.048)	(0.503)	(1.117)	(1.331)	(0.814)
College FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1116267	1116267	364985	364985	751282	751282
Mean of Dep. Var.	18.28	41.61	11.86	33.53	21.40	45.54

Notes: $***$ p<0.01, $**$ p<0.05, and $*$ p<0.1. The variable *SAT* is an indicator that takes on a value of one if college *c* had adopted the SAT as a requirement for all applications by year *t*. Columns 1-2 look at all elite colleges, Columns 3-4 restrict to the elite private, Columns 5-6 restrict to elite public. The outcome variable in Columns 1, 3, and 5 is an indicator that takes on a value of one if a student grew up in a family in the bottom 50% of the parental income distribution. The outcome variable in Columns 2, 4, and 6 is an indicator that takes on a value of one if the student grew up in a family in percentiles 20-80 of such distribution. All outcome variables are multiplied by 100 so that we can directly interpret the coefficient estimates as percentage point differences. All regressions include college and year fixed effects. Standard errors are clustered at the college level.

Table B3: Impacts of Introducing the SAT on Students' Socioeconomic Backgrounds at Elite Colleges, Allowing For School-Group-Year Specific Trends

	All			Private		
	(1) Bottom 20%	(2) Top 20%	(3) Black	(4) Bottom 20%	(5) Top 20%	(6) Black
SAT	0.579*	0.466	-0.0389	-0.0810	-0.738	-0.155*
	(0.320)	(0.721)	(0.0726)	(0.185)	(1.137)	(0.0791)
College FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
College Group X Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1116267	1116267	1116267	364985	364985	364985
Mean of Dep. Var.	4.704	53.68	0.953	3.366	63.10	0.894

Notes: $***$ p<0.01, $**$ p<0.05, and $*$ p<0.1. The variable *SAT* is an indicator that takes on a value of one if college *c* had adopted the SAT as a requirement for all applications by year *t*. Columns 1-3 look at all elite colleges and Columns 4-6 restrict to the elite private. The outcome variable in Columns 1 and 4 is an indicator that takes on a value of one if a student grew up in a family in the bottom 20% of the parental income distribution. The outcome variable in Columns 2 and 5 is an indicator that takes on a value of one if the student grew up in a family in the top 20% of such distribution. The outcome variable in Columns 3 and 6 is an indicator that takes on a value of one if the student is Black. All outcome variables are multiplied by 100 so that we can directly interpret the coefficient estimates as percentage point differences. All regressions include college and year fixed effects, as well as college group and year interactions. Standard errors are clustered at the college level.