UNIVERSITY OF CAMBRIDGE
Faculty of Economics

Cambridge Working Papers in Economics

Cambridge Working Papers in Economics: 2056

# DO ELITE UNIVERSITIES PRACTISE MERITOCRATIC ADMISSIONS? EVIDENCE FROM CAMBRIDGE

Debopam Bhattachary
(University of Cambridge)

Renata Rabovic
(University of Cambridge)

22 June 2020

The merit-vs-diversity balance in university-admissions remains a controversial issue. Statistical analysis of these problems is jeopardized by applicant characteristics observed by admission-officers but unobserved by researchers. Using administrative microdata from the two-stage Cambridge admission-process, we compare post-entry exam-scores of directly admitted *h*-type students with *g*-types entering via the "pool"–a second-round clearing-mechanism. Better performance by the latter implies higher admission-standards for *g*-types, irrespective of the unobservability problem. We find strong evidence of higher admission-standards for males in STEM/Economics, and a weak one for private-school applicants. The gender-gap weakens over time for a cohort, and is non-evident in Law/Medicine.

# Do Elite Universities Practise Meritocratic Admissions? Evidence from Cambridge*

Debopam Bhattacharya†
University of Cambridge

Renata Rabovic
University of Cambridge

22 June 2020.

**Abstract**

The merit-vs-diversity balance in university-admissions remains a controversial issue. Statistical analysis of these problems is jeopardized by applicant characteristics observed by admission-officers but unobserved by researchers. Using administrative microdata from the two-stage Cambridge admission-process, we compare post-entry exam-scores of directly admitted $h$-type students with $g$-types entering via the "pool" – a second-round clearing-mechanism. Better performance by the latter implies higher admission-standards for $g$-types, irrespective of the unobservability problem. We find strong evidence of higher admission-standards for males in STEM/Economics, and a weak one for private-school applicants. The gender-gap weakens over time for a cohort, and is non-evident in Law/Medicine.

*The opinions expressed in this article are solely those of the authors, and do not reflect the views of the Faculty of Economics, the University of Cambridge and its colleges or the European Research Council.

†The authors acknowledge financial support from the European Research Council via grant number 681565, EDWEL. They also express deep gratitude to the Admissions Research Steering Group, Dr Alexa Horner and the Central Admissions Office at Cambridge for access to the data and Dr Julia Shvets for helpful discussions. Address for correspondence: debobhatta@gmail.com

# 1 Introduction

Top universities around the world are routinely criticized for encouraging social elitism in their undergraduate admissions, and not contributing adequately to inter-generational mobility. On the other hand, the universities' own attempts at promoting diversity are viewed by others as undermining academic merit for catering to populist political demand. Another common criticism is that under-prepared students admitted via affirmative action cannot cope with the workload, which hurts their ultimate career-prospects (Sander 2004, Arcidiacono et al 2016). These issues are frequently covered by the news media, discussed by politicians, and have even resulted in high-profile lawsuits. In the United Kingdom, much of this debate concerns undergraduate admissions to Oxford and Cambridge, collectively called 'Oxbridge'. Unlike the top US colleges, they are primarily state-funded, and are often criticized for the high share of students from independent, i.e. fee-charging, private schools which typically cater to relatively wealthier households, as opposed to publicly funded state schools that are free to attend. For example, a highly cited 2018 Sutton Trust report revealed that between 2015-17, Oxford and Cambridge "recruit(ed) more students from eight top schools than almost 3,000 other UK state schools put together" (BBC 7 Dec, 2018) – six of these eight being expensive, independent schools. On the other hand, the fraction of offers made to female applicants has been steadily rising in both universities, and has been close to or higher than 50% in recent years.[1] In their admission statements, both universities claim to be looking *solely* for academic promise irrespective of social background.[2] Given this context, an interesting question is whether and to what extent do the aggregate admission statistics result from merit-based entry.

In this paper, we formally investigate the extent of meritocracy in Cambridge admissions by asking whether all applicants are held to the same bar of academic promise as measured by expected future exam performance, with social/demographic background playing a role only insofar as they predict academic merit. To answer this question, we use recent applicant-level administrative micro-data from Cambridge, matched with post-admission performance of admitted students from different socio-demographic groups in Cambridge's blindly marked, internal exams. Had the admission-bar been higher for one group, the marginal – but not necessarily the average – entrants from that group should perform better than the marginal entrants from the favoured group. This

[1] https://www.ox.ac.uk/sites/files/oxford/Admissions%20Report%202019.pdf

https://www.undergraduate.study.cam.ac.uk/files/publications/ug_admissions_statistics_2019_cycle_0.pdf

[2] ox.ac.uk/about/facts-and-figures/admissions-statistics and undergraduate.study.cam.ac.uk/applying/what-are-we-looking-for

idea, going back to Becker 1957, has been used by researchers to test for fair decision-making in various applied settings including college admissions, e.g. Bertrand et al 2010. However, detecting the marginal applicant is impossible when admissions are based on many separate indicators of academic merit, some of which are used by admission officers but unobserved by researchers (e.g. the quality of confidential reference letters), leading to the so-called "infra-marginality" problem. Furthermore, the common practice of predicting future academic performance for *applicants* by regressing the outcomes of *admitted* students on their pre-entry characteristics and test-scores suffers from classic sample selection bias, typically biasing estimated coefficients towards zero. These problems apply to many other settings where discrimination is a concern, such as law-enforcement and legal sentencing, c.f. Arnold et al 2018.

In the present paper, our strategy to get around the infra-marginality problem is to use the two-staged nature of Cambridge's admission process, explained below in details. The key feature of this process that is of interest to us is that a subgroup of applicants enter Cambridge through the 'pool'. Students placed in the pool are those perceived to be less meritorious than the first-round direct admits, but worth keeping in contention. The data can identify this subgroup. The applicants perceived to be the best among the pooled may eventually be admitted, depending on how many spots are available after the first round of admissions. Therefore, comparing the post-entry exam performance of *pooled* admits of social group $g$ with that of the *direct* admits of group $h$ can reveal whether the criteria used to rank applicants in the first round are statistically consistent with a meritocratic goal. In particular, if the former perform systematically better in post-admission exams, then that suggests that $g$-types face a higher bar of academic merit for direct admissions than $h$-types, whether or not the analyst observes all of the relevant applicant characteristics observed by the admission officers.

**Short Literature Review**: Holzer and Neumark 2000 provide a broad economic perspective on affirmative action policies in various real-life settings; Arcidiacono et al 2015 summarize the existing empirical literature on affirmative action in (mostly US) college admissions with a focus on applicant-race. Bhattacharya et al 2017 analyzed the question of meritocratic admissions using a small dataset for a single subject at a different UK university; their approach was *not* outcome-based, and addressed the infra-marginality problem using the assumption that observables and unobservables were positively associated. Educational sociologists in the UK have documented lower application success rates at the most selective (Russell Group) universities for ethnic minority and state-school students with similar observable qualifications as their ethnically white and private school counterparts, c.f. Boliver 2013 and Zimdars et al 2009. These latter studies are somewhat descriptive, and do not consider the role of unobservables, nor any post-admission outcomes. In contrast to them, the present paper uses an *outcome-based* test (viz. post-entry exam performance) in the tradition of Becker 1957, and an institutional detail – as opposed to an assumption – to overcome the infra-marginality problem caused by unobservables. In addition, we document some interesting empirical contrasts between STEM versus competitive non-STEM subjects, and the progression of performance-gaps through the years for a cohort which, to our knowledge, are novel in the literature, and throw new light on the question of meritocracy in university admissions.

## 2 Cambridge Admissions

In the UK, high-school leavers at age 18+ apply to university to study a specific subject such as Mathematics, History, Law etc. Cambridge and Oxford are among the most prestigious and selective among UK universities, and a student can apply to one but not both in the same year. At Cambridge (similar to Oxford), a student applies through one of the constituent colleges which conducts assessment for that applicant. Each college has an intended range for the number of places for each subject in which it admits, e.g. 10-12 spots for Mathematics, 7-8 spots for Economics etc. The exact assessment procedure varies by subject, but the general procedure consists of the following steps. Applications are initially assessed on the basis of past indicators of academic ability, including an initial entrance exam in some but not all subjects. The most promising ones are then invited to a further in-house assessment – usually consisting of an academic interview and, in some subjects, a written test – and the rest are rejected. Applicants who appear strongest on the basis of all available information, including the in-house assessments and school-leaving board exams e.g. International Baccalaureate (IB) or GCSE/A-levels, are admitted directly by the college they had applied to.[3] A subset of the remaining students who are deemed relatively strong by the application college are placed in a common, university-wide 'pool'. If a college has not filled all its places after the first round, then it can admit one or more students from the common pool in the second round.

During each academic year following admissions, all admitted students attend common, university-wide lectures for each paper (i.e. 'course' in US terminology) that constitutes their subject, and small-group teaching sessions called 'supervisions', some of which are taught by academics affiliated

[3] Typically, A-level and IB results are not published by the time of assessments. In these cases, the offers are conditional upon reaching a certain minimum standard in those exams.

with their own college and some by those in other colleges. All students sit a common, end-of-year exam called the 'Tripos' in these papers at the end of each year of their study. These exams are assessed centrally at the university-level, and marked blindly. We use the percentage of marks obtained at the first year Tripos, standardized by subject, as our primary 'outcome' of interest. We then use performance in later years to examine the longer term validity of our main conclusions.[4]

## 2.1 Econometric Model and Identification Strategy

Let the two groups under consideration be denoted by $g$ and $h$, e.g. $g$ can be males and $h$ females, or $h$ can be state-school educated and $g$ privately educated, etc. Denote by $A_g$ the academic ability of a $g$-type applicant, as inferred by admission officers. The admission decision can be summarized via three thresholds $g_2 < g_3 < g_1$ such that if $A_g > g_1$, then the applicant is admitted directly; if $A_g < g_2$ then s/he is rejected straight away; and if $g_2 < A_g < g_1$, then the candidate is put in the pool. Finally, if $g_3 < A_g < g_1$, then the candidate is admitted from the pool, provided there are unfilled spots after direct admissions in the first round. For $h$-type applicants, denote the analogous quantities by $A_h$ and $h_1, h_2, h_3$ respectively.

Following admissions, the annual exam performance (Tripos score) is then generated by $Y_g = A_g + \varepsilon_g$ for $g$-type and $Y_h = A_h + \varepsilon_h$ for $h$-type respectively, where $\varepsilon_g$ and $\varepsilon_h$ are stochastic noise terms that affect exam scores over and above $A_g$ and $A_h$. We will assume that admission officers' inference is correct on average, i.e. $E(\varepsilon_g|A_g) = 0$ and $E(\varepsilon_h|A_h) = 0$. This may be justified as follows. Suppose the admission officers observe the set of characteristics $X_g$ (some components of which may be unobservable to the researcher) for each group $g$, and from this infer the ability $A_g = E(Y_g|X_g)$. Then $\varepsilon_g = Y_g - E(Y_g|X_g)$, implying by definition that $E(\varepsilon_g|X_g) = 0$, and therefore $E(\varepsilon_g|A_g) = 0$, since $A_g$ is solely a function of $X_g$. In particular, this implies that with $F_{A_g}(\cdot)$ denoting the marginal distribution of $A_g$, any set $C$ with $F_{A_g}(\cdot)$-positive probability, we have that:

$$E(\varepsilon_g|A_g \in C) = \int_{a \in C} \underbrace{E(\varepsilon_g|A_g = a)}_{=0} dF_{A_g}(a)\, da = 0. \tag{1}$$

This implication will be used below.

**Detecting Sign of Threshold Differences**: The econometrician observes the distributions

[4] Our analysis of meritocracy is thus at the aggregate university level, combining all colleges together; in our empirical work, we will investigate the robustness of our findings to the inclusion of college-fixed effects, where relevant.

of Tripos scores of all entrants and, in particular, those of $g$-type admits entering via the pool: $Y_g = A_g + \varepsilon_g | g_3 < A_g < g_1$, and of Non-Pooled (i.e. directly admitted) $h$-type admits: $Y_h = A_h + \varepsilon_h | A_h \geq h_1$. Therefore, the observable average Tripos score of Pooled $g$-type admits satisfies

$$\begin{aligned} E[Y_g | g_3 < A_g < g_1] &\equiv E\left[A_g + \varepsilon_g | g_3 < A_g < g_1\right] \\ &= E\left[A_g | g_3 < A_g < g_1\right] + \underbrace{E\left[\varepsilon_g | g_3 < A_g < g_1\right]}_{=0 \text{ by } (1)} \\ &= E\left[A_g | g_3 < A_g < g_1\right] < g_1, \end{aligned} \tag{2}$$

while the observable average Tripos Score of Non-Pooled $h$-type admits satisfies

$$\begin{aligned} E\left[Y_h | A_h \geq h_1\right] &\equiv E\left[A_h + \varepsilon_h | A_h \geq h_1\right] \\ &= E\left[A_h | A_h \geq h_1\right] + \underbrace{E\left[\varepsilon_h | A_h \geq h_1\right]}_{=0 \text{ by } (1)} \\ &= E\left[A_h | A_h \geq h_1\right] \geq h_1. \end{aligned} \tag{3}$$

It follows from (2) and (3) that

$$g_1 - h_1 > E\left[A_g + \varepsilon_g | g_3 < A_g < g_1\right] - E\left[A_h + \varepsilon_h | A_h \geq h_1\right]. \tag{4}$$

The RHS of (4) therefore provides a lower bound on the difference in thresholds $g_1 - h_1$. In particular, if the average Tripos score for pooled $g$-type admits is (weakly) higher than that of non-pooled $h$-type admits, i.e.

$$E\left[A_g + \varepsilon_g | g_3 < A_g < g_1\right] \geq E\left[A_h + \varepsilon_h | A_h \geq h_1\right], \tag{5}$$

then $g_1 > h_1$. In particular, if $\Pr(g_3 < A_g < g_1) > 0$, and $\Pr(A_h > h_1) > 0$ – corresponding to the likely scenario that perceived ability is continuously distributed – then even *equality* of mean Tripos scores, i.e. $E\left[A_g + \varepsilon_g | g_3 < A_g < g_1\right] = E\left[A_h + \varepsilon_h | A_h \geq h_1\right]$ will also imply $g_1 > h_1$.

In interpreting the above inequalities, we do not and cannot distinguish between the possibilities that any observed deviations from meritocratic admissions arises either (i) because admission officers have systematically biased beliefs against $g$-types but use the same bar of expected performance for everyone, or (ii) because they have correct beliefs on average but use a systematically higher threshold for $g$-types as part of an affirmative action plan. Note also that reversal of inequality (5) i.e. $E\left[A_g + \varepsilon_g | g_3 < A_g < g_1\right] < E\left[A_h + \varepsilon_h | A_h \geq h_1\right]$ is consistent with both $g_1 > h_1$ and $g_1 < h_1$, and is therefore inconclusive.

**Risk and Stochastic Dominance**: Suppose admission-officers are risk averse and base the admission-decision on $B_g \equiv E(U(Y_g)|X_g)$ for a concave, increasing $U(\cdot)$, and with $B_g > g_1$ implying direct admission etc. as above. Define $\varepsilon_g = U(Y_g) - B_g$, $\varepsilon_h = U(Y_h) - B_h$ and note that $E[\varepsilon_g|g_3 < B_g < g_1] = 0 = E[U(Y_h)|B_h \geq h_1]$ by the same logic as above. Now, if the distribution of $Y_g|g_3 < B_g < g_1$ first order stochastic dominates (FOSD) that of $Y_h|B_h \geq h_1$ (as found in the empirics reported below), then $E[U(Y_g)|g_3 < B_g < g_1] \geq E[U(Y_h)|B_h \geq h_1]$ for *all* increasing $U(\cdot)$, and therefore,

$$g_1 - h_1 > E[U(Y_g)|g_3 < B_g < g_1] - E[U(Y_h)|B_h \geq h_1] \geq 0. \tag{6}$$

So if FOSD holds, then the conclusion of higher admission standards for $g$-types remains robust to risk-aversion considerations.

## 3 Data and Key Variables

In this study, we utilize anonymized administrative micro-data from the Cambridge Admissions Office, containing pre-entry qualifications and characteristics for 47804 applicants, including the applicant's sex and type of school but not the reference letters and student's statement-of-purpose. We also observe performance on past exams and subject-specific admissions/aptitude tests, if any. We do not observe any indicator of interview performance obtained at admission-assessments. Among those who eventually enter, we also observe their Tripos percentage score in each available year of study. Specifically, our dataset consists of students who applied to study Economics, Engineering, Mathematics, Natural Sciences, Law, or Medicine between 2013 – 2017. Out of those, 8877 were accepted. The final first-year sample, for whom we have both pre-entry information and post-entry exam scores, contains 8354 observations.

At the end of each year, students in each subject take standardized exams in multiple papers which are blindly marked; the only identifying information an examiner observes is the candidate's centrally assigned 5-digit registration number on the script. For years 1, 2, and 3, we observe aggregate scores expressed as percentages of the total. In the first year, most papers are compulsory, while optional papers are more common in later years of study. For the regression analysis below, we standardize these variables (i.e. subtract mean, divide by standard deviation) by subject.

For the school-type part of the analysis, three school categories are considered. *Maintained* schools are state schools funded by the government, and therefore free of charge. *Independent* schools are private schools that charge fees, and typically enrol children from higher socioeconomic status households. The category *Other and overseas* is used for schools outside the UK.

### 3.1 Summary Statistics

The detailed summary statistics are reported in Appendix Table A.2. The first-year sample consists of 63.91% of males. The majority of students were from the UK (71.83%), of which 46% and 32% came from maintained and independent schools, respectively. Nearly 20% of the first year sample was admitted from the pool. The sample-sizes are roughly equal across five years.

The admission success-rates and first year raw average scores by subject are displayed in Appendix Table A.1. The overall success rates were nearly equal between males and females (18.89% vs 18.04% resp.), with some heterogeneity across subjects. Males outperformed females on average in the first-year exams in all subjects. The largest differences were in Mathematics and Engineering, where females scored 5 and 7 percentage points lower than males on average, respectively. In terms of school-type, students from independent schools had a higher success rate (28.7% vs 20.65%) overall relative to those from maintained schools. In respect of the first year average exam score, independent school admits slightly outperformed their maintained school counterparts but were outperformed by admits from overseas schools.

## 4 Gender Results

We first apply the methods outlined in Section 2.1 to the above datasets with the groups $g$ and $h$ denoting males and females, respectively. In Figure 1, Panel A, we plot the cumulative distribution function of first year Tripos percentage scores for four subgroups of admitted students, viz. pooled male, non-pooled male, pooled female and non-pooled female, broken up by ESTEM (STEM plus Economics) versus non-ESTEM, i.e. Medicine and Law. In ESTEM, we see clear evidence that the non-pooled, i.e. directly admitted males distribution first-order stochastically dominate the rest, followed by pooled males, non-pooled females and, finally pooled females. The fact that pooled males have stochastically higher Tripos-scores than non-pooled i.e. directly admitted female entrants *throughout the distribution* suggests a higher admission bar for males, c.f. equations (4) and (6) above. In contrast, pooled females are dominated by both the pooled and the non-pooled males. On the other hand, in the non-ESTEM subjects, the performance distribution of pooled males and non-pooled females can be seen to be very similar. The same pattern is also observed within each subject comprising ESTEM and non-ESTEM, these subject-specific graphs are reported in the Appendix Fig A.1.

The graphical evidence above is then corroborated by regression analysis on the combined sample of pooled male and non-pooled female entrants to check whether the observed gender gaps are statistically significant. We do this in Table 1, Panel A, Column (1), where we regress Tripos percentage (standardized across subjects) on a dummy indicating pooled males. The positive, statistically highly significant coefficient 0.139 on the dummy variable implies that compared to non-pooled female applicants, the pooled males score an average of 0.14 standard deviations higher, which represents the lower bound on $g_1 - h_1$ (c.f. (4)), the male-female difference in admission-thresholds.

There are three key challenges in interpreting this estimate as an indication of lower admission standards for females. Firstly, males could apply to subjects that are more selective in the first round of admissions than females. Secondly, the effect could be driven by college specific characteristics, if for example, pooled males were typically admitted to a different set of colleges than non-pooled females. Thirdly, the effect could stem from decentralized admissions if males applied disproportionately to colleges that happen to be more selective in the first round of admissions. We alleviate these concerns by including fixed effects for subject, offer college, and apply college. In Column (4), the estimate shrinks to 0.101 with p-value less than 1%. Columns (5) and (6) confirm the previous finding that these results are being driven entirely by ESTEM. Within ESTEM itself, Economics and Mathematics had the largest gender-gap, while all the others were positive and with p-values below 5%; for brevity, we do not report the subject-specific results here; Appendix Fig A.1 reports the subject-specific graphs.

**Persistence over Time**: To understand the nature of the differences found in first year results, we investigate whether the gap persists as the students progress through the years. As the last examination period that we observe is Spring 2018, second- and third-year Tripos results are available solely for the earlier cohorts of students. Therefore, we re-perform the above exercise using these shorter samples and compare the results to the original sample used above; Table 1, Panel B, shows that the first-year gap is very similar in the three samples for ESTEM subjects. The key substantive finding from Panel B is that as the students progressed through the years, the first year gender gap appears to fall. For example, when judged by third-year Tripos performance in ESTEM subjects (Columns (3) and (4) in Table 1, Panel B) the gap appears to have shrunk considerably (0.167 vs -0.049 standard deviations). This narrowing could either reflect better acclimatization of female students to exam conditions, or could be a result of efficient sorting into optional papers which are typically offered in later years. We also performed the same exercise for non-ESTEM subjects and found that the estimates remain statistically insignificant in later years.

### 4.1 Possible Mechanisms

**Pre-Admission Qualifications:** We first investigate whether and to what extent can the observed gender difference be attributable to specific pre-admission qualifications.

To do this, we first use the subsample of enrolled students (pooled males plus non-pooled females) who have taken GCSE or IGCSE, a common, compulsory board exam taken at the age of 16 in the UK. For each exam, a pupil receives grades $A^*$, A, B,...,U, where $A^*$ is the highest and U is the lowest. We use the number of $A^*$s as a measure of prior ability and redo our analysis. Column (1) of Table 2 shows that in this subsample, the gap between pooled males and non-pooled females based on their first-year performance in the Cambridge Tripos remains statistically significant, as in the entire sample. However, in respect of the pre-entry $A^*$ count, Column (2) shows that the non-pooled females in fact *dominate* pooled males.

Next, we consider the Sixth Term Examination Papers in Mathematics (STEP) test. After receiving a conditional acceptance in Maths, a student is typically required to take parts 2 and 3 of this test, and eventual admission depends on that performance. The STEP 2 and STEP 3 scores of non-pooled females are statistically indistinguishable from pooled males (c.f. Columns (3) – (5) in Table 2).

The above evidence suggest that our observed gender gaps cannot be attributable to pre-admission qualification and preparation to the extent revealed by standardized tests. They also suggest that differential impact of post-admission teaching methods and/or any implicit bias in (blind) internal assessment may play a role in producing the observed gender-gap. In other words, our measure of merit, i.e. the Tripos score, is to be interpreted as inclusive of the post-entry academic experience of admitted individuals, as in any outcome-based test of fair decisions.

**Peer Effects:** The gap between pooled males and non-pooled females could be driven by peer effects. In our sample, only 36.09% of students are females. It is conceivable that females were affected negatively by environments with fewer female classmates. To test this, we investigate whether the gap in first-year performance was lower in subject-year-college combinations with a higher share of females. We do not find any evidence supporting this hypothesis (c.f. Column (6) in Table 2).

**The Effect of Being Pooled:** One alternative explanation of our finding is that the knowledge of being pooled prompted the pooled *male* students to exert more effort relative to students who were directly admitted.[5] We investigate this through a difference-in-differences strategy, using a reform introduced in 2015. Effectively, before 2015 and not after, pooled admits who were eventually admitted by the college they applied to knew that they had been pooled, and these serve as the treatment group. Pooled males who were admitted by a different college than the one they had applied to serve as our control group, since this group knew that they had been pooled both before and after the reform. The statistically insignificant estimate of *Same college* $\times$ *Year = 2013* in Column (7) of Table 2 shows that the parallel trend assumption holds for the pre-reform period. Column (8) shows that the knowledge of being pooled does not affect the performance of pooled males in first year exams. Therefore, it is unlikely that our findings are driven by the effect of being pooled.

Three other robustness checks include:

(i) **Women-Only Colleges:** Cambridge has three colleges that admit only women; these colleges were established later than most others, and therefore have had slightly different history and trajectories. Our gender results remain robust to including college fixed effects, and separately, dropping the women-only colleges from the sample.

(ii) **Other Background Controls:** Another possible concern is false attribution, e.g. the gender-gap stems from school-type if non-pooled females come mainly from state schools whereas pooled males come mainly from private schools. To check this, we included dummies for school-type (independent, maintained, and other and overseas) and domicile category (EU excluding the UK, the UK, and other and overseas) as additional controls.

(iii) **Year Effects:** We included fixed effects for apply year and exam year to ensure that the scores are comparable across years.

Table 3 shows that the performance-gap between pooled males and non-pooled females is robust to all three issues discussed above.

**Alternative Identification Strategy**: Lastly, we consider a small subset of pooled students who were not picked by other colleges, and eventually admitted by the same college that had placed them into the pool. These students may be viewed, albeit somewhat crudely[6], as 'marginal' admits in the sense that they were considered weaker than the direct admits in the first round

[5] We are grateful to Julia Shvets for discussions on this point.

[6] A negligibly small number are pooled while the deciding college is waiting for additional information on them.

by the pooling college and also not picked by other colleges from the pool, but did eventually get admitted. Column (5) of Table 3 shows that, conditional on subject and offer college fixed effects, the males in this subgroup scored 0.142 standard deviations higher in first-year exams than females in this subgroup. This result lends further support to our finding that, judging by the first year Tripos performance, the admission threshold for females is lower than that for males. The number of such students is small (342), hence we did not break this up by subjects; but since the overall difference is large, we conjecture that it would be higher still if we considered ESTEM students alone.

## 5 School-Type Results

The political debate around Oxbridge admissions is primarily centred on school-type rather than gender, i.e. whether the admission process enhances social diversity and mobility by admitting sufficiently many students from maintained, i.e. state-funded and free schools. To investigate this issue from the perspective of meritocracy, we apply the methods of Section 2.1 using as group $g$ the candidates from independent and overseas schools; the two groups are combined because both of them attract candidates from relatively more affluent backgrounds. We use the label *Other schools* for this combined group. This group is then compared with group $h$ consisting of candidates from maintained UK schools.

Plotting the CDFs of first year Tripos scores for pooled and non-pooled subgroups within these two categories (Figure 1, Panel B) shows that the non-pooled dominate the pooled, within and across school-types. A t-test for equality of the means between pooled others and non-pooled (i.e. directly admitted) maintained yields a p-value of 0.7 with the estimated mean being slightly higher for the former. This implies the inequality $g_1 - h_1 > 0$ (see discussion following equation (5)). Column (3) in Table 4 shows that the mean-difference increases to 0.118 with a small p-value once fixed effects for subject and offer college are included. However, it becomes insignificant once we include fixed effects for apply college c.f. Column (4). Further, Columns (5) – (8) show that the gap remains statistically insignificant in later years.

The above findings suggest that while there is some suggestive evidence of lower admission-standards for maintained school applicants, the strength of this evidence is weaker than the gender results.

# 6 Conclusion

Analyzing merit versus diversity balance in university admissions is difficult due to applicant characteristics observed by admission-officers but unobservable to analysts, leading to the well-known infra-marginality problem. We address this problem in the context of undergraduate admissions at Cambridge University, using 5 years' (2013-17) administrative micro-data on applicants, matched with blindly marked, end-of-year exam scores for the admitted students, which we take as our measure of merit. To overcome the infra-marginality issue, we utilize the two-staged feature of Cambridge admissions, involving students who were directly admitted in a first round vis-a-vis those not directly admitted in the first round but admitted in a second round clearing mechanism called the 'pool'. If students of (demographic) type $g$ admitted from the pool performed better in post-entry exams than $h$-types directly admitted in the first round, then we can conclude that entry-bar for direct admissions was higher for $g$-types, irrespective of the unobservability problem. Applying this idea to our data, we find strong evidence that in STEM fields and economics, male applicants faced a higher admission threshold. This finding is resilient to a variety of robustness checks, and are suggestive of a genuine underlying regularity. We find no evidence to suggest that pre-admission standardized test scores are any worse for female applicants, or that the relative scarcity of female students has a detrimental effect on performance, e.g. via lower morale and confidence. The gender-gap becomes less apparent as students advanced through the years, and is not detected in *any* year of study in the non-STEM but highly competitive subjects of Law and Medicine.

The evidence for school-type is less conclusive; there is at most weak evidence that the admission bar was lower for applicants from maintained (i.e. state-funded and free) schools, who on average are economically less privileged than those attending fee-charging private schools.

## References


[1] Arcidiacono, P., Aucejo, E.M. and Hotz, V.J., 2016. University differences in the graduation of minorities in STEM fields: Evidence from California. American Economic Review, 106(3), pp.525-62.

[2] Arcidiacono, P., Lovenheim, M. and Zhu, M., 2015. Affirmative action in undergraduate education. Annual Review of Economics, 7(1), pp.487-518.

[3] Arnold, D., Dobbie, W. and Yang, C.S., 2018. Racial bias in bail decisions. The Quarterly Journal of Economics, 133(4), pp.1885-1932.

[4] Becker, G. 1957. The Economics of Discrimination. University of Chicago Press.

[5] Bertrand, M., Hanna, R. and Mullainathan, S., 2010. Affirmative action in education: Evidence from engineering college admissions in India. Journal of Public Economics, 94(1-2), pp.16-29.

[6] Bhattacharya, D., Kanaya, S. and Stevens, M., 2017. Are university admissions academically fair?. Review of Economics and Statistics, 99(3), pp.449-464.

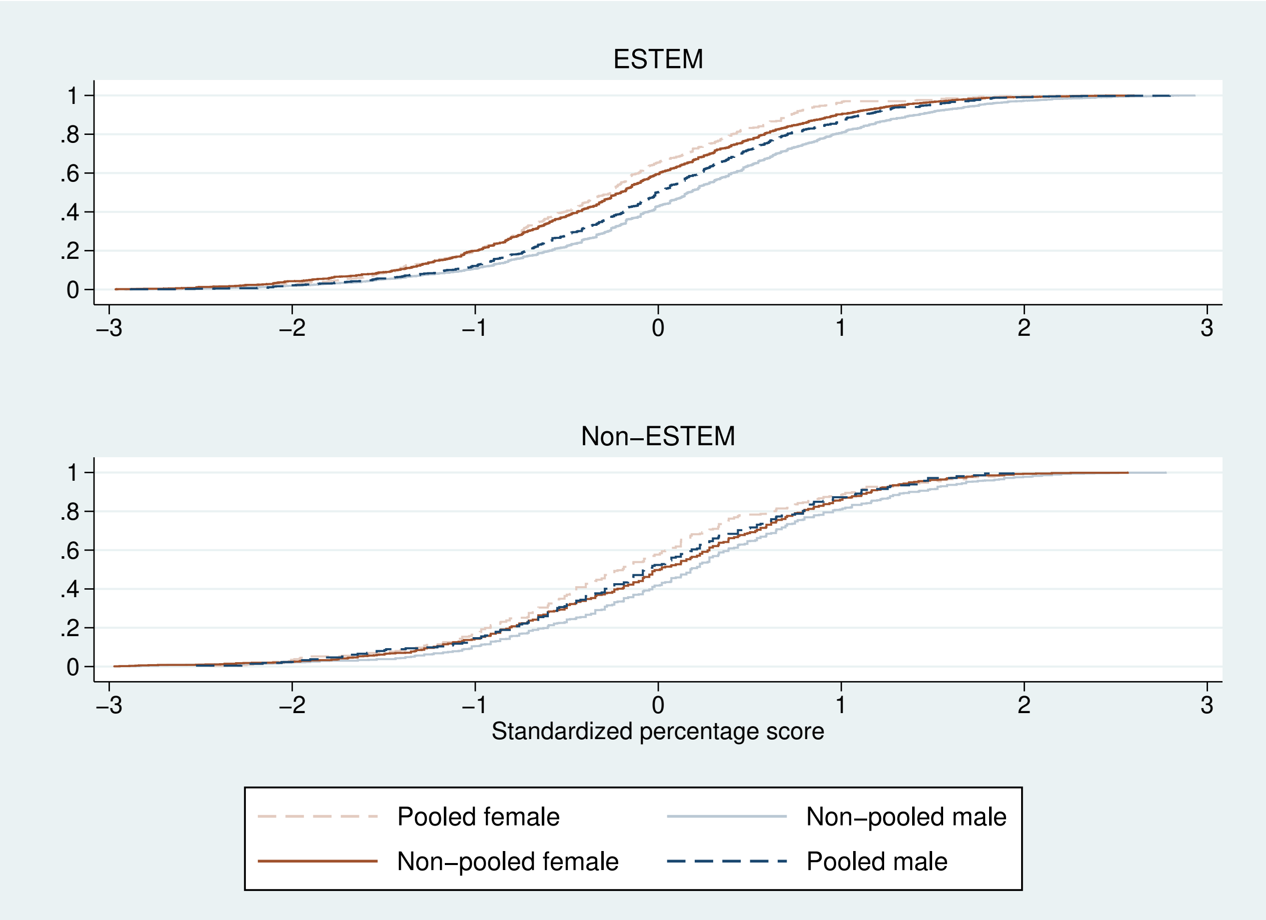
[7] Boliver, V., 2013. How fair is access to more prestigious UK universities?. The British Journal of Sociology, 64(2), pp.344-364.

[8] Holzer, H. and Neumark, D., 2000. Assessing affirmative action. Journal of Economic Literature, 38(3), pp. 483-568.

[9] Sander, R.H., 2004. A systemic analysis of affirmative action in American law schools. Stanford Law Review, 57, pp. 367-483.

[10] Zimdars, A., Sullivan, A. and Heath, A., 2009. Elite higher education admissions in the arts and sciences: Is cultural capital the key?. Sociology, 43(4), pp.648-666.

Panel A. Gender




Panel B. School-type

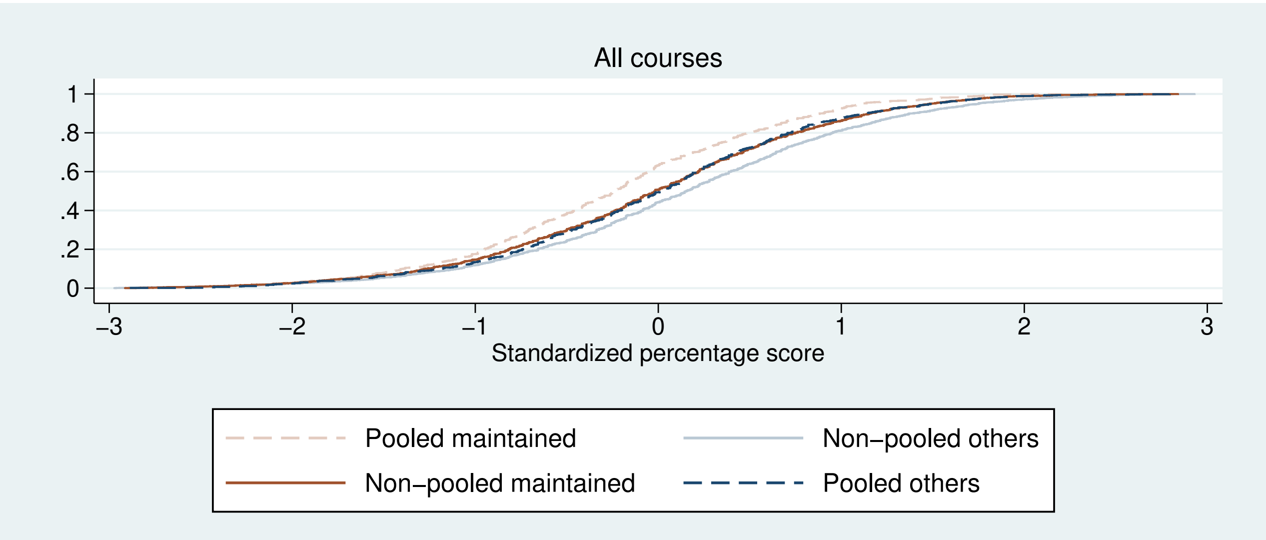



Figure 1: First-year Tripos percentage scores by pool status and gender and school type.

*Note.* The graph shows the cumulative distribution function of first-year Tripos percentage scores for different subgroups of students. In Panel A, the functions are plotted separately for ESTEM and non-ESTEM subjects, whereas in Panel B the two sets of subjects are combined.

Table 1: The gap between pooled males and non-pooled females.

*Panel A. First year*

| | All subjects | | | | ESTEM | Non-ESTEM |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Pooled male | 0.139 | 0.185 | 0.251 | 0.101 | 0.169 | -0.088 |
| | (0.049) | (0.056) | (0.045) | (0.051) | (0.061) | (0.095) |
| Observations | 3245 | 3245 | 3245 | 3245 | 2214 | 1031 |
| Subject FE | | ✓ | ✓ | ✓ | ✓ | ✓ |
| Offer college FE | | | ✓ | ✓ | ✓ | ✓ |
| Apply college FE | | | | ✓ | ✓ | ✓ |

*Panel B. Later years: ESTEM*

| | Second-year sample | | Third-year sample | |
|---|---|---|---|---|
| | Year 1<br>(1) | Year 2<br>(2) | Year 1<br>(3) | Year 3<br>(4) |
| Pooled male | 0.154 | 0.081 | 0.167 | -0.049 |
| | (0.067) | (0.069) | (0.077) | (0.095) |
| Observations | 1641 | 1641 | 1154 | 1154 |
| Subject FE | ✓ | ✓ | ✓ | ✓ |
| Offer college FE | ✓ | ✓ | ✓ | ✓ |
| Apply college FE | ✓ | ✓ | ✓ | ✓ |

*Note.* Each column reports the results from a different OLS regression. *Pooled male:* a dummy variable that equals one for pooled males and zero for non-pooled females. Standard errors clustered at the offer college level (Columns (1) and (2) in Panel A) and robust standard errors (the rest of the columns in Panel A and all columns in Panel B) are reported in parentheses. *Panel A.* Sample: pooled males and non-pooled females for whom first-year Tripos scores are available. Dependent variable: the standardized score obtained in first-year Tripos exams. The sets of subjects used to obtain the corresponding results are indicated in the column header. *Panel B.* First-, second-, and third-year samples consist of students in ESTEM subjects for whom first-, second-, and third-year Tripos scores are available, respectively. Dependent variable: indicated in the column header, where *Year 1*, *2*, and *3* stand for standardized scores obtained in first-, second-, and third-year Tripos exams, respectively.

Table 2: Gender gap: mechanisms.

| | A* count | | STEP | | | Peer effects | Being pooled | |
|---|---|---|---|---|---|---|---|---|
| | Year 1 (1) | A* count (2) | Year 1 (3) | STEP 2 (4) | STEP 3 (5) | Year 1 (6) | Year 1 (7) | Year 1 (8) |
| Pooled male | 0.118 (0.066) | -0.372 (0.053) | 0.331 (0.183) | -0.024 (0.117) | 0.033 (0.117) | 0.105 (0.113) | | |
| Share female | | | | | | -0.025 (0.215) | | |
| Pooled male × Share female | | | | | | 0.243 (0.310) | | |
| Same college | | | | | | | -0.341 (0.217) | -0.276 (0.114) |
| Same college × Year=2013 | | | | | | | 0.126 (0.274) | |
| Same college × Year=2015 | | | | | | | 0.090 (0.224) | |
| Same college × Year=2016 | | | | | | | -0.072 (0.272) | |
| Same college × Year=2017 | | | | | | | 0.175 (0.288) | |
| Same college × Post-reform | | | | | | | | 0.001 (0.124) |
| Observations | 1661 | 1661 | 296 | 296 | 296 | 2214 | 992 | 992 |
| Subject FE | ✓ | ✓ | | | | ✓ | | |
| Offer college FE | ✓ | | ✓ | | | ✓ | | |
| Apply college FE | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | | |
| Apply year FE | | | | | | | ✓ | ✓ |

*Note.* Each column reports the results from a different OLS regression. The multicolumn headers represent potential mechanisms. Sample for *A* count*: pooled males and non-pooled females in ESTEM subjects for whom the number of A*s obtained in the (I)GCSE exam is available. Sample for *STEP*: pooled males and non-pooled females who applied to Mathematics and for whom STEP 2 and STEP 3 scores are available. Sample for *Peer effects*: pooled males and non-pooled females in ESTEM subjects for whom first-year Tripos scores are available. Sample for *Being pooled*: pooled males for whom first-year Tripos scores are available. Dependent variable: indicated in the column header, where *Year 1*, *A* count*, and *STEP 2* and *3* stand for the standardized score obtained in first-year Tripos exams, the standardized A* count obtained in (I)GCSE exams, and the standardized STEP 2 and 3 scores. *Pooled male:* a dummy variable that equals one for pooled males and zero for non-pooled females. *Share female*: for each student, we calculate the share of females in his/her subject-year-college combination (including that student). *Same college:* a dummy variable that equals one for pooled males admitted by the college they applied to and zero for pooled males admitted by another college. *Year:* application year. *Post-reform:* a dummy variable that equals one for the years 2015, 2016, and 2017; and zero for the years 2013 and 2014. Robust standard errors (Columns (1) – (6)) and standard errors clustered at the offer college level (Columns (7) and (8)) are reported in parentheses.

Table 3: Gender gap: robustness checks.

| | Main (1) | Year FE (2) | Mixed (3) | Controls (4) | Marginal (5) |
|---|---|---|---|---|---|
| Pooled male | 0.169 | 0.170 | 0.173 | 0.178 | |
| | (0.061) | (0.061) | (0.061) | (0.061) | |
| Marginal male | | | | | 0.142 |
| | | | | | (0.105) |
| Observations | 2214 | 2214 | 2093 | 2214 | 342 |
| Subject FE | ✓ | ✓ | ✓ | ✓ | ✓ |
| Offer college FE | ✓ | ✓ | ✓ | ✓ | ✓ |
| Apply college FE | ✓ | ✓ | ✓ | ✓ | |

*Note.* Each column reports the results from a different OLS regression. Sample: in Columns (1) – (4), the initial sample consists of pooled males and non-pooled females in ESTEM subjects for whom first-year Tripos scores are available. In Column (5), the sample consists of students in all subjects taken back from the pool by the college they applied to. Dependent variable: the standardized score obtained in first-year Tripos exams. *Pooled male:* a dummy variable that equals one for pooled males and zero for non-pooled females. *Marginal male:* a dummy variable that equals one for marginal males and zero for marginal females. *Main:* main specification. *Year FE:* with fixed effects for apply year and exam year. *Mixed:* only mixed gender colleges. *Controls:* additional controls for school-type (independent, maintained, and other and overseas) and domicile category (EU excluding the UK, UK, other and overseas). Robust standard errors are reported in parentheses.

Table 4: The gap between non-pooled students from maintained schools and pooled students from other schools.

| | First-year sample | | | | Second-year sample | | Third-year sample | |
|---|---|---|---|---|---|---|---|---|
| | Year 1 (1) | Year 1 (2) | Year 1 (3) | Year 1 (4) | Year 1 (5) | Year 2 (6) | Year 1 (7) | Year 3 (8) |
| Pooled others | 0.014 | 0.010 | 0.118 | 0.033 | -0.033 | -0.071 | -0.065 | -0.092 |
| | (0.046) | (0.047) | (0.065) | (0.054) | (0.062) | (0.064) | (0.078) | (0.087) |
| Observations | 3982 | 3982 | 3982 | 3982 | 3013 | 3013 | 2136 | 2136 |
| Subject FE | | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Offer college FE | | | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Apply college FE | | | | ✓ | ✓ | ✓ | ✓ | ✓ |

*Note.* Each column reports the results from a different OLS regression. First-, second-, and third-year samples consist of students for whom first-, second-, and third-year Tripos scores are available, respectively. Sample: non-pooled students from maintained schools and pooled students from other schools. Dependent variable: indicated in the column header, where *Year 1, 2,* and *3* stand for standardized scores obtained in first-, second-, and third-year Tripos exams, respectively. *Pooled others:* a dummy variable that equals one for pooled candidates from other schools and zero for non-pooled candidates from maintained schools. Standard errors clustered at the offer college level (Columns (1) – (2)) and robust standard errors (Columns (3) – (8)) are reported in parentheses.

# Online Appendix for "Do Elite Universities Practise Meritocratic Admissions? Evidence from Cambridge"

Debopam Bhattacharya
University of Cambridge

Renata Rabovič
University of Cambridge

June 20, 2020

## Appendix A Additional Tables and Figures

Table A.1: Admission success-rates and first-year outcomes.

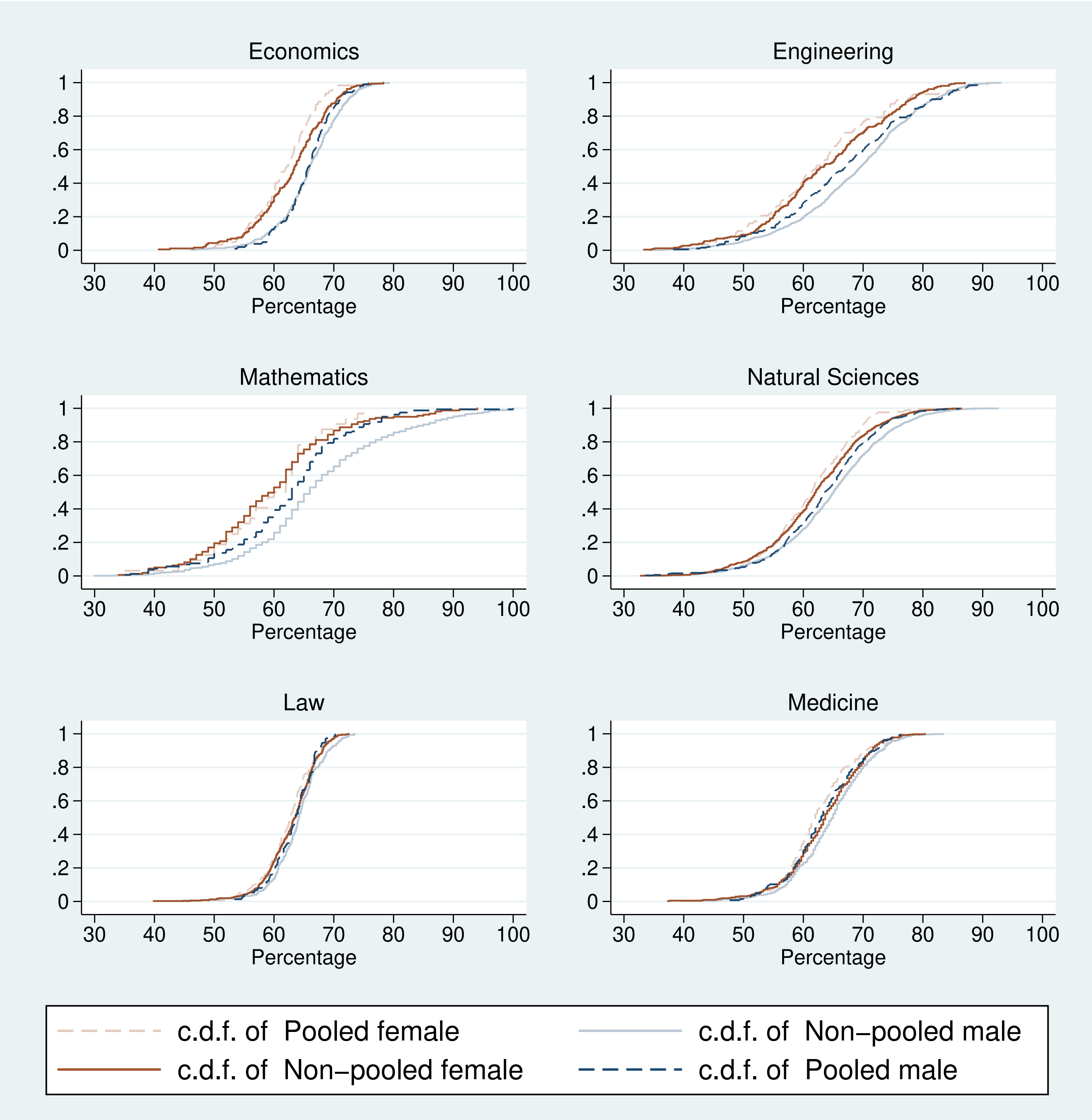
| Variable | Gender | | School-type | | |
|---|---|---|---|---|---|
| | Female | Male | Independent | Maintained | Other & overseas |
| **Sucess-rate** | | | | | |
| *All subjects* | 18.04 | 18.89 | 28.71 | 20.65 | 11.12 |
| Economics | 13.78 | 14.06 | 20.19 | 18.15 | 7.67 |
| Engineering | 18.36 | 14.30 | 22.91 | 16.97 | 9.82 |
| Law | 19.99 | 20.84 | 31.34 | 23.02 | 15.53 |
| Mathematics | 12.21 | 20.29 | 33.75 | 14.97 | 16.59 |
| Medicine | 16.06 | 22.13 | 31.32 | 26.59 | 4.81 |
| Natural Sciences | 21.41 | 22.87 | 33.33 | 23.73 | 13.23 |
| **First-year score** | | | | | |
| *All subjects* | 62.26 | 65.63 | 64.42 | 63.36 | 66.64 |
| Economics | 62.60 | 65.70 | 64.56 | 64.45 | 65.18 |
| Engineering | 62.99 | 68.23 | 66.17 | 65.59 | 69.91 |
| Law | 62.80 | 63.79 | 63.07 | 63.15 | 63.39 |
| Mathematics | 59.36 | 66.34 | 64.37 | 62.96 | 70.18 |
| Medicine | 63.24 | 64.40 | 64.15 | 63.29 | 66.27 |
| Natural Sciences | 61.70 | 64.50 | 64.01 | 62.44 | 64.81 |

*Note. Success-rate:* Percent accepted out of initial applications. *First-year score*: Raw percentage of total. The table provides success-rates and mean first-year Tripos percentage scores for all six subjects bunched together and for each of them separately.

Table A.2: Summary statistics: first-year admits.

| Variable | Share |
|---|---|
| **Subject** | |
| Economics | 8.92 |
| Engineering | 16.93 |
| Law | 11.02 |
| Mathematics | 14.08 |
| Medicine | 15.15 |
| Natural Sciences | 33.90 |
| **Gender** | |
| Female | 36.09 |
| Male | 63.91 |
| **School-type** | |
| Independent | 32.00 |
| Maintained | 46.11 |
| Other and overseas | 21.89 |
| **Domicile category** | |
| EU (excluding UK) | 9.77 |
| UK | 71.83 |
| Other and overseas | 18.31 |
| **Pool** | |
| Pooled | 21.00 |
| Non-pooled | 79.00 |
| **Application year** | |
| 2013 | 19.60 |
| 2014 | 20.35 |
| 2015 | 20.24 |
| 2016 | 20.34 |
| 2017 | 19.48 |

*Note.* The table reports percentage shares of the respective categories using the first-year sample of 8354 students.




*Note.* Based on the first-year sample. *Pooled:* admitted from the pool. The graph shows the cumulative distribution function of first-year Tripos percentage scores for four subgroups of students – pooled females, non-pooled females, pooled males, and non-pooled males – by subject.

Figure A.1: First-year Tripos percentage scores by gender, pool status, and subject.