Do Elite Universities Pick Sports to Pick Students? Athletic Admissions and SES Targeting*

Ahmed El Fatmaoui
University of Wisconsin-La Crosse

Tyler Ransom
University of Oklahoma, IZA & GLO

December 9, 2025

Abstract

This study investigates the extent to which U.S. universities strategically use athletic admissions to shape the socioeconomic status (SES) of their student bodies. Using a novel dataset linking NCAA roster data to neighborhood characteristics, we document substantial SES segregation across sports and universities. More selective institutions, particularly elite private universities, allocate up to 30% of enrollment to athletes who typically come from higher-SES backgrounds than their non-athlete peers. However, contrary to popular belief, we find that elite institutions enroll similarly wealthy athletes across all sports. Estimates of our structural model of sports bundle choice reveal that this SES homogeneity across sports limits universities' ability to systematically choose sports offerings to target higher-SES students. Counterfactual analyses demonstrate that athletic enrollment caps would create additional seats for non-athletes but require complementary policies to meaningfully impact socioeconomic mobility.

JEL Classification: C35, I23, I24, L83

Keywords: higher education; college athletics; income segregation; university admissions

*Corresponding author: Tyler Ransom, Department of Economics, University of Oklahoma, 322 CCD1, 308 Cate Center Dr, Norman, OK 73072. E-mail: ransom@ou.edu.

We thank Michael Dinerstein and seminar participants at the University of Oklahoma and Washington University in St. Louis for helpful comments. The following individuals provided excellent research assistance: Muheb Akbar, Mahamat Nour Bourma, Joseph Bufogle, Taiwei Chang, Stephen Cromwell, Boris Ganchev, Orin Imtiaz, Stephanie Jansson, Katie King, Ryan McAlister, Grant Naberhaus, Carter Pettigrew, Sam Quick, Samantha Siemer, Katy Yut, and Xinyu Zhang.

1 Introduction

Elite colleges have an outsized influence on economic mobility in the United States (Chetty et al., 2020; Chetty, Deming, and Friedman, 2023). However, as we show below, these institutions also enroll a disproportionate number of athletes, reserving up to 30% of seats for varsity athletes compared to 1–3% at public flagship universities. The current system of athletic preferences in admissions consistently favors wealthy students: athletes at every selectivity tier tend to come from higher-socioeconomic (SES) backgrounds than their non-athlete peers. This effectively reduces opportunities for socioeconomic mobility.

While previous literature has identified potential institutional benefits to enrolling athletes—including alumni donations and engagement (Clotfelter, 2003; Meer and Rosen, 2009a,b), increased visibility (Ehrenberg, 2000; Pope and Pope, 2009, 2014) and social network benefits (Rivera, 2016; Amornsiripanitch et al., 2023)—a critical question remains unexplored: do universities strategically offer specific sports as a means of enrolling higher-SES students? Elite universities disproportionately offer expensive niche sports like squash, fencing, and sailing that draw from wealthy populations, are rarely offered at the high school level, and have a scant following on campus. This pattern makes the question of SES-based targeting particularly important given that universities retain substantial discretion over which sports to offer and that higher-SES backgrounds predict greater future donor capacity.

In this paper, we answer this question using a structural model of universities' sport bundle choices. We analyze roster data from 1,013 NCAA institutions covering over 385,000 athletes. We measure athlete SES through neighborhood characteristics linked via publicly available rosters. Our estimates reveal that universities do *not* strategically select sports based on athlete SES. Instead, facility complementarities and institutional factors seem to drive these decisions. Our main finding stems from the fact that income segregation within sports largely mirrors segregation across universities: the richest players of any sport tend to attend the richest universities. This then limits universities' ability to use sports as a means of SES targeting.

Our data set combines several sources and focuses on all NCAA institutions for academic year 2019–20.[1] We collect roster data from the athletic websites and measure the SES of athletes using the characteristics of their high school and its ZIP code.[2] We also collect information about universities and sports offered from the Integrated Postsecondary Education Data System (IPEDS), the Equity in Athletics Disclosure Act (EADA), and Chetty et al. (2020).

We use our data to uncover several novel descriptive facts. First, the typical elite university enrolls *more athletes* than the typical public flagship. For example, the top 32 private universities educate 1.8% of all students, but enroll nearly 6% of all athletes, while the 53 public flagship universities educate nearly 20% of all students but enroll just 8% of all athletes. Second, elite universities achieve greater athletic enrollment by offering many more niche sports rather than expanding commonly-offered sports.[3] Third, athletes at every selectivity tier come from higher-income backgrounds than their non-athlete peers. Fourth, SES segregation patterns are similar within both lower-SES sports (football, basketball) and higher-SES sports (fencing, lacrosse). This implies substantial homophily in athlete and university pairings.

Our study contributes to the following literatures: admissions preferences for athletes (Bowen and Levin, 2003; Espenshade, Chung, and Walling, 2004; Arcidiacono, Kinsler, and Ransom, 2022, 2024); the privileged backgrounds of NCAA athletes (Thompson, 2019; Garthwaite et al., 2025; Lewis, 2020; Hextrum, 2021; Gladwell, 2024); the human capital content of sports (Ransom and Ransom, 2018; Heckman, Loughlin, and Tian, 2025); and the returns to elite university attendance (Zimmerman, 2019; Chetty et al., 2020; Chetty, Deming, and Friedman, 2023; Barrios-Fernández, Neilson, and Zimmerman, 2024). However, these literatures have taken athletic admissions as given and have not considered whether universities strategically choose their athletic offerings.

We address this gap by endogenizing athletic enrollment, focusing on the intensive margin of athletic decision-making (i.e., which specific sports to offer) rather than the extensive

[1] We focus on this time frame because it precedes the COVID-19 pandemic, substantial conference realignment in Division I, and recent changes to name, image and likeness (NIL) and other compensation regulations.

[2] See Garthwaite et al. (2025) for a similar dataset on a subset of NCAA member institutions.

[3] Niche sports include those rarely offered at the high school level and typically requiring specialized facilities or training, such as fencing, squash, sailing, water polo, rowing, and skiing.

margin of whether to offer athletics at all. We are the first to systematically measure how athlete SES backgrounds vary across sports and institutional selectivity, and to estimate universities' preferences for different sports bundles.[4]

We develop and estimate a characteristic-based utility model (Lancaster, 1971) where universities choose sport bundles of up to 41 different sports to maximize utility based on bundle characteristics like athlete numbers and SES, bundle composition including facility complementarities that capture economies of scope, and university-specific preferences that vary across NCAA divisions. Our model accounts for how universities trade off different bundle attributes such as athlete SES, financial profitability, and economies of scope in facility use.

Our structural estimation approach connects with several other papers in industrial organization (IO) and in applying IO methods to the US market for higher education. Our bundled choice model incorporates methods from Manski and Sherman (1980); Train, McFadden, and Ben-Akiva (1987) and Gentzkow (2007). While much of the recent literature in this area focuses on equilibrium models of strategic competition among universities (Epple, Romano, and Sieg, 2006; Fu, 2014; Epple et al., 2017; Fillmore, 2023; Cook, 2025), we take a single-agent approach that allows us to focus on the institutional decision-making process for sports offerings without solving for market-wide equilibrium responses. While this may lead to some bias in our estimates, it allows for more transparent identification and provides a useful benchmark for future equilibrium analyses.

We identify the model's parameters through variation in bundle choices across universities within NCAA divisions (e.g., Division I, II, III) and estimate separate conditional logit models for each NCAA division using choice-based sampling (McFadden, 1978; Davis et al., 2019). This is because the choice set has a prohibitively large $2^{41} \approx 2.2$ trillion bundle alternatives. Our ability to include unobservable preference heterogeneity is limited by the extreme degree of serial correlation in choices over time. Thus, we allow for preference

[4]We focus on the intensive margin for two reasons. First, there is substantially more variation in *which* sports universities offer than in *whether* they offer sports at all—approximately 1,300 of the 2,267 four-year institutions in the United States are members of the NCAA or NAIA (National Center for Education Statistics, 2023; National Collegiate Athletic Association, 2024; National Association of Intercollegiate Athletics, 2024). Second, data limitations prevent us from modeling the extensive margin: credibly simulating a no-athletics counterfactual would require comprehensive data on admissions, enrollment, and post-graduation outcomes (including donation behavior) for both athletes and non-athletes.

heterogeneity in the form of institution type fixed effects and student income characteristics interacted with bundle attributes. This allows preferences to vary systematically across different types of institutions within each division.

Our model estimates imply that elite universities do not place much weight on athlete SES when choosing which sports to bundle together. We find some evidence that Division II universities prefer bundles with more athletes in the top 5% of income, but fewer in the top 1% of income. Facility complementarities and institutional factors drive sport offering decisions much more than athlete SES. This is likely driven by the high degree of income homophily between athletes and universities—the richest universities can find rich athletes even among lower-SES sports, so this limits the importance of athlete SES in the decision process.

Findings in the prior literature regarding the value of a seat at an elite university (Zimmerman, 2019; Chetty, Deming, and Friedman, 2023) motivate us to consider how universities would respond to a cap in athlete enrollment. While no one has publicly advocated for such a policy in the United States, it is worth noting that nearly all other countries do not tie athletics to higher education—effectively implementing a cap of zero. Economic logic suggests that if elite university seats generate substantial private and social returns, then enrolling students with lower academic credentials—including recruited athletes—may fall short of social optimality. A cap on athlete enrollment is tightly connected to current policy proposals to ban legacy and donor admissions preferences because both are unrelated to an applicant's academic preparation. The latter have been passed by several states and even apply to private institutions in California (Newsom, 2024). Many of these policy proposals have gained momentum in the wake of the *Students For Fair Admissions v. Harvard* ruling (Students for Fair Admissions, 2023), which ended admissions preferences based on race or ethnicity. This has created an environment that is questioning of any admissions preferences unrelated to academic achievement (Arcidiacono, Kinsler, and Ransom, 2023).

We validate our model by showing that it fits the data well in several dimensions of both targeted and untargeted moments. We then use the model to explore the effects of a hypothetical policy that caps athlete enrollment at 5% of all students at each university in the top three tiers of selectivity (124 universities in total). We choose 5% because it is more in-line with athletic enrollment levels at public flagship universities. Consistent with our model's estimates, we show that the SES of elite university athletes under the restricted allocations would not be very different from the levels observed in the data. We also show that there is a wide variety of responses in terms of which sports would be cut if athlete enrollment were forced downward to such a degree. Elite university responses to the COVID-19 shock are consistent with our model's predictions.

Finally, we compare the effects of two different ways that universities could choose to satisfy the 5% enrollment cap. If elite universities were to keep their current enrollment levels and reduce their athletic enrollments, this would result in an extra 6–10% of seats for non-athletes. Alternatively, elite universities could keep their current athletic enrollment levels and increase their non-athletic enrollments. We show that this would require a massive amount of expansion, on the order of 2.7 times the current enrollment.

Ultimately, if elite universities were to have their athletic enrollments capped, the impact on socioeconomic mobility would depend on how they fill the newly vacated seats. As Chetty, Deming, and Friedman (2023) show, athletes at elite universities do not do as well as their academically outstanding peers in terms of extreme post-college success (earning in the top 1% of income or working at a prestigious firm). This implies that enrollment caps on athletics would need to be paired with other admissions policies in order to achieve the intended results: otherwise, vacated athlete seats may go to legacies or other well-off applicants, which would negligibly impact economic mobility.

2 Data

We combine data from several different sources to create a dataset that combines university characteristics and the characteristics of the sports that each university offers, including the socioeconomic status of its athletic rosters. Our final dataset consists of a cross-section of 1,013 universities with complete data as of the 2019–20 academic year.[5] We choose this time period specifically because it predates the COVID-19 pandemic, substantial conference

[5] Online Appendix A contains full details on our sample selection and data cleaning process. Online Appendix Table A.2 contains complete details about each of our data sources, as well as which variables come from which sources.

realignment in Division I, and changes to name, image, and likeness (NIL) rules that have significantly altered the landscape of college athletics (National Collegiate Athletic Association, 2021).

2.1 University-level data

Our university-level dataset includes information on all NCAA Division I, II, and III institutions as of the 2019–2020 academic year. We collect general institutional characteristics from the Integrated Postsecondary Educational Data System (IPEDS), accessed through the R package `rscorecard` (Skinner, 2023). These data include undergraduate enrollment, student demographics, admissions rates, graduation rates, and financial information. We also utilize data from Opportunity Insights (Chetty et al., 2020), which includes institution-level statistics on the parental income distribution and intergenerational mobility outcomes for undergraduate students.

We also collect data on the number of student-athletes at each institution from College-Factual.com. Data from the NCAA includes a list of all sports officially sponsored by each NCAA institution in the 2019–20 academic year. We supplement this with data on athletic championships from the NCAA Championships Summary. Furthermore, we collect data on the start and end dates of each sponsored sport at each university.

To understand the financial dimensions of college sports, we incorporate data from the Equity in Athletics Data Analysis (EADA) database, which contains detailed information on athletic revenues and expenses by institution and sport.

2.2 Athlete-level data

The core of our athlete-level dataset comes from athletic rosters published on university websites for the 2019–2020 academic year. We collect this information for all available NCAA institutions. From these rosters, we extract each athlete's high school, which serves as our primary source regarding their socioeconomic background.

We link each high school to demographic and socioeconomic characteristics using data from the Common Core of Data (CCD) for public schools and the Private School Survey

(PSS) for private institutions, both administered by the National Center for Education Statistics (NCES).[6] These datasets provide information on school enrollment, racial composition, student-to-faculty ratios, and whether the school is public or private.

To develop more detailed measures of socioeconomic status, we match each high school's ZIP code to tax data from the Internal Revenue Service (IRS) and demographic information from the American Community Survey (ACS).[7] The ZIP code-level variables include average wage and salary income, educational attainment, and income distribution metrics.

In total, our athlete-level dataset contains information on 385,268 student-athletes enrolled at 1,013 NCAA-member universities during the 2019–2020 academic year. For each athlete, we observe the university attended, their sport, the high school they graduated from, and the characteristics of their high school and its surrounding community (as defined by ZIP code). We only observe neighborhood and high school characteristics for domestic athletes.

One relevant institutional feature that we do not observe is athletic scholarship status. Athletic scholarship policies vary substantially across NCAA divisions and sports and may influence recruitment patterns and athlete SES composition. Division I and II institutions generally offer athletic scholarships, while Division III institutions and the Division-I Ivy League prohibit them entirely. We do not observe individual scholarship status in our data. These institutional differences provide useful context for interpreting the descriptive patterns below, and we partially account for them by estimating our model separately for each NCAA division.

3 Descriptive analysis

Our descriptive analysis focuses on four key questions. First, how segregated is athletic enrollment by college selectivity tier? Second, how different are sport offerings across tiers? Third, how does the SES of athletes compare to that of non-athletes, and how does this vary

[6] We access the CCD data via the `educationdata` R package (Tyagi, Ueyama, and The Urban Institute, 2022). The PSS data come directly from the NCES website.

[7] We access the ACS data using the `tidycensus` R package (Walker, Herman, and Eberwein, 2023). The IRS data come directly from the IRS website.

by tier? Fourth, how does athlete SES vary by sport and selectivity tier?

We divide institutions into the following seven familiar selectivity tiers: Ivy Plus (Chetty et al., 2020; Chetty, Deming, and Friedman, 2023); Elite Liberal Arts College (LAC); Other Elite Private; Public Flagship; Mid-tier Public; Mid-tier Private; and Other Non-selective.[8]

For ease of exposition, we also divide sports into four groups: football, standard, regional, and niche. Standard sports include sports commonly offered at all public high schools (e.g., basketball, baseball/softball, track & field, tennis, swimming, golf). Regional sports are offered at many public high schools in certain regions (e.g., lacrosse, field hockey, beach volleyball, ice hockey). Niche sports include sports that are rarely offered at the high school level (e.g., gymnastics, water polo, fencing, squash, skiing, sailing). For a complete list of sport groups, see Online Appendix A.2.1.

3.1 Elite universities disproportionately enroll athletes

To answer our first question, Table 1 summarizes athletic enrollment, selectivity, and total enrollment by each of the seven tiers among the 1,013 NCAA-member universities in our sample. Column 2 shows that, while 5.8% of all students at these universities are NCAA athletes, there is dramatic variation across the tiers. While just 2.4% of public flagship students are athletes, this rate is over 10% at private universities, over 13% among Ivy Plus universities, and over 30% at Elite LACs.

The final three columns show that the universities with the lowest admit rates have the highest rates of athletic enrollment. The Ivy Plus and Elite LAC institutions educate almost 6% of all athletes despite enrolling less than 2% of all students. In contrast, public flagships educate 8% of all athletes but more than 10 times as many students.

[8] Ivy Plus includes the eight Ivy League universities plus Stanford, MIT, Duke, and University of Chicago (Chetty et al., 2020). Elite LACs include highly selective liberal arts colleges: "Little Ivies" such as Williams, Amherst, Middlebury, Bowdoin, Colby, Bates, Connecticut College, and Wesleyan plus other top LACs (e.g. Swarthmore, Haverford, Hamilton, Colgate, Davidson, Kenyon, Oberlin, Vassar, Carleton, Macalester, Pomona-Pitzer, Claremont McKenna-Harvey Mudd-Scripps). Other Elite Private includes highly selective private universities such as Georgetown, Northwestern, Washington University in St. Louis, Rice, Vanderbilt, Carnegie Mellon, Notre Dame, and others. Public Flagships include the 53 highest-ranked public flagship universities. Mid-tier Private and Mid-tier Public respectively represent less-selective private and public four-year institutions. Other Non-selective includes remaining institutions with low admission standards.

3.2 Elite universities disproportionately offer regional and niche sports

While Table 1 shows that elite universities enroll more athletes, it is important to know the composition of the sports offered. In Figure 1, we plot the average number of different types of sports offered by universities within each tier, using the broad groupings described above.

Consistent with Table 1, the figure shows that typical Ivy Plus and Elite LAC universities offer many more sports than universities in the lower tiers. However, the additional sports are concentrated in the regional and niche groups—the typical Ivy Plus university only offers about one additional standard sport than the typical public flagship university does. Rather, the average Ivy Plus university offers six regional and six niche sports, while the average public flagship offers two regional and one niche.

3.3 Athletes come from more-advantaged backgrounds than non-athletes in every tier

To answer our third question, we use our athlete-level data to compare athletes' SES to their non-athlete peers. Our key assumption is that an athlete's high school ZIP code is a reasonable proxy of their household SES.[9] Our income comparisons to non-athletes use data from Chetty et al. (2020) which is based on individual tax returns (IRS Form 1098-T) of the parents of students who apply for federal financial aid. While these measures may seem to be incompatible, we reconcile them by looking at quantiles of the underlying income distribution instead of levels of income. For other measures of SES, we do not have data on non-athletes so we present statistics for athletes only.

[9] See Garthwaite et al. (2025) for a similar approach. The primary threat to this assumption is athlete recruitment to elite sports academies such as IMG Academy, Montverde Academy, and Oak Hill Academy. These academies function as professional training facilities and concentrate in sports with professional pipelines like basketball and football. If lower-SES athletes disproportionately attend these academies—which are often located in affluent ZIP codes—our measures would overstate athletes' true SES and potentially obscure important differences between these sports and higher-SES sports like lacrosse, rowing, or tennis. That said, attendance at elite private academies may itself constitute a form of SES accrual in the sense of Jack (2019): exposure to institutional resources and elite peer networks confers advantages distinct from family background, which partially offsets this measurement concern. Regardless, even among basketball and football players in our sample, attendance rates at specialized academies are quite low at less than one percent each. Moreover, these specialized academies tend to come from lower-income ZIP codes than other private schools, which likely mitigates any bias.

Figure 2 compares the (enrollment-weighted) proportion of athletes and non-athletes who come from various quantiles of the income distribution, separately for each selectivity tier. The columns of the figure are in increasing order of income level, moving from top quintile on the left to top percentile on the right. Our results show that athlete income is tightly correlated with selectivity tier. Across all quantiles, the Ivy Plus has the highest proportion of high-income athletes, followed closely by Elite LACs. Other Elite Private comes next, followed by Public Flagship. For non-athletes, trends are roughly similar, although Elite LACs enroll fewer high-income non-athletes than universities that are Other Elite Private or Public Flagships. The Ivy Plus enrolls athletes that are most similar to their non-athlete peers in terms of income background, while the opposite is true for the Elite LACs. About 1 in 5 Ivy Plus athletes come from the top 1% of the income distribution, compared to around 1 in 6 non-athletes. For Elite LACs, almost 1 in 5 athletes come from the top 1% compared to just 1 in 25 non-athletes. Online Appendix Figure C.1 presents overall athlete and non-athlete proportions pooled across all tiers. It shows a similar pattern to the above.

We also measure athlete SES in other ways: attending a private high school; attending a high school outside the university's state; and the proportion of household heads in the high school ZIP code of the athlete that have bachelor's degrees or higher. On each of these measures, we observe trends across tiers that are similar to those in Figure 2. For example, Ivy Plus athletes have the highest rates of private high school attendance (over 50%; see Online Appendix Figure C.2),[10] the lowest rates of in-state high school attendance (less than 12%; C.3), and the second-highest rates of coming from households whose head has a bachelors degree or higher (just under 50%; Elite LAC is highest by a tiny margin; C.4). By contrast, racial and ethnic patterns differ markedly from these SES trends. We document in Online Appendix Figure C.5 that the racial and ethnic composition of athletes' high schools is quite homogeneous across tiers.[11]

[10]Across all four-year universities, the rate of students having graduated from private high schools is around 11% (National Center for Education Statistics, 2021).

[11]Online Appendix Table C.1 shows the characteristics of high schools that did and did not show up in our athlete roster data. High schools that send athletes to universities in our sample tend to have higher ZIP code incomes, larger enrollment, higher shares of white and Asian students, and lower shares of Black and Hispanic students.

3.4 Heterogeneity in athlete SES across tiers persists within standard sports

The results in the previous two subsections bring us to our fourth question. If more-selective universities tend to enroll higher-SES athletes, is this because they disproportionately offer regional and niche sports that tend to be played only by high-income high school students? Or is it the case that standard sports like football, basketball, track & field and soccer are also income-segregated across selectivity tiers?

Figure 3 repeats Figure 2 but explores heterogeneity by sport instead of selectivity tier. We sort the rows of the figure by the proportion of athletes in that sport who come from the top 1%. For ease of interpretation, we include only a subset of all possible sports. Unsurprisingly, niche sports such as squash, sailing, fencing, water polo, rowing and lacrosse top the list. Across all universities in our sample, over 1 in 4 squash athletes come from the top 1%, while nearly 1 in 10 lacrosse athletes do. At the other end of the spectrum, standard sports like football, wrestling and softball have much lower rates of top 1% income. This ordering by sport does not change very much if we instead look at less-extreme cutoffs such as the top 5% or top 10% of income.

While Figure 3 might seem to indicate that income segregation of athletes across tiers is driven by disparate rates of offering regional and niche sports, it may also be the case that standard sports are also income-segregated. Figure 4 compares rates of athletes being from the top 5% of the income distribution by selectivity tier and sport group. On the left is a subset of standard sports. On the right is the six sports with the highest overall rates of top 1% representation from Figure 3, which are either regional or niche according to our definition. While the rates are higher in the right hand graph, the two graphs show similar patterns in terms of income segregation across selectivity tiers. This implies that athlete income segregation is not solely driven by the composition of the sports that universities offer.

This within-sport income segregation has an important implication: even if elite universities dropped traditionally lower-SES sports like football and basketball, this would be unlikely to substantially affect their overall athlete SES. Within each sport, elite universities already recruit the highest-SES players, so income segregation across tiers is driven by within-sport sorting rather than differential sport offerings.

Several figures in the online appendix illustrate that this phenomenon is not limited to the top 5% of income. For example, rates of attending private or out-of-state high schools, or ZIP code educational attainment, follow nearly identical trends. See Online Appendix Figures C.6–C.9 for complete details.

3.5 Decomposing SES gaps: sport composition versus within-sport sorting

The visual patterns documented above suggest that within-sport sorting plays a dominant role in explaining athlete SES gaps across tiers. To quantify the relative importance of the two channels—which sports universities offer (*composition effect*) versus which athletes within each sport they recruit (*within-sport sorting effect*)—we formally decompose the total SES gap.

For each selectivity tier g, we decompose the deviation from the roster-size-weighted average athlete SES as:

$$\overline{Y}_g - \overline{Y} = \underbrace{\sum_s (w_{gs} - w_s)\overline{Y}_s}_{\text{composition effect}} + \underbrace{\sum_s w_{gs}(\overline{Y}_{gs} - \overline{Y}_s)}_{\text{within-sport sorting effect}} \tag{1}$$

where w_{gs} is the share of athletes in tier g playing sport s, $\overline{Y}_{gs}$ is the roster-size-weighted mean SES for sport s in tier g, and w_s and $\overline{Y}_s$ are the corresponding pooled values across all tiers (weighted by total roster sizes).

Table 2 presents the decomposition results using the share of athletes from ZIP codes in the top 5% of the national income distribution. Online Appendix Table C.2 contains full results for all of our SES measures. For top 5% income representation, within-sport sorting accounts for over three-quarters of the observed SES gaps across selectivity tiers, while differences in sport portfolios account for less than one-quarter. The sorting effect tends to be larger at private institutions (87–98%), although it is lower in the Ivy Plus tier (79%). The sorting effect is also much larger at Public Flagships (87%) compared to Mid-tier

Publics (76%).

The decomposition reveals that within-sport sorting dominates sport composition in explaining athlete SES gaps across selectivity tiers. This raises the question of whether universities strategically select sports based on athlete SES (the remaining one-quarter margin of our decomposition), or whether other factors like facility complementarities and institutional traditions drive these decisions. We address this with a structural model of sport bundle choices.

4 Model, identification and estimation

4.1 A bundle choice model of university sport offerings

We now proceed with our model of universities' choices over sport offerings. The model endogenizes the scale and composition of sport offerings and accounts for facility and other complementarities across sports, as well as the SES background of participating athletes.

We adopt a characteristic-based utility function in the tradition of Lancaster (1971). However, our model abstracts from several important features of university athletics. We do not model strategic interactions among universities within conferences, nor do we account for conference-level spillover effects that might influence sport adoption decisions. Additionally, we treat university characteristics as exogenous and focus on a single cross-section rather than a dynamic adjustment process.[12]

University i in NCAA division d chooses from a set of feasible bundles indexed by j in order to maximize utility. The choice set $\mathcal{J}_i$ contains all feasible combinations of 41 different men's and women's sports. Utility is a function of observable bundle characteristics X_{ij} and F_j, observable university characteristics W_i, and i.i.d. unobservable taste shocks as follows.

$$U_{i(d)j} = X_{ij}\beta_d + (W_i \cdot X_{ij})\gamma_d + F_j\delta_d + (W_i \cdot F_j)\phi_d + \varepsilon_{i(d)j}, \tag{2}$$

[12]This is primarily due to data limitations. We do not observe key variables (athlete SES, sport expenses, etc.) in the distant past even though we do observe historical take-up rates of sports. Moreover, the regulatory environment has changed dramatically over time such that separating preferences from regulatory constraints in the distant past would be challenging (Ehrenberg, 2000).

where X_{ij} includes characteristics of the bundle (e.g. log number of athletes, measures of athlete SES). F_j includes measures of the composition of bundle j (i.e. whether football is offered; number of sports in the bundle that are standard offerings, regional offerings, or niche offerings, separately for men and women) as well as composition complementarities (e.g. number of men's regional sports interacted with the presence of football) and facility complementarities meant to account for economies of scope in facility usage.[13] We measure facility complementarities as the number of additional sports beyond the first that share common infrastructure, with separate measures for each facility type (e.g. indoor wooden courts, outdoor turf fields, aquatic centers, etc.) to allow for heterogeneous complementarity effects across facility categories.

A key challenge in our approach is that the bundle characteristics X_{ij} (such as athlete SES measures) are only observed for universities' actual choices, not for hypothetical bundles they could have chosen but did not. For instance, we observe the income distribution of athletes on Harvard's current sports teams, but not what this distribution would be if Harvard hypothetically offered women's bowling. We address this by developing a systematic approach to impute these characteristics for non-chosen bundles based on sport-specific and university-specific patterns. Specifically, we estimate two-way fixed effects models that decompose each characteristic (e.g., athlete SES, roster size) into an institution-specific component and a sport-specific component. This allows us to predict what Harvard's women's bowling program would look like by combining Harvard's institutional effect with bowling's sport-specific characteristics. We further refine these predictions by incorporating conference-level variation when available, recognizing that sport characteristics may differ systematically across athletic conferences. This approach allows for university-specific unobserved heterogeneity, which makes our imputed values more realistic. Full details are provided in Online Appendix B.1.

Our model includes preference heterogeneity across NCAA divisions and universities by allowing the parameters to vary by division and by including observable university characteristics as preference interaction terms. W_i includes institution type fixed effects as well as the percentage of undergraduates who are from the top 10% of the income distribution

[13]Online Appendix A.2.1 details our definitions of standard, regional, and niche sports.

(Chetty et al., 2020). We provide complete specification details later on when discussing estimation of the model.

Equation (2) represents universities' preferences over the characteristics of their offered bundles of sports. By explicitly modeling the utility-maximizing choice process that underlies these decisions, we recover preference parameters that are interpretable as universities' willingness to trade off different bundle characteristics. We can then use these estimated preference parameters to predict the impacts of hypothetical policies that would restrict the number of athletes that certain universities could enroll.

4.2 Identification

Identification of the parameters in equation (2) requires three key components: (*i*) sufficient variation in observed bundle choices; (*ii*) correct specification of feasible choice sets; and (*iii*) appropriate distributional assumptions for the error terms.

The division-specific main effect parameters β_d and δ_d are identified from variation in bundle choices across universities within each division. Universities with similar observable characteristics W_i but different observed choices reveal their preferences over bundle attributes X_{ij} and F_j. Our key identifying assumption is that the $\varepsilon_{i(d)j}$'s are i.i.d. conditional on observed university characteristics W_i and bundle characteristics (X_{ij}, F_j).

The division-specific interaction parameters γ_d and ϕ_d are identified from cross-university variation in characteristics W_i within each division. For instance, more selective universities or those enrolling more high-income students may place greater weight on sports with higher-SES athletes, while less selective institutions may prioritize sports with stronger facility complementarities to minimize infrastructure costs. The interaction terms $(W_i \cdot X_{ij})$ and $(W_i \cdot F_j)$ capture these preference differences, provided there is sufficient variation in W_i within divisions.

We specify university i's feasible choice set $\mathcal{J}_i$ based on three types of constraints. First, we impose NCAA division-specific requirements: Division I Football Bowl Subdivision (FBS) schools must offer at least 10 sports total, including football, with at least 5 men's sports and 6 women's sports; non-FBS Division I schools must offer at least 8 sports with at least 4 each for men and women; and Division I Football Championship Subdivision (FCS) schools must include football.[14] Second, we enforce Title IX gender balance constraints by limiting the number of men's sports to exceed women's sports by at most $2 + 4 \times \text{share\_men}_i$ based on each university's undergraduate gender composition. Third, we apply geographic and institutional constraints, such as excluding skiing programs for universities in unsuitable climates and ensuring single-gender institutions can only choose sports for their enrolled gender. We exclude from $\mathcal{J}_i$ any bundle violating any of these three sets of constraints. As validation of our approach, each university's observed choice is in its feasible set.

While the above sources of variation allow us to identify the parameters in equation (2), our cross-sectional framework limits our ability to identify unobservable preference heterogeneity (e.g., through a latent class model or random coefficients). Though such approaches are feasible, they would require strong parametric assumptions for the mixing distributions that we prefer to avoid. We instead opt for a specification that includes many interaction terms to capture observable preference heterogeneity across divisions, institution types, and student SES characteristics.

Our primary reason for conducting cross-sectional analysis is that universities rarely make dramatic year-to-year changes to their sport portfolios due to facility investments, coaching contracts, and other institutional factors. This serial correlation limits the identifying variation that would be available in panel data models. We address this cross-sectional limitation in our counterfactual analysis by leveraging complete historical data on sport offerings, which allows us to identify which sports universities consider to be "traditional" and hence less likely to drop under hypothetical athletic enrollment cuts.

4.3 Estimation

Direct estimation of the discrete choice model in equation (2) is computationally infeasible given that universities face choice sets with close to $2^{41} \approx 2.2$ trillion possible sport bundle combinations. We therefore employ choice-based sampling (McFadden, 1978; Davis et al., 2019) by randomly drawing 249 feasible bundles for each university from its feasible choice set $\mathcal{J}_i$, along with the actually chosen bundle.

[14] In practice, minimum sport requirements are larger than these, but we have aggregated some sports together (e.g. swimming and diving, track and field and cross-country) that the NCAA keeps separate.

Our sampling approach involves two stages. First, we construct a common set of $\approx$ 10,000 representative bundles that capture the key dimensions of variation in sport offerings. Second, for each university, we randomly sample 249 feasible bundles from this common set, plus the chosen bundle (details in Online Appendix B.2). This two-stage approach ensures both computational tractability and sufficient coverage of the relevant choice space.

Under the assumption that $\varepsilon_{i(d)j}$ follows an i.i.d. Type I Extreme Value distribution, the probability that university i chooses bundle j from feasible set $\mathcal{J}_i$ takes the following conditional logit form:

$$P_{i(d)j} = \frac{\exp(X_{ij}\beta_d + (W_i \cdot X_{ij})\gamma_d + F_j\delta_d + (W_i \cdot F_j)\phi_d)}{\sum_{k\in\mathcal{J}_i} \exp(X_{ik}\beta_d + (W_i \cdot X_{ik})\gamma_d + F_k\delta_d + (W_i \cdot F_k)\phi_d)} \tag{3}$$

We estimate separate conditional logit models for NCAA Division I, Division II, and Division III universities, as these divisions face substantially different constraint environments and likely exhibit different preference structures. The choice-based sampling approach yields consistent parameter estimates under standard regularity conditions (McFadden, 1978).

We now detail the specification of equation (2). The vector X_{ij} includes the natural log of total athletes in bundle j, an indicator for whether j is monetarily profitable, and athlete SES measures: percentages of athletes in j who respectively come from domestic and international private high schools, and percentages whose high school ZIP code income respectively falls in the top 20%, 10%, 5%, and 1% of the national distribution. All of these percentages are weighted by the roster sizes of each component sport.

The vector F_j captures bundle composition and complementarity effects. Sport composition variables include an indicator for football, as well as counts of standard, regional, and niche sport offerings for men and women separately. Sport complementarities include interactions between football and sport category counts, as well as pairwise interactions between sport categories.

Facility complementarity variables are defined as $\text{facility\_comp}_{fj} = 1[\text{facility\_count}_{fj} > 1] \times (\text{facility\_count}_{fj} - 1)$ where $\text{facility\_count}_{fj} = \sum_{s\in S_f} 1[\text{sport}_s \in \text{bundle}_j]$ counts sports in bundle j using facility type f. This captures economies of scope in facility usage across different infrastructure types (indoor courts, outdoor turf fields, aquatic centers, etc.). We consider eight facility types: track and field complexes (shared by track and field, cross country and football), indoor courts (basketball and volleyball), aquatic centers (swimming, diving and water polo), outdoor fields (soccer, lacrosse, field hockey and football), racquet courts (tennis and squash), ice rinks (ice hockey), mat rooms (wrestling and gymnastics), and baseball/softball diamonds. All facility types include both men's and women's varieties of each underlying sport.

University characteristics W_i include heterogeneity group fixed effects (pooling some selectivity tiers to ensure sufficient variation within divisions) and the percentage of undergraduates from the top 10% of the income distribution (Chetty et al., 2020).[15] We interact the top 10% income share with all bundle characteristics, while heterogeneity group fixed effects are interacted with the bundle characteristics X_{ij} and the sport category counts in F_j. For the heterogeneity group fixed effects, we designate Mid-tier Public as the reference category.

In all, we estimate 116 parameters each for Divisions I and II, and 129 for Division III. The number of universities in our sample belonging to these divisions is respectively 343, 277 and 393. Differences in the number of parameters by division arise due to limited identifying variation in the heterogeneity groups (e.g., there are few public flagships competing outside of Division I) and cases where particular sport groups are nearly universally adopted within a division (e.g., football in Division I). In these cases, we constrain the corresponding coefficients to zero.

We use the Stata command `cmclogit` to estimate our conditional logit model via maximum likelihood. Our conditional logit objective function is globally concave in parameters, which guarantees that standard optimization routines converge to the unique global maximum.

[15]We create five heterogeneity groups by pooling selectivity tiers with sparse representation in certain divisions: (1) Elite Private (combining Ivy Plus, Elite LACs, and Other Elite Private); (2) Public Flagship; (3) Mid-tier Private; (4) Mid-tier Public; and (5) Other Non-selective. We reassign institutions to ensure adequate sample sizes as follows: Elite Private schools in Division II join Mid-tier Private; Public Flagships outside Division I join Mid-tier Public; and Other Non-selective Division I schools join Mid-tier Public.

5 Estimation results and model validation

We now discuss the most pertinent parameter estimates, as well as assessing the validity of the model.

5.1 Discrete choice parameter estimates

Due to the large number of estimated parameters (361 in all), we opt for verbally discussing coefficients of interest. Moreover, as individual logit coefficients are difficult to interpret on their own, we discuss sign and significance without detailing exact magnitudes. Online Appendix Table C.3 contains the complete set of parameter estimates and standard errors.

One of our key research questions is whether universities strategically choose sport offerings in order to achieve a higher-SES student body, and whether this behavior differs by selectivity tier or level of student body income. Our measures of SES are private high school status and likelihood of coming from ZIP codes with top incomes.

For Elite Private universities, we find mixed evidence regarding preferences for bundles with higher shares of athletes from top income percentiles across divisions. Division I Elite Private institutions demonstrate a strong positive preference for athletes from the top 5% of income, while showing no significant preference for the top 1%. In contrast, Division III Elite Private institutions show a significant negative preference for athletes from the top 1% of income.[16] This pattern is consistent with Figure 4, which shows that Elite Private institutions already attract high rates of top-income athletes across different sets of sports. This leaves little remaining scope for income-based strategic differentiation through sport selection.

Among D-II universities, we find significant preferences for bundles with top-income athletes. There is a positive and significant coefficient on the percentage of athletes from the top 5%, but a large and negative coefficient on the percentage of athletes from the top 1%. This suggests that D-II institutions prefer 'upper-middle-class' rather than ultra-wealthy athletes. Division III institutions tend to show no significant preferences based on income, although there is some heterogeneity based on tier.

[16] In almost all cases, the standard errors on the coefficients associated with athlete SES are quite small.

Beyond income, we find limited evidence of preferences for athletes from private high schools. Among D-I universities, the main effect is negative and not statistically significant, and the interaction for Public Flagships is negative and significant. In Division III, the main effect remains small and insignificant, with a significant negative interaction for Elite Privates. One exception is that D-III universities have positive preferences for athletes from international private high schools, while for D-I the effect is small and negative.

For the other bundle characteristics such as sport and facility complementarities, we find heterogeneous effects by division and tier. Many of these coefficients are large and statistically significant, while some are negative. For example, court and mat complementarities are positive for D-I and D-III but field and ice complementarities both show opposing patterns across divisions. These patterns imply that facility constraints and non-pecuniary factors are important components of universities' sport offering decisions.

Our estimates of the interactions between student body income and many of the bundle characteristics show that income moderates several relationships. Many complementarity and athlete income variables include significant interactions with overall student body income. This evidence points to campus culture—as measured by income—being an important factor governing which sports are offered.

In summary, while we do find evidence of SES-based strategic selection through sport offerings, this behavior varies considerably by division and institutional tier. The most pronounced effects occur at the Division II level rather than among the most elite institutions.

5.2 Model validation

We now turn to validating our estimated model. By the fundamental properties of maximum likelihood estimation (MLE), our model will, by construction, perfectly replicate the observed sample means of covariates explicitly included in the likelihood function—the so-called "targeted moments." For example, the average characteristics of the model-predicted choices will match the average characteristics of the observed choices for the log number of athletes, the bundle profitability rates, the percentage of athletes from the top 1% of income, etc. This perfect fit for targeted moments reflects the optimization criterion inherent in MLE rather than true model validity.

The critical test of our model's explanatory power therefore lies in its ability to reproduce untargeted moments—empirical patterns not directly incorporated into the estimation procedure. These include take-up rates of individual sports, co-occurrences of specific sports, and other distributional features of the covariates (e.g. the 90th percentile as opposed to the average of log athletes), among others.

We compute the model-predicted chosen bundle characteristics by computing weighted averages of various variables that are not directly included in the estimation. For a given variable Z,

$$\overline{Z}^{pred} = \frac{1}{N} \sum_{i} \left[\sum_{j \in \mathcal{J}_i} \hat{P}_{ij} \cdot Z_{ij} \right] \tag{4}$$

where i indexes universities and j indexes bundles as in equation (2), and where $\hat{P}$ refers to equation (3) evaluated at the estimated parameter values. Here, we do the computation over all $\approx$ 10,000 common bundles that are feasible for each university, rather than the choice-based sampled set of 250.

In this setup, Z can be any variable observed in our data that is not included in our model. In Figure 5, we compare $\overline{Z}^{pred}$ with $\overline{Z}$ where Z is a set of indicators of whether each of the 41 sports in our data is present in the bundle. We depict these comparisons visually for ease of presentation. Tan-colored bars represent the average take-up rate of each sport in our estimation subsample of the data, while navy-colored bars represent model-predicted take-up rates. Overall, the model does a decent job of matching the take-up rate of nearly every single sport. Niche sports in the far tail tend to be over-predicted, as well as aquatics, women's field hockey, women's ice hockey and men's volleyball. Men's wrestling is underpredicted. Online Appendix Figures C.10–C.13 show versions of Figure 5 specific to each division, as well as specific to elite universities (Ivy Plus, Elite LAC, and Other Elite Private tiers pooled together). The pictures are highly similar to Figure 5.

We also include several other fit plots for untargeted moments in the online appendix. These include metrics for the number of sports in the bundle, the fraction of athletes in the bundle who come from top income levels, and the cost of the bundle (i.e. total expenses). See Online Appendix Figures C.14–C.17. Many of these not only fit perfectly on average (by construction) but also fit well across the entire distribution. We also show in Online Appendix Figures C.18 and C.19 that our model fits the co-occurrence of sports offerings reasonably well.

Overall, our model fits the data well even on moments not targeted by the estimation procedure. This suggests our discrete choice model has successfully identified the salient aspects of preferences. Having established model validity, we now move on to counterfactual exercises which rely on the model's ability to predict behavior in alternative settings.

6 Counterfactual simulations and policy implications

6.1 Counterfactual simulations

We use our estimated model to predict how university sport offerings would change under two scenarios related to our research question: (*i*) how would sport offerings change if elite universities were forced to cap their athletic enrollment to 5% of total enrollment?; and (*ii*) how would these changes differ if universities were additionally required to prioritize their longest-offered sports? The latter is intended to improve credibility by rooting choices in tradition, effectively capturing switching costs. In both scenarios, we assume that universities would neither increase their capacity nor adjust their athletic recruiting behavior. We also hold fixed external regulations such as Title IX and NCAA bylaws.

Our thinking in constructing these two scenarios is as follows. Since members of the Ivy Plus, Elite LAC, and Other Elite Private tiers are the most sought-after universities, admissions preferences for athletes constitute the largest opportunity cost from a talent allocation standpoint. We target a 5% threshold because it is more in line with athlete enrollment shares at public universities. While these universities typically have much larger enrollments, 5% is still a large enough number for most elite private universities to field several sports teams for both genders.

We implement these scenarios using the counterfactual bundle generation approach described in Online Appendix B.3. For each university, we generate up to 350 bundles that are feasible with respect to regulations and geographical constraints. Our bundle generation algorithm sequentially adds sports taking into account length of time offered, gender balance under Title IX, and geographical feasibility. We then use the parameter estimates of equation (2) and the formula in equation (4) to calculate the expected characteristics of the counterfactually chosen bundle.

To quantify the impact of each scenario, we present two pieces of evidence. First, we look at the income profiles of athletes in the counterfactually chosen bundles and compare them to those of the chosen bundle in the data. Second, we consider how the take-up rates of individual sports would change in the counterfactual scenarios. This sheds light on which specific sports would be likely to be dropped if top universities' athletic enrollment were restricted.

Figure 6 shows the percentage of athletes in the chosen bundle who come from ZIP codes in the top percentiles of income and compares it to the predicted bundles in the two counterfactuals. Consistent with the results in Section 5.1, we predict that there would be little change in the SES status of athletes in either counterfactual scenario. This is because there are not very strong preferences for athlete SES in the first place, owing to the fact that within-sport income segregation closely matches across-institution income segregation. Even traditionally lower-SES sports like football and basketball attract high-income players at elite universities as shown in Figure 4.

In Figure 7, we present model-predicted and counterfactual sport take-up rates among elite universities where our counterfactual constraint binds. For this exercise, we focus on the tradition-weighted athlete cap, as these provide more plausible predictions. The figure shows that all sports would see sizable cuts, which is not surprising, given that the 5% cap on athlete enrollment implies a more than 50% reduction (and in some cases, much larger) in the number of athletes on campus. We predict large reductions in track & field, soccer, golf, lacrosse, and squash, but very little change in fencing, women's skiing, or women's bowling. In contrast, Online Appendix Figure C.20 shows what would happen without the tradition weighting. This figure predicts massive shifts to niche sports and away from track & field, softball, baseball, and swimming & diving.

It is interesting to note that the shock of the COVID-19 pandemic provides external validity of our model's predictions. As a reminder, we conduct our analysis on data for the

2019–20 academic year, prior to the COVID shock. While the shock is not exactly the same policy, it induced several elite universities—including Brown, Dartmouth, and Stanford—to publicly consider cutting or actually cut some of their sports teams. Brown initially demoted 11 teams to club status while eventually reinstating five (Miller, 2022). Stanford had planned to also cut 11 teams before alumni intervention and fundraising encouraged the decision to be reversed (Rubin, 2021). A similar story unfolded at Dartmouth, where five teams were set to be cut before being eventually reinstated (Dartmouth College, 2021). Many of the sports that were proposed to be cut are the sports that we predict would be most likely to be cut, e.g. swimming and diving, golf, track and field, rowing, and squash.[17]

6.2 Policy implications

Recent research (Chetty, Deming, and Friedman, 2023) has questioned whether, in terms of a social optimum, there are too many wealthy students admitted to top universities in the United States. The 5% enrollment cap we consider in this paper is one way to limit enrollment of the wealthy because athletes tend to come from higher income backgrounds than non-athletes.

While our counterfactuals show minimal change in athlete SES composition, the policy's value lies in reallocating seats from athletes to potentially higher-merit non-athletes. The answer of whether too many wealthy students are enrolled at top universities hinges on the outcomes of the students after college. Chetty, Deming, and Friedman (2023) document, for athletes in particular, that there are zero-to-negative effects of having attended as an athlete on the likelihood of earning a top income or working at a prestigious firm. In contrast, being academically excellent is much more correlated with these outcomes.

In Table 3, we compare two ways that universities could comply with the 5% cap we consider. First, they could cut athlete enrollment, as we have shown in the previous subsection. Second, they could expand non-athlete enrollment.[18] Panel B shows that cutting athlete

[17] Our predictions are not infallible. Brown ended up actually cutting its men's fencing and women's skiing teams, which our model predicts would be unlikely to happen. This is because fencing has a particularly long history as a college sport. Likewise, our model predicts broad cuts to soccer teams, but these were never publicly considered.

[18] The athlete shares in both panels are not exactly 5% due to how we set up our counterfactual. Namely, bundles with athlete shares below 5% are feasible, and there is some small probability of choosing them due

enrollment without changing overall capacity would result in 9.6% more non-athletes at Ivy Plus universities, 36% more at Elite LACs, and 6% more at Other Elite Privates. On the whole, there would be 8.2% additional non-athletes at all universities in these tiers. Panel C shows what would happen if athlete admissions stayed the same, but non-athlete seats were expanded. This would require a monumental expansion of seats, on the order of 150% (i.e. 2.5x). This is because these universities have much higher athlete-to-non-athlete ratios than the proposed 1-to-20 ratio.

Whether our proposed 5% policy would improve socioeconomic mobility at the most elite universities depends on who universities would replace their athletes with. As Chetty, Deming, and Friedman (2023) show, marginal effects on extreme positive outcomes are small for legacies and those with strong non-academic ratings. Thus, our policy would likely need to be coupled with some other admissions policies in order to produce the intended effect.

7 Conclusion

In this paper, we investigate whether universities strategically choose the types of sports they offer in order to enroll greater numbers of high-SES athletes. We find that they do not, mainly because there is a high degree of income homophily between athletes and universities. Instead, we find that facility complementarities and other institutional factors drive the decision over which sports to offer. Our paper is the first to document the socioeconomic backgrounds of all NCAA athletes and compare them to those of non-athletes.

We conduct counterfactual simulations of our model to show that a cap on athletic enrollment at elite universities would do little to alter the SES background of their athletes, but would dramatically shift the types of sports offered. However, athletic enrollment caps would open up more seats to non-athletes. The ultimate impact of these additional seats on the outcomes of elite university students would depend on how the university chooses to fill the vacancies. Thus, a policy that restricts athletic enrollment at elite universities would need to be paired with other admissions policies to achieve maximum success.

to the logit preference shocks. In practice, this minor discrepancy makes no substantive difference to our results.

There are several limitations to our research. First, our cross-sectional analysis ignores costs of switching out sports (e.g., building new facilities, demolishing old facilities, hiring new coaching staffs, etc.). Second, we abstract from strategic interactions among competing universities. Third, our analysis is on data prior to dramatic changes to NCAA rules, including allowing athletes to monetize their name, image and likeness (NIL) or receive direct payments from their affiliated institutions. However, it is unclear the degree to which this affects the decisions of elite universities. Finally, as we do not have data on donations, it is impossible for us to quantify the future donor value of an athlete compared to a non-athlete.

Most of the discourse on the issue of athletic enrollment focuses on admissions preferences for athletes and ignores the fact that universities can choose their level of athletic enrollment. Our research fills this gap by pointing out that these admissions preferences are downstream of the university's chosen intensity and composition of sport offerings. This implies that policies that restrict athletic enrollment will automatically limit the scope of these controversial admissions preferences (Chappell and Kennedy, 2019).

However, any reappraisal of elite university admissions should consider universities' broader institutional objectives. Prior research suggests that sports play a critical role in fulfilling multiple objectives simultaneously—allowing universities to "craft a class" and secure donations (Karabel, 2005; Golden, 2006; Stevens, 2007; Meer and Rosen, 2009a), and providing consumption amenities and alumni networks (Rivera, 2016; Jacob, McCall, and Stange, 2018). While athletes themselves may not do as well after college (Chetty, Deming, and Friedman, 2023), there may be spillover effects on the socioeconomic mobility of non-athletes. This could be a fruitful path for future research.

References

Amornsiripanitch, Natee, Paul Gompers, George Hu, Will Levinson, and Vladimir Mukharlyamov. 2023. "No Revenge for Nerds? Evaluating the Careers of Ivy League Athletes." Working Paper 31753, National Bureau of Economic Research.

Arcidiacono, Peter, Josh Kinsler, and Tyler Ransom. 2022. "Legacy and Athlete Preferences at Harvard." *Journal of Labor Economics* 40 (1):133–156.

———. 2023. "What the Students for Fair Admissions Cases Reveal about Racial Preferences." *Journal of Political Economy Microeconomics* 1 (4):615–668.

———. 2024. "Divergent: The Time Path of Legacy and Athlete Admissions at Harvard." *Journal of Human Resources* 59 (3):653–683.

Barrios-Fernández, Andrés, Christopher Neilson, and Seth Zimmerman. 2024. "Elite Universities and the Intergenerational Transmission of Human and Social Capital." Discussion Paper 17252, IZA Institute of Labor Economics. URL https://docs.iza.org/dp17252.pdf.

Bowen, William G. and Sarah A. Levin. 2003. *Reclaiming the Game: College Sports and Educational Values*. Princeton, New Jersey: Princeton University Press.

Chappell, Bill and Merrit Kennedy. 2019. "U.S. Charges Dozens Of Parents, Coaches In Massive College Admissions Scandal." *NPR* URL https://www.npr.org/2019/03/12/702539140/u-s-accuses-actresses-others-of-fraud-in-wide-college-admissions-scandal. Accessed March 23, 2019.

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

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

Clotfelter, Charles T. 2003. "Alumni Giving to Elite Private Colleges and Universities." *Economics of Education Review* 22 (2):109–120.

Cook, Emily E. 2025. "Market Structure and College Access in the US." Working paper, Texas A&M University.

Dartmouth College. 2021. "Five Varsity Teams Reinstated, External Reviews Planned." URL https://home.dartmouth.edu/news/2021/01/five-varsity-teams-reinstated-external-reviews-planned.

Davis, Donald R., Jonathan I. Dingel, Joan Monras, and Eduardo Morales. 2019. "How Segregated Is Urban Consumption?" *Journal of Political Economy* 127 (4).

Ehrenberg, Ronald G. 2000. *Tuition Rising: Why College Costs So Much*. Cambridge, Massachusetts: Harvard University Press.

Epple, Dennis, Richard Romano, Sinan Sarpça, and Holger Sieg. 2017. "A General Equilibrium Analysis of State and Private Colleges and Access to Higher Education in the U.S." *Journal of Public Economics* 155:164–178.

Epple, Dennis, Richard Romano, and Holger Sieg. 2006. "Admission, Tuition, and Financial Aid Policies in the Market for Higher Education." *Econometrica* 74 (4):885–928.

Espenshade, Thomas J., Chang Y. Chung, and Joan L. Walling. 2004. "Admission Preferences for Minority Students, Athletes, and Legacies at Elite Universities." *Social Science Quarterly* 85 (5):1422–1446.

Fillmore, Ian. 2023. "Price Discrimination and Public Policy in the US College Market." *Review of Economic Studies* 90 (3):1228–1264.

Fu, Chao. 2014. "Equilibrium Tuition, Applications, Admissions, and Enrollment in the College Market." *Journal of Political Economy* 122 (2):225–281.

Garthwaite, Craig, Nicole Holz, Jordan Keener, and Matthew Notowidigdo. 2025. "Who Profits From Amateurism? Rent-Sharing in Modern College Sports." Working paper, Northwestern University. URL https://users.nber.org/~notom/research/GKNO_Manuscript_mar2025.pdf.

Gentzkow, Matthew. 2007. "Valuing New Goods in a Model with Complementarity: Online Newspapers." *American Economic Review* 97 (3):713–744.

Gladwell, Malcolm. 2024. *Revenge of the Tipping Point: Overstories, Superspreaders, and the Rise of Social Engineering*. Little, Brown and Company.

Golden, Daniel. 2006. *The Price of Admission: How America's Ruling Class Buys Its Way into Elite Colleges – and Who Gets Left Outside the Gates*. New York: Crown Publishers.

Heckman, James J, Colleen P Loughlin, and Haihan Tian. 2025. "The Benefits of Scholastic Athletics." Working Paper 34046, National Bureau of Economic Research.

Hextrum, Kirsten. 2021. *Special Admission: How College Sports Recruitment Favors White Suburban Athletes*. Rutgers University Press.

Jack, Anthony Abraham. 2019. *The Privileged Poor: How Elite Colleges Are Failing Disadvantaged Students*. Harvard University Press.

Jacob, Brian, Brian McCall, and Kevin Stange. 2018. "College as Country Club: Do Colleges Cater to Students' Preferences for Consumption?" *Journal of Labor Economics* 36 (2):309–348.

Karabel, Jerome. 2005. *The Chosen: The Hidden History of Admission and Exclusion at Harvard, Yale, and Princeton*. New York: Houghton Mifflin.

Lancaster, Kelvin. 1971. *Consumer Demand: A New Approach*. New York: Columbia University Press.

Lewis, Michael. 2020. *Playing to Win*. Audible Originals. URL https://www.amazon.com/Playing-to-Win/dp/B08DL7ZJDX. Audible Audiobook—Original recording.

Manski, Charles F. and Leonard Sherman. 1980. "An Empirical Analysis of Household Choice among Motor Vehicles." *Transportation Research Part A: General* 14A (5–6):349–366.

McFadden, Daniel. 1978. "Modelling the Choice of Residential Location." In *Spatial Interaction Theory and Planning Models*, edited by A. Karlqvist, L. Lundqvist, F. Snickars, and J. Weibull. Amsterdam: North-Holland, 75–96.

Meer, Jonathan and Harvey S. Rosen. 2009a. "Altruism and the Child Cycle of Alumni Donations." *American Economic Journal: Economic Policy* 1 (1):258–286.

———. 2009b. "The Impact of Athletic Performance on Alumni Giving: An Analysis of Microdata." *Economics of Education Review* 28 (3):287–294.

Miller, Nicholas. 2022. "Athletes, Coaches, Administrators Reflect on Turbulent Two Years in Brown Athletics." *The Brown Daily Herald* URL https://www.browndailyherald.com/article/2022/05/athletics-retrospective-commencement-2022.

National Association of Intercollegiate Athletics. 2024. "2025-2026 NAIA Institutions." Tech. rep. URL https://www.naia.org/schools/files/2025-2026_NAIA_Institutions.pdf. Accessed December 3, 2025.

National Center for Education Statistics. 2021. "Type of High School Last Attended by Parents' Highest Education Level." National Postsecondary Student Aid Study: 2016 Undergraduates (NPSAS:UG). Generated using NCES PowerStats on September 16, 2021. Table identifier: xrjrxt.

———. 2023. "Characteristics of Degree-Granting Postsecondary Institutions." Tech. rep., U.S. Department of Education, Institute of Education Sciences. URL https://nces.ed.gov/programs/coe/indicator/csa. Accessed December 3, 2025.

National Collegiate Athletic Association. 2021. "National Collegiate Athletic Association v. Alston." *U.S.* 594:69. No. 20-512.

———. 2024. "NCAA Member Schools." URL https://www.ncaa.com/schools-index/0. Accessed December 3, 2025.

Newsom, Gavin. 2024. "California Bans Legacy and Donor Preferences in Admissions at Private, Nonprofit Universities." Press release, Office of the Governor of California. URL https://www.gov.ca.gov/2024/09/30/california-bans-legacy-and-donor-preferences-in-admissions-at-private-nonprofit-unive

Pope, Devin G. and Jaren C. Pope. 2009. "The Impact of College Sports Success on the Quantity and Quality of Student Applications." *Southern Economic Journal* 75 (3):750–780.

———. 2014. “Understanding College Application Decisions: Why College Sports Success Matters.” *Journal of Sports Economics* 15 (2):107–131.

Ransom, Michael R and Tyler Ransom. 2018. “Do High School Sports Build or Reveal Character? Bounding Causal Estimates of Sports Participation.” *Economics of Education Review* 64:75–89.

Rivera, Lauren A. 2016. *Pedigree: How Elite Students Get Elite Jobs*. Princeton, New Jersey: Princeton University Press.

Rubin, Jeremy. 2021. “Stanford to Reinstate All 11 Discontinued Varsity Sports.” *The Stanford Daily* URL https://www.stanforddaily.com/2021/05/18/stanford-to-reinstate-all-11-discontinued-varsity-sports/. Accessed December 6, 2021.

Skinner, Benjamin. 2023. *rscorecard: A Method to Download Department of Education College Scorecard Data*. URL https://CRAN.R-project.org/package=rscorecard. R package version 0.26.0.

Stevens, Mitchell L. 2007. *Creating a Class: College Admissions and the Education of Elites*. Cambridge, Massachusetts: Harvard University Press.

Students for Fair Admissions. 2023. “Students for Fair Admissions, Inc. v. President and Fellows of Harvard College.” *U.S.* 600:181. No. 20-1199.

Thompson, Derek. 2019. “The Cult of Rich-Kid Sports.” *The Atlantic* URL https://www.theatlantic.com/ideas/archive/2019/10/cult-rich-kid-sports/599394/.

Train, Kenneth E., Daniel L. McFadden, and Moshe Ben-Akiva. 1987. “The Demand for Local Telephone Service: A Fully Discrete Model of Residential Calling Patterns and Service Choices.” *RAND Journal of Economics* 18 (1):109–123.

Tyagi, Erika, Kyle Ueyama, and The Urban Institute. 2022. *educationdata: Retrieve Records from the Urban Institute’s Education Data Portal API*. URL https://CRAN.R-project.org/package=educationdata. R package version 0.1.3.

Walker, Kyle, Matt Herman, and Kris Eberwein. 2023. *tidycensus: Load US Census Boundary and Attribute Data as ‘tidyverse’ and ‘sf’-Ready Data Frames*. URL https://CRAN.R-project.org/package=tidycensus. R package version 1.4.4.

Zimmerman, Seth D. 2019. “Elite Colleges and Upward Mobility to Top Jobs and Top Incomes.” *American Economic Review* 109 (1):1–147.

Figures and Tables

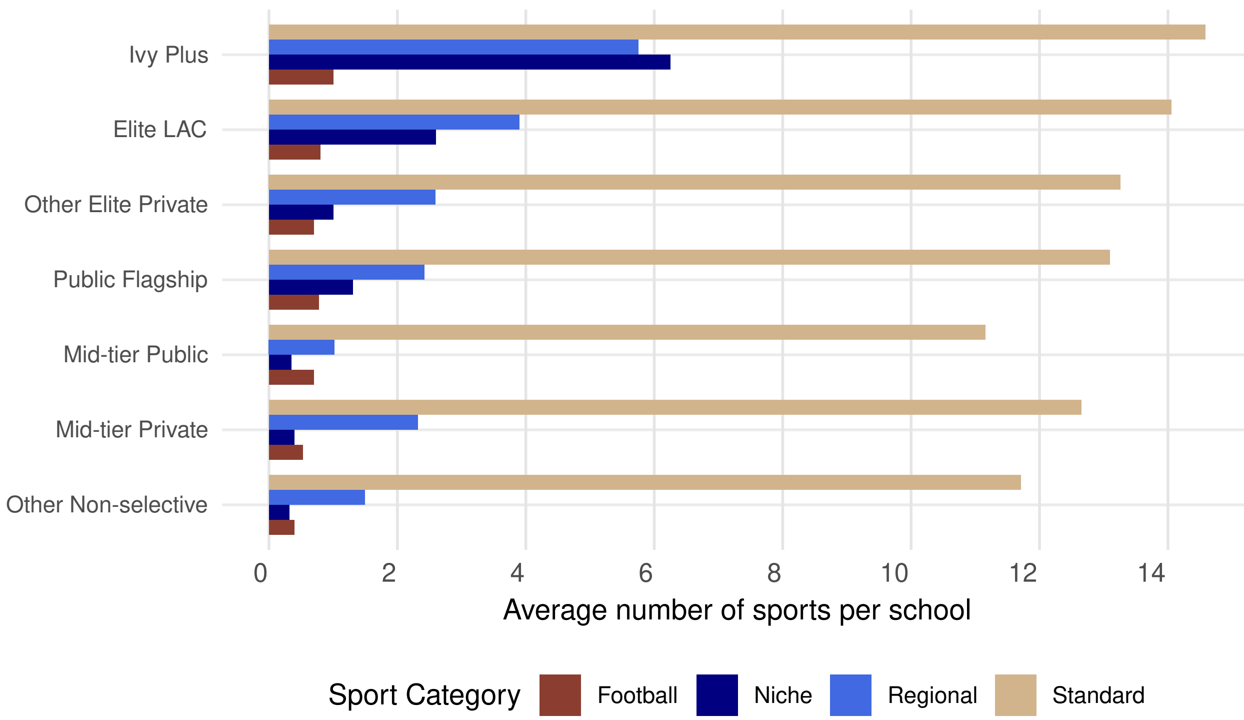
Table 1: University Athletic Enrollment by Selectivity Tier

Selectivity Tier	Institutions	Athletes (% of Students)	Admission Rate (%)	Share of all Students (%)	Share of all Athletes (%)
Ivy Plus	12	13.1	7.1	1.2	2.8
Elite LAC	20	30.7	19.2	0.6	3.0
Other Elite Private	92	10.0	36.6	6.6	11.2
Public Flagship	53	2.4	56.8	19.7	8.0
Mid-tier Public	350	3.7	73.9	51.2	32.1
Mid-tier Private	417	14.3	74.1	15.5	37.8
Other Non-selective	69	5.7	77.2	5.3	5.2
Total	1,013	5.8	67.1	100.0	100.0

Source.—Authors' calculations using data from IPEDS, College Scorecard, and CollegeFactual.

Notes.—Admission rates are enrollment-weighted averages within tier. Overall share columns show the percentage of total undergraduate students and athletes enrolled in each selectivity tier (probability mass). Ivy Plus includes the eight Ivy League universities plus Stanford, MIT, Duke, and University of Chicago (Chetty et al., 2020). Elite LACs include highly selective liberal arts colleges: several NESCAC schools (Amherst, Bates, Bowdoin, Colby, Connecticut College, Hamilton, Middlebury, Wesleyan, and Williams) plus other top LACs (Swarthmore, Haverford, Colgate, Davidson, Kenyon, Oberlin, Vassar, Carleton, Macalester, Pomona-Pitzer, Claremont McKenna-Harvey Mudd-Scripps). Other Elite Private includes highly selective private universities such as Georgetown, Northwestern, Washington University in St. Louis, Rice, Vanderbilt, Carnegie Mellon, Notre Dame, and others. Public Flagships include highly ranked and selective public flagship universities. Mid-tier Private and Mid-tier Public respectively represent less-selective private and public four-year institutions. Other Non-selective includes remaining institutions with low admission standards.

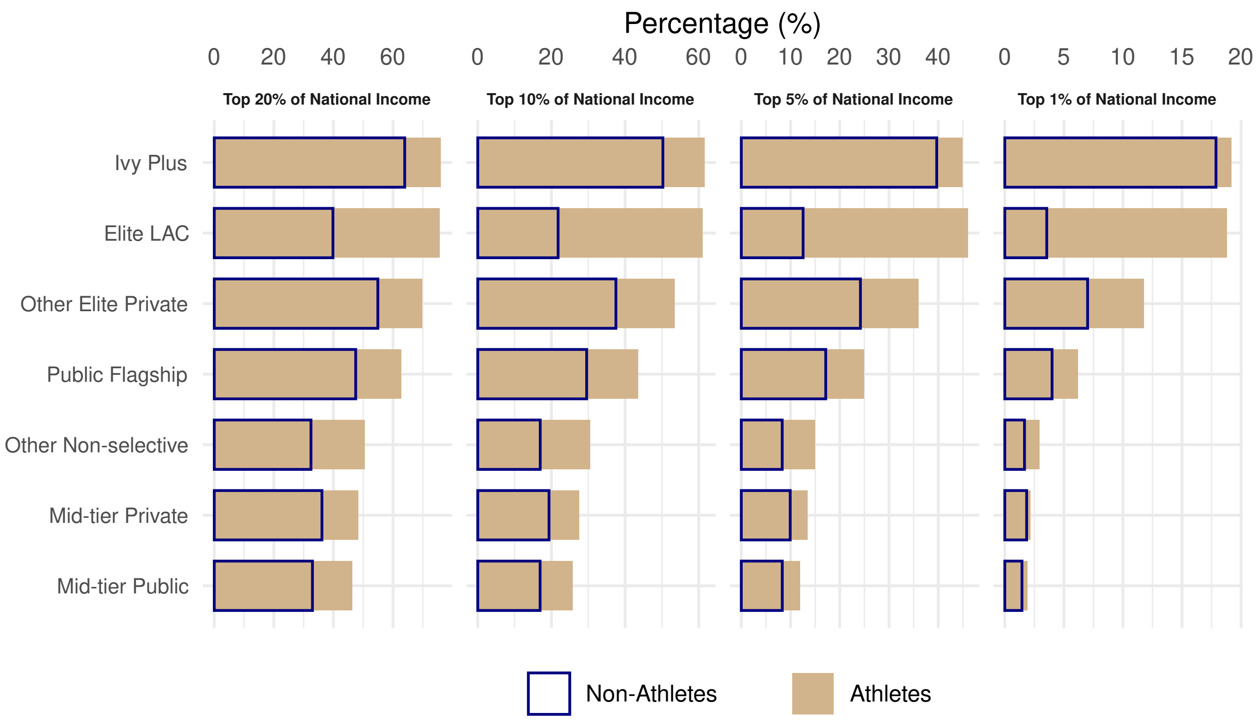
Figure 1: Sport Offerings by Sport Category and Selectivity Tier



Source.—Authors' calculations from NCAA directory data linked to university characteristics.

Notes.—This figure plots the average number of sports in each sports group among universities in the corresponding selectivity tier. We group sports into the following categories: Football; standard (i.e., typically offered at most high schools); Regional (i.e., only offered regionally at the high school level); and Niche (i.e., rarely offered at the high school level and/or typically requiring specialized facilities or training). See Online Appendix A.2.1 for a complete description of sport groups.

Figure 2: Representation of Athletes and Non-Athletes in Upper Tails of Income Distribution, by Selectivity Tier



Source.—Authors' calculations from NCAA roster data linked to high school ZIP code characteristics and university characteristics.

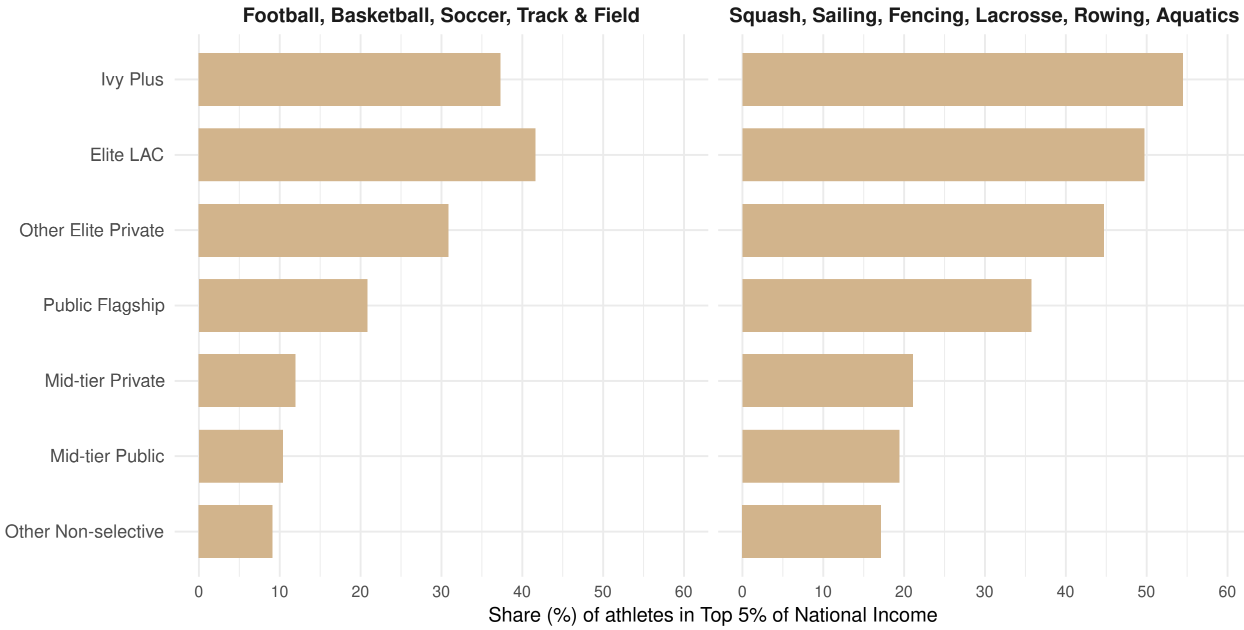
Notes.—This figure plots the likelihood of an athlete or non-athlete to come from a ZIP code in the upper percentiles of the national income distribution. Percentages are weighted by the total undergraduate enrollment of each institution in the respective tier. See the notes to Table 1 for a description of the selectivity tiers.

Figure 3: Percentage of Athletes from ZIP Codes in Upper Income Percentiles by Sport, Across All Selectivity Tiers

Source.—Authors' calculations from NCAA roster data linked to high school ZIP code characteristics and university characteristics.

Notes.—This figure plots the likelihood of an athlete to come from a ZIP code in the upper percentiles of the national income distribution conditional on appearing on the roster for the listed sport. For expositional ease, we include only a subset of all possible sports.

Figure 4: Rates of Athletes Originating from Top 5% of Income by Selectivity Tier and Sport Group



Source.—Authors' calculations from NCAA roster data linked to high school ZIP code characteristics and university characteristics.

Notes.—This figure plots the likelihood of an athlete in the given sport groups to come from a ZIP code in the top 5% of the national income distribution, separately by selectivity tier.

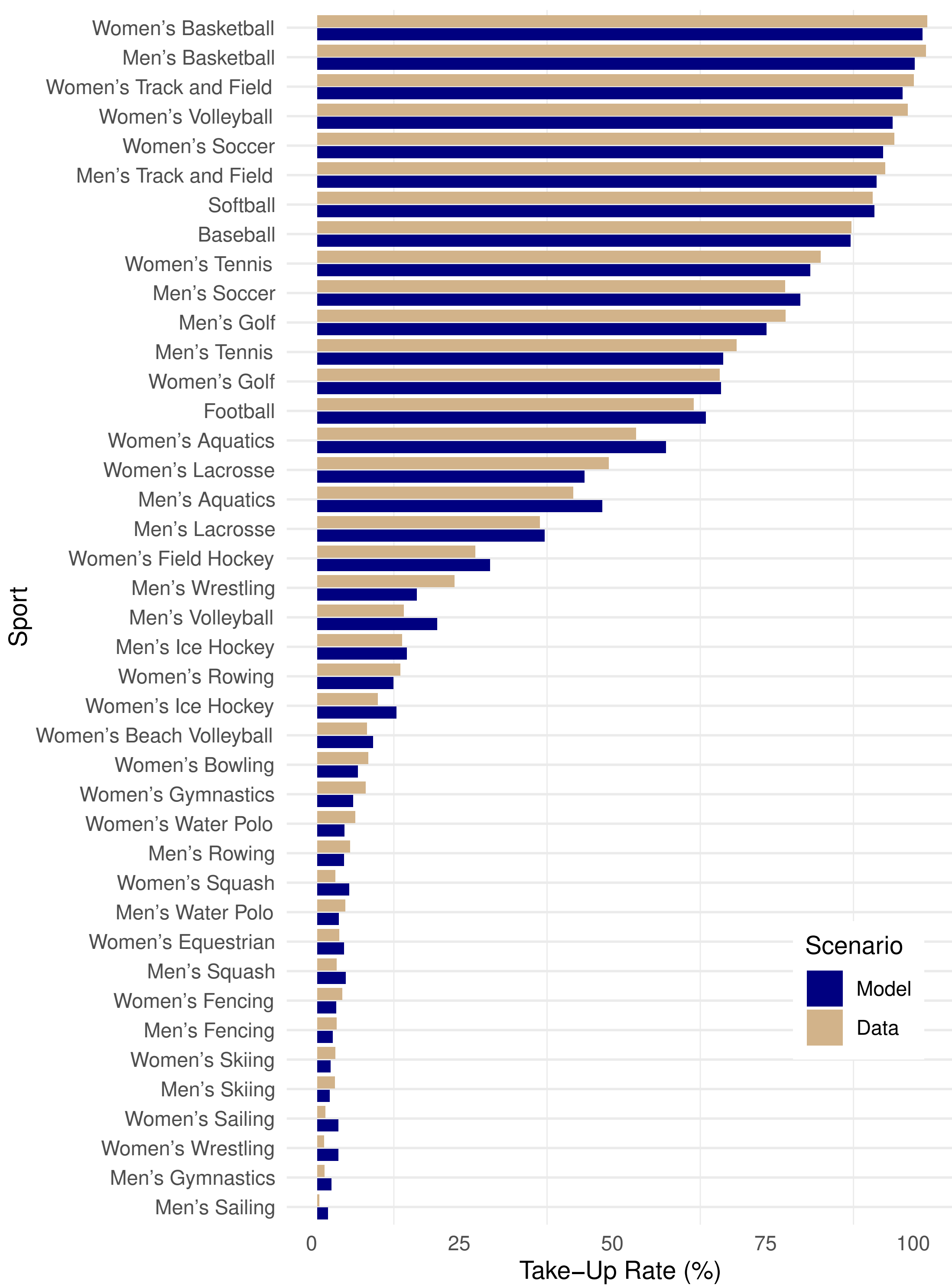
Table 2: Decomposition of Top 5% Income Gaps Across Selectivity Tiers

	Observed...		Percentage of gap due to...	
Selectivity Tier	Rate (%)	Gap (pp)	Composition	Sorting
Ivy Plus	45.2	27.7	21.0	79.0
Elite LAC	44.7	27.3	12.6	87.4
Other Elite Private	34.8	17.3	10.7	89.3
Public Flagship	24.6	7.2	12.9	87.1
Mid-tier Private	13.3	-4.2	1.9	98.1
Mid-tier Public	11.2	-6.2	24.5	75.5
Other Non-selective	9.9	-7.6	8.6	91.4

Sources.—Authors' calculations from NCAA roster data and IRS ZIP-code income data.

Notes.—This table decomposes the gap in athlete SES (measured by rate of Top 5% ZIP-code income) between each selectivity tier and the roster-size-weighted average across all tiers. The *composition effect* shows what the gap would be if all tiers had the same sport portfolios (weighted by athletes) but tier-specific within-sport athlete SES remained unchanged. The *sorting effect* shows what the gap would be if all tiers had tier-specific sport portfolios but within-sport athlete SES was equalized to the pooled mean. Percentages show each effect's contribution to the total observed gap and sum to 100%. Percentages may not sum exactly to 100% due to rounding.

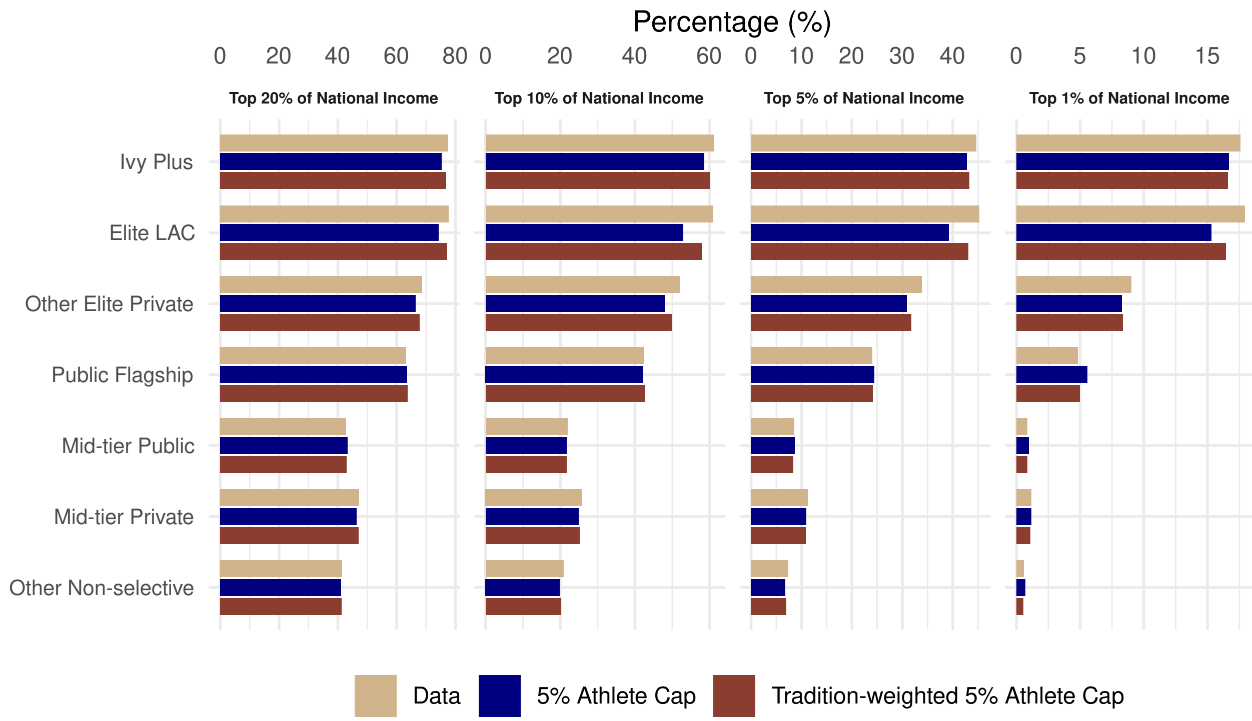
Figure 5: Model Fit of Untargeted Moments



Source.—Authors' calculations from comparing data on university sport offerings with predicted probabilities of sport offerings.

Notes.—This figure plots the actual share of universities offering each sport against the model-predicted shares. Sample includes all universities in our sample.

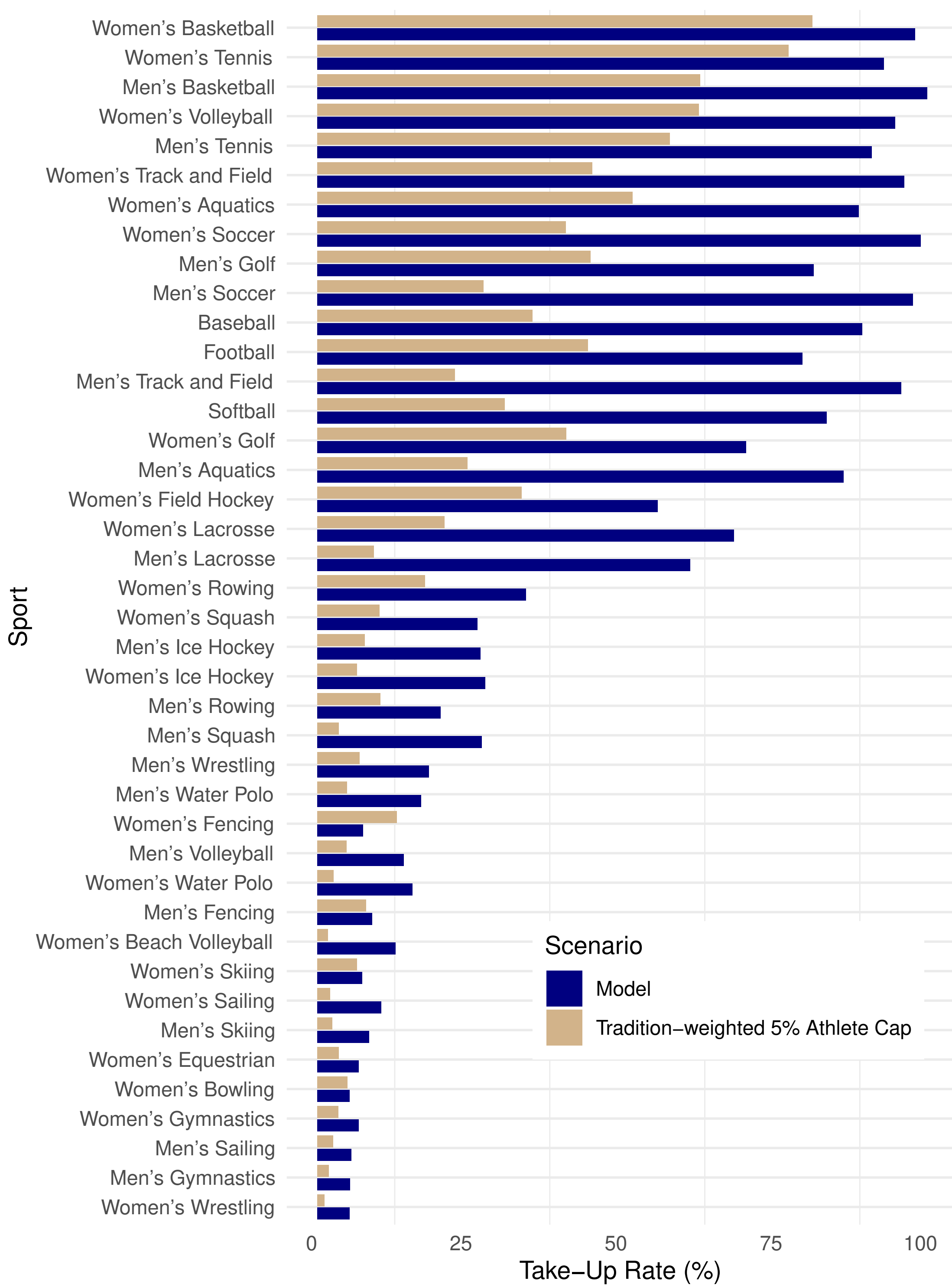
Figure 6: Counterfactual Representation of Athletes in Upper Tails of Income Distribution, by Selectivity Tier



Source.—Authors' calculations from comparing data on predicted probabilities of sport offerings in different scenarios.

Notes.—This figure plots the percentage of athletes in a university's chosen bundle to come from a ZIP code in the upper percentiles of the national income distribution. Percentages are *not* weighted by the total undergraduate enrollment of each institution in the respective tier, but weighting makes little differences to the depicted rates. The rates in this figure differ slightly with those of Figure 2 due to the current figure's focus on the bundle as the unit of analysis (as opposed to individual athletes in Figure 2).

Figure 7: Counterfactual Sport Take-up Rates at Elite Universities



Source.—Authors' calculations from comparing data on predicted probabilities of sport offerings in different scenarios.

Notes.—This figure plots the model-predicted share of universities offering each sport against the counterfactual predicted shares. Sample includes all universities in the Ivy Plus, Elite LAC, and Other Elite Private tiers.

Table 3: Athletic Enrollment Policy Analysis: Current Status and Reform Scenarios

Selectivity Tier	Total Enrollment	Athletes	Athlete Share (%)	Non-Athletes	Athletes Eliminated	Required Expansion (%)	%Δ Non-Athlete Seats
Panel A. Current Athletic Enrollment Patterns							
Ivy Plus	88,970	11,622	13.1	77,348	—	—	—
Elite LAC	40,607	12,481	30.7	28,126	—	—	—
Other Elite Private	466,791	46,456	10.0	420,335	—	—	—
Total (Elite Tiers)	596,368	70,559	11.8	525,809	—	—	—
Panel B. Meet Athlete Cap with Reduced Athlete Enrollment							
Ivy Plus	88,970	4,192	4.7	84,778	7,430	—	+9.6
Elite LAC	40,607	2,349	5.8	38,258	10,132	—	+36
Other Elite Private	466,791	21,060	4.5	445,731	25,396	—	+6
Total (Elite Tiers)	596,368	27,601	4.6	568,767	42,958	—	+8.2
Panel C. Meet Athlete Cap with Expanded Non-Athlete Enrollment							
Ivy Plus	246,641	11,622	4.7	235,019	—	+177.2	+203.8
Elite LAC	215,763	12,481	5.8	203,282	—	+431.3	+622.8
Other Elite Private	1,029,669	46,456	4.5	983,213	—	+120.6	+133.9
Total (Elite Tiers)	1,492,073	70,559	4.7	1,421,514	—	+150.2	+170.3

Source.—Authors' calculations using IPEDS data and athletic enrollment cap simulation.

Notes.—Panel A shows status-quo enrollment at elite institutions where athlete shares of enrollment exceed those at less selective institutions. Panel B shows outcomes from capping athletic enrollment at $\approx 5\%$ (comparable to less selective institutions). Panel C shows enrollment expansion required to maintain current athlete numbers while achieving target athlete shares. Athletes Eliminated shows positions cut in Panel B. Required Expansion shows percentage enrollment increase needed in Panel C. %Δ Non-Athlete shows the percentage change in non-athlete enrollment relative to Panel A.

Online Appendix

A Sample Selection, Data Sources and Processing Steps

A.1 Sample selection

Table A.1 presents our sequential sample selection process. We begin with the universe of 1,084 NCAA Division I, II, and III institutions active in Academic Year (AY) 2019–2020. At each step, we exclude institutions missing data from the specified source, conditional on having complete data from all previous sources.

We first exclude 40 institutions with missing roster data. These are primarily smaller institutions without comprehensive athletic websites or those with rosters that could not be reliably scraped. From the remaining 1,044 institutions, we exclude 2 institutions with missing data on athlete socioeconomic status. We further exclude 27 institutions missing EADA financial data. Institutions exempt from EADA reporting include military academies (e.g. U.S. Air Force Academy, U.S. Coast Guard Academy) and some small private colleges that do not receive federal financial aid. Finally, we exclude 2 institutions missing from the Chetty et al. (2020) mobility data. Our final analysis sample contains 1,013 institutions, representing 93.5% coverage of all NCAA institutions.

Table A.1: Sample Selection

Selection criterion	Institutions dropped	Resultant universities
All NCAA institutions (as of 2019–2020 AY)	0	1,084
Drop missing roster data	40	1,044
Drop missing athlete SES data	2	1,042
Drop missing EADA financial data	27	1,015
Drop missing Chetty mobility data	2	1,013

Source.—NCAA Sports Sponsorship and Participation (2020), university athletic department websites, EADA database, IPEDS, and Chetty et al. (2020).

A.2 Data sources and processing

We now detail the different data sources we use, as well as specific data processing steps. A detailed summary of each data source is included in Online Appendix Table A.2 at the end of this section.

We first outline the aggregation of specific sports into more tractable categories. We then detail the process we follow for utilizing AI assistance in collecting data on sport sponsorship history for each university. Finally, we discuss how we adjust roster sizes to ensure that scraped roster websites and CollegeFactual.com sources on total athletic enrollment match.

A.2.1 Sport Aggregation and Categorization

Aggregation from NCAA administrative data The NCAA administrative data contains 77 distinct sports categories. We aggregate these into 65 combined sport categories using a crosswalk that maps related sports into common groups.

- **Track and Field**: We combine Men's and Women's Indoor Track, Outdoor Track, and Cross Country into "Men's Track and Field" and "Women's Track and Field" categories.

- **Aquatics**: We combine Swimming, Diving, and Swimming & Diving into "Men's Aquatics" and "Women's Aquatics" categories.

This aggregation is necessary because many institutions treat these related sports as part of a single program with shared coaching staffs, facilities, and budgets.

The crosswalk is also necessary because of different levels of aggregation in sports across our different datasets (NCAA administrative data, EADA survey data, and university athletics websites). For example, some sources treat swimming and diving as separate sports, while others combine the two.

Final sport selection We further reduce these 65 sports to 41 sports for our final analysis. This reduction involves the following steps.

- Removing extremely rare sports with minimal participation across universities

- Retaining only sports with sufficient observations to enable reliable parameter estimation
- Excluding sports that fewer than a threshold number of institutions offer

The following specific sports remain:

- Men's & Women's Basketball
- Men's & Women's Soccer
- Men's & Women's Track and Field
- Men's & Women's Tennis
- Men's & Women's Golf
- Men's & Women's Aquatics (Swimming & Diving)
- Baseball & Softball
- Men's & Women's Volleyball
- Men's & Women's Lacrosse
- Men's & Women's Wrestling
- Football
- Men's & Women's Ice Hockey
- Men's & Women's Rowing
- Men's & Women's Fencing
- Men's & Women's Skiing
- Men's & Women's Squash
- Men's & Women's Gymnastics

- Men's & Women's Water Polo
- Men's & Women's Sailing
- Women's Field Hockey
- Women's Bowling
- Women's Beach Volleyball
- Women's Equestrian

For convenience, we also categorize sports according to tiers of Standard (i.e., typically offered at most high schools), Regional (i.e., typically offered only regionally at the high school level), and Niche (i.e., rarely offered at the high school level and typically requiring specialized facilities or training), as follows:

- **Football**: Football
- **Women's Standard**: Basketball, Softball, Volleyball, Soccer, Track and Field, Tennis, Golf, Aquatics
- **Men's Standard**: Basketball, Baseball, Soccer, Track and Field, Tennis, Golf, Aquatics
- **Women's Regional**: Lacrosse, Field Hockey, Ice Hockey, Beach Volleyball, Rowing
- **Men's Regional**: Lacrosse, Wrestling, Volleyball, Ice Hockey
- **Women's Niche**: Bowling, Gymnastics, Water Polo, Fencing, Equestrian, Squash, Skiing, Wrestling, Sailing
- **Men's Niche**: Rowing, Water Polo, Fencing, Squash, Skiing, Gymnastics, Sailing

We categorize women's rowing as Regional due to its popularity at the college level which is on par with other women's regional sports.

A.2.2 Collecting historical data on sport offerings

Our counterfactual simulations rely on leveraging variation in universities' historical precedence of sports offerings. To construct this historical dataset, we implement a novel data collection approach using Perplexity's artificial intelligence-assisted research to systematically gather information on university athletic program evolution.

AI-assisted data collection methodology We employ Perplexity's "Deep Research" function to conduct comprehensive searches across hundreds of web sources for each university in our sample. This approach systematically queries multiple databases, university archives, athletic department websites, conference records, and historical documents to compile information on:

- University establishment dates and institutional history
- NCAA membership join dates and divisional classification changes over time
- NCAA Division progression (transitions between Divisions I, II, and III)
- Sport-specific program inception dates for all 41 sports in our analysis
- Program discontinuation and reinstatement dates
- Conference membership history and athletic affiliation changes (e.g. moving from NAIA to NCAA)

For each university-sport combination, the AI system searches through university archives, conference historical records, NCAA databases, news articles, and official athletic department publications. This methodology allows us to systematically collect historical data that would be prohibitively expensive to gather through manual research across our sample of 1,000+ institutions and 41 sport categories.

Imputation of sport inception years Our comprehensive dataset contains 44,321 observations at the university-sport group level (1,081 universities $\times$ 41 sports). While our data collection yields sport inception dates for 13,146 observations (29.7% of all possible university-sport combinations), the remaining 4,271 university-sport pairs where the sport is currently offered require imputation of inception years. Note that 26,904 observations represent university-sport combinations where the sport is never offered and thus do not require inception date imputation.

We implement a hierarchical imputation strategy that leverages institution- and sport-specific patterns, proceeding through increasingly general matching criteria until all missing values are assigned. We categorize universities into establishment eras (Pre-1850, 1850-1899, 1900-1919, 1920-1939, 1940-1959, 1960-1979, 1980-1999, and 2000+) and enrollment size categories (Very Small [<1K], Small [1K-3K], Medium [3K-8K], Large [8K-15K], and Very Large [15K+]).

The imputation hierarchy proceeds as follows, with the number of observations imputed at each level shown in parentheses:

- **Level 1: Division-Era-Sport-Enrollment-Conference (1,061 observations)** We impute inception years using median values from institutions within the same NCAA division, establishment era, offering the same sport, same enrollment size category, and same athletic conference, providing the most restrictive matching criteria.

- **Level 2: Division-Era-Sport-Enrollment (2,517 observations)** When Level 1 cells contain insufficient observations, we remove the conference restriction while maintaining division, era, sport, and enrollment size category matching criteria.

- **Level 3: Division-Era-Sport (560 observations)** For remaining missing values, we use median inception dates among institutions in the same NCAA division and establishment era offering the same sport, removing enrollment size restrictions.

- **Level 4: Division-Sport-Enrollment (47 observations)** We impute using median inception dates for institutions in the same NCAA division offering the same sport within the same enrollment size category, pooling across establishment eras.

- **Level 5: Division-Sport (62 observations)** We use median inception dates for the same sport within the university's current NCAA division, removing both era and enrollment restrictions.

- **Level 6: Era-Sport-Enrollment (21 observations)** We impute using median inception dates for institutions from the same establishment era within the same enrollment size category offering the same sport, pooling across NCAA divisions.
- **Level 7: Era-Sport (0 observations)** We use median inception dates for institutions from the same establishment era offering the same sport, removing enrollment size restrictions. However, our data contain no cases applying to this approach.
- **Level 8: Sport-Enrollment (1 observation)** Imputation based solely on sport type and enrollment size category, pooling across divisions and eras.
- **Level 9: Sport-Only (2 observations)** For the final cases, we use the overall median inception date for the specific sport across all institutions.

The imputed inception years range from 1837 to 2027, with a median inception year of 1978—the year that NCAA Divisions I-A and I-AA were created. Our hierarchical imputation approach ensures that imputed values reflect the most relevant institutional and temporal context available while maintaining sufficient sample sizes for reliable median calculations within each grouping category.

Construction of historical sport variables Using the cleaned and imputed historical data, we construct two key variables for our analysis:

- Sport program age: Years since sport inception at each institution as of 2019–20. This variable is only defined for sports that have ever been offered.
- Program discontinuation history: The data allow us to identify programs that were offered historically but subsequently discontinued, which we can use in our counterfactual predictions.

A.2.3 Athletic roster size adjustment

To ensure consistency between our sport-specific roster counts and institution-level athletic participation totals, we implement an iterative roster size adjustment procedure. This process addresses discrepancies between the sum of sport-specific roster sizes (taken from the scraped university athletic websites) and the total number of athletes reported in administrative data (collected via the CollegeFactual.com website).

Handling missing roster data For sports where institution-specific roster sizes are missing, we implement a hierarchical imputation approach:

1. First, we use NCAA average roster sizes by division, subdivision, and sport taken directly from the NCAA's Sports Sponsorship and Participation Rates Report of December 1, 2021.

2. These reference values are used to impute missing roster counts while preserving the sport-division-subdivision specific patterns in team size.

Iterative adjustment algorithm Our adjustment procedure consists of the following steps:

1. For each institution, we merge the NCAA average roster sizes for each sport based on division and subdivision classifications as described above.

2. We calculate the ratio between the institution's reported total athletes (from CollegeFactual) and the sum of sport-specific roster counts (from scraped university athletic websites)

3. We apply adjustment factors to bring the sum of sport-specific roster counts in line with reported institution totals:

$$\text{adjustment factor} = \frac{\text{Total reported athletes}}{\text{Sum of sport-specific roster counts}}$$

4. To avoid extreme adjustments, we bound the adjustment factors between 0.67 and 1.5.

5. We iteratively apply these adjustment factors (up to three iterations) to achieve convergence.

This approach ensures that our sport-specific roster size estimates are consistent with institution-level totals while preserving the relative size differences across sports within divisions and subdivisions.

Table A.2: Data Sources and Variable Descriptions

Dataset	ID Variables	Source	Description
Athletic Rosters	University ID, Sport ID, Athlete Name	University Athletic Department Websites	Individual-level data on NCAA athletes including name, sport, academic year, hometown, and high school for 2019-20 academic year
CCD	NCES School ID, ZIP Code	National Center for Education Statistics	Public secondary school characteristics including location, enrollment, student demographics, and staffing for 2017-18 academic year
PSS	NCES School ID, ZIP Code	National Center for Education Statistics	Private secondary school characteristics including location, enrollment, student demographics, religious affiliation for 2017-18 academic year
IRS SOI	ZIP Code	Internal Revenue Service	ZIP code level tax return data including income distribution, number of returns, and types of income for tax year 2017
ACS	ZCTA	U.S. Census Bureau	ZIP Code Tabulation Area (ZCTA) level demographic and socioeconomic characteristics from 2013-2017 5-year estimates
IPEDS	University ID	National Center for Education Statistics	University-level enrollment, financial, and institutional characteristics for 2019-20 academic year
NCAA Directory	University ID	NCAA	Sports offered and athletic division/conference membership for 2019-20 academic year
EADA	University ID, Sport ID	U.S. Department of Education Office of Postsecondary Education	Sport-specific revenues, expenses, participation, and coaching staff data at university level for 2019-20 academic year

Continued on next page

Table A.2 – continued from previous page

Dataset	ID Variables	Source	Description
Opportunity Atlas	University ID	Opportunity Insights	University-level data on student and parent income distributions, student outcomes, and mobility rates based on 1980-1991 birth cohorts
Championship History	University ID, Sport ID	NCAA Championships Summary dated June 22, 2020	Running total of NCAA team and individual championships for every NCAA sponsored event
Sport Sponsorship History	University ID, Sport ID	`perplexity.ai` Deep Research tool	Years of sponsorship for every sport officially sponsored by each university at the intercollegiate varsity level

Notes: CCD refers to Common Core of Data; PSS refers to Private School Survey; IRS SOI refers to Statistics of Income; ACS refers to American Community Survey; IPEDS refers to Integrated Postsecondary Education Data System; EADA refers to Equity in Athletics Data Analysis. ZCTA refers to ZIP Code Tabulation Area, which is the Census Bureau's statistical equivalent of ZIP codes. Opportunity Atlas data were constructed using population-level federal income tax returns and other administrative data as described in Chetty et al. (2020).

B Data Preparation for Structural Estimation

This appendix section details the steps we take to prepare our data for structural estimation. There are three separate issues that we must resolve. First, we need to calculate the observable characteristics of bundles that are not chosen. Second, we need to choose a feasible sample of bundles on which to estimate the model. Third, we need to follow a slightly different process for generating bundles for our counterfactual simulations.

Section B.1 details the steps we take for obtaining the characteristics of non-chosen bundles. Section B.2 describes how we obtain feasible bundles for each university to include in estimation of our structural model. Section B.3 details our process for constructing a data set of counterfactual bundles which we use to characterize university responses to hypothetical policies.

B.1 Computing Expected Characteristics of Non-Chosen Sport Bundles

A key challenge in any discrete choice model is to characterize the attributes of non-chosen alternatives. For example, in an occupational choice model, the researcher may need to predict what the worker's wage would be in several non-chosen occupations in order to estimate how sensitive workers' choices are to wage differences across occupations.

We describe how we use two-way fixed effects models to predict, e.g., what a university's non-chosen sports might look like (in terms of revenues, expenses, roster size, and athlete SES). These models rely on the assumption that institution-level unobservables would have the same effect on the attributes of non-chosen sports as they do for chosen sports.

B.1.1 Two-Way Fixed Effects Models

For each sport characteristic Y (e.g. log expenses, log revenues, log number of athletes, percentage of athletes from domestic private high schools, etc.), we estimate two sets of two-way fixed effects models. For percentage variables, we convert the percentages to real numbers using the log odds transformation, $Y = \log(p/(1-p))$ where p is the percentage expressed as a probability.

1. **First Step**: Sport $\times$ Conference Fixed Effects, where i indexes universities, s indexes sports, and c indexes athletic conferences.

$$Y_{i(c)s} = \alpha_i + \gamma_{sc} + \varepsilon_{i(c)s} \tag{B.1}$$

   where α_i represents institution fixed effects, γ_{sc} represents sport-by-conference fixed effects, and $\varepsilon_{i(c)s}$ is an error term.

2. **Second Step**: Institution and Sport Fixed Effects

$$Y_{is} = \alpha_i + \phi_s + \varepsilon_{is} \tag{B.2}$$

   where α_i represents institution fixed effects, ϕ_s represents sport fixed effects, and ε_{is} is an error term.

We apply appropriate inverse transformations to recover the original scale for variables expressed in logs or log odds units.

B.1.2 Merged Prediction Approach

To maximize prediction accuracy and data coverage, we implement a two-step prediction approach:

1. Generate predictions using the Sport $\times$ Conference models, which capture conference-specific patterns in sport characteristics

2. Generate predictions using the standard TWFE (institution + sport) models

3. Use the Sport $\times$ Conference predictions when available

4. Fill in any missing values from the Sport $\times$ Conference predictions with the standard TWFE predictions

This approach allows us to leverage the precision of conference-specific estimates while ensuring complete coverage across all institution-sport combinations.

B.2 Bundle Generation and Feasibility Analysis

In this subsection, we describe how we create a common set of $\approx$ 10,000 bundles for estimation from which we sample 250 feasible bundles for each university. We include every unique bundle observed in the data as part of the 10,000 and supplement these with constructed bundles that result in sufficient identifying variation for estimation of our model's parameters.

B.2.1 Observed Bundle Identification

We create a university-sport matrix by identifying all unique combinations of sports chosen by universities in our sample. Each row represents a university, and each column represents a sport, with binary indicators (0/1) showing whether the university offers that sport. This process yields a unique "bundle ID" for each university—a string of 41 binary digits representing their chosen set of sports.

B.2.2 Synthetic Bundle Generation

To create a comprehensive choice set for our discrete choice model, we generate synthetic bundles that universities could feasibly adopt but that no university in our data has chosen. To structure this process, we expand upon several archetypal bundles of sports:

1. **Saturated bundle**: We create a hypothetical option where the university offers all sports in our dataset.

2. **Women's saturated bundle**: We design a bundle containing all women's sports but no men's sports.

3. **Standard-heavy bundles**: We generate multiple variations ($n = 500$) of bundles containing standard men's and women's sports, with low probabilities of including regional or niche sports.

4. **Standard and Regional bundles**: We create several subtypes including:

    - Balanced standard and regional bundles ($n = 500$): all standard and regional sports for both men and women, with low probabilities of including niche sports

- Women's standard and regional bundles ($n = 500$): all women's standard and regional sports, with minimal men's sports
- Men's heavy standard and regional bundles ($n = 500$): all men's standard and regional sports, with moderate probabilities of women's sports

5. **Standard, Regional, and Niche bundles**: We develop several subtypes including:
   - Balanced bundles ($n = 500$): include core sports like basketball, baseball/softball, soccer, volleyball, and track and field for both genders, with high probabilities of other standard and regional sports and moderate probabilities of niche sports
   - Women's focused bundles ($n = 500$): prioritize core women's sports with high probabilities of additional women's sports and minimal men's sports
   - Men's heavy bundles ($n = 500$): emphasize core men's sports with high probabilities of additional men's sports and moderate probabilities of women's sports
6. **Symmetric bundles** ($n = 500$ for each type): We construct bundles that maintain gender symmetry, always including both men's and women's versions of the same sport (e.g., Men's Basketball and Women's Basketball), with three subtypes:
   - Symmetric standard bundles
   - Symmetric regional bundles
   - Symmetric niche bundles
7. **Asymmetric bundles** ($n = 500$): We create bundles that intentionally break gender symmetry by including only one gender's version of a sport for multiple sports, with biases that sometimes favor men's sports and sometimes favor women's sports. These bundles provide key identifying variation for the facility complementarity parameters in the utility function.
8. **Women's-only bundles** ($n = 500$): We design bundles containing only women's sports with varying probabilities based on sport categories.
9. **Random bundles**: We generate three subtypes with different gender balances:

- Women-focused random bundles ($n = 500$): assign high probabilities for women's sports, low probabilities for men's sports
- Gender-balanced random bundles ($n = 500$): use equal probabilities for men's and women's sports
- Men-focused random bundles ($n = 500$): set high probabilities for men's sports, low probabilities for women's sports

10. **Football variations**: For a subset of bundles, we create alternative versions with football status flipped (added or removed).

We assign each synthetic bundle a unique bundle ID and combine them with observed bundles to form the complete choice set. We remove duplicate bundles, retaining only the first occurrence of each unique bundle ID. Our final dataset contains approximately 10,000 unique bundles per university, including both observed and synthetic bundles.

B.2.3 Bundle Characteristic Calculation

For each generated bundle, we calculate key characteristics:

- **Financial metrics**: Total expenses and total revenues
- **Athletic participation**: Total athletes, gender distributions, and sport counts
- **Socioeconomic indicators**: Averages of athlete characteristics (e.g., percentage from domestic private high schools, top-tier income levels, etc.), weighted by roster size

B.2.4 Bundle Feasibility Constraints

We implement three tiers of institutional constraints to determine whether a synthetic bundle is in fact feasible for a given university:

1. **NCAA Division Requirements**:
   - Division I: Minimum number of total sports and minimum number of sports for each gender

- FCS and FBS Subdivisions: Larger minimum number of total sports; must offer football

2. **Title IX Compliance**: Ratio of men's to women's sports closely matches ratio of male to female undergraduates

3. **Geographical constraints**: Prohibit skiing in states without suitable terrain or climate

Bundles meeting all constraints are marked as feasible and included in the university's choice set for subsequent estimation. This results in an unbalanced panel from which we sample 250 remaining bundles for each university. Some universities end up with fewer than 250 bundles in estimation.

B.3 Counterfactual Bundle Generation

For our counterfactual exercises, we follow a similar procedure as in Section B.2, except we only include bundles that are feasible under the corresponding counterfactual scenario.

For our counterfactual policy simulations, we implement a forward-building algorithm that constructs feasible sport bundles under the given policy constraints. The algorithm incorporates data on historical sport offerings and pays attention to Title IX gender balancing constraints.

The two policy scenarios we consider are: a 5% enrollment cap at elite universities that limits total athletes to 5% of undergraduate enrollment; and a similar policy that prioritizes historically significant sports in the bundle generation algorithm. For each scenario, we generate ≈ 350 unique bundles per university. We incorporate institution-specific constraints such as geographic restrictions (e.g., prohibiting skiing programs in unsuitable climates) and single-sex institutional requirements.

Each candidate bundle undergoes feasibility testing against NCAA divisional requirements, Title IX compliance thresholds, and geographic constraints, as well as the 5% enrollment cap. We discard bundles that are duplicates or that fail to meet all of the constraints.

To generate variation in bundle sizes, we randomly scale each university's athlete allotment by a factor drawn from a uniform distribution between 0.3 and 1.5. This prevents all generated bundles from clustering at exactly the policy cap and ensures sufficient variation in bundle characteristics for prediction.

C Supporting Figures and Tables

Table C.1: Summary Statistics for High Schools, by School Sector and Athlete-Sending Status

	Schools that did not send an athlete					Schools that sent an athlete				
	N	Mean	SD	Min	Max	N	Mean	SD	Min	Max
A. Private High Schools										
Enrollment	6,812	125.79	268.11	1.00	7701.00	6,095	312.21	358.33	1.00	4789.00
Share white	6,812	65.08	33.30	0.00	100.00	6,095	67.79	28.21	0.00	100.00
Share African American	6,812	15.87	25.27	0.00	100.00	6,095	12.46	20.12	0.00	100.00
Share Hispanic American	6,812	11.42	19.43	0.00	100.00	6,095	9.62	15.82	0.00	100.00
Share Asian American	6,811	3.83	11.34	0.00	100.00	6,095	5.56	10.20	0.00	100.00
Student to teacher ratio	6,812	10.06	30.16	0.22	2073.68	6,095	10.62	25.87	0.20	1500.00
Top 100 private high school	6,812	0.00	0.00	0.00	0.00	6,095	1.62	12.64	0.00	100.00
Catholic	6,812	4.54	20.81	0.00	100.00	6,095	20.53	40.39	0.00	100.00
Other religious	6,812	60.79	48.83	0.00	100.00	6,095	51.78	49.97	0.00	100.00
Non-religious	6,812	34.67	47.60	0.00	100.00	6,095	27.69	44.75	0.00	100.00
ZIP code average salary	6,134	56.83	31.96	11.49	436.53	5,843	67.68	47.25	17.79	974.30
Top 20% ZIP income	6,134	27.24	44.52	0.00	100.00	5,843	38.68	48.71	0.00	100.00
Top 10% ZIP income	6,134	14.79	35.50	0.00	100.00	5,843	24.17	42.81	0.00	100.00
Top 5% ZIP income	6,134	7.42	26.21	0.00	100.00	5,843	14.15	34.86	0.00	100.00
Top 1% ZIP income	6,134	1.68	12.85	0.00	100.00	5,843	4.00	19.61	0.00	100.00
Pct ZIP Advanced Degree	6,523	11.42	8.63	0.00	75.34	6,082	14.38	10.55	0.00	61.59
Pct ZIP Bachelor's Degree	6,523	18.25	8.99	0.00	100.00	6,082	20.92	9.44	0.00	100.00
Large city	1,981	16.51	37.13	0.00	100.00	674	16.02	36.71	0.00	100.00
Small or midsize city	1,981	15.90	36.58	0.00	100.00	674	14.24	34.98	0.00	100.00
Suburb	1,981	31.50	46.46	0.00	100.00	674	34.27	47.50	0.00	100.00
Town	1,981	10.60	30.79	0.00	100.00	674	13.35	34.04	0.00	100.00
Rural	1,981	25.49	43.59	0.00	100.00	674	22.11	41.53	0.00	100.00
B. Public High Schools										
Enrollment	9,218	231.26	391.99	1.00	14319.00	15,998	896.01	780.15	1.00	12033.00
Share white	9,218	47.06	34.35	0.00	100.00	15,998	58.93	32.65	0.00	100.00
Share African American	9,218	16.59	24.85	0.00	100.00	15,998	14.20	22.78	0.00	100.00
Share Hispanic American	9,218	26.18	29.58	0.00	100.00	15,998	18.70	23.95	0.00	100.00
Share Asian American	9,218	1.78	4.93	0.00	100.00	15,998	3.38	7.30	0.00	99.15
Student to teacher ratio	8,346	15.25	17.56	0.00	665.00	15,810	16.29	8.42	0.00	434.00
Share free or reduced price lunch	8,217	60.24	26.40	0.00	100.00	14,888	47.43	24.69	0.00	100.00
ZIP code average salary	8,611	49.39	21.71	17.49	298.89	15,590	55.21	29.43	15.68	588.73
Top 20% ZIP income	8,611	17.87	38.31	0.00	100.00	15,590	24.55	43.04	0.00	100.00
Top 10% ZIP income	8,611	8.00	27.13	0.00	100.00	15,590	12.25	32.78	0.00	100.00
Top 5% ZIP income	8,611	3.04	17.18	0.00	100.00	15,590	5.79	23.36	0.00	100.00
Top 1% ZIP income	8,611	0.50	7.05	0.00	100.00	15,590	1.25	11.11	0.00	100.00
Pct ZIP Advanced Degree	9,188	8.59	6.84	0.00	66.00	15,968	10.08	7.92	0.00	62.50
Pct ZIP Bachelor's Degree	9,188	15.46	8.20	0.00	100.00	15,968	16.98	8.49	0.00	100.00
Large city	9,218	17.56	38.05	0.00	100.00	15,998	11.33	31.69	0.00	100.00
Small or midsize city	9,218	13.60	34.28	0.00	100.00	15,998	9.95	29.94	0.00	100.00
Suburb	9,218	23.24	42.24	0.00	100.00	15,998	27.60	44.70	0.00	100.00
Town	9,218	13.00	33.63	0.00	100.00	15,998	15.32	36.02	0.00	100.00
Rural	9,218	32.60	46.88	0.00	100.00	15,998	35.80	47.94	0.00	100.00

Sources.—Authors' calculations from the following datasets: NCES Private School Universe Survey (PSS), NCES Common Core of Data (CCD), and Internal Revenue Service zip-code level income data.

Notes.—Sample includes all public and private high schools in the PSS and CCD with positive enrollment, valid ZIP codes, and complete demographic data.

Table C.2: Decomposition of SES Gaps Across Selectivity Tiers: All Measures

	Observed...		Percentage of gap due to...	
Selectivity Tier	Rate (%)	Gap (pp)	Composition	Sorting
A. Parent Income: Top 20%				
Ivy Plus	76.4	24.7	22.1	77.9
Elite LAC	75.0	23.3	17.1	82.9
Other Elite Private	68.7	17.1	12.6	87.4
Public Flagship	62.6	11.0	8.6	91.4
Mid-tier Private	48.0	-3.7	-5.6	105.6
Mid-tier Public	44.8	-6.9	30.0	70.0
Other Non-selective	43.7	-8.0	5.1	94.9
B. Parent Income: Top 10%				
Ivy Plus	61.9	29.8	20.9	79.1
Elite LAC	60.1	28.0	14.4	85.6
Other Elite Private	52.2	20.0	11.2	88.8
Public Flagship	43.1	11.0	9.3	90.7
Mid-tier Private	27.4	-4.7	-1.7	101.7
Mid-tier Public	24.6	-7.5	26.7	73.3
Other Non-selective	23.6	-8.5	7.4	92.6
C. Parent Income: Top 5%				
Ivy Plus	45.2	27.7	21.0	79.0
Elite LAC	44.7	27.3	12.6	87.4
Other Elite Private	34.8	17.3	10.7	89.3
Public Flagship	24.6	7.2	12.9	87.1
Mid-tier Private	13.3	-4.2	1.9	98.1
Mid-tier Public	11.2	-6.2	24.5	75.5
Other Non-selective	9.9	-7.6	8.6	91.4
D. Parent Income: Top 1%				
Ivy Plus	19.5	15.3	18.8	81.2
Elite LAC	18.2	14.0	11.4	88.6
Other Elite Private	11.1	6.9	11.7	88.3
Public Flagship	6.0	1.8	22.2	77.8
Mid-tier Private	2.1	-2.1	4.8	95.2
Mid-tier Public	1.8	-2.4	25.2	74.8
Other Non-selective	1.4	-2.8	11.8	88.2
E. Private High School Attendance				
Elite LAC	39.6	21.2	6.6	93.4
Ivy Plus	36.0	17.6	12.3	87.7
Other Elite Private	27.9	9.5	5.5	94.5
Public Flagship	18.6	0.2	-89.3	189.3
Mid-tier Private	17.3	-1.1	-5.2	105.2
Other Non-selective	16.0	-2.4	13.5	86.5
Mid-tier Public	13.3	-5.1	9.7	90.3
F. International High School Attendance				
Ivy Plus	10.3	2.8	70.0	30.0
Public Flagship	10.3	2.8	1.3	98.7
Mid-tier Public	8.8	1.2	-73.6	173.6
Other Elite Private	6.7	-0.8	-98.8	198.8
Other Non-selective	6.6	-0.9	-33.8	133.8
Mid-tier Private	6.5	-1.0	-15.3	115.3
Elite LAC	4.5	-3.0	-63.8	163.8

Sources.—Authors' calculations from NCAA roster data, IRS ZIP-code income data, and NCES school data.

Notes.—This table decomposes the gap in athlete SES across multiple measures between each selectivity tier and the roster-size-weighted average across all tiers. For income measures (Panels A–D), athlete SES is defined by the share of athletes from ZIP codes in the specified percentile of the national income distribution. Panel E shows the share of athletes who attended private high schools. Panel F shows the share who attended international high schools. The *composition effect* shows what the gap would be if all tiers had the same sport portfolios (weighted by roster size) but tier-specific within-sport athlete SES remained unchanged. The *sorting effect* shows what the gap would be if all tiers had tier-specific sport portfolios but within-sport athlete SES was equalized to the pooled mean. Percentages show each effect's contribution to the total observed gap and sum to 100%. Percentages may not sum exactly to 100% due to rounding.

Table C.3: Complete Conditional Logit Estimates, by NCAA Division

	D-I	D-II	D-III
Log athletes	-7.482***	-0.345	-0.309
	(2.062)	(2.587)	(2.763)
Log athletes × Pct all students top 10% inc	0.205**	0.216	0.098
	(0.094)	(0.179)	(0.118)
Elite Private × Log athletes	3.829		7.589***
	(3.207)		(2.841)
Public Flagship × Log athletes	8.468***		
	(2.691)		
Mid-tier Private × Log athletes	1.633	-3.308***	3.313**
	(2.062)	(1.064)	(1.465)
Other Non-selective × Log athletes		-0.667	1.257
		(2.053)	(1.897)
Profitable	-0.664	0.622	1.444**
	(0.564)	(0.432)	(0.651)
Profitable × Pct all students top 10% inc	0.025	-0.010	0.024
	(0.025)	(0.024)	(0.021)
Elite Private × Profitable	-0.764		-2.668***
	(1.097)		(0.946)
Public Flagship × Profitable	-0.617		
	(0.968)		
Mid-tier Private × Profitable	0.553	-0.140	-1.757***
	(0.650)	(0.449)	(0.642)
Other Non-selective × Profitable		0.098	-1.040
		(0.904)	(0.853)
Pct athletes int'l private HS	-0.169**	0.101*	0.467***
	(0.071)	(0.057)	(0.149)

Continued on next page

Table C.3 – continued from previous page

	D-I	D-II	D-III
Pct athletes int'l private HS × Pct all students top 10% inc	0.007**	0.005	0.001
	(0.003)	(0.003)	(0.006)
Elite Private × Pct athletes int'l private HS	0.302**		-0.333
	(0.146)		(0.244)
Public Flagship × Pct athletes int'l private HS	0.019		
	(0.106)		
Mid-tier Private × Pct athletes int'l private HS	0.108	-0.188***	-0.268
	(0.081)	(0.057)	(0.167)
Other Non-selective × Pct athletes int'l private HS		-0.142	-0.178
		(0.093)	(0.189)
Pct athletes domestic private HS	-0.132	0.073	-0.244
	(0.114)	(0.099)	(0.173)
Pct athletes domestic private HS × Pct all students top 10% inc	0.003	-0.010	0.025***
	(0.004)	(0.006)	(0.006)
Elite Private × Pct athletes domestic private HS	0.021		-0.492*
	(0.186)		(0.275)
Public Flagship × Pct athletes domestic private HS	-0.319**		
	(0.155)		
Mid-tier Private × Pct athletes domestic private HS	-0.013	-0.021	-0.262
	(0.127)	(0.105)	(0.177)
Other Non-selective × Pct athletes domestic private HS		-0.204	-0.111
		(0.202)	(0.179)
Pct athletes top 20% inc	-0.216*	-0.102	0.045
	(0.113)	(0.131)	(0.158)
Pct athletes top 20% inc × Pct all students top 10% inc	0.013**	0.007	0.002
	(0.006)	(0.008)	(0.007)
Elite Private × Pct athletes top 20% inc	-0.577**		0.020

Continued on next page

Table C.3 – continued from previous page

	D-I	D-II	D-III
	(0.275)		(0.291)
Public Flagship × Pct athletes top 20% inc	-0.153		
	(0.167)		
Mid-tier Private × Pct athletes top 20% inc	-0.107	0.053	-0.043
	(0.152)	(0.121)	(0.153)
Other Non-selective × Pct athletes top 20% inc		0.324	-0.131
		(0.230)	(0.190)
Pct athletes top 10% inc	0.375**	-0.104	-0.143
	(0.154)	(0.146)	(0.218)
Pct athletes top 10% inc × Pct all students top 10% inc	-0.011	-0.009	-0.017*
	(0.007)	(0.008)	(0.009)
Elite Private × Pct athletes top 10% inc	0.326		0.309
	(0.344)		(0.352)
Public Flagship × Pct athletes top 10% inc	0.085		
	(0.245)		
Mid-tier Private × Pct athletes top 10% inc	0.364*	0.255*	0.526***
	(0.208)	(0.139)	(0.201)
Other Non-selective × Pct athletes top 10% inc		-0.340	0.270
		(0.278)	(0.232)
Pct athletes top 5% inc	0.144	0.314**	-0.175
	(0.159)	(0.146)	(0.187)
Pct athletes top 5% inc × Pct all students top 10% inc	-0.021***	-0.006	0.007
	(0.007)	(0.006)	(0.007)
Elite Private × Pct athletes top 5% inc	0.863***		0.048
	(0.304)		(0.298)
Public Flagship × Pct athletes top 5% inc	0.298		
	(0.223)		

Continued on next page

Table C.3 – continued from previous page

	D-I	D-II	D-III
Mid-tier Private × Pct athletes top 5% inc	0.262	-0.080	-0.164
	(0.197)	(0.133)	(0.173)
Other Non-selective × Pct athletes top 5% inc		-0.314	-0.046
		(0.442)	(0.208)
Pct athletes top 1% inc	-0.289	-1.327*	-0.012
	(0.335)	(0.682)	(0.302)
Pct athletes top 1% inc × Pct all students top 10% inc	0.022**	0.054**	0.008
	(0.009)	(0.026)	(0.012)
Elite Private × Pct athletes top 1% inc	-0.573		-0.794*
	(0.400)		(0.430)
Public Flagship × Pct athletes top 1% inc	-0.151		
	(0.335)		
Mid-tier Private × Pct athletes top 1% inc	-0.222	-0.757	-0.406
	(0.402)	(0.615)	(0.272)
Other Non-selective × Pct athletes top 1% inc		0.868	-0.316
		(1.484)	(0.659)
Football		5.163**	-1.722
		(2.182)	(2.188)
Football × Pct all students top 10% inc		-0.358**	-0.158
		(0.149)	(0.100)
Elite Private × Football			
Public Flagship × Football			
Mid-tier Private × Football			
Other Non-selective × Football			

Continued on next page

Table C.3 – continued from previous page

	D-I	D-II	D-III
Women's Standard count	1.992***	0.392	0.402
	(0.420)	(0.461)	(0.579)
Men's Standard count	2.553***	0.881	2.555***
	(0.592)	(0.652)	(0.722)
Women's Regional count	2.406***	-1.048	-0.935
	(0.865)	(1.238)	(0.990)
Men's Regional count	1.086	-2.642**	-1.881*
	(1.069)	(1.274)	(1.063)
Women's Niche count	1.558	0.433	-1.063
	(0.981)	(0.998)	(1.204)
Men's Niche count	1.042	-2.049	0.876
	(1.457)	(1.688)	(1.547)
Football × Women's Standard count		-0.404	0.046
		(0.260)	(0.293)
Football × Men's Standard count		-0.305	-0.019
		(0.240)	(0.275)
Football × Women's Regional count		-0.089	-0.225
		(0.367)	(0.288)
Football × Men's Regional count		-0.137	0.134
		(0.405)	(0.307)
Football × Women's Niche count		0.517	0.481
		(0.516)	(0.490)
Football × Men's Niche count		1.789*	0.422
		(0.998)	(0.787)
Women's Standard count × Men's Standard count	-0.247***	-0.151*	-0.335***
	(0.075)	(0.079)	(0.091)

Continued on next page

Table C.3 – continued from previous page

	D-I	D-II	D-III
Women's Standard count × Women's Regional count	-0.341***	-0.033	0.018
	(0.108)	(0.165)	(0.143)
Women's Standard count × Men's Regional count	0.011	0.530***	0.338**
	(0.139)	(0.186)	(0.149)
Women's Standard count × Women's Niche count	-0.250*	-0.395**	-0.086
	(0.132)	(0.175)	(0.171)
Women's Standard count × Men's Niche count	-0.028	0.287	-0.228
	(0.179)	(0.302)	(0.257)
Men's Standard count × Women's Regional count	-0.005	0.024	-0.113
	(0.071)	(0.101)	(0.114)
Men's Standard count × Men's Regional count	-0.359***	-0.253*	-0.282**
	(0.110)	(0.141)	(0.140)
Men's Standard count × Women's Niche count	0.017	0.148	0.044
	(0.087)	(0.143)	(0.167)
Men's Standard count × Men's Niche count	-0.370**	-0.401	-0.442*
	(0.146)	(0.261)	(0.262)
Women's Regional count × Men's Regional count	0.217	0.219	0.178
	(0.140)	(0.163)	(0.109)
Women's Regional count × Women's Niche count	-0.011	0.010	-0.239
	(0.137)	(0.247)	(0.206)
Women's Regional count × Men's Niche count	≈0.000	-0.585	-0.025
	(0.227)	(0.425)	(0.289)
Men's Regional count × Women's Niche count	-0.055	0.047	0.225
	(0.189)	(0.305)	(0.216)
Men's Regional count × Men's Niche count	0.063	0.120	0.184
	(0.285)	(0.461)	(0.331)
Women's Niche count × Men's Niche count	-0.064	0.308	0.058

Continued on next page

Table C.3 – continued from previous page

	D-I	D-II	D-III
	(0.180)	(0.211)	(0.245)
Elite Private × Women's Standard count	-0.653**		-0.554
	(0.317)		(0.395)
Public Flagship × Women's Standard count	-0.352		
	(0.244)		
Mid-tier Private × Women's Standard count	-0.227	0.333*	0.110
	(0.208)	(0.183)	(0.233)
Elite Private × Men's Standard count	-0.250		0.527
	(0.292)		(0.350)
Public Flagship × Men's Standard count	-0.341*		
	(0.206)		
Mid-tier Private × Men's Standard count	0.067	0.735***	0.361*
	(0.194)	(0.153)	(0.194)
Elite Private × Women's Regional count	≈0.000		-1.551***
	(0.395)		(0.392)
Public Flagship × Women's Regional count	-0.315		
	(0.294)		
Mid-tier Private × Women's Regional count	0.009	0.596**	-0.683***
	(0.266)	(0.250)	(0.250)
Elite Private × Men's Regional count	0.042		-0.137
	(0.427)		(0.402)
Public Flagship × Men's Regional count	0.185		
	(0.313)		
Mid-tier Private × Men's Regional count	0.047	0.587**	0.122
	(0.310)	(0.268)	(0.257)
Elite Private × Women's Niche count	-0.373		-0.396
	(0.357)		(0.484)

Continued on next page

Table C.3 – continued from previous page

	D-I	D-II	D-III
Public Flagship × Women's Niche count	-0.278		
	(0.267)		
Mid-tier Private × Women's Niche count	-0.160	0.734**	-0.594
	(0.250)	(0.318)	(0.370)
Elite Private × Men's Niche count	0.688		1.069
	(0.546)		(0.706)
Public Flagship × Men's Niche count	0.486		
	(0.452)		
Mid-tier Private × Men's Niche count	0.712	-0.457	0.840
	(0.434)	(0.501)	(0.637)
Other Non-selective × Women's Standard count		-0.085	0.174
		(0.325)	(0.296)
Other Non-selective × Men's Standard count		0.613**	-0.210
		(0.292)	(0.228)
Other Non-selective × Women's Regional count		-0.032	-0.750**
		(0.471)	(0.351)
Other Non-selective × Men's Regional count		0.434	-0.138
		(0.447)	(0.343)
Other Non-selective × Women's Niche count		0.101	0.198
		(0.598)	(0.445)
Other Non-selective × Men's Niche count		-0.532	
		(1.065)	
Women's Standard count × Pct all students top 10% inc	-0.025	-0.007	0.022
	(0.017)	(0.032)	(0.026)
Men's Standard count × Pct all students top 10% inc	-0.024	-0.039	-0.027
	(0.022)	(0.046)	(0.032)
Women's Regional count × Pct all students top 10% inc	-0.073**	-0.044	-0.013

Continued on next page

Table C.3 – continued from previous page

	D-I	D-II	D-III
	(0.031)	(0.077)	(0.045)
Men's Regional count × Pct all students top 10% inc	-0.085**	-0.056	0.029
	(0.038)	(0.084)	(0.049)
Women's Niche count × Pct all students top 10% inc	-0.086**	-0.009	0.011
	(0.037)	(0.068)	(0.049)
Men's Niche count × Pct all students top 10% inc	-0.077	0.081	0.006
	(0.047)	(0.094)	(0.055)
Football × Women's Standard count × Pct all students top 10% inc		0.016	0.014
		(0.018)	(0.012)
Football × Men's Standard count × Pct all students top 10% inc		0.011	-0.001
		(0.016)	(0.012)
Football × Women's Regional count × Pct all students top 10% inc		0.025	0.011
		(0.021)	(0.011)
Football × Men's Regional count × Pct all students top 10% inc		0.023	0.001
		(0.025)	(0.013)
Football × Women's Niche count × Pct all students top 10% inc		-0.001	-0.013
		(0.032)	(0.016)
Football × Men's Niche count × Pct all students top 10% inc		-0.084*	-0.001
		(0.048)	(0.020)
Women's Standard count × Men's Standard count × Pct all students top 10% inc	0.001	0.002	0.001
	(0.002)	(0.005)	(0.004)
Women's Standard count × Women's Regional count × Pct all students top 10% inc	0.005	≈-0.000	-0.002
	(0.003)	(0.010)	(0.006)
Women's Standard count × Men's Regional count × Pct all students top 10% inc	0.009*	-0.008	-0.006
	(0.004)	(0.012)	(0.006)
Women's Standard count × Women's Niche count × Pct all students top 10% inc	0.009*	0.006	0.003
	(0.004)	(0.011)	(0.006)

Continued on next page

	D-I	D-II	D-III
Women's Standard count × Men's Niche count × Pct all students top 10% inc	0.004	-0.017	≈-0.000
	(0.005)	(0.016)	(0.008)
Men's Standard count × Women's Regional count × Pct all students top 10% inc	0.005**	0.001	0.006
	(0.002)	(0.006)	(0.005)
Men's Standard count × Men's Regional count × Pct all students top 10% inc	0.002	0.006	≈-0.000
	(0.004)	(0.009)	(0.006)
Men's Standard count × Women's Niche count × Pct all students top 10% inc	0.002	-0.007	-0.004
	(0.003)	(0.008)	(0.006)
Men's Standard count × Men's Niche count × Pct all students top 10% inc	0.007	0.017	0.004
	(0.004)	(0.015)	(0.007)
Women's Regional count × Men's Regional count × Pct all students top 10% inc	-0.007*	-0.005	-0.001
	(0.004)	(0.009)	(0.004)
Women's Regional count × Women's Niche count × Pct all students top 10% inc	-0.002	-0.003	0.003
	(0.003)	(0.013)	(0.006)
Women's Regional count × Men's Niche count × Pct all students top 10% inc	0.003	0.023	0.007
	(0.005)	(0.016)	(0.007)
Men's Regional count × Women's Niche count × Pct all students top 10% inc	0.005	0.006	0.001
	(0.004)	(0.017)	(0.007)
Men's Regional count × Men's Niche count × Pct all students top 10% inc	-0.002	0.005	-0.007
	(0.006)	(0.019)	(0.009)
Women's Niche count × Men's Niche count × Pct all students top 10% inc	≈0.000	-0.009	-0.005
	(0.003)	(0.011)	(0.005)
Track & Field complementarity	0.808*	0.795	0.830*
	(0.415)	(0.500)	(0.468)
Court complementarity	1.371***	0.258	1.831***
	(0.357)	(0.332)	(0.271)
Aquatic complementarity	-1.340***	-0.525	-0.343

Continued on next page

Table C.3 – continued from previous page

	D-I	D-II	D-III
	(0.328)	(0.388)	(0.282)
Field complementarity	-0.740***	0.037	1.215***
	(0.193)	(0.244)	(0.261)
Racquet complementarity	-0.505*	-0.339	-0.241
	(0.281)	(0.361)	(0.302)
Ice complementarity	-0.285	-1.412	1.877***
	(0.799)	(1.024)	(0.561)
Mat complementarity	1.103***	0.082	1.356**
	(0.385)	(0.936)	(0.596)
Diamond complementarity	0.662*	0.853*	1.256***
	(0.359)	(0.436)	(0.414)
Track & Field complementarity × Pct all students top 10% inc	-0.035**	-0.002	-0.025
	(0.014)	(0.037)	(0.020)
Court complementarity × Pct all students top 10% inc	-0.030***	0.025	-0.022**
	(0.011)	(0.021)	(0.011)
Aquatic complementarity × Pct all students top 10% inc	0.032***	-0.013	0.036***
	(0.009)	(0.023)	(0.010)
Field complementarity × Pct all students top 10% inc	0.021***	0.055***	0.027**
	(0.006)	(0.016)	(0.010)
Racquet complementarity × Pct all students top 10% inc	0.014	-0.013	0.031***
	(0.008)	(0.024)	(0.010)
Ice complementarity × Pct all students top 10% inc	-0.004	0.083	0.003
	(0.021)	(0.055)	(0.022)
Mat complementarity × Pct all students top 10% inc	-0.020*	-0.071	-0.033
	(0.011)	(0.071)	(0.024)
Diamond complementarity × Pct all students top 10% inc	-0.046***	-0.008	-0.039**
	(0.012)	(0.029)	(0.017)

Continued on next page

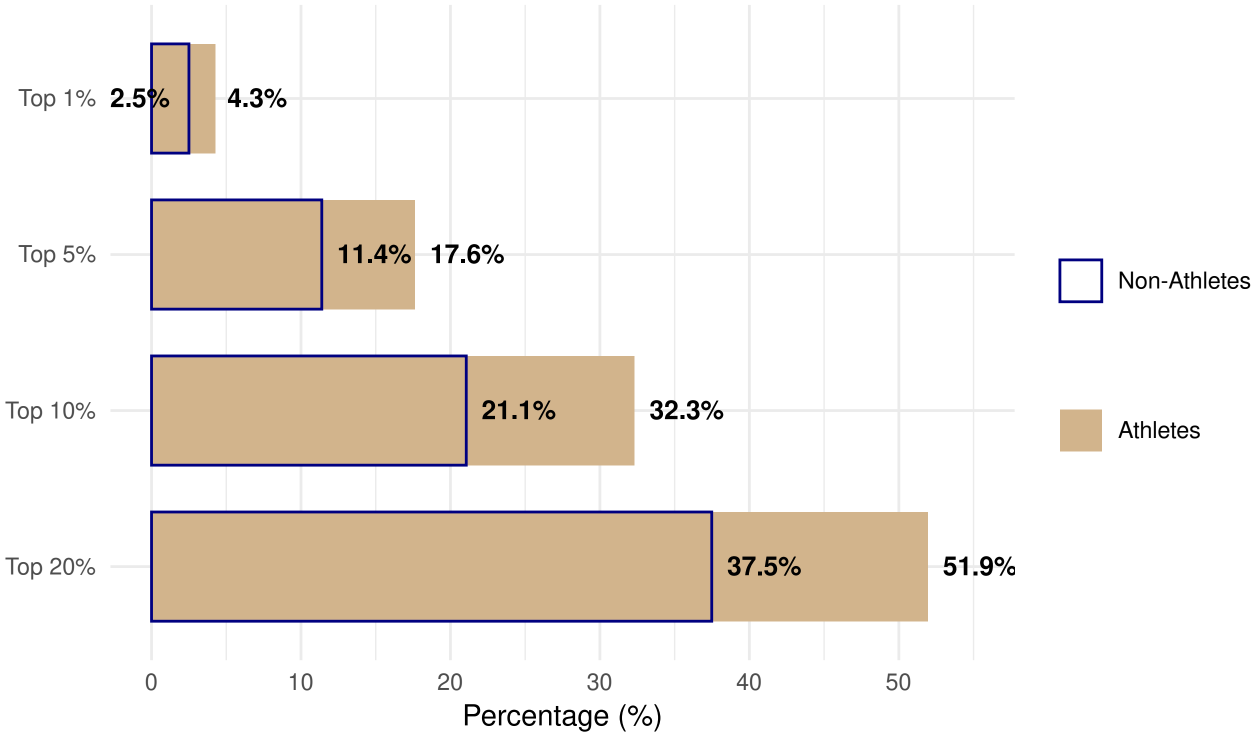
Table C.3 – continued from previous page

	D-I	D-II	D-III
Log likelihood	-1,281.07	-867.51	-1,079.64
Universities	343	277	393
Observations	85,750	69,250	98,005

NOTES.—Standard errors below each estimate in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Mid-tier Public is the omitted tier. Sports count variables are counts of sports in the groupings defined in the Appendix. Complementarity variables measure economies of scope in facility usage. Each variable equals the number of additional sports (of either gender) beyond the first that use the same type of facility. These include Track (track & field and football); Wooden Courts (basketball, volleyball); Aquatics (swimming, diving, water polo); Fields (soccer, lacrosse, field hockey, football); Racquet Courts (tennis, squash); Ice (hockey); Mats (wrestling, gymnastics); and Diamonds (baseball, softball).

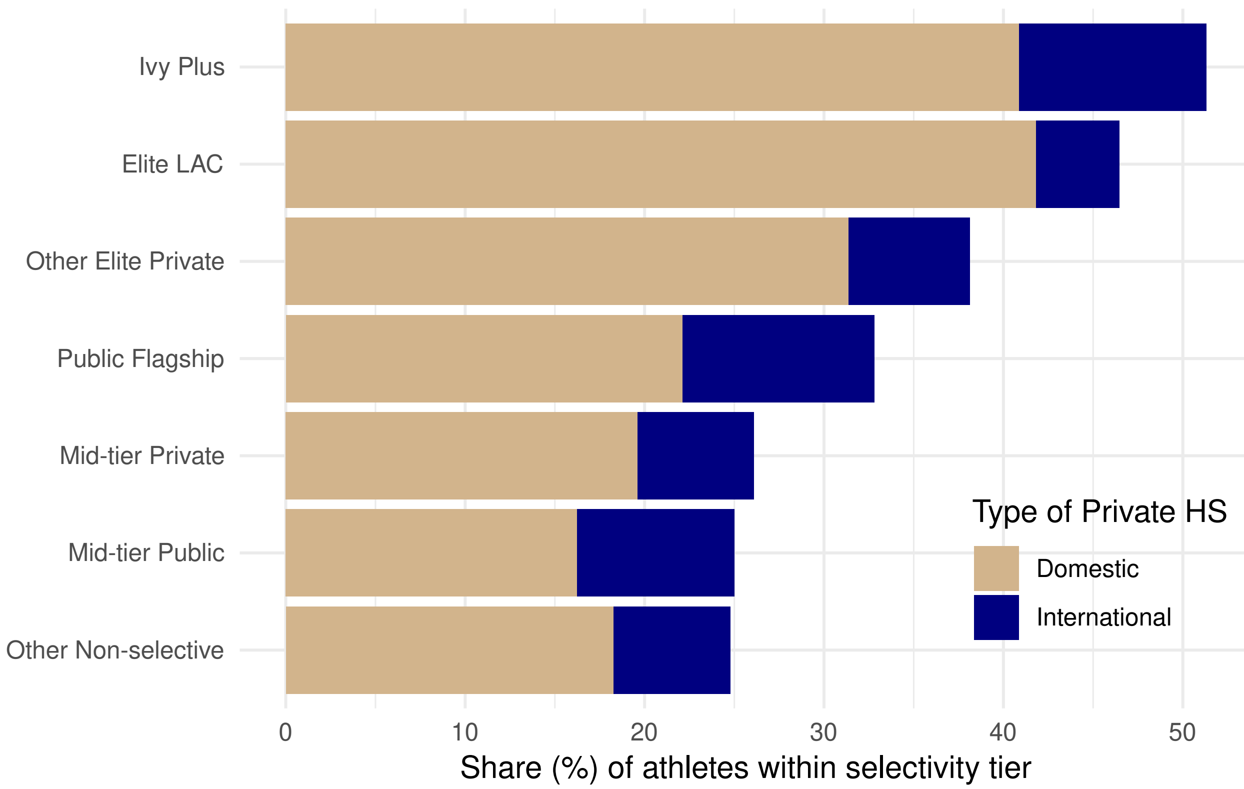
Figure C.1: Representation of Athletes and Non-Athletes in Upper Tails of Income Distribution Across All Selectivity Tiers



Source.—Authors' calculations from NCAA roster data linked to high school ZIP code characteristics and university characteristics.

Notes.—This figure plots the likelihood of an athlete or non-athlete to come from a ZIP code in the upper percentiles of the national income distribution. See the notes to Table 1 for a description of the selectivity tiers.

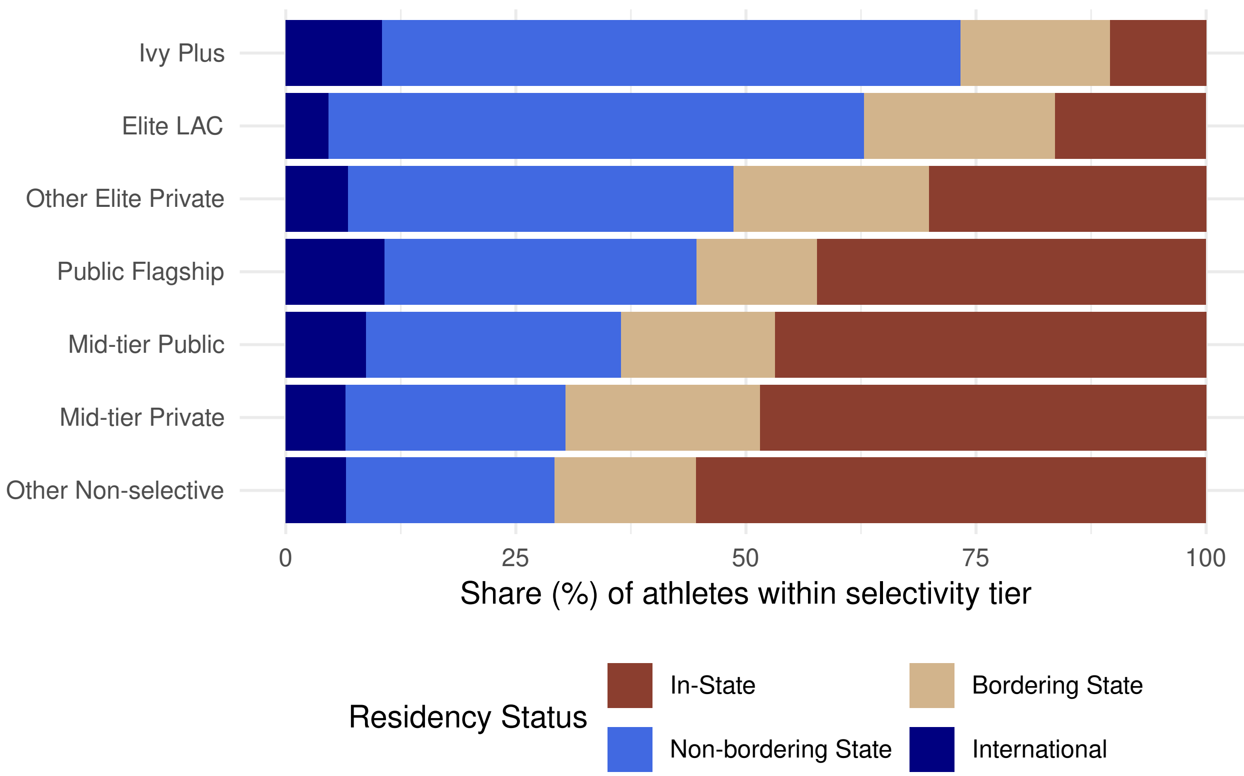
Figure C.2: Share of Athletes Attending Private High School, by Selectivity Tier



Source.—Authors' calculations from NCAA roster data linked to high school and university characteristics.

Notes.—This figure plots the likelihood of an athlete to have attended either a domestic or international private high school by selectivity tier. See the notes to Table 1 for a description of the selectivity tiers.

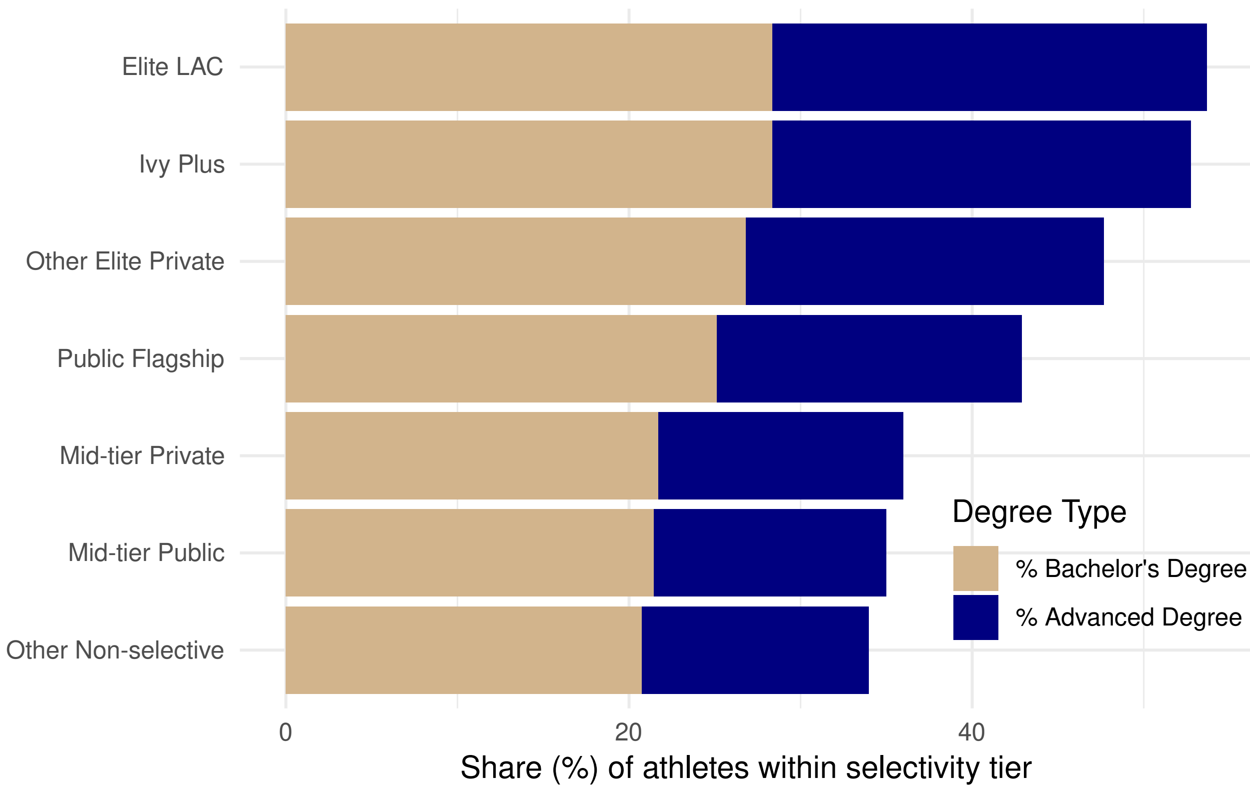
Figure C.3: Athlete Residence by Selectivity Tier



Source.—Authors' calculations from NCAA roster data linked to high school and university characteristics.

Notes.—This figure plots the likelihood of an athlete to have attended high school in-state, a state bordering the university, a state not bordering the university, or internationally. See the notes to Table 1 for a description of the selectivity tiers.

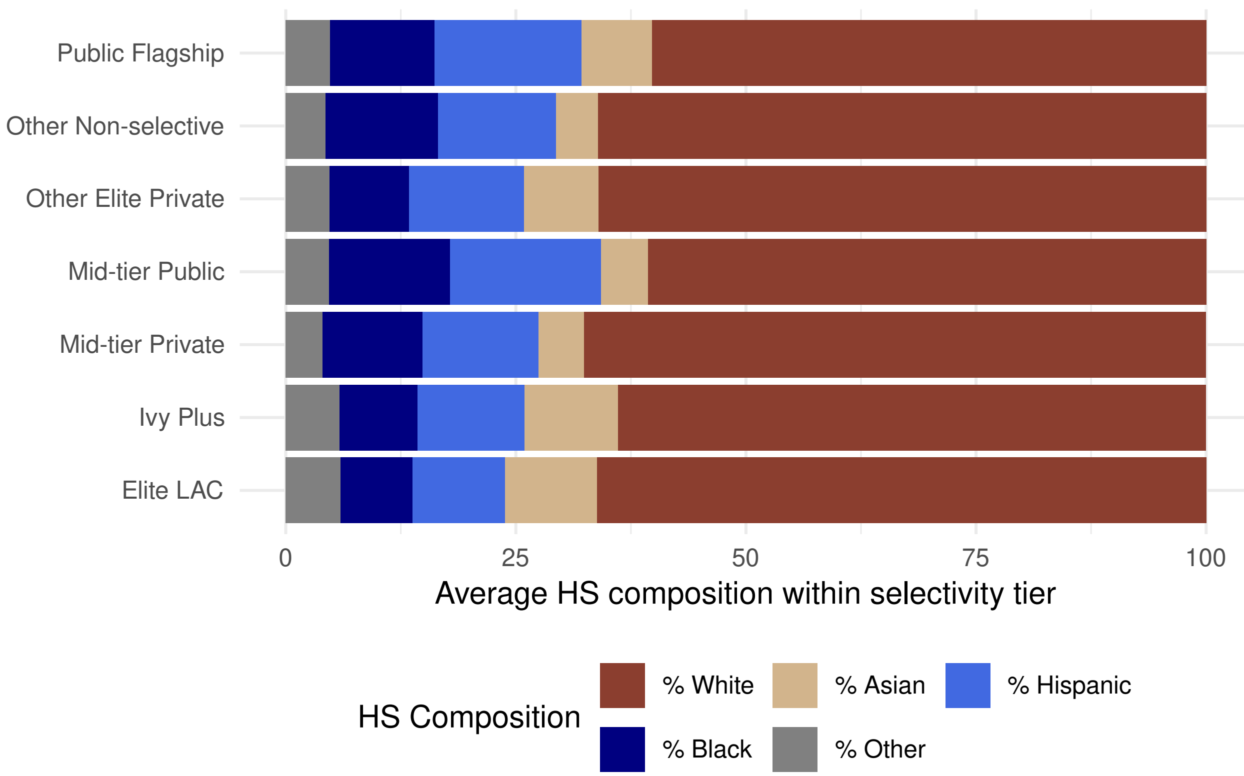
Figure C.4: Athlete HS ZIP Code Education Rates by Selectivity Tier



Source.—Authors' calculations from NCAA roster data linked to high school and university characteristics.

Notes.—This figure plots the percentage of residents in the ZIP code of an athlete who have completed a bachelor's or advanced degree. See the notes to Table 1 for a description of the selectivity tiers.

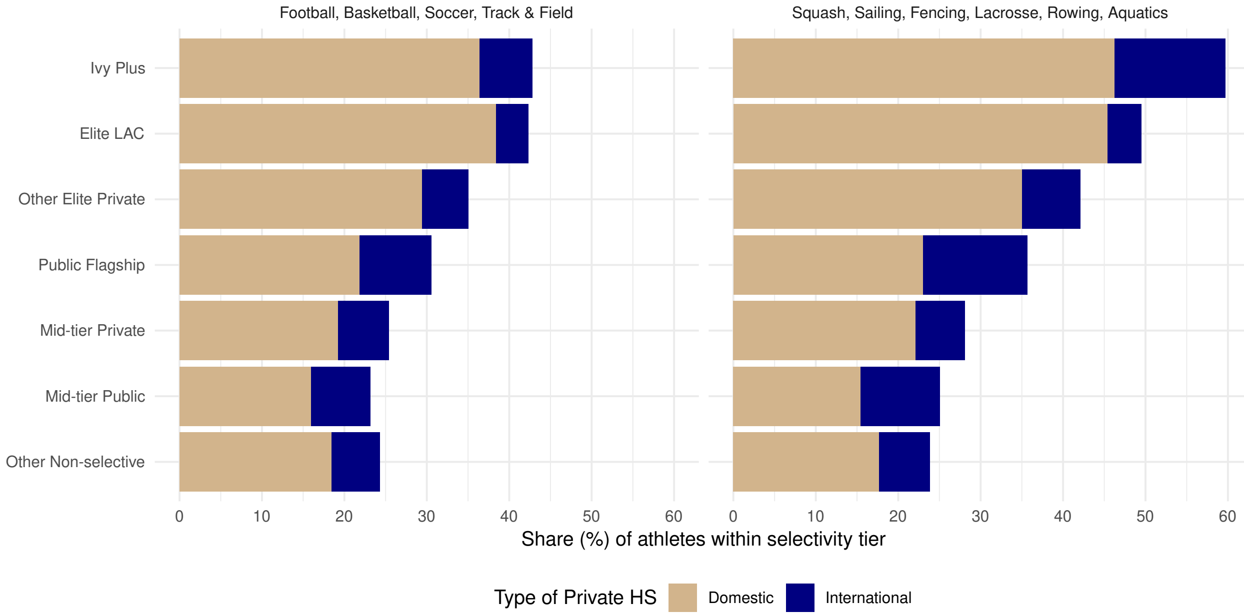
Figure C.5: Athlete HS Race/Ethnicity by Selectivity Tier



Source.—Authors' calculations from NCAA roster data linked to high school and university characteristics.

Notes.—This figure plots the racial/ethnic distribution of an athlete's high school. See the notes to Table 1 for a description of the selectivity tiers.

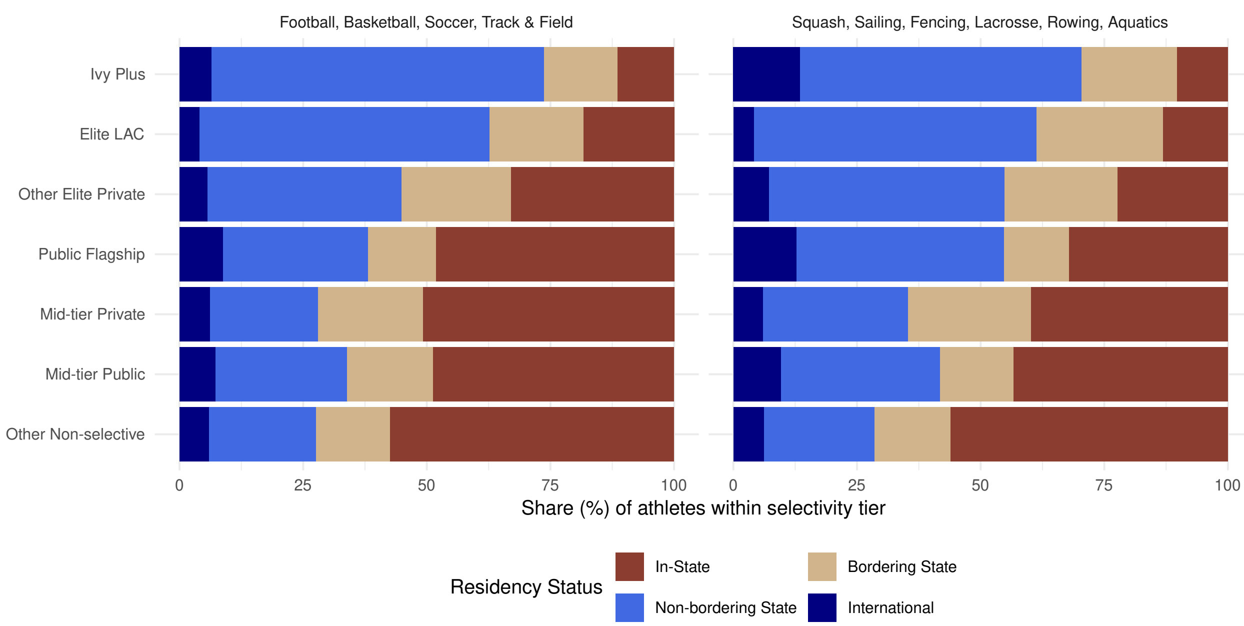
Figure C.6: Share of Athletes Attending Private High School, by Selectivity Tier and Sport Group



Source.—Authors' calculations from NCAA roster data linked to high school and university characteristics.

Notes.—This figure plots the likelihood of an athlete to have attended either a domestic or international private high school, conditional on being on the roster for Football or Basketball, or the roster for Soccer or Track & Field. See the notes to Table 1 for a description of the selectivity tiers.

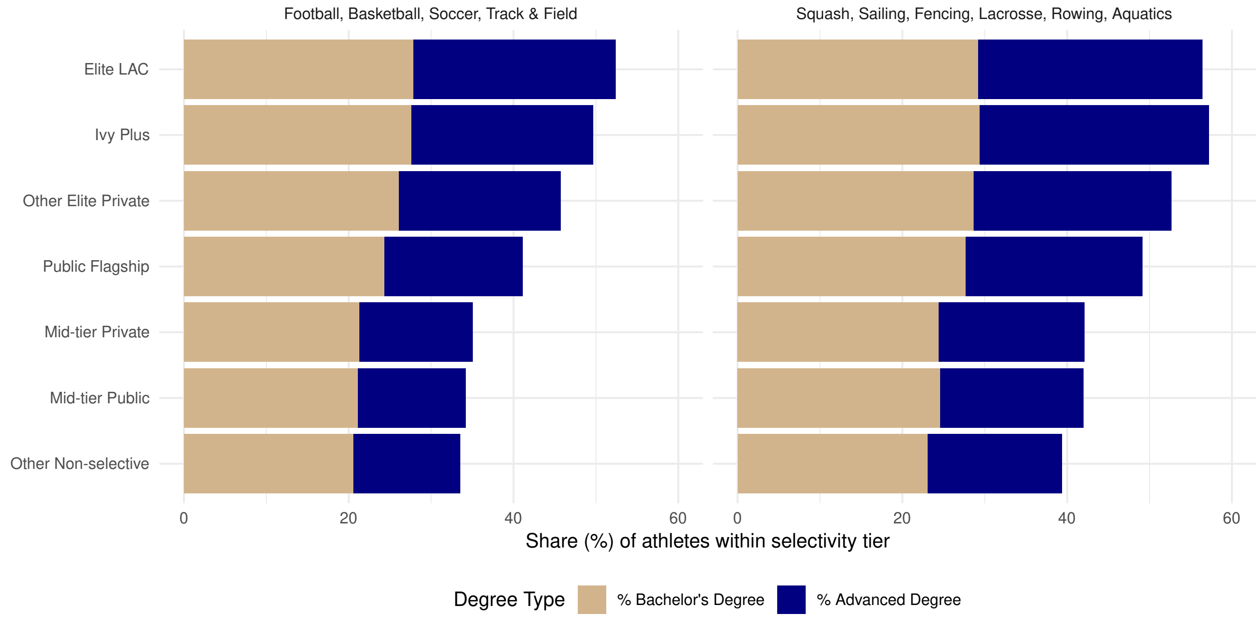
Figure C.7: Athlete Residence by Selectivity Tier and Sport Group



Source.—Authors' calculations from NCAA roster data linked to high school and university characteristics.

Notes.—This figure plots the likelihood of an athlete to have attended high school in-state, a state bordering the university, a state not bordering the university, or internationally, conditional on being on the roster for Football or Basketball, or the roster for Soccer or Track & Field. See the notes to Table 1 for a description of the selectivity tiers.

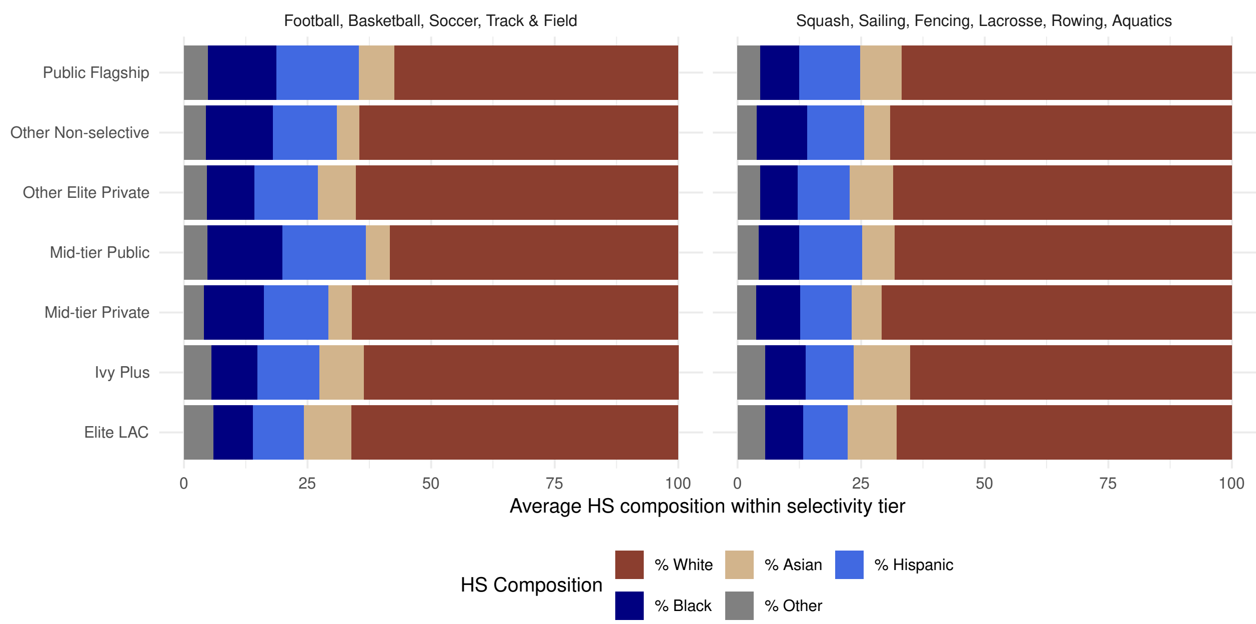
Figure C.8: Athlete HS ZIP Code Education Rates by Selectivity Tier and Sport Group



SOURCE.—Authors' calculations from NCAA roster data linked to high school and university characteristics.

NOTES.—This figure plots the percentage of residents in the ZIP code of an athlete who have completed a bachelor's or advanced degree, conditional on the athlete being on the roster for Football or Basketball, or the roster for Soccer or Track & Field. See the notes to Table 1 for a description of the selectivity tiers.

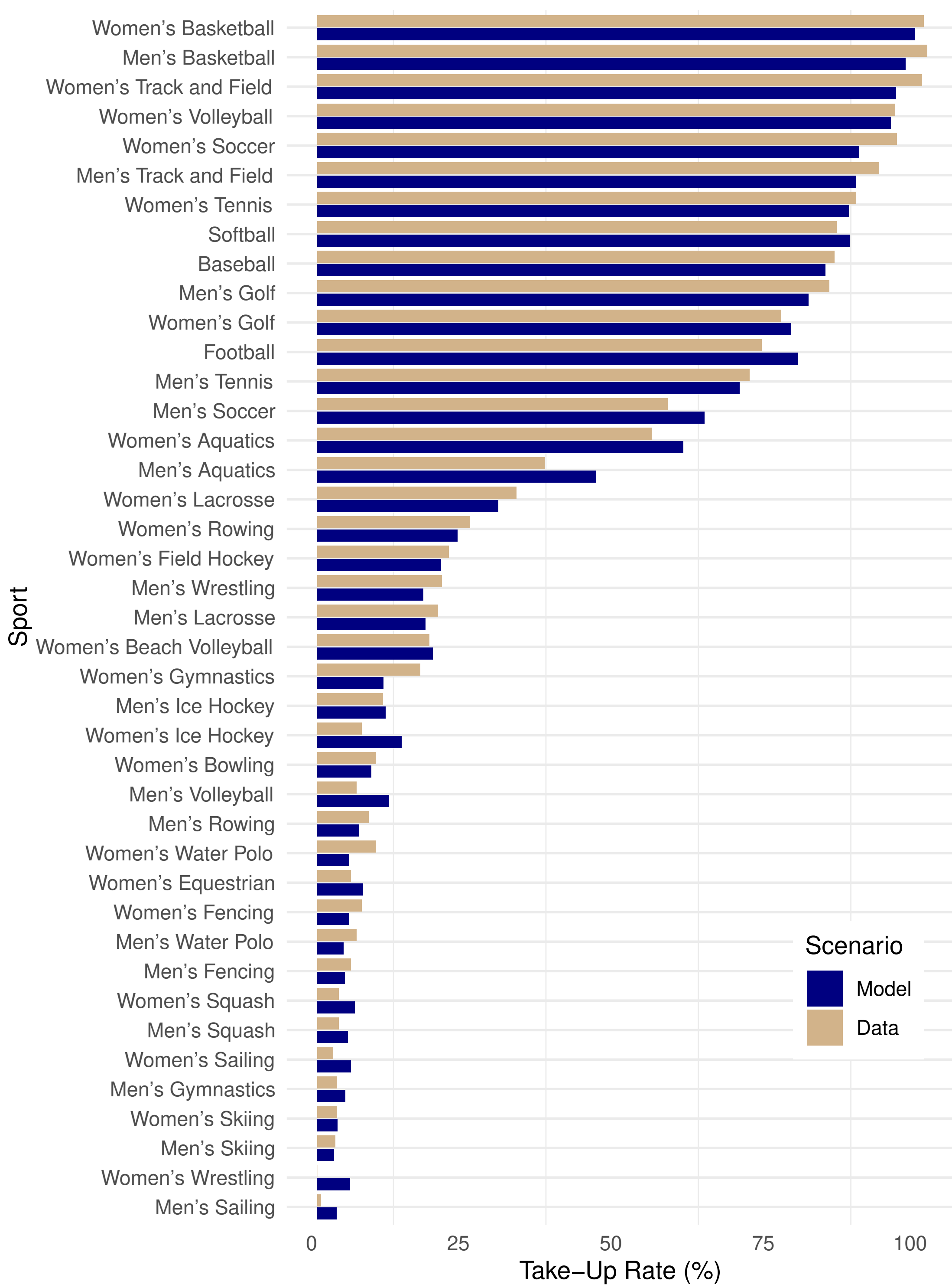
Figure C.9: Athlete HS Race/Ethnicity by Selectivity Tier and Sport Group



SOURCE.—Authors' calculations from NCAA roster data linked to high school and university characteristics.

NOTES.—This figure plots the racial/ethnic distribution of an athlete's high school, conditional on being on the roster for Football or Basketball, or the roster for Soccer or Track & Field. See the notes to Table 1 for a description of the selectivity tiers.

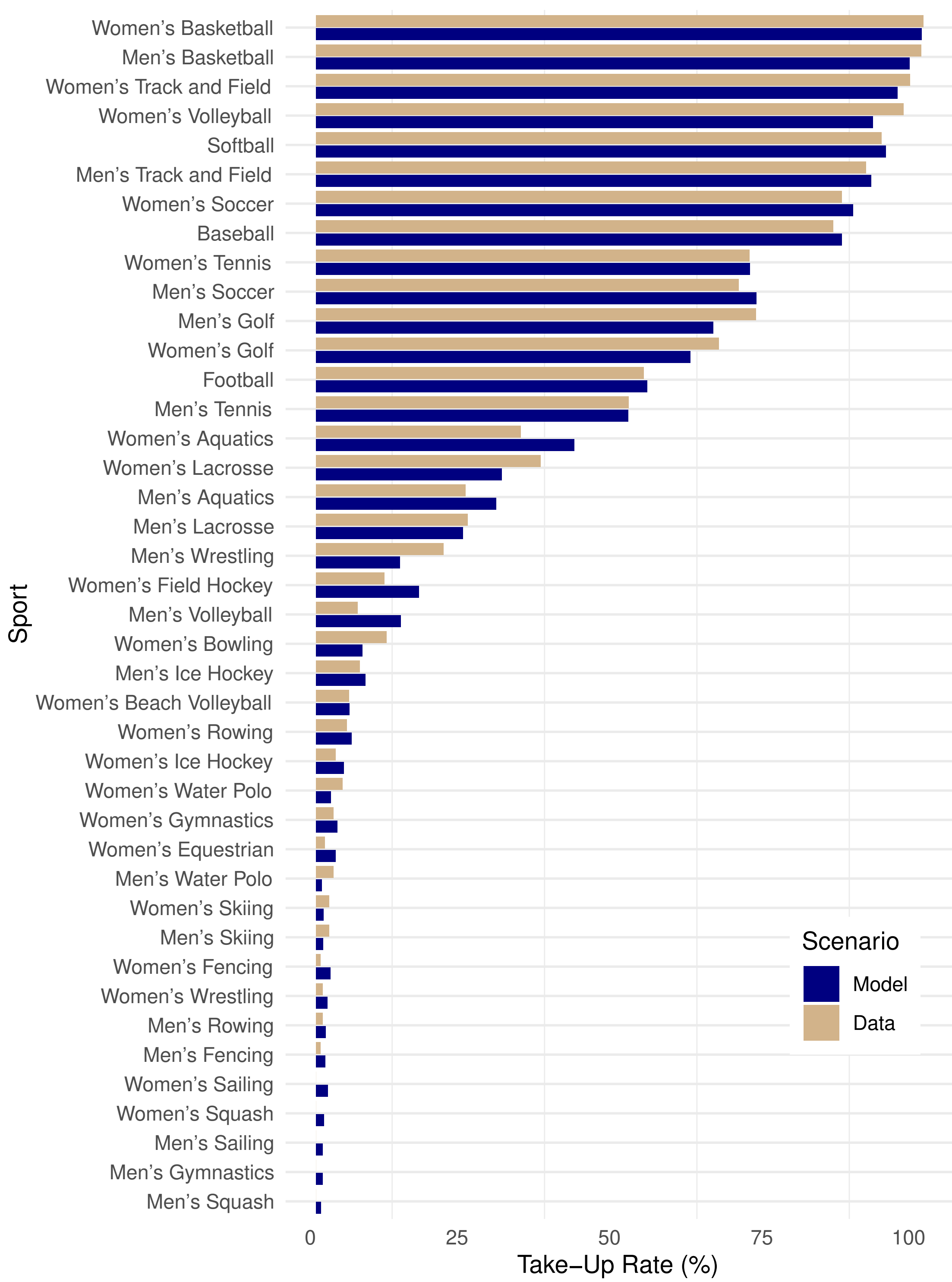
Figure C.10: Model Fit of Untargeted Moments, NCAA Division I Universities



Source.—Authors' calculations from comparing data on university sport offerings with predicted probabilities of sport offerings.

Notes.—This figure plots the actual share of universities offering each sport against the model-predicted shares. Sample includes all NCAA Division I universities in our sample.

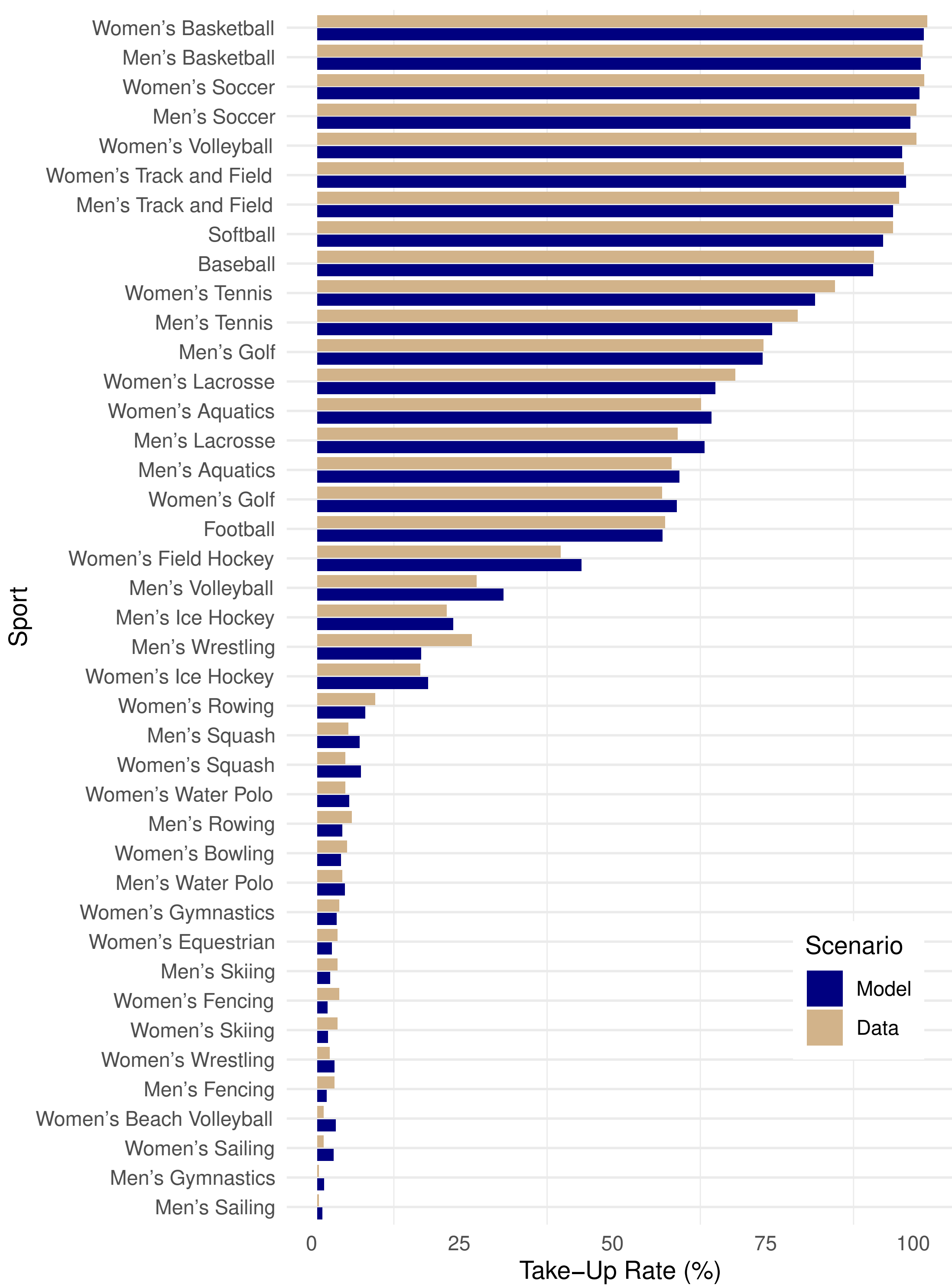
Figure C.11: Model Fit of Untargeted Moments, NCAA Division II Universities



Source.—Authors' calculations from comparing data on university sport offerings with predicted probabilities of sport offerings.

Notes.—This figure plots the actual share of universities offering each sport against the model-predicted shares. Sample includes all NCAA Division II universities in our sample.

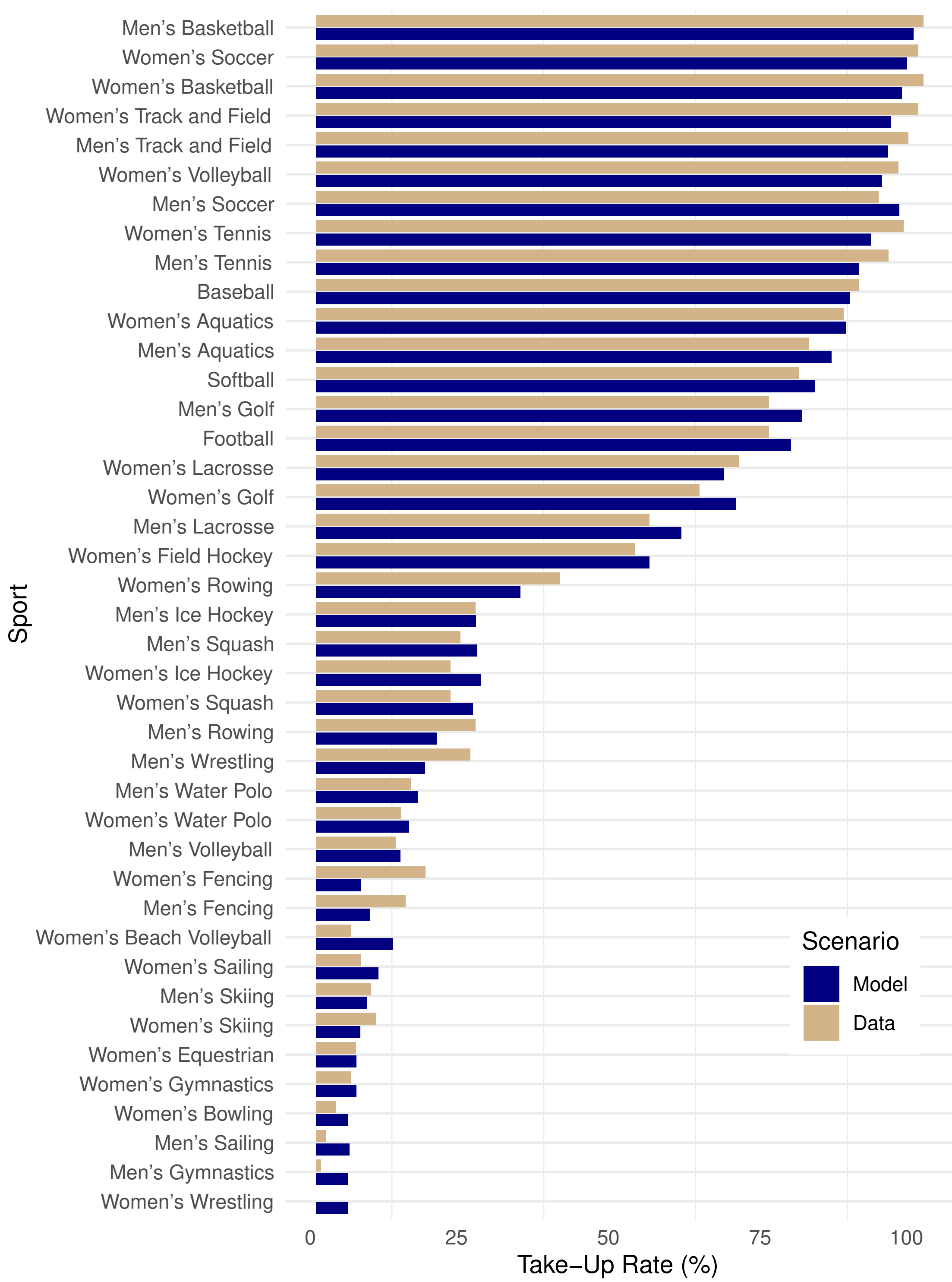
Figure C.12: Model Fit of Untargeted Moments, NCAA Division III Universities



Source.—Authors' calculations from comparing data on university sport offerings with predicted probabilities of sport offerings.

Notes.—This figure plots the actual share of universities offering each sport against the model-predicted shares. Sample includes all NCAA Division III universities in our sample.

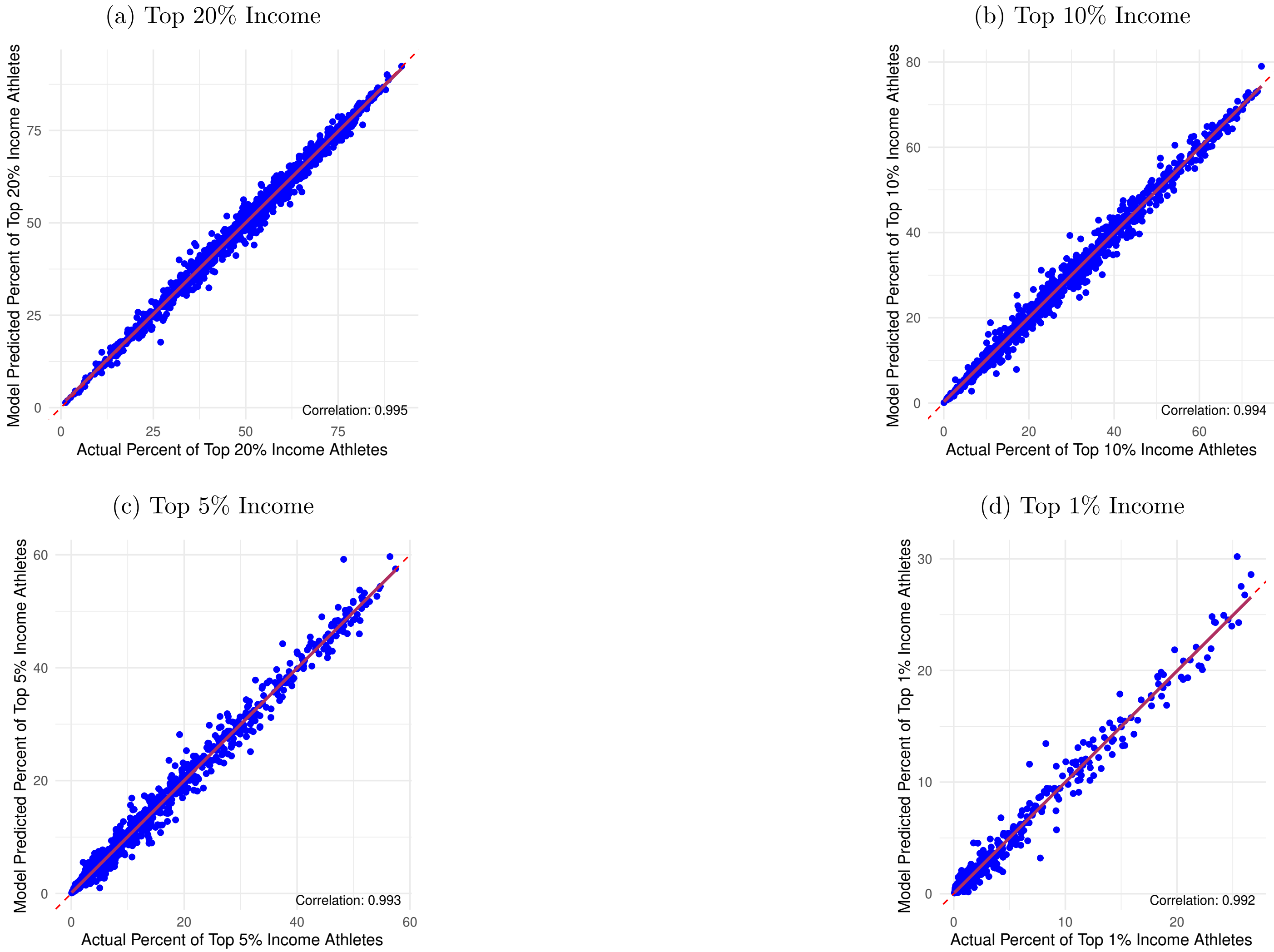
Figure C.13: Model Fit of Untargeted Moments, Top-Tier Universities



Source.—Authors' calculations from comparing data on university sport offerings with predicted probabilities of sport offerings.

Notes.—This figure plots the actual share of universities offering each sport against the model-predicted shares. Sample includes Ivy Plus, Elite LAC, and Other Elite Private universities only.

Figure C.14: Athlete Rates of Top Income Representation in Chosen Bundle, Data vs. Model

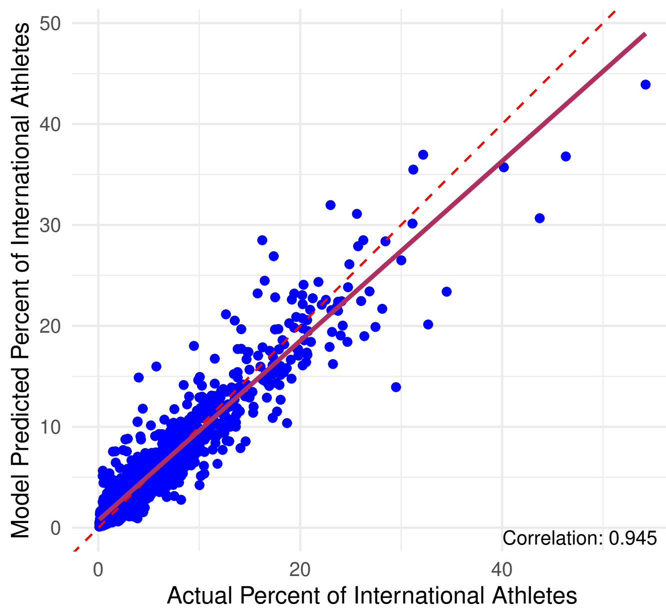


Source.—Authors' calculations from comparing data and model predictions.

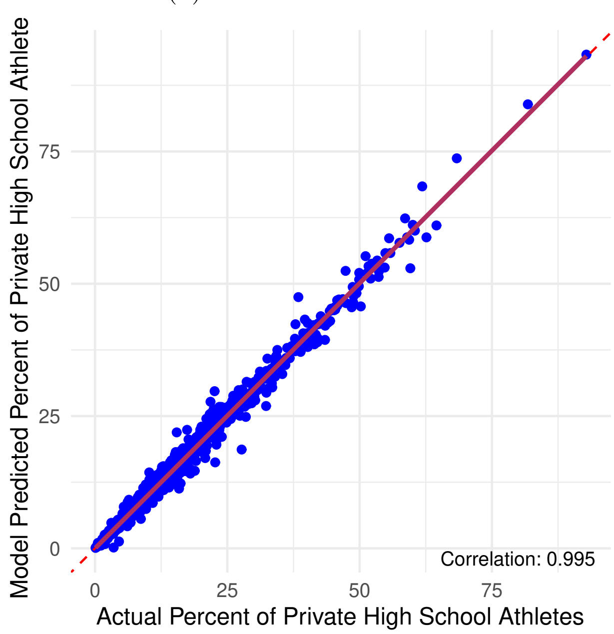
Notes.—This figure plots the data and model-predicted values of the given variable for all universities in our sample. Each university is a blue point. The dashed red line is the 45° line. The solid red line is the best fit line through the blue points.

Figure C.15: Athlete Rates of Private HS Origination in Chosen Bundle, Data vs. Model

(a) International Private HS



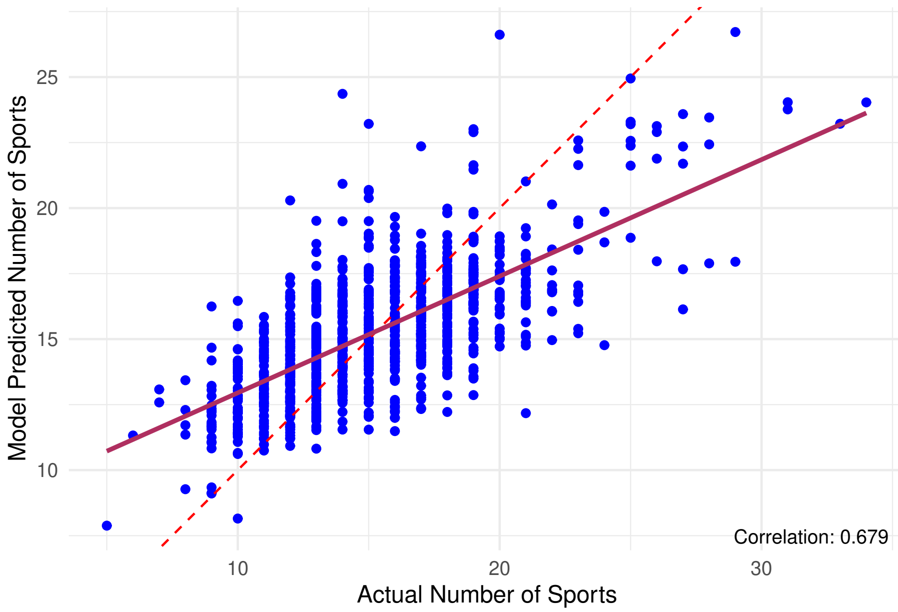
(b) Domestic Private HS



Source.—Authors' calculations from comparing data and model predictions.

Notes.—This figure plots the data and model-predicted values of the given variable for all universities in our sample. Each university is a blue point. The dashed red line is the 45° line. The solid red line is the best fit line through the blue points.

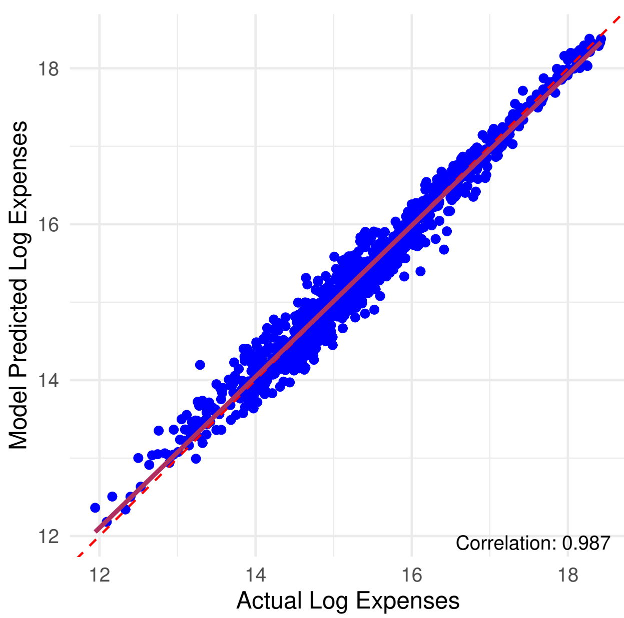
Figure C.16: Total Number of Sports Offered in Chosen Bundle, Data vs. Model



SOURCE.—Authors' calculations from comparing data and model predictions.

NOTES.—This figure plots the data and model-predicted values of the given variable for all universities in our sample. Each university is a blue point. The dashed red line is the 45° line. The solid red line is the best fit line through the blue points.

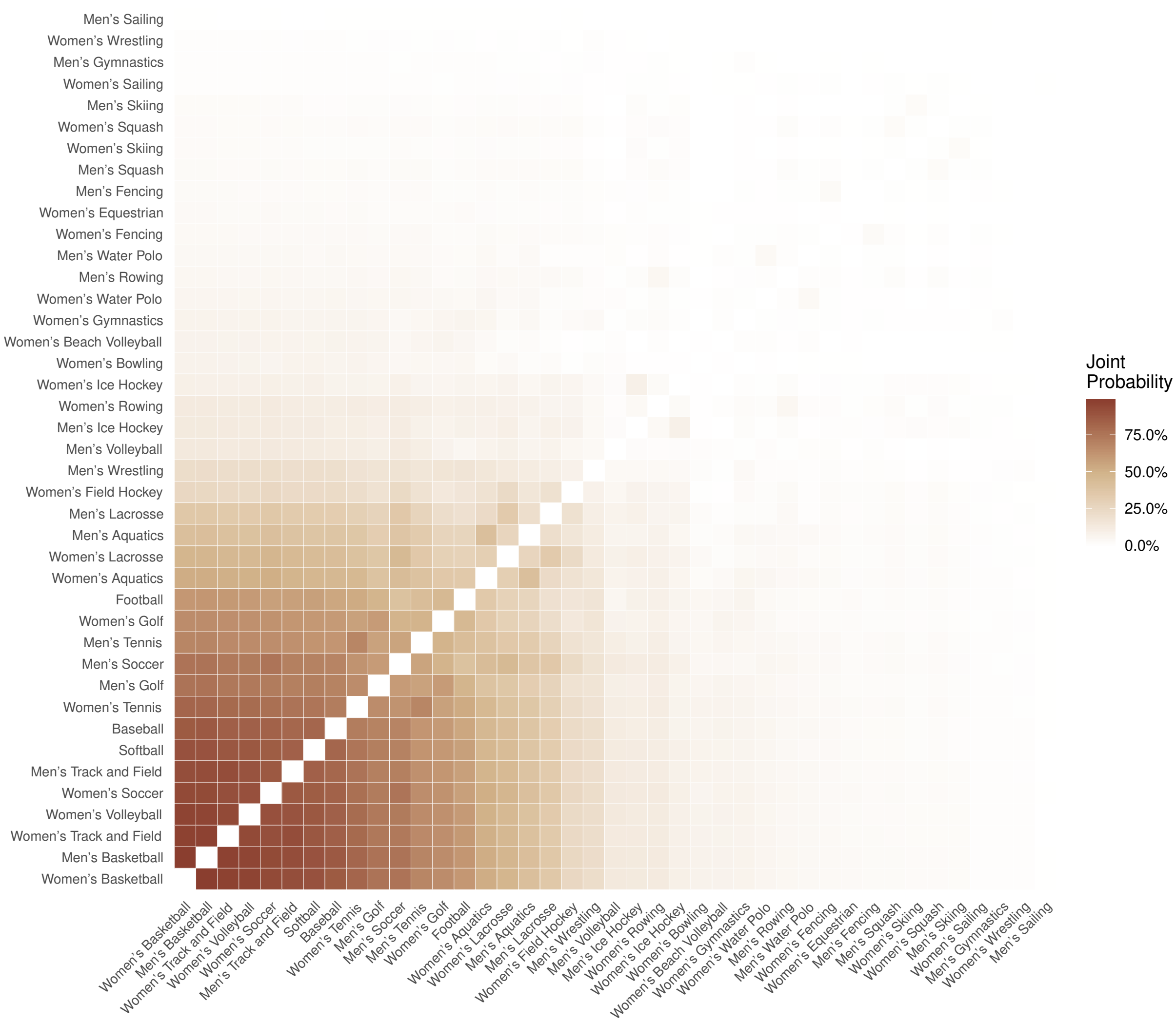
Figure C.17: Log Expenses of Chosen Bundle, Data vs. Model



SOURCE.—Authors' calculations from comparing data and model predictions.

NOTES.—This figure plots the data and model-predicted values of the given variable for all universities in our sample. Each university is a blue point. The dashed red line is the 45° line. The solid red line is the best fit line through the blue points.

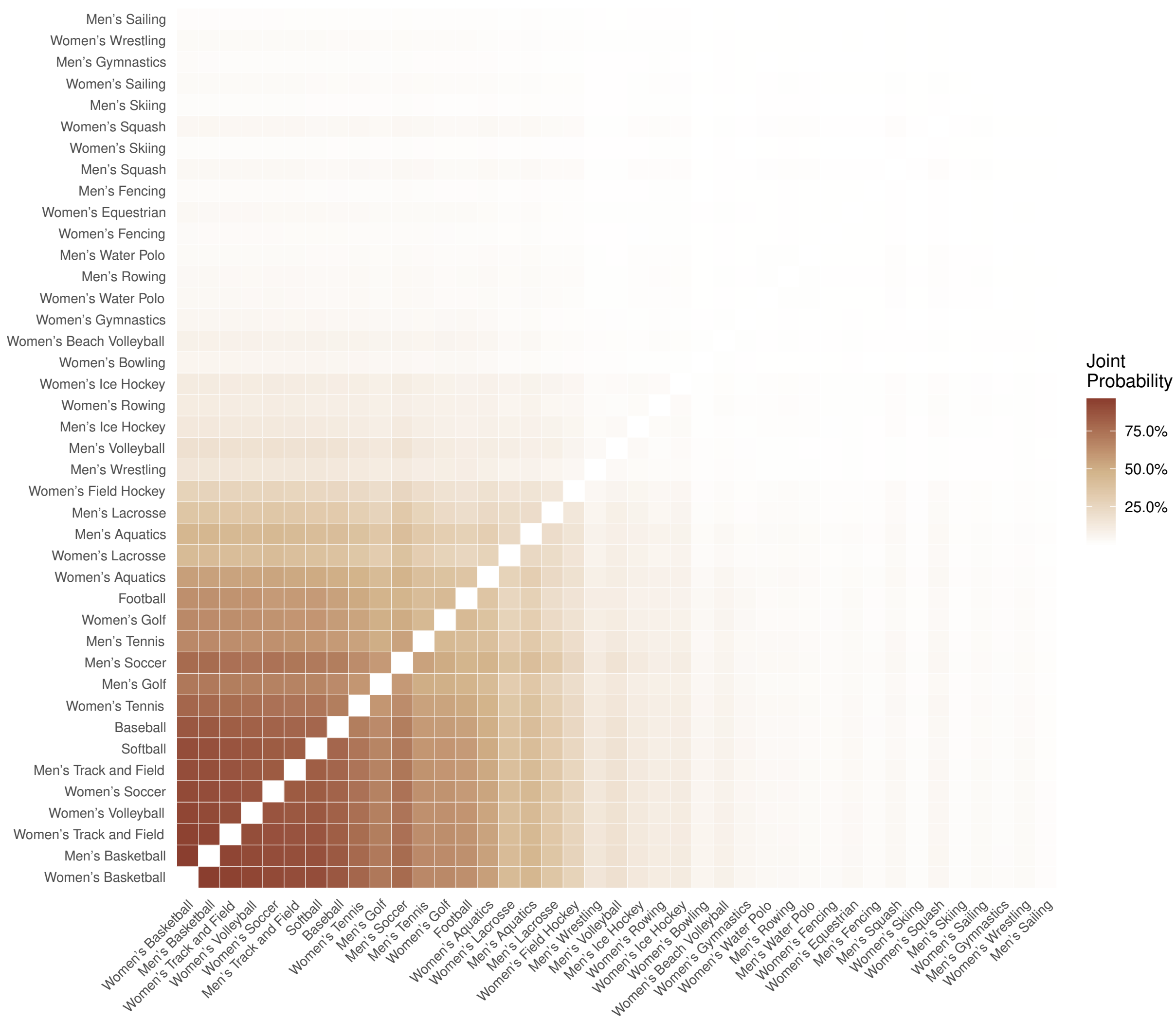
Figure C.18: Sport Co-occurrence rates in Chosen Bundle in Data



SOURCE.—Authors' calculations from data on university sport offerings.

NOTES.—This figure plots the joint likelihood of different sports being found in the chosen bundle in the data.

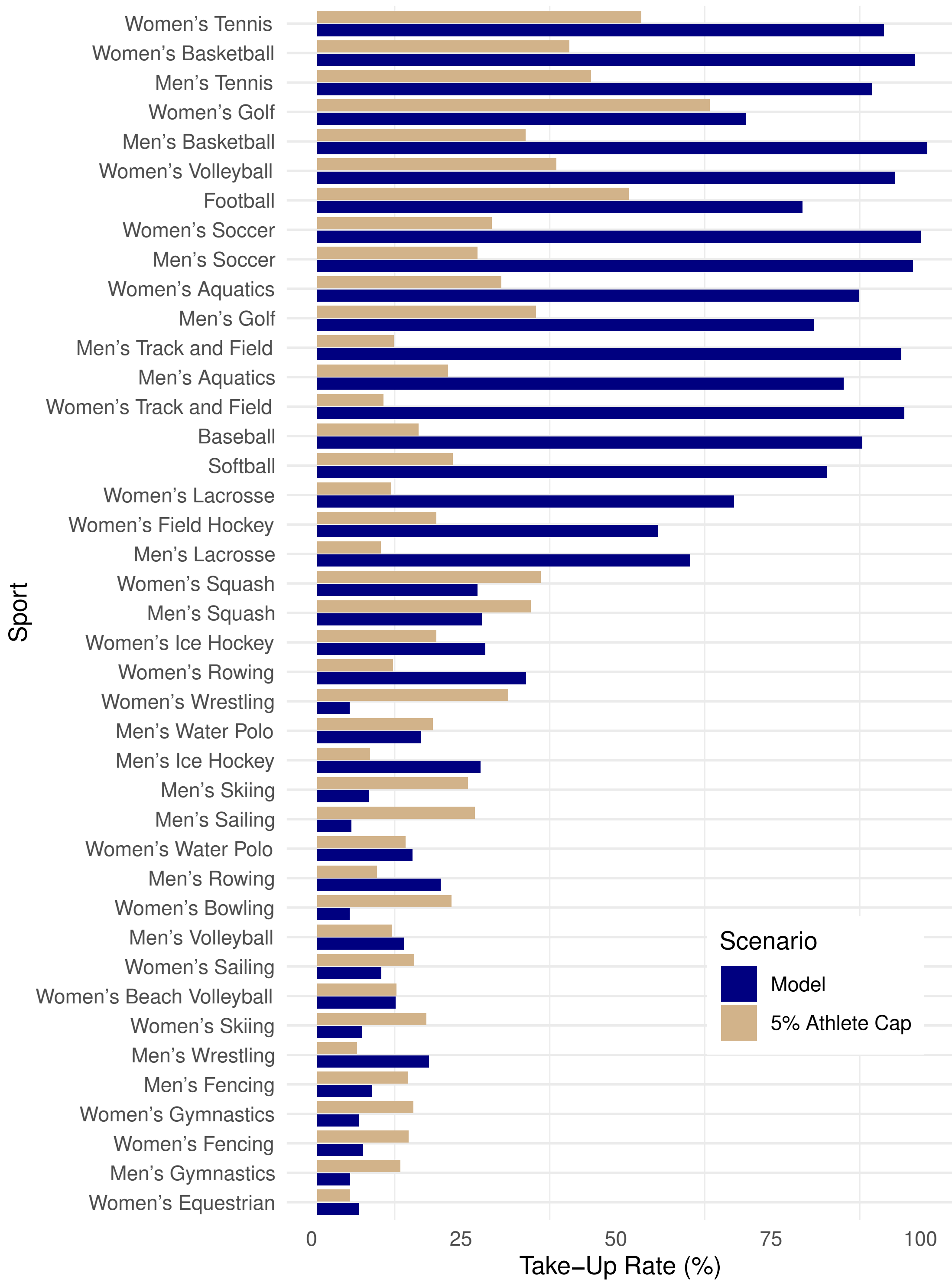
Figure C.19: Sport Co-occurrence rates in Model-Predicted Chosen Bundle



Source.—Authors' calculations from model-predicted probabilities of sport offerings.

Notes.—This figure plots the model-predicted joint likelihood of different sports being found in the same bundle.

Figure C.20: Counterfactual Sport Take-up Rates at Elite Universities, Non-Tradition-Weighted



SOURCE.—Authors' calculations from comparing data on predicted probabilities of sport offerings in different scenarios.

NOTES.—This figure plots the model-predicted share of universities offering each sport against the counterfactual predicted shares. Sample includes all universities in the Ivy Plus, Elite LAC, and Other Elite Private tiers.