The Institution of Merit: A Study of Chinese College Admissions*

Jimmy Chan† Zhi Wang‡ Shunyi Zhao§

February 13, 2020

Abstract

We empirically estimate a model of Chinese college admissions to study factors behind the geographical distribution of admissions at elite universities in China. We find that elite Chinese universities have remarkably similar preferences. Contrary to popular beliefs, there is little evidence that elite Chinese universities favor home-province applicants. Rather, they consistently prefer students from richer provinces even though college tuition fees in China are fixed by the government at a low and affordable level and have little impact on admissions. One possibility is that universities prefer students with higher ability and students from richer provinces have higher ability because such provinces spend more on pre-college education.

1 Introduction

It is commonly accepted that places at elite colleges should be allocated primarily by merit and not price, and students, regardless of their social

*For helpful comments, we thank Sambuddha Ghosh, Xican Xi, Ruli Xiao, Daniel Xu, Jin Yan, as well as participants of various seminars and conferences. We are grateful to Deng Pan and Wenzhang Zhang for sharing data on the distribution of test scores. Guangrui Chen, Hao Chen, Kaiyang Chen, Yunyun Chu, Qingyuan Dong, Hanzhi Huang, Wan Li, Sijie Lin, Xinyi Long, Xinshan Ren, Yangfan Sun, Feiluona Tu'erhong, Juan Peng, Huijie Wang, Haoyang Wu, Yuhe Wu, Shijie Xu, Xifan Xue, Qingyuan Zhang, Xueqing Zhang, Dezhi Zhu helped us collect data. The research is supported by the Research Grants Council of Hong Kong (GRF no. 14623518).

†The Chinese University of Hong Kong. Email: jimmyhingchan@cuhk.edu.hk

‡School of Economics, Fudan University, and Shanghai Institute of International Finance and Economics. Email: wangzhi@fudan.edu.cn

§University of Pennsylvania. Email: shzhao@sas.upenn.edu

backgrounds, should have fair and equal opportunities to succeed. In the words of Rawls (1971),

"...those who are at the same level of talent and ability, and have the same willingness to use them, should have the same prospects of success regardless of their initial place in the social system, that is, irrespective of the income class into which they are born."[1]

While in an ideal society these two goals—admissions by merit and by fair and equal opportunity—need not be in conflict, in reality, the social situation that one is born into has a large impact on one's chance of academic success, and there is little consensus as to whether the same standard of merit should apply to applicants with very different backgrounds. In the US, affirmative action has long been controversial. There are on-going lawsuits about whether the admissions criteria at elite colleges discriminate against White and Asian Americans.

In this paper, we focus on the case of China. Much like in the US, the allocation of places at the top universities is a highly contentious subject in China, although the concern is geographic rather than ethnic. The Chinese college market is very different from the US's. Most of the top Chinese universities are public, and their tuition fees are set by the government at an affordable level. Matching between students and universities is centralized. Applicants submit a preference list instead of applying to individual universities. Admission decisions are largely based on the National College Entrance Examination.

For a millennium, the imperial civil examination was the gateway to officialdom and a source of legitimacy to the ruling elites. The National College Entrance Examination plays a similar role in present times.[2] Ideally, it would be best to apply the same standard to the entire country. But in reality, education funding and the demand for quality differ significantly across provinces. The current practice is thus a compromise; it maintains a uniform structure but gives some freedom to provinces and universities. While the same subjects are tested in every province, the contents and standards of the exams may vary. Students do not compete across provinces. Instead, each university assigns an admission quota to each province. Within a province, universities are required to admit applicants with the highest test scores, but they enjoy considerable autonomy in choosing quotas.

There were over eight hundred public universities in China in 2017. At the top were universities belonging to Project 211 and Project 985, govern-

[1]Rawls is referring to all advantageous social positions, not just university places.

[2]It is common to refer to the student who scores the highest in the exam as a *zhuang yuan*, the highest honor in the imperial civil examination.

ment programs that provide extra resources to the top universities.[3] Project 211 contains 112 universities. The more prestigious Project 985 includes 39 of the Project-211 universities.[4] A common grievance against the current system is that universities in Beijing and Shanghai are biased in favor of local applicants. Since there are more Project-211 and Project-985 universities in Shanghai and Beijing, even if all universities are similarly biased, a higher fraction of places in the top universities will go to applicants in Beijing and Shanghai.[5] Figure 1 shows the fraction of exam takers admitted by a Project-211 or Project-985 university between 2013 and 2015 in the 27 provinces for which we have data. The top two performers are Qinghai, a remote and sparsely populated province bordering Tibet, and Tianjin, a provincial-level city near Beijing. Beijing and Shanghai are ranked third and fourth, respectively. About 11.75 percent of the exam takers in the top four provinces are admitted by a Project-211 university, which is almost double the national average. At the other extreme are mostly poorer provinces in central China. Only about 3.37 percent of the exam takers in the bottom five provinces are admitted by a Project-211 university.

Our objective is to understand the factors behind the current allocation of places. Do universities prefer students from their home provinces? Or do students prefer universities in their home provinces? Since different provinces use different exams, we cannot directly compare test scores across provinces to decide whether a university is favoring applicants from its home province. In Section 4, we model the Chinese college admission system as a quota-setting game. We show that in equilibrium each university should allocate its places so that, to the university, the value of the marginal student (i.e., the applicant with the lowest test score among the admitted) is the same in every province from which the university admits students. Thus, a university's cutoff scores reveal its preferences over applicants from different provinces. The intuition is straightforward. If a university finds the value of the marginal student from province A to be higher than that of a student from province B, then it can increase its utility by transferring a place from province B to province A. While different universities may value the same student differently, there is no reason why they should systematically favor one non-home province over another non-home province. This suggests that we can use the average difference in cutoff scores (i.e., the scores of the marginal students)

[3]The name of Project 211 comes from its objective to promote "the Top 100 Chinese universities in the 21st Century." The name of Project 985 refers to "May 1998," the month the program was launched.

[4]Every Project-985 university is also a part of Project 211.

[5]There are 26 Project-211 universities in Beijing and 10 in Shanghai. The corresponding numbers for Project-985 universities are eight and four, respectively.

between two provinces among universities located in a third province to identify the average value of the students from a certain province to universities outside of the province. The home bias of a university can then be backed out by comparing the value a university assigns to the home province's students to the value universities in other provinces assign to the same students.

We apply our methodology to the cutoff data of 107 Project-211 universities from 2013-2016 in Section 7.[6] We find that universities in our sample have remarkably similar preferences. There is little difference in preferences between higher- and lower-ranked universities and between universities in richer and poorer provinces. Universities that are higher ranked or under provincial control are more likely to be home biased. But overall the magnitude of the bias is very small. Contrary to the popular belief, there is no evidence that universities in Beijing and Shanghai favor home-province students. Rather, universities systematically prefer students from richer provinces. The tendency explains not only the high admission rates for Beijing and Shanghai students, but also the low admission rates for the students of the poorer provinces in China.

To evaluate the quantitative importance of our findings, we follow Akyol and Krishna (2017) and use the admission quotas from 2013 through 2015 to estimate a logit model for the demand for university places. In each province, we recover from the data the average utility the students of that province would receive from attending a particular university. An important finding is that students prefer universities closer to home, which explains why a high fraction of students attend universities in their home provinces. We conduct counterfactual analysis with the estimated student and university preferences in Section 8. The unequal geographical representation at elite Chinese universities is mainly caused by the universities' systematic preference for students from richer provinces. Eliminating universities' systematic preferences largely equalizes the distribution of places. By contrast, removing the home bias of the universities has almost no impact on the geographical distribution of places. Eliminating students' preference for universities closer to home significantly reduces the fraction of students attending universities in their home provinces, but further increases the fraction of Project-211 uni-

[6]There are 112 universities in Project 211. Among them, four universities have two separate campuses, and each is operated independently. Therefore, we treat each campus as an independent university. These universities are: North China Electric Power University (Beijing and Baoding), China University of Geosciences (Beijing and Wuhan), China University of Petroleum (Beijing and Shandong), and China University of Mining and Technology (Beijing and Xuzhou). In addition, the medical schools of Peking University, Fudan University, and Shanghai Jiao Tong University are treated as independent universities as they all independently recruit students.

versity places going to students from Beijing and Shanghai, as more Beijing and Shanghai students are willing to attend universities in other provinces.

Why do Chinese universities prefer students from richer provinces? Unlike US colleges, Chinese universities do not engage in price discrimination and do not rely on alumni donations. One plausible explanation is that students from richer provinces are perceived to be of better academic quality on average. As richer provinces spend more on pre-college education, their students may be better trained.[7] Alternatively, the university preference may reflect a bias against poorer provinces. But if that is the case, the bias is widespread and not confined to universities in richer provinces. Regardless of the reason for this preference, our results suggest that replacing the current quota system by a new standardized exam for the whole country might not lead to a more equal geographical distribution of places, as the new test scores are still likely to be positively correlated with income. We discuss the policy implications of our results in Section 9.

2 Literature

Our basic approach is similar to recent works that adopt a structural approach to study the US college market (Epple, Romano, and Sieg (2006); Fu (2014)).[8] Like them, we assume that universities prefer students with higher ability and choose admission policies to maximize utility, while students choose their favorite universities among those that would admit them.

The details of our model, however, are very different from theirs, as the structure of the Chinese college market is very different from the US's. First, price discrimination is important in the US market. Second, the US college market is decentralized. Students must apply to individual colleges, which apply their own standards to evaluate applicants. Epple, Romano, and Sieg (2006) and Fu (2014) explicitly incorporate these features into their models.

By contrast, in China, college tuitions are set by the government and not by universities. Admissions are centralized and are based on test scores within a province. The only major decision a university needs to make is to allocate its capacity among provinces. The simple structure means that our model is more transparent and easier to estimate than those of Epple, Romano, and Sieg (2006) and Fu (2014).

In Epple, Romano, and Sieg (2006) and Fu (2014), students' prefer-

[7] In fact, pre-college education spending explains the universities' provincial preference marginally better than GDP per capita.

[8] For a review of the earlier literature, see Ehrenberg (2004).

ences for college may vary with ability and income.[9] Since we do not have individual-level data, we follow Akyol and Krishna (2017) and use cutoff scores to estimate a logit demand system.[10]

Whether the same admission standard should be applied to applicants from very different backgrounds is at the heart of the debate about affirmative action in US college admissions. Fryer and Loury (2005) provide an overview of the main issues. Bowen and Bok (1998) address the long-term consequences of affirmative action. Chan and Eyster (2003) show that the gap in standardized test scores between the marginal majority and minority students reflects a college's preference for under-represented minority students.[11] We apply this idea to identify universities' systematic preferences for applicants from different provinces.

There is a growing body of literature on the Chinese college market. Chen and Kesten (2017) examine the properties of the matching mechanism currently in use in China.[12] Du and Zhong (2018) use cutoff scores of Project-211 universities to estimate students' preferences under the assumption that quotas are exogenous.[13] Gao, Huang, White and Zhong (2018) study how the top two universities in China set provincial quotas to compete for the best students. As far as we know, our paper is the first structural model of the Chinese college market. Our model encompasses both the demand and supply sides of the market. We conduct counterfactual analysis to understand which factor drives the current geographical distribution of college places.

3 Background

In 2017, there were about 800 public universities and 400 private ones in China. Among public universities, 117 are under the supervision of the Ministry of Education, with the rest under the supervision of the education bureaus of the provinces where the universities are located. Tuition fees at public universities are set by the government. In 2017, the average tuition was about 6,000 RMB per year, which is affordable to most Chinese families. Admissions to Project-211 universities are very competitive. Most students

[9] Abdulkadiroğlu, Agarwal and Pathak (2007) and Hastings, Kane and Staiger (2009) use micro data to estimate students' preferences over colleges.

[10] Avery, Glickman, Boxby and Metrick (2012) use a similar approach to derive a ranking of US colleges.

[11] Also see Fryer, Loury and Yuret (2008).

[12] For more on the matching mechanism used in China, see Wu and Zhong (2014, 2016), Bo, Liu, Shiu, Song and Zhou (2018), and Lien, Zheng and Zhong (2017).

[13] They find that while school quality is the most important determinant of student preferences, there is evidence that students prefer local universities.

admitted by a Project-211 university would enroll.

The college admission process is the largest centralized system in the world. In 2017, nine million high-school graduates in 31 provinces competed for four million college places. At the heart of the process is the National College Entrance Examination. Chinese high-school education is divided into two streams: arts and science. For each stream, the exam includes the same subjects in every province. Students are required to take the exam in their home province (i.e., the province of their household registration). Before 2000, the same exam papers were used in all provinces. Since 2001, provinces have been allowed to design their own curricula and exam papers.[14] In 2013, 16 provinces used their own exam papers, and the other 15 adopted exam papers developed by the Ministry of Education. The ministry has since partially recentralized curriculum design. By 2016, 26 provinces had adopted a common national curriculum. Nevertheless, the exam papers used in these provinces may still differ because there are three sets of exam questions for each subject and provinces can choose the set that suits the standard of their students. Since grading is coordinated at the provincial level, grading standards may differ among provinces that use the same set of exam questions. See Appendix B for more details.

Admissions are divided into general admissions (*tong kao tong zhao*) and special admissions. The former, which is the focus of our study, accounts for 70 percent of the places at the universities in our sample. The remaining places are divided among a variety of special programs. Some target applicants from a particular background; others require military or social service upon graduation.[15] General admissions are based on a point system. A student's points are essentially her test score in the National College Entrance Examination, plus bonus points for special achievements or minority background.

The quota-setting process is decentralized. Each year, every university submits an admission plan to the supervising government department for approval. The admission plan details the number of places in each major that are allocated to each province through each channel. Admission plans typically do not change drastically from one year to the next. But universities

[14]Before 2000, the setting of exam questions was overseen by the Ministry of Education.

[15]As shown in Table A1 in the appendix, general admissions take up about 70 percent of the places in the universities in our sample. The admissions of special programs account for another 10 percent. The other 20 percent of the total admissions are used for the admissions of the autonomous admission program, the exam-exempt programs, the pre-college program, and the admissions allocated to Hong Kong, Macau, and Taiwan students, and those allocated to foreign students.

are free to fine-tune the plans.[16] The central government may set broad policy goals, but the universities can decide how to meet them.

The approved admission plans are announced before the National College Entrance Examination in June. Students submit preference lists after learning their scores. In each province, universities are divided into three tiers, with places in the first tier allocated first, followed by places in the second and the third. All Project-211 universities belong to the first tier. The places in each tier are allocated by the so-called Shanghai Mechanism, a variant of the delayed-acceptance mechanism (Chen and Kesten, 2016), whereby a student's preference list is divided into multiple choice bands, each containing up to three to five universities. Within a choice band, students are matched to universities through a standard delayed-acceptance mechanism. The assignments are made permanent at the end of a choice band, and unmatched students will move on to the next choice band, in which the same process recurs. For every university on the preference list, an applicant may list a number of preferred majors. In addition, she may indicate that she is willing to accept majors not on the list. In practice, most applicants to top universities use this option to maximize their chance of admittance.

4 Model

In this section we introduce a model that captures the main features of the current general admission process. The model provides a foundation for our empirical work.

Preferences. Let $\mathcal{I}$ denote a set of m provinces. In each province i there is a continuum of applicants of size λ_i. Each applicant l in province i is characterized by a test score $s_{i,l} \in S_i \equiv [\underline{s}_i, \overline{s}_i] \subseteq \Re$. The provincial distribution of s_i, denoted by G_i, is smooth, and the density $g_i(s_i)$ is strictly positive for all $s_i \in S_i$.

There is a set $\mathcal{J}$ of n universities. Each university j has a capacity K^j. There is a fixed cost to recruit students from a province.[17] A university may choose not to recruit from a province from which it cannot attract good applicants. In our sample, some lower-ranked universities in poorer provinces do not recruit students from provinces far from the university. Let $C(j) \subseteq \mathcal{I}$ denote the set of provinces from which university j admits students, and $D(i) \subseteq \mathcal{J}$ the set of universities that recruits students from province i.

[16] The quota distribution is not fixed. In our data, they change slightly from year to year. See Section 6.2.2.

[17] Universities often hold recruiment talks in the top high schools of a province to attract better applicants.

Throughout, we take $C(j)$ and $D(i)$ as exogenous.

Normalize the reservation utility for not attending any university to zero. The utility an applicant l in province i receives from attending university j is

$$u_{i,l}^{j} = d_i^j + \epsilon_{i,l}^{j},$$

where d_i^j is a constant that measures the popularity of university j in province i, and $\epsilon_{i,l}^{j}$ is a random component that represents an applicant's idiosyncratic taste for university j.[18] We assume that $\epsilon_{i,l}^{j}$ is identically and independently distributed across l and j, and its distribution, denoted by F_i, is smooth and has strictly positive density everywhere on $\Re$.

We assume that in each province the number of students who prefer university j to the outside option is greater than the total capacity of all universities.

Assumption 1 *For any province i and university $j \in D(i)$,*

$$\lambda_i F_i\left(\epsilon_i^j \geq -d_i^j\right) \geq \sum_j K^j.$$

Assumption 1 ensures that any university j can always admit enough students to fill its capacity in any province.

A university j receives a utility of $\alpha_i(s_i) + \gamma_i^j$ from admitting an applicant with score s_i from province i. The objective of a university is to maximize its total utility. The function $\alpha_i(s_i)$ captures the common value to all universities of an applicant's desirable traits, including academic and non-academic ability, as well as any systematic bias in favor of or against a particular province. We assume that $\alpha_i(s_i)$ is continuous, strictly increasing in s_i and goes to infinity as s_i goes to $\overline{s}_i$. The last property ensures that in equilibrium university j admits students in every province $i \in C(j)$. The variable γ_i^j measures university j's idiosyncractic preference for an applicant from province i. For example, a university may prefer applicants from its home province because it receives funding from the provincial government to support them. Any bias or disagreement about the average ability of the applicants from a province will also be absorbed into γ_i^j.

Admission Process. The admission process consists of two stages. In the first stage, each university j independently chooses a set of positive quotas $q^j = \left\{q_i^j\right\}_{i \in C(j)}$with $\sum_{i \in C(j)} q_i^j = K^j$, where q_i^j is the number of places assigned to province i. In the second stage, applicants apply to universities.

[18]Alternatively, $\epsilon_{i,l}^{j}$ can be interpreted as a decision error as in a quantal response model (McKelvey and Palfrey 1995).

The preferences of the universities and students, as well as the distributions $\{G_i, F_i\}_{i\in\mathcal{I}}$, are common knowledge among universities and students.

In the second stage, since there is no aggregate uncertainty, each applicant in equilibrium would be able to perfectly predict which universities he could gain acceptance to. Hence, we can, without loss of generality, assume that each student applies to at most one university and enrolls upon acceptance.[19]

The application process is as follows. The applicants know their own test scores and preferences. Upon learning $q_i = \left\{q_i^j\right\}_{j\in D(i)}$, each simultaneously applies to one university or chooses the outside option. Because applicants do not compete across provinces, the application process in each province can be treated separately. An application strategy is a measurable function $\phi_i : S_i \times \Re^n \to C(i) \cup \{\emptyset\}$ where $\phi_i(s_i, u_{i,l})$ denotes either a university chosen by an applicant l or the outside option. Let F denote the distribution of $u_{i,l}$. Under ϕ_i, the number of applicants who apply to university j and score more than x is

$$H_i^j(\phi_i, x) \equiv \lambda_i \int_{s_i\geq x} \int_{u_{i,l}} \mathbf{I}_j(\phi_i(s_i, u_{i,l}))\, dF(u_i)\, dG_i(s_i),$$

where

$$\mathbf{I}_j(\phi_i(s_i, u_{i,l})) = \begin{cases} 1 & if \quad \phi_i(s_i, u_{i,l}) = j \\ 0 & if \quad \phi_i(s_i, u_{i,l}) \neq j \end{cases}.$$

Define

$$c_i^j\left(\phi_i, q_i^j\right) \equiv \begin{cases} \min\left(s_i'|H_i^j(\phi_i, s_i') = q_i^j\right) & if \quad \left\{s_i'|H_i^j(\phi_i, s_i') = q_i^j\right\} \neq \emptyset \\ -\infty & if \quad \left\{s_i'|H_i^j(\phi_i, s_i') = q_i^j\right\} = \emptyset \end{cases}.$$

We call c_i^j the cutoff score of university j in province i. Universities admit applicants with the highest test scores. If all applicants follow ϕ_i, then an applicant to university j is admitted if and only if he scores higher than $c_i^j\left(\phi_i, q_i^j\right)$. An applicant l with test score s_i receives $u_{i,l}^j$ if $\phi_i(s_i, u_{i,l}) = j$ and $s_i \geq c_i^j\left(\phi_i, q_i^j\right)$; otherwise, he receives zero.

Given cutoff scores $c_i = \left\{c_i^j\right\}_{j\in D(i)}$, the set of universities that are feasible for an applicant with score s_i is

$$J(c_i, s_i) \equiv \left\{j \in D(i)\,|s_i \geq c_i^j\right\}.$$

The best response for an applicant l in province i is

$$br(s_i, u_{i,l}; c_i) = \begin{cases} j & if \quad u_i^j \geq \max\left(\max_{k\in J(c_i,s_i)} u_i^k, 0\right) \\ \emptyset & if \quad \max_{k\in J(c_i,s_i)}\left(u_i^k\right) < 0 \end{cases}.$$

[19]Hence, in equilibrium applicants apply only to universities that are superior to the outside option.

We say that an application strategy ϕ_i is an equilibrium in a province i given quotas $q_i = \left\{q_i^j\right\}_{j \in D(i)}$ if $\phi_i(s_i, u_{i,l}) = br(s_i, u_{i,l}; c_i(\phi_i))$ for all s_i and $u_{i,l}$.

Proposition 1 *In any province $i \in \mathcal{I}$, given any q_i with $q_i^j \in (0, K^j]$ for any university $j \in D(i)$, the application game in province i has a unique equilibrium ϕ_i^*. In equilibrium, for each university j,*

$$H_i^j\left(\phi_i^*, c_i^j(\phi_i^*)\right) = q_i^j.$$

Furthermore, $c_i^j\left(\phi_i^*, q_i^j\right)$ strictly increases in d_i^j and strictly decreases in d_i^k for any $k \neq j$.

All proofs are in the appendix. The argument for existence is standard. Uniqueness follows from the strict monotonicity of $H_i^j\left(\phi_i, c_i^j\right)$ with respect to c_i^j. Assumption 1 ensures that every university is sufficiently popular in each province i that there is always enough demand to fill any feasible quota in equilibrium. The last part of Proposition 1 says that when a university becomes more popular in a province, it attracts more high-scoring applicants from other universities, pushing up its own cutoff score, while lowering those of the other universities.

The equilibrium defines within each province a matching between applicants and universities. The matching is stable in the sense that if one applicant strictly prefers a university to his own match, then his test score must be lower than any applicant admitted by this university.[20] Azevedo and Leshno (2016) show that, in a general setting that includes the application stage of our model as a special case, there is a unique stable matching that corresponds to a market equilibrium defined by cutoff scores. Thus, Proposition 1 can be taken as a corollary of the result of Azevedo and Leshno (2016).

In the first stage, universities set quotas, taking the applicants' equilibrium behavior as given. Let $c_i^*(q_i)$ denote the equilibrium in Proposition 1. The fraction of applicants with score s_i who optimally choose university j is equal to

$$\pi_i(j, s_i, J(c_i^*, s_i)) \equiv \begin{cases} \Pr\left[u_i^j \geq \max\left(\max_{k \in J\left(c_i^*, s_i\right)} u_i^k, 0\right)\right] & if \quad j \in J(c_i^*, s_i), \\ 0 & if \quad j \notin J(c_i^*, s_i). \end{cases} \tag{1}$$

Since F_i has strictly positive density everywhere, $\pi_i(j, s_i, c_i^*) > 0$ whenever $j \in J(c^*, s_i)$. The objective of each university is to choose $q^j = \left\{q_i^j\right\}_{i \in C(j)}$ to

[20]See Azevedo and Leshno (2016) for a formal definition of stability is this setting.

maximize

$$v^j\left(q^j, q^{-j}\right) = \sum_{i\in\mathcal{I}} \int_{s_i \geq c_i^{j*}(q_i)} \left(\alpha_i\left(s_i\right) + \gamma_i^j\right) \pi_i\left(j, s_i, J\left(c_i^*, s_i\right)\right) dG\left(s_i\right) \tag{2}$$

subject to the constraint that $\sum_{i\in C(j)} q_i^j = K^j$. Formally, $\{C(j)\}_{j\in\mathcal{J}}$, $\{v^j\}_{j\in\mathcal{J}}$, and $\{K^j\}_{j\in\mathcal{J}}$ define a quota-setting game. A vector $(q^1, ...q^n)$ is an equilibrium if it is a Nash equilibrium of the game.

Proposition 2 *There exists a unique equilibrium $q^* = (q_1^*, ...q_n^*)$ in the quota-setting game. Furthermore, for each university j there is a constant ψ^j such that for each province $i \in C(j)$*

$$\alpha_i\left(c_i^{j*}\left(q_i^*\right)\right) + \gamma_i^j = \psi^j. \tag{3}$$

The uniqueness of equilibrium follows from the fact that the number admitted in a province is strictly decreasing in cutoff score. If there were two distinct equilibria, then some universities in one equilibrium would have a higher cutoff score in every province in one equilibrium than the other. But this would imply that these universities as a group admit different numbers of applicants in the two equilibria.

The left-hand side of (3) is the utility university j receives from a marginal applicant in province i. In equilibrium, each university j must derive the same utility from the marginal applicant in every province in $C(j)$. Otherwise, it can gain by reallocating quotas from a province with a lower marginal utility to one with a higher marginal utility.

The equilibrium condition (3) allows us to distinguish whether it is a university which prefers the applicants from a certain province or the applicants from the province who prefer the university. Starting with an equilibrium q^*, if university j has a stronger idiosyncratic preference for applicants from province i (i.e., γ_i^j increases), then the utility its receives from the marginal student will increase. It will transfer more places to province i from other provinces until the marginal utilities are equalized again.

Alternatively, if university j has become more popular in province i (i.e., d_i^j becomes larger), its cutoff score will go up (Proposition 1), which, in turn, will raise the value of the marginal student from province i. University j will shift places from other provinces to province i until the marginal utilities are equalized.

In both cases university j assigns more places to province i, but the cutoff scores react differently. In the first case, the new equilibrium cutoff score decreases in province i and increases in other provinces, whereas in the second case the cutoff score increases in every province.

Remarks. The current model assumes that, given the quotas, universities can perfectly foresee the cutoff scores. Suppose the preferences of the applicants in province i are affected by some stochastic factor δ_i whose realization is common knowledge among the applicants but not observed by the universities. Let $\left\{\widetilde{c}_i^j\left(q_i, \delta\right)\right\}_{j \in D(i)}$ denote the unique set of market-clearing cutoff scores given δ_i and q_i. Following the same argument in Proposition 2, it is straightforward to show that (3) will hold in expectation. That is, in the equilibrium of the quota-setting game

$$E_\delta\left[\alpha_i\left(\widetilde{c}_i^j\left(q_i^*, \delta\right)\right)\right]+\gamma_i^j$$

will be the same in every province $i \in C(j)$.
Some Chinese universities may be required by the provincial governments to admit a minimum number of applicants from the home province. This type of numerical constraint can be easily incorporated into the university's maximization problem. A version of (3) will continue to hold in equilibrium. The shadow cost of any numerical constraints university j faces in province i will be absorbed into γ_i^j.

5 Empirical Methodology

5.1 University Side Estimation

We apply the first-order condition (3) to identify the preferences of the universities. To implement the idea, we need some additional assumptions.

While each university may be biased in favor of students from its home province, it should not favor one non-home province over another non-home province. The following assumption formalizes this intuition and decomposes γ_i^j into a home-bias component ρ^j and a zero-mean random component ξ_i^j.[21]

Assumption 2 *For each university j and each province i,*

$$\gamma_i^j=\rho^j I_i^j+\xi_i^j \tag{4}$$

where I_i^j is an indicator function that equals one when school j is in province i, ρ^j is a constant, and ξ_i^j is a random component with zero mean that is identically and independently distributed across provinces and universities.

[21] Note that random component ξ_i^j is known to the universities and students; they are stochastic only to the econometrician.

Thus far, the only restriction on α_i is that it strictly increases in s_i. To apply the model to the data, we need to pin down the functional form of α_i. In the data, the distribution of test scores varies both across provinces and over time. While a student's raw test score may be affected by the difficulty of the exam questions, his rank should be relatively more stable. We therefore first turn the test scores into percentile ranks. We then map each percentile rank into the value of a standard-normal random variable of the same percentile rank to obtain a *transformed* test score. Let Φ^{-1} denote the inverse distribution function of a standard normal distribution. For a student with percentile rank r, the transformed test score is $\Phi^{-1}(r)$. By construction, the transformed test score follows a standard normal distribution, and the transformed score of a student depends only on his percentile rank and not his raw test score conditional on rank. Finally, we assume that α_i is linear in the transformed test score with a slope of one in every province i.

Assumption 3 *For each province i,*

$$\alpha_i(s_i) = a_i + s_i, \tag{5}$$

where s_i is the transformed test score.

Since all our empirical work is done with the transformed test scores, we sometimes refer to s_i as the test score. To motivate Assumption 3, suppose that, in every province i, the unobserved "value" of a student, denoted by α_i, follows a normal distribution with mean $\overline{\alpha}_i$ variance σ^2, and that s_i is equal to α_i plus a noise term that is also normally and identically distributed with variance ξ^2. Then s_i will be normally distributed with variance $\sigma^2+\xi^2$, and the expectation of α_i conditional on s_i is

$$E(\alpha_i|s_i) = \frac{\xi^2}{\sigma^2+\xi^2}\overline{\alpha}_i + \frac{\sigma^2}{\sigma^2+\xi^2}s_i.$$

Assumption 3 can be obtained by normalizing $\sigma^2+\xi^2$ to one and assuming that

$$\alpha_i(s_i) = \frac{E(\alpha_i|s_i)}{\xi^2} = \frac{\overline{\alpha}_i\xi^2}{\sigma^2} + s_i.$$

Note that Assumption 3 implicitly assumes that both σ^2 and ξ^2 are constant across provinces. If they are not, then $E(\alpha_i|s_i)$ will be linear but have different slopes in different provinces. In Section 6.3, we show that Assumption 3 fits the data well.

Under Assumption 3, any difference in the average pre-college education quality in different provinces will be captured by the difference in the provincial constants. If universities are systematically biased in favor of students from a particular province, then the effect of this bias will also be reflected by the provincial constant. In the following, we refer to a_i as the average value of province i's students.

Assumptions 2 and 3 are our main identification assumptions. Let c_i^{j*} denote the cutoff (measured by the transformed score) of university j in province i. The first-order condition (3) implies that for any university j and any provinces $i, k \in C(j)$,

$$\alpha_i\left(c_i^{j*}\right) + \gamma_i^j = \alpha_k\left(c_k^{j*}\right) + \gamma_k^j. \tag{6}$$

Substituting (5) and (4) into (6) and rearranging terms, we have for any university j and any provinces $i, k \in C(j)$,

$$c_i^{j*} - c_k^{j*} = a_k - a_i + \gamma_k^j - \gamma_i^j. \tag{7}$$

In equilibrium, the difference in the cutoffs of university j in provinces i and k depends only on a_k, a_i, and the idiosyncratic preferences of university j. The students' preferences over universities affect a university's cutoffs but not the difference in the cutoffs between any two provinces. If a university becomes more popular among students, the equilibrium cutoffs will go up by the same amount in every province, leaving the cutoff gap between any two provinces unchanged.

An important implication of Equation (7) is that $a_1, ..., a_m$ can be identified up to a constant without fully specifying the student side of the model. By Equation (7) and Assumption 2,

$$E\left[c_k^{j*} - c_i^{j*}\right] = a_i - a_k$$

for any university j and any non-home provinces i and k from which the university recruits. Since university j is unbiased between any two non-home provinces, the fact that it is willing to accept a student with a lower cutoff score in non-home province k than in non-home province i indicates that the average value of province k's students is higher than that of province i's.

Let $D^*(i,k)$ denote the set of universities outside of province i or k that admits students from both. In our sample, the following condition is satisfied.

Condition 1 *$D^*(i,k)$ is non-empty for any provinces $i, k \in \mathcal{I}$.*

Condition 1 means that there is always an unbiased university to compare the students in any two provinces. For any two provinces i and k, $\sum_{j \in D^*(i,k)} \left(\xi_i^j - \xi_k^j\right) / \left|D^*(i,k)\right|$ goes to zero as $\left|D^*(i,k)\right|$ goes to infinity.[22]

[22]Let $|X|$ denote the cardinality of set X.

Hence, we can unbiasedly and consistently estimate $a_i - a_k$ by the mean estimator:

$$\frac{1}{|D^*(i,k)|}\sum_{j\in D^*(i,k)}\left(c_k^{j*}-c_i^{j*}\right). \tag{8}$$

Select one of the provinces as a base and refer to it as province 0. Normalize, a_0 to zero. We can estimate the value of non-base province i by

$$\widetilde{a}_i = \frac{1}{|D^*(i,0)|}\sum_{j\in D^*(i,0)}\left(c_0^{j*}-c_i^{j*}\right).$$

If all universities belonged to every $D^*(i,k)$ for any provinces i and k, then the mean estimators (8) would be transitive. For any three provinces i, k and l, the difference between $\widetilde{a}_i$ and $\widetilde{a}_l$ would be equal to the sum of the difference between $\widetilde{a}_i$ and $\widetilde{a}_j$ and the difference between $\widetilde{a}_j$ and $\widetilde{a}_l$. In this case, the choice of the base province does not affect the difference between $\widetilde{a}_i$ and $\widetilde{a}_k$.

However, in our data, slightly different sets of universities belong to different $D^*(i,k)$'s. As there is a Project-211 university in every province, at least two universities do not belong to each $D^*(i,k)$. Furthermore, a small number of universities do not admit students from every province. Under Assumptions 2 and 3, the mean estimators (8) will still be transitive asymptotically as the number of universities becomes large, but in a finite sample, the difference between $\widetilde{a}_i$ and $\widetilde{a}_k$, which are determined solely by the cutoffs of universities in $D^*(i,0)\cap D^*(k,0)$, would change with the choice of the base province.

There is no economic reason why a specific province should be chosen as the base. To avoid choosing a base arbitrarily, we adopt an alternative estimator of a_i that minimizes the idiosyncratic preferences of the universities. Denote the set of out-of-province universities that recruits from province i by $D^*(i)$, and the set of non-home provinces from which university j admits students by $C^*(j)$. Consider the minimization problem P1:

$$\min_{\{\widehat{a}_i\}_{i\in\mathcal{I}},\left\{\gamma_i^j\right\}_{i\in\mathcal{I},j\in D^*(i)}}\sum_{i\in\mathcal{I},j\in D^*(i)}\left(\gamma_i^j\right)^2$$

subject to (7) for any university j and any provinces $i,k\in C^*(j)$. Any solution $\{\widetilde{a}_i\}_{i\in\mathcal{I}}$ to this minimization problem must satisfy the first-order conditions: for all $i\in\mathcal{I}$,

$$\sum_{j\in D^*(i)}\sum_{k\in C^*(j)/i,}\frac{(\widehat{a}_i-\widehat{a}_k)}{|C^*(j)|} = \sum_{j\in D^*(i)}\sum_{k\in C^*(j)/i,}\frac{\left(c_k^{j*}-c_i^{j*}\right)}{|C^*(j)|}. \tag{9}$$

It is straightforward to see that if $\{\widehat{a}_i\}_{i\in\mathcal{I}}$ is a solution to P1, then any $\{\widehat{a}_i + k\}_{i\in\mathcal{I}}$ for any constant k is also a solution. Select some province as province 0 and normalize a_0 to 0.[23] Define the estimators $\{\widehat{a}_i\}_{i\in\mathcal{I}}$ as the solution to P1 with $\widehat{a}_0 = 0$.

Proposition 3 *Given Condition 1, there exists a unique solution $\{\widehat{a}_i\}_{i\in\mathcal{I}}$ to P1 with $\widehat{a}_0 = 0$. Furthermore, for each $i \in \mathcal{I}/\{0\}$, $\widehat{a}_i$ converges to a_i almost surely as $\min_{i,k\in\mathcal{I}} |D^*(i,k)|$ goes to infinity.*

Each university $j \in D^*(i)$ compares the students in province i to the students in $|C^*(j)| - 1$ other non-home provinces where it recruits students. By Assumptions 2, on average, the cutoff gap between province i and these provinces should be equal to the value gap between the students in province i and these provinces. That is,

$$\sum_{k\in C^*(j)/i,} (a_i - a_k) = \sum_{k\in C^*(j)/i,} E\left[c_k^{j*} - c_i^{j*}\right].$$

Equation (9) effectively assigns a weight of $1/|C^*(j)|$ to each comparison made by university j.[24] In the Appendix, we derive $\widehat{a}_i$ as a least-squares estimator. Since the difference between any $(\widehat{a}_i - \widehat{a}_k)$ is not affected by the choice of the base province, $\{\widehat{a}_i\}_{i\in\mathcal{I}/\{0\}}$ is more appealing conceptually than $\{\widetilde{a}_i\}_{i\in\mathcal{I}/\{0\}}$.

In general, $\{\widehat{a}_i\}_{i\in\mathcal{I}}$ will be different from the mean estimators $\{\widetilde{a}_i\}_{i\in\mathcal{I}}$ (with the same province 0) unless all universities belong to every $D^*(i,k)$ for any provinces i and k. However, in our data the difference between the two sets of estimators is negligible, as both $\widehat{a}_i$ and $\widetilde{a}_i$ are consistent and most of the universities in our sample recruit in all non-home provinces.[25]

Let $i^*(j)$ denote the home province of university j. Substituting (4) into (7) and rearranging terms, we have for every $k \in C^*(j)$

$$\rho^j = a_k - a_{i^*(j)} + c_k^{j*} - c_{i^*(j)}^{j*} + \xi_k^j - \xi_{i^*(j)}^j.$$

Intuitively, ρ^j is positive, meaning that the university j is biased in favor of home-province students, if its cutoff score in the home province is too

[23]Note that which province is selected as province 0 has no impact on the difference between any $\widehat{a}_i$ and $\widehat{a}_k$.

[24]By contrast, the mean estimator assigns a weight of one to every comparison to the base province.

[25]The difference becomes larger when, as a robustness test, we exclude a university from valuing the students in both home and neighboring provinces (See Section 7.1).

low relative to its cutoff scores in non-home provinces. Replacing $\{a_i\}_{i\in\mathcal{I}}$ by $\{\widehat{a}_i\}_{i\in\mathcal{I}}$, we can unbiasedly estimate ρ^j, the home bias of university j, by

$$\widehat{\rho}^j = \frac{1}{|C^*(j)|}\sum_{k\in C(j)}\left(\widehat{a}_k - \widehat{a}_{i^*(j)} + c_k^{j*} - c_{i^*(j)}^{j*}\right).$$

In Section 7.1.1, we explore how $\widehat{a}_i$ and $\widehat{\rho}^j$ correlate with provincial and university characteristics to shed more light on the preferences of the universities.

5.2 Student Side Estimation

The university side of the model is identified separately from the student side. Nevertheless, we need to model the student side in order to quantify the effect of the universities' preferences on admission outcomes. We assume that ξ_i^j follows a type-1 generalized extreme value distribution and model the students' demand by a logit demand system. In each province i, the number of students choosing j given cutoffs c_i is

$$\pi\left(j, s_i, J\left(c_i, s_i\right)\right) = \frac{\exp\left(d_i^j\right)}{1+\sum_{k\in J(c_i,s_i)}\exp\left(d_i^k\right)}. \tag{10}$$

We observe in the data the cutoff vector c_i^* and the quota vector q_i^* for each province i. Market clearing implies that in each province i and for each university $j \in D(i)$,

$$q_i^{j*} = \lambda_i \int_0^1 \frac{\exp\left(d_i^j\right)}{1+\sum_{k\in J\left(c_i^*,s_i\right)}\exp\left(d_i^k\right)} d\Phi\left(s_i\right). \tag{11}$$

In each province i, there are $|D(i)|$ equations and $|D(i)|$ unknowns. Following Akyol and Krishna (2017), we can uniquely numerically solve for $\left\{d_i^j\right\}_{j\in D(i)}$.

6 Data sources and descriptive statistics

6.1 Data sources

Our analysis focuses on 107 of the 119 universities in the Project-211 program, excluding universities (three military academies, one art academy, and five medical schools) that do not admit students primarily based on test scores, and universities in Tibet and Xinjiang.[26] Thirty-seven of the 107 universities belong to the more prestigious Project-985 program.[27] The geographic distribution of the universities is very unequal (Table 1). About 30 percent of the universities are in Beijing and Shanghai. While every province has at least one Project-211 university, nearly half do not have a Project-985 university. In 2013, the universities in our sample admitted through general admissions a total of about 349,000 freshmen, or 4.8 percent of the total exam takers, from the 27 provinces for which we have data on the distribution of test scores.[28]

Our main analysis relies on three sets of data: cutoff scores from 2013 through 2016, admission quotas from 2013 through 2015, and test-score distributions from 2013 through 2016. We briefly describe the data sources below. Details are provided in the appendix. Stream-specific cutoff scores in each province were collected from the universities' websites. When they are not available, we use figures from two publications that contain cutoff scores of major Chinese universities.[29] Enrollment data were downloaded from gaokao.chsi.com.cn, a website affiliated with the Ministry of Education (MOE), which provides enrollment numbers by major, student provincial origin, and admission channel. We also downloaded admission plans, when available, from universities' websites. The quotas in the admissions plan were very close to the actual enrollment numbers, indicating that admissions were mostly carried out according to plan. Test score distributions for each stream in 27 provinces and provincial-level cities were obtained through various sources, including provincial education examination authority websites and college guides. We do not have the test score distributions for Shaanxi,

[26]There are 112 Project-211 universities according to the official website of the Ministry of Education of the People's Republic of China. (http://old.moe.gov.cn//publicfiles/business/htmlfiles/moe/moe_94/201002/82762.html). Some universities have multiple campuses that admit students independently. For our analysis, each independent campus is treated as a separate university.

[27]Table A3 in the appendix presents the list of the 107 universities in our sample. Among the 107 universities in our sample of analysis, 83 are under the control of the central government and the other 24 are under provincial control. Among the 83 universities under central control (including all Project-985 universities), 71 are under the control of the Ministry of Education, and the other 12 are under the control of other central ministries and commissions. Most of the public universities in China are under provincial control (700 out of 817).

[28]Table A2 in the appendix presents more information on numbers of exam takers and admissions from 2013 through 2016.

[29]These two publications are *Quanguo Zhongdiandaxue Luqufenshuxian 2017* (which contains the cutoff scores in 2015 and 2016) and *Quanguo Gaoxiao Luqufenshuxian Tongji 2015* (which contains the cutoff scores in 2013 and 2014), published by Beijing Institute of Technology Press.

Gansu, Xinjiang, and Tibet.

6.2 Descriptive statistics

6.2.1 Summary statistics

In our sample, each university on average admits students through general admissions from 26 of the 27 provinces for which we have data.[30] The provinces from which a university does not recruit are usually far away from the university. On average, each university admits 95 science-stream and 21 arts-stream students per province per year. Admissions are highly selective. The average cutoff percentiles for science and arts streams are 91 and 97, respectively. See Table 2.

As described in the Introduction, the distribution of places across provinces is unequal (Figure 1). Another salient feature of the distribution of places is that a large fraction of places is allocated to students from the home provinces of the universities. Define the normalized admission rate of a province to a university as the fraction of places of a university going to a province, divided by the share of the exam takers of the province. The rate is one if a province's share of places at a university is proportional to its share of exam takers. Figure 2 depicts the average normalized admission rate as a function of distance between the university and province. The admission rate to universities in the same province is very high (13.7 and 18.3 for the science and arts streams, respectively). Overall, about 30 percent of all general-admission places are allocated to the home province. Among non-home provinces, more places are assigned to provinces that are closer to the university.

6.2.2 Admissions Trend

From 2013 to 2014, the total number of general-admission places in our sample decreased by 7.2 percent. The decline was mainly caused by a shift of places from general admissions to special admissions programs. From 2014 to 2015, the total number of general-admisisons places increased by 1.1 percent. There is a total of 2,889 university-province pairs in our sample. The average change in provincial quotas is -3 percent between 2013 and 2014 and 2.9 percent between 2014 and 2015. There is no clear shift in admissions patterns between 2013 and 2015. Of the 2,889 university-province pairs, 255 pairs had their quotas increased in both 2014 and 2015, while 704 pairs had

[30]On average, each university recruits from 26 of the 27 provinces for the science stream and 21 of the 27 provinces for the arts stream. Figure A1 shows the frequencies of the number of provinces in which each university recruits.

their quotas reduced in both years. The rank correlation coefficients of each province's share of exam takers admitted by a Project-211 university between any two years between 2013 and 2015 ranges from 0.77 to 0.86. The cutoff percentiles are also highly correlated between years (Table A4 in the appendix). For the 2,889 university-province pairs, the average change in the cutoff percentiles was 0.7 percent and 0.3 percent for the science and arts streams, respectively.

6.2.3 Raw score distribution

The distributions of raw test sores are hump-shaped, but non-normal to various degrees. For example, Figure 3 plots the distributions of raw test scores in Yunnan province and Henan province in 2014 and 2015. The dashed curve in each chart represents a normal distribution with the same mean and variance as the raw test score. In general, the difference in distribution across provinces is bigger than that across years. See Figure 3. For our analysis, we transform the raw test scores into ones with standard normal distribution using the procedure described in Section 4.

6.3 Relationship between transformed cutoff scores in different provinces

Rearranging (7), we have, for any pair of provinces i and k and any university j that is located in neither province but recruits in both,

$$c_i^{j*} = (a_k - a_i) + c_k^{j*} + (\xi_k^j - \xi_i^j). \tag{12}$$

Thus, our model predicts that, for any two provinces, plotting the transformed cutoffs of a university located in neither province in one province against its transformed cutoff in the other province should result in a straight line with a slope of one.

Panel A of Figure 4 plots the transformed cutoffs in Yunnan against the cutoffs in Henan. Note that the plot is indeed close to a straight line with a slope of one. That all universities have lower cutoffs in Yunnan than in Henan suggests that all universities prefer Yunnan students to Henan students.

To test whether this relationship holds in general, we separately regress c_i^{j*} on c_k^{j*} for each pair of provinces i and k for each year of our data. The regression results show that the estimated coefficients on c_k^{j*} are centered close to 1. For 40% of the regressions, we cannot reject the null hypothesis that the coefficient on c_k^{j*} is equal to 1 at the 10% significance level. In addition, we compare the R-squared of these unrestricted regressions with the R-squared of the regressions that restrict the coefficient of c_k^{j*} to be one. Figure 5 shows that the density plot of the R-squared of the unrestricted model largely overlaps that of the restricted model. The average R-squared's for the restricted regressions are 0.77 and 0.75 for the science and arts stream, respectively. The results indicate that the pattern in Panel A of Figure 4 holds broadly.

In our model, two universities that have the same cutoff score for a given province may not admit the same number of students from the province, as one university may be more popular among the students than the other. Thus, while our model predicts a linear relationship between cutoff scores, it *does not* predict any similar relationship between the number of admitted students. Panel B of Figure 4 plots the normalized admission rates of Yunnan against the normalized admission rates of Henan. The dots below the 45-degree line represent universities that admit proportionally more from Henan than from Yunnan. Although all universities prefer Yunnan students to Henan students, not all recruit proportionally more from Yunnan than Henan.

7 Estimation results

7.1 University-side estimation

7.1.1 Systematic value

Since there is no clear trend between years (Section 6.2.2), we separately estimate the systematic value of the students of each province for each year and each stream from 2013 through 2016 using the method presented in Section 4.1. Detailed results are presented in Table A5. The model fits the data very well. The correlation between the actual cutoff difference and predicted cutoff difference ranges from 0.8 to 0.9. The estimates for each stream are highly consistent over the years, with correlations between any two years ranging from 0.95 to 0.99 (Panel A of Table A6).

Panel A of Figure 6 displays the estimated systematic values of the provinces ($\widehat{a}$) averaged over years and streams. Panel B of Figure 6 shows that $\widehat{a}$ is positively correlated with the fraction of exam takers in a province admitted by a 211-Program university. Interestingly, richer provinces tend to rank higher in $\widehat{a}$ than in the fraction of exam takers admitted. For example, Shanghai and Beijing, the two richest Chinese provinces, are ranked first and second in $\widehat{a}$ but fourth and third in the fraction of exam takers admitted. Guangdong, the sixth richest, is ranked sixteenth in $\widehat{a}$ and last in admission rate. As we shall see in the next section, this may be due to the existence of better outside options for students in richer provinces.

Figure 7 shows that there is a strong positive relationship between the estimated average value of students from each province and the province's GDP per capita.[31] Since $\widehat{a}$ are estimated using the cutoff scores of non-home-province universities, the relationship is not caused by universities in richer provinces favoring home-province students.

To test the robustness of our estimates, we re-estimate the systematic values with four different subsamples of universities: Project-985 universities, non-Project-985 universities, universities in provinces with below-median GDP per capita, and universities in provinces with above-median GDP per capita. The estimates do not change significantly in any case, suggesting that university preferences do not correlate significantly with ranking or location.

We assume that universities do not systematically discriminate between non-home provinces. One may worry that some universities may prefer students from neighboring provinces that share similar cultures and dialects. To address this concern, we re-estimate the systematic values using the cutoffs of universities in non-home and non-neighboring provinces.[32] The estimates remain largely unchanged.[33]

The preference for richer provinces is unlikely to be imposed by the central government. If anything, the Ministry of Education has in recent years tried to shift more places to poorer provinces. In 2016, the Ministry of Education released a document instructing Project-211 universities and universities in fourteen wealthy provinces and provincial cities to admit more students from poorer provinces, resulting in public protests in several cities adversely affected by the plan.

What then explains the university preferences? As Chinese universities charge the same tuition fee to all students and alumni donations are not an important source of school finance, there is no financial reason to favor applicants from richer provinces. It is possible that universities are prejudiced against students from poorer provinces. But in that case, the prejudice is widespread and not limited to universities in richer provinces.

[31]We obtain each province's GDP per capita from the website of the Natual Bureau of Statistics of China (www.stats.gov.cn/).

[32]That is, the cutoff of a university in a certain province is used in the estimation of the systematic values only if the university is neither in the province itself nor in any province that shares a border with the province.

[33]We also conduct robustness checks that 1) exclude the university-province with cutoff outliers; 2) exclude the top five universities (Peking University, Tsinghua University, Fudan University, Shanghai Jiaotong University, and Zhejiang University). The results remain robust.

A more plausible explanation is that universities perceive students from richer provinces to be better educated. Provincial pre-college education spending is highly correlated with provincial per-capita GDP, and parents in richer provinces are also likely to spend more on their children's education. We regress the estimated $\widehat{a}$ on various provincial characteristics. Table 3 presents the regression results. Provincial GDP per capita and educational spending size are both positively associated with a province's systematic value. Educational spending, however, seems to be the more important of the two: it alone explains nearly 80 percent of the variation in the systematic values (column 2). Controlling for education spending, per-capita GDP is no longer significant (column 3). The coefficient for ethnic minority population is positive and significant, indicating that universities tend to favor provinces with a larger minority population. The fraction of a birth cohort that finishes high school measures the selectivity of high-school education in a province. As high schools become less selective, the overall quality of students may decline. That the coefficient for this variable is negative is consistent with the hypothesis that universities prefer students with higher quality.[34]

7.1.2 Home bias

We separately estimate each university j's home bias (ρ^j) for each year and each stream, using cutoff data from 2013 through 2016. The estimates are similar across years, with correlations between 0.3 and 0.8 (Panel B of Table A6). Figure A2 plots the distributions of the estimates, $\widehat{\rho}^j$, averaged over these four years. The mean of the distribution is -0.017 for science and -0.037 for arts, suggesting that overall there is a slight bias *against* home-province students.

To shed light on the factors that determine a university's home bias, we regress the average home-bias estimates of each university on university characteristics. Table 4 presents the results. Higher-ranked universities (i.e., the

[34] Pre-college educational spending per student is calculated by dividing the total government spending on pre-college education (including primary schools, middle schools and high schools) by the total enrollment in primary, middle and high schools. Data on per-student government spending on pre-college education are from the Ministry of Education. Data on total enrollment of primary, middle and high schools are from the National Bureau of Statistics. We calculate the fraction of high-school graduates in a birth cohort as the ratio of the number of high-school admissions to the number of primary-school graduates for the same cohort. For example, for the incoming class of 2010, we use the ratio of the number of high-school admissions in 2010 to the number of pupils who graduated from primary school in 2007 (for Shanghai, we use the number of pupils graduating from primary school in 2006 as the denominator as Shanghai students spend only five years in primary schools) to measure the percent of the cohort admitted to a grammar high school. The data are obtained from the National Bureau of Statistics of China.

Project-985 universities) and those under provincial control are more likely to be biased in favor of home students. Admissions at Project-985 universities may be sufficiently selective that the universities can compromise the admission standard for home-province students without significantly affecting the overall quality of the student body. Compared to universities under the Ministry of Education, universities under local control receive a larger fraction of their funding from their provincial governments and, hence, may be required to admit more home-province students. Interestingly, contrary to the popular belief, universities in Beijing and Shanghai, as a group, seem to favor out-of-province students.

Our model assumes a constant marginal value of home-province students. In reality, the marginal value could be decreasing in the number of home students, and the small home bias that we find may indicate only that given the current number of students attending home-province universities, the marginal gain from admitting more home-province students is small.[35] This may explain the negative overall bias of Beijing and Shanghai universities, as the number of Beijing and Shanghai students attending Beijing and Shanghai universities is particularly large.

7.2 Student-side estimation

For each province i, we estimate a logit demand system to recover d_i^j, the average utility for attending university j, for each stream and each year. While the numerical estimates fluctuate over time, the rank order of the universities within a province is fairly stable. If one university is more popular than another university among students in a certain province in 2013, then it is likely to be more popular in 2014.[36]

To understand the effect of distance on student preferences, we run the following regression by pooling data from 2013 through 2015:

$$\widehat{d_{it}^j} = \beta log(dist_i^j + 1) + \eta_i + \kappa^j + \tau_t + \xi_{it}^j \tag{13}$$

where $dist_i^j$ is the distance between university j and province i's provincial capital city; η_i are the provincial fixed effects that capture the average value of attending a university in our sample (relative to the outside option); κ^j are the university fixed effects that capture the average value of university j; τ_t are the year fixed effects that capture any time trend; and ξ_i^j is an

[35] Similarly, universities may have to meet numerical targets for home-province students, but the constraints are currently non-binding.

[36] The average Kendall's τ between the rank orders of universities in any two years is about 80 percent. See Figure A3 in the appendix.

error term. The coefficient β is negative and significant (Columns 1 and 3, Table 5), showing that, controlling for university and province fixed effects, students prefer universities closer to home.

To further investigate the factors that determine the preferences of the students, we replace the university dummies with a list of university characteristics (Columns 2 and 4 of Table 5).[37] Students prefer universities in Project 985, universities whose graduates earn higher income, universities in larger cities, and universities in cities with lower housing cost.[38] The distance coefficient remains negative and significant.

Since the value of the outside option in every province is normalized to zero, the attractiveness of the outside option for students in a particular province is captured by the provincial fixed effect. In particular, a larger provincial fixed effect implies a worse outside option. Figure 8 plots the provincial fixed effects. Shanghai, Beijing, Jiangsu, Guangdong, and Zhejiang are among the provinces with the best outside options. Instead of attending a lower-ranked Project-211 university, students in these rich provinces may opt to study abroad or attend a local non-Project-211 university. This explains why these provinces underperform in sending students to Project-211 universities (Section 7.1.1).

8 Counterfactual Analysis

We calibrate our model to evaluate the quantitative significance of our findings. Using (7) to replace (3) and (10) to replace (1), we obtain a parametric version of our theoretical model. The model is characterized by $\{C(j)\}_{j\in\mathcal{J}}$, $\{a_i\}_{i\in\mathcal{I}}$, $\{\xi_i^j\}_{i\in C(j),j\in\mathcal{J}}$, $\{\rho^j\}_{j\in\mathcal{J}}$, $\{d_i^j\}_{i\in C(j),j\in\mathcal{J}}$, $\{K^j\}_{j\in\mathcal{J}}$ and $\{\lambda_i\}_{i\in\mathcal{I}}$. A set of cutoffs and quotas, $\{c_i^j\}_{i\in C(j),j\in\mathcal{J}}$ and $\{q_i^j\}_{i\in C(j),j\in\mathcal{J}}$, is an equilibrium if there exists a set of marginal utilities, $\{\psi^j\}_{j\in\mathcal{J}}$, such that the following

[37] Including a home-province dummy does not significantly change the results.

[38] The average monthly salary of graduates during the first five years after entering the labor force is calculated by IPIN.com based on data on a sample of 40 million young workers who had recently graduated from college. See http://edu.sina.com.cn/gaokao/2015-06-16/1031473427.shtml. The hedonic quality-controlled home price index of the city in which the university is located as of June 2013 and it is obtained from the Hang Lung Center for Real Estate at Tsinghua University. The distance from university to province is measured by the line distance between the city where the university is located and the capital city of the province.

equations are satisfied:

$$a_i + c_i^{j*} + \xi_i^j + \rho^j I_i^j = \psi^j, \ \forall j \in \mathcal{J}, i \in C(j); \tag{14}$$

$$q_i^j = \lambda_i \int \frac{exp(d_i^j)}{1 + \sum_{k \in J(C(j), s_i)} exp(d_i^k)} d\Phi(s_i) \ \forall j \in \mathcal{J}, i \in C(j); \tag{15}$$

$$\sum_{i \in C(j)} q_i^j = K^j \quad \forall j \in \mathcal{J}. \tag{16}$$

Eq. (14) requires that each university receives constant utility from the marginal students in each province from which it recruits. Eq. (15) requires that the number of students choosing a university in a province is equal to the quota assigned to the province. Eq. (16) requires that the sum of a university's quotas is equal to the capacity of the university. Given a set of parameters, we can numerically solve for the equilibrium, which is unique by Proposition 2.

The number of exam takers, $\{\lambda_i\}_{i \in \mathcal{I}}$, and the capacities of the universities, $\{K^j\}_{j \in \mathcal{J}}$, are directly observed from the data for each year from 2013 through 2015. To establish a benchmark, we set the universities' idiosyncractic preferences, $\{\xi_i^j\}_{i \in C(j), j \in \mathcal{J}}$, to zero and use the estimates $\{\widehat{a}_i\}_{i \in \mathcal{I}}$, $\{\widehat{\rho}^j\}_{j \in \mathcal{J}}$, and $\left\{\widehat{d}_i^j\right\}_{i \in C(j), j \in \mathcal{J}}$ from Section 7 to simulate the allocation of places in each year from 2013 through 2015. Panel A of Figure 9 compares the predicted distribution of places of Project-211 universities with the actual distribution. As the figure shows, the model predicts the distribution of places well.[39]

We carry out three counterfactual experiments. First, we set $\{\rho^j\}_{j \in \mathcal{J}}$ to zero to tease out the effect of home bias. Then, we shut down the students' preferences for universities near home by replacing each $\widehat{d}_i^j$ in the first counterfactual experiment by

$$\widehat{d}_i^{j*} = \widehat{d}_i^j - \widehat{\beta} \log(dist_i^j + 1)$$

where $\widehat{\beta}$ is the estimate for β from the regression equation (13). Finally, we eliminate the systematic preferences of the universities by replacing each $\widehat{a}_i$ in the second counterfactual experiment with zero.

Panels B through D of Figure 9 present the distribution of places across provinces in the counterfactual experiments. For illustration purposes, we average the predictions from 2013 through 2015. The hollow bars represent the simulated distribution of university places in each of the three counterfactual experiments. The solid bars represent the actual distribution. Table

[39] The corresponding numbers are reported in Table A8 in the appendix.

6 shows the average simulated normalized admission rate to universities in the home provinces of the students.

The experiments show that the current unequal distribution of places across provinces is mainly driven by the systematic preferences of the universities. Eliminating home bias has little effect on either the geographical distribution of places or the fraction of places allocated to home-province students. Many students attend universities in their home provinces because they prefer universities close to home. Eliminating this preference reduces the average normalized admission rate to the home province of a university from 14.73 to 3.9. However, as more students from Beijing and Shanghai are willing to attend universities outside of their home provinces, the fraction of Project-211 places allocated to Beijing and Shanghai students further increases (Figure 9 Panel C). By contrast, eliminating the systematic preference of universities for rich-province students largely equalizes the geographical distribution of places (Figure 9 Panel D).[40]

9 Conclusion

College admissions in China are centralized at the provincial level. Universities set provincial admission quotas and admit students from each province largely based on their test scores. In this paper, we argue that the provincial quotas are the outcome of a competition between universities. While it is the universities that set the quotas, in equilibrium the quotas reflect both universities' preferences over students and students' preferences over universities. We develop a methodology to estimate the universities' preferences using the universities' cutoff scores in each province. We find that the preferences of the top Chinese universities are surprisingly similar. Regardless of location and rank, universities prefer students from richer provinces. Contrary to common belief, there is little evidence that universities prefer home-province students. Rather it is the students who choose to study close to home.

One plausible reason that universities prefer students from richer provinces is that richer provinces spend more on pre-college education. If this is the case, then the long-term solution to the unequal distribution of places is to raise the education spending of poorer provinces. Alternatively, since students prefer to study close to home, raising the funding of universities in poorer provinces would improve the educational opportunities of the students in these provinces. In the shorter term, the central government may compel universities to allocate more quotas to poorer provinces. However,

[40]The remaining difference across provinces is caused by the part of the student preferences unrelated to distance.

such a policy would likely meet resistance from richer provinces. Replacing the current quota system with a uniform college entrance exam might not improve the situation as the new standardized test scores are likely to correlate positively with income.

Compared to many countries, the college admission process in China is arguably more equitable and meritocratic. It is centralized and test-score-based. Tuitions are low and affordable. Nevertheless, students from richer provinces enjoy a substantial edge over students from poorer provinces. The lesson is that, in an unequal society, trade-offs between meritocracy and equal representation are inevitable. Policymakers, instead of trying in vain to devise an admission process that is both meritocratic and fair, should decide what is the proper tradeoff. This paper assumes that all students in a province receive the same expected utility from attending a university. To better understand the welfare tradeoff, future research should explicitly model how the productivity gain from attending a university varies with a student's test scores and other characteristics.

10 Appendix

Proof of Proposition 1.

We first establish existence. Fix a province i and provincial quotas q_i. For each university j, define $\underline{c}_i^j$ implicitly by

$$\lambda_i \int_{s_i \geq \underline{c}_i^j} F_i\left(\epsilon_i^j \geq -d_i^j\right) dG_i\left(s_i\right) = \sum_l K^l,$$

and define $\overline{c}_i^j$ implicitly by

$$\lambda_i \int_{s_i \geq \overline{c}_i^j} dG_i\left(s_i\right) = q_i^j.$$

Assumption 1 implies that $\underline{c}_i^j$ is well defined. Intuitively, at the cutoff $\underline{c}_i^j$, university j can fill up its capacity in province i even when all other universities assign all their capacities to province i.

Let c_i^{-j} denote the province i's cutoffs of all universities in $D\left(i\right)$ other than university j's. Define $h_i = \left\{h_i^j\right\}_{j \in D(i)} : \prod_{j \in \mathcal{D}(i)} \left[\underline{c}_i^j, \overline{c}_i^j\right] \rightarrow \prod_{j \in \mathcal{D}(i)} \left[\underline{c}_i^j, \overline{c}_i^j\right]$ where

$$h_i^j\left(c_i\right) \equiv \arg \min_{c_i^j \in \left[\underline{c}_i^j, \overline{c}_i^j\right]} \left(\lambda_i \int_{s_i} \pi_i\left(j, s_i, J\left(c_i^{j'}, c_i^{-j}, s_i\right)\right) dG_i\left(s_i\right) - q_i^j\right)^2. \quad (17)$$

Since $\pi_i\left(j, s_i, J\left(c_i^{j'}, c_i^{-j}, s_i\right)\right)$ strictly decreases in $c_i^{j'}$, h_j is a function. Since

$$\int_{s_i} \pi_i\left(j, s_i, J\left(c_i^{j'}, c_i^{-j}, s_i\right)\right) dG_i\left(s_i\right)$$

is continuous in c_i^{-j}, h_j is continuous by the Theorem of Maximum. It follows from the Kakutani's Fixed Point Theorem that there exists a fixed point c_i^* such that $h_i\left(c^*\right) = c^*$.

We need to prove that for each university $j \in D\left(i\right)$,

$$\lambda_i \int_{s_i} \pi_i\left(j, s_i, J\left(c_i^*, s_i\right)\right) dG_i\left(s_i\right) = q_i^j \tag{18}$$

By the definition of $\overline{c}_i^j$,

$$\lambda_i \int_{s_i} \pi_i\left(j, s_i, J\left(\overline{c}_i^j, c_i^{-j*}, s_i\right)\right) dG_i\left(s_i\right) \leq q_i^j.$$

Since h_j is continuous, if

$$\lambda_i \int_{s_i} \pi_i\left(j, s_i, J\left(c_i^{j^*}, c_i^{-j*}, s_i\right)\right) dG_i\left(s_i\right) > q_i^j$$

then there would exist some $c_i^{j'} \in (c_i^{j*}, \overline{c}_i^j]$ such that

$$\lambda_i \int_{s_i} \pi_i\left(j, s_i, J\left(c_i^{j'}, c_i^{-j*}, s_i\right)\right) dG_i\left(s_i\right) = q_i^j,$$

which contradicts the supposition that c_i^{j*} minimizes

$$\left(\lambda_i \int_{s_i} \pi_i\left(j, s_i, J\left(c_i^j, c_i^{-j*}, s_i\right)\right) dG_i\left(s_i\right) - q_i^j\right)^2.$$

Hence, for each $j \in D\left(i\right)$,

$$\lambda_i \int_{s_i} \pi_i\left(j, s_i, J\left(c_i^*, s_i\right)\right) dG_i\left(s_i\right) - q_i^j \leq 0. \tag{19}$$

Suppose that for some university j

$$\lambda_i \int_{s_i} \pi_i\left(j, s_i, J\left(c_i^*, s_i\right)\right) dG_i\left(s_i\right) - q_i^j < 0.$$

Then it must be that $c_i^{j*} = \underline{c}_i^j$. Note that for any $s_i \geq \underline{c}_i^j$

$$\pi_i\left(j, s_i, J\left(\underline{c}_i^j, c_i^{-j*}, s_i\right)\right) \geq F_i\left(\epsilon_i^j \geq -d_i^j\right) - \sum_{k \neq j} \pi\left(k, s_i, J\left(\underline{c}_i^j, c_i^{-j*}, s_i\right)\right). \quad (20)$$

Inequality (20) says that for students who is qualified for university j, the ones choosing it must be greater than the ones who prefer it to the outside option, minus the students who ends up choosing other universities.

It follows from (20) that

$$\begin{aligned}
& \lambda_i \int_{s_i} \pi_i\left(j, s_i, J\left(\underline{c}_i^j, c_i^{-j*}, s_i\right)\right) dG_i(s_i) \\
\geq \; & -\lambda_i \sum_{k \neq j} \int_{s_i \geq \underline{c}_i^j} \pi_i\left(k, s_i, J\left(\underline{c}_i^j, c_i^{-j*}, s_i\right)\right) dG_i(s_i) \\
& +\lambda_i \int_{s_i \geq \underline{c}_i^j} F_i\left(\epsilon_i^j \geq -d_i^j\right) dG_i(s_i) \\
= \; & \sum_{k \neq j} \left(K^k - \lambda_i \int_{s_i \geq \underline{c}_i^j} \pi_i\left(k, s_i, J\left(\underline{c}_i^j, c_i^{-j*}, s_i\right)\right) dG_i(s_i) \right) + K^j.
\end{aligned}$$

The last equality follows from the definition of $\underline{c}_i^j$.

By rearranging terms, we have

$$\begin{aligned}
& \lambda_i \int_{s_i} \pi_i\left(j, s_i, J\left(\underline{c}_i^j, c_i^{-j*}, s_i\right)\right) dG_i(s_i) - K^j \\
& \qquad \geq \sum_{k \neq j} \left(K^k - \lambda_i \int_{s_i \geq \underline{c}_i^j} \pi_i\left(k, s_i, J\left(\underline{c}_i^j, c_i^{-j*}, s_i\right)\right) dG_i(s_i) \right). \quad (21)
\end{aligned}$$

Since, by supposition,

$$\lambda_i \int_{s_i} \pi_i\left(j, s_i, J\left(\underline{c}_i^j, c_i^{-j*}, s_i\right)\right) dG_i(s_i) < q_i^j,$$

the left-hand side of (21) is strictly negative. Hence, (21) can hold only if

$$K^k - \lambda_i \int_{s_i} \pi_i\left(k, s_i, J\left(\underline{c}_i^j, c_i^{-j*}, s_i\right)\right) dG_i(s_i) < 0$$

for some $k \neq j$, which contradicts (19). This establishes existence.

We now prove uniqueness. Suppose c_i^* satisfies (18). Consider any $c_i \neq c_i^*$. Let

$$\begin{aligned} J^+ &= \left\{ j \in D(i) : c_i^{j*} > c_i^j \right\}; \\ J^- &= \left\{ j \in D(i) : c_i^{j*} < c_i^j \right\}. \end{aligned}$$

By supposition, either J^+ or J^- is non-empty. Suppose J^+ is non-empty. For all s_i,

$$J^+ \cap J(c_i^*, s_i) \subseteq J^+ \cap J(c_i, s_i)$$

with the inclusion strict when $s_i \in \left(c_i^j, c_i^{j*}\right)$ for some j. It follows that for all s_i

$$\begin{aligned} \sum_{j \in J^+} \pi\left(j, s_i, J\left(c_i^*, s_i\right)\right) &= \Pr\left[\max_{j \in J^+ \cap J\left(c_i^*, s_i\right)} u_i^j \geq \max\left(\max_{k \in J\left(c_i^*, s_i\right)/J^+} u_{i,k}, 0\right)\right] \\ &\leq \Pr\left[\max_{j \in J^+ \cap J(c_i, s_i)} u_i^j \geq \max\left(\max_{k \in J(c_i, s_i)/J^+} u_{i,k}, 0\right)\right] \\ &= \sum_{j \in J^+} \pi\left(j, s_i, J\left(c_i, s_i\right)\right), \end{aligned}$$

with the inequality strict when $s_i \in \left(c_i^j, c_i^{j*}\right)$ for some j. Thus, the set of universities in J^+ admit strictly more students under c_i than under c_i^*. By a similar argument, the set of universities in J^- admits strictly fewer students under c_i. Hence, any $c_i \neq c_i^*$ does not satisfy (18).

Proof of Proposition 2. Let $c_i^*(q_i)$ denote the equilibrium in Proposition **1**. Given q^{-j}, university j's constrained maximization problem is to maximize (2) subject to $\sum_{i \in \mathcal{I}} q_i^j = K^j$. Since α_i is smooth, $v^j\left(q^j, q^{-j}\right)$ is continuously differentiable in q^j with

$$\frac{\partial v^j}{\partial q_i^j} = \alpha_i\left(c_i^{j*}\left(q_i^j, q_i^{-j}\right)\right) + \gamma_i^j.$$

Let $br_j\left(q^{-j}\right) = \left\{br_i^j\left(q^{-j}\right)\right\}_{i \in C(j)}$ denote a best response of university j. Since we assume that $\lim_{s_i \to \infty} \alpha_i(s_i) = \infty$, the best response is an interior solution. The Kuhn-Tucker conditions imply that for each province $i \in C(j)$

$$\alpha_i\left(c_i^{j*}\left(br_i^j\left(q^{-j}\right), q_i^{-j}\right)\right) + \gamma_i^j = \psi^j \tag{22}$$

for some constant ψ^j. Since $\alpha_i(s_i)$ strictly increases in s_i, a higher ψ^j corresponds to a higher cutoff score in every province. As the number of admitted students strictly decreases in the cutoff score, there is only one set of quotas that can simultaneously satisfy (22) and the capacity constraint. Hence, university j's best response is unique. Since v^j is continuous in q^{-j}, $br_j(q^{-j})$ is continuous q^{-j} by the Theorem of Maximum. It then follows from the Kakutani's fixed-point theorem that a Nash equilibrium exists.

Suppose, by way of contradiction, that there exist two distinct Nash equilibria q^* and q^{**}. Let $\{c_i^{j*}\}_{i\in\mathcal{I},j\in\mathcal{D}(i)}$ and $\{c_i^{j**}\}_{i\in\mathcal{I},j\in\mathcal{D}(i)}$ denote the equilibrium cutoffs corresponding to q^* and q^{**}, respectively. From (22), if for some j and i, $c_i^{j*} > c_i^{j*}$ then $c_k^{j*} > c_k^{j*}$ for all $k \in C(j)$. We can therefore partition all universities into three groups. Let J^+ and J^- denote, respectively, universities that have higher and lower cutoffs in equilibrium q^*. By supposition, either J^+ or J^- is non-empty. Suppose J^+ is non-empty. By the argument in Proposition1, the universities in J^+ must, as a group, admit more applicants under equilibrium q^{**} than under equilibrium q^*, which contradicts the supposition that both q^* and q^{**} are equilibria. By the same logic, J^- cannot be non-empty.

Proof of Proposition 3

Since equations (9) are the first-order conditions of a constrained minimization, any solution to the minimization problem must satisfy (9). To show that $\{\widehat{a}_i\}_{i\in\mathcal{I}/\{0\}}$ is unique, it is sufficient to show that if $\{\widehat{a}_i\}_{i\in\mathcal{I}}$ and $\{a_i^*\}_{i\in\mathcal{I}}$ are both solutions to the constrained minimization problem then $\widehat{a}_i - a_i^*$ must be the same for all i. Suppose that $\widehat{a}$ and a^* are both solutions to the minimization problem. Since the number of provinces is finite, there must exist a province such that $\widehat{a}_i - a_i^* \geq \widehat{a}_k - a_k^*$ for all $k \in \mathcal{I}$. Consider equation (9) for this province i. The right-hand side of (9) does not depend on a_i. Since $\widehat{a}_i - \widehat{a}_k \geq a_i^* - a_k^*$ for all $k \in \mathcal{I}$, the left-hand side of (9) is weakly smaller when $\{\widehat{a}_i\}_{i\in\mathcal{I}}$ is replaced by $\{a_i^*\}_{i\in\mathcal{I}}$. Given Condition 1, (9) can hold for $\{a_i^*\}_{i\in\mathcal{I}}$ only when for all $k, i \in \mathcal{I}$, $\widehat{a}_k - \widehat{a}_i = a_k^* - a_i^*$.

We now prove the second part of the proposition. Substituting (7) into the right-hand side of (9), we have

$$\sum_{j\in D^*(i)} \sum_{k\in C^*(j)/i,} \frac{\left(\widehat{a}_i^j - \widehat{a}_k^j\right)}{|C^*(j)|} = \sum_{j\in D^*(i)} \sum_{k\in C^*(j)/i,} \frac{(a_i - a_k)}{|C^*(j)|} + \sum_{j\in D^*(i)} \sum_{k\in C^*(j)/i,} \frac{\gamma_i^j - \gamma_k^j}{|C^*(j)|}.$$

The second summation term on the right-hand side converges to 0 as $\sum_{j\in D^*(i)} (|C^*(j)| - 1)$ goes to infinity. Let $x^* = \min_{i\in\mathcal{I}} \left(\sum_{j\in D^*(i)} (|C^*(j)| - 1)\right)$. For each $i \in \mathcal{I}$,

$$\sum_{j\in D^*(i)} \sum_{k\in C^*(j)/i,} \frac{\lim_{x^*\to\infty} (\widehat{a}_i - \widehat{a}_k)}{|C^*(j)|} = \sum_{j\in D^*(i)} \sum_{k\in C^*(j)/i,} \frac{(a_i - a_k)}{|C^*(j)|}. \tag{23}$$

We know that $\lim_{x^*\to\infty} \widehat{a}_i = a_i$ for each $i \in \mathcal{I}$ satisfy (7) for all $i \in \mathcal{I}$. By our earlier argument, this is the only set of limits that satisfy (7) for all $i \in \mathcal{I}$.

References

[1] Abdulkadiroğlu, A., Agarwal, N., Pathak, P. A. The Welfare Effects of Coordinated Assignment: Evidence from the NYC High School Match. mimeo, 2017.

[2] Akyol, S., Krishna, K. Preferences, Selection, and Value Added: A Structural Approach. *European Economic Review*, 2017, 91:89-117.

[3] Avery, C., Glickman, M., Hoxby, C., Metrick, A. A Revealed Preference Ranking of U.S. Colleges and Universities. *The Quarterly Journal of Economics*, 2013, 128(1), 425-467.

[4] Bo, S., Liu, J., Shiu, J., Song, Y., Zhou, S. "Admission Mechanisms and the Mismatch between Colleges and Students: Evidence from a Large Administrative Dataset from China." *Economics of Education Review*, 2018, 68, 27-37.

[5] Bowen, William G., and Derek C. Bok. The Shape of the River: Long-term Consequences of Considering Race in College and University Admissions. Princeton University Press, 1998.

[6] Chan, J., Eyster, E. Does Banning Affirmative Action Lower Student Quality? *American Economic Review*, 2003, 93(3): 858-872.

[7] Chen, Y., Kesten, O. Chinese College Admissions and School Choice Reforms: A Theoretical Analysis. *Journal of Political Economy*, 2014, 125(1): 99-139.

[8] Du, W., Zhong, X. College, Major or Location: Analyzing Student Preference over College Based on Matching Theory. *China Journal of Economics*, 2018, 5(1), 1-25 (in Chinese).

[9] Ehrenberg, R. G. Econometric Studies of Higher Education. *Journal of econometrics*, 2004, 121(1): 19-37.

[10] Epple, D., Romano, R., Sieg, H. Admission, tuition, and financial aid policies in the market for higher education. *Econometrica*, 2006, 74(4): 885-928.

[11] Fryer, Roland G., and Glenn C. Loury. Affirmative Action and Its Mythology. *The Journal of Economic Perspectives,* 2005, 19, no. 3: 147-62. http://www.jstor.org/stable/4134977.

[12] Fryer, R., Loury, G., Yuret, T. An Economic Analysis of Color-Blind Affirmative Action. *Journal of Law, Economics, and Organization*, 2008, 24 (2) :319-355.

[13] Fu, C. Equilibrium tuition, applications, admissions, and enrollment in the college market. *Journal of Political Economy*, 2014, 122(2): 225-281.

[14] Gao, X., Huang, Y., White, A., Zhong, X. "Competition in matching markets: Evidence from college admission in China's top two universities." mimeo, 2018.

[15] Hastings, J., Kane, T., Staiger, D. Heterogeneous preferences and the efficacy of public school choice. *NBER Working Paper No.w12145*, 2008.

[16] Azevedo, Eduardo M., Leshno, Jacob D. A Supply and Demand Framework for Two-Sided Matching Markets. *Journal of Political Economy, 2016,* 124, no. 5: 1235-1268.

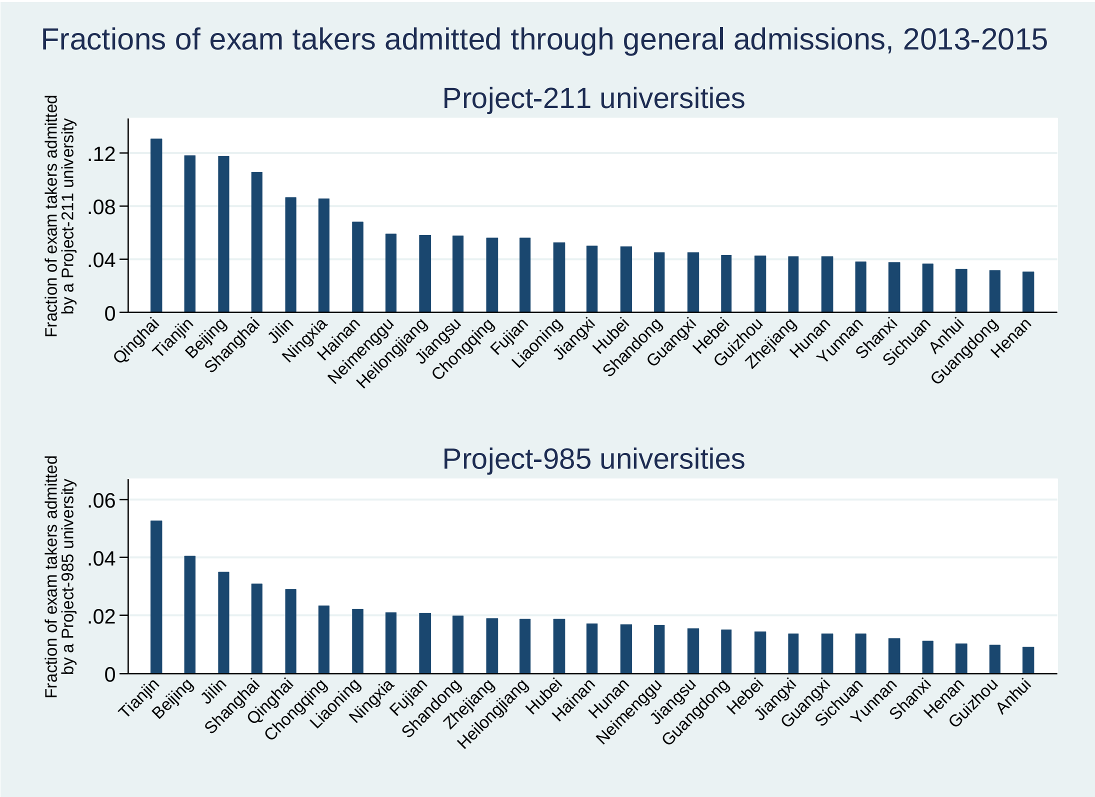
[17] Lien, J.W., Zheng, Jie, Zhong, X. Ex-ante fairness in the Boston and serial dictatorship mechanisms under pre-exam and post-exam preference submission. *Games and Economic Behavior*, 2017, 101, 98-120.

[18] McKelvey, R., Palfrey, T. Quantal Response Equilibria for Normal Form Games. *Games and Economic Behavior*, 1995, Volume 10, Issue 1: 6-38.

[19] Rawls, J. 1921-2002. A Theory of Justice. Cambridge, Mass. :Belknap Press of Harvard University Press, 1971.

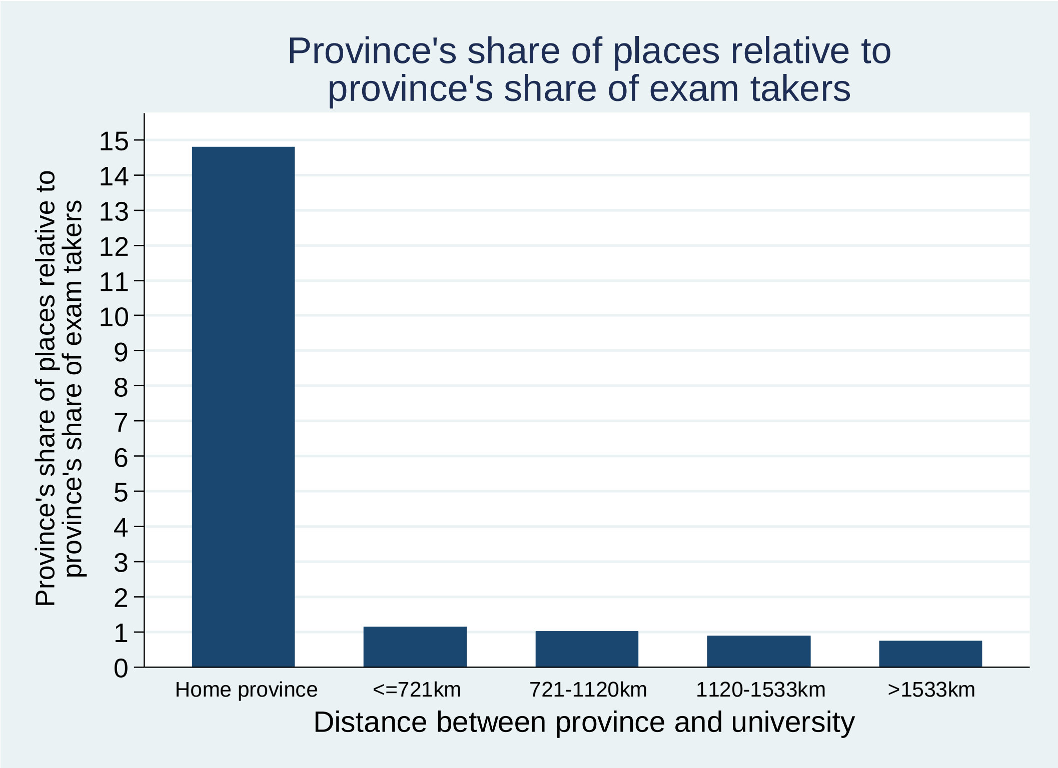
[20] Wu, B. and Zhong, X. Matching mechanisms and matching quality: Evidence from a top university in China. *Games and Economic Behavior*, 2014, 84(0): 196-215.

Figure 1: Geographical distribution of places



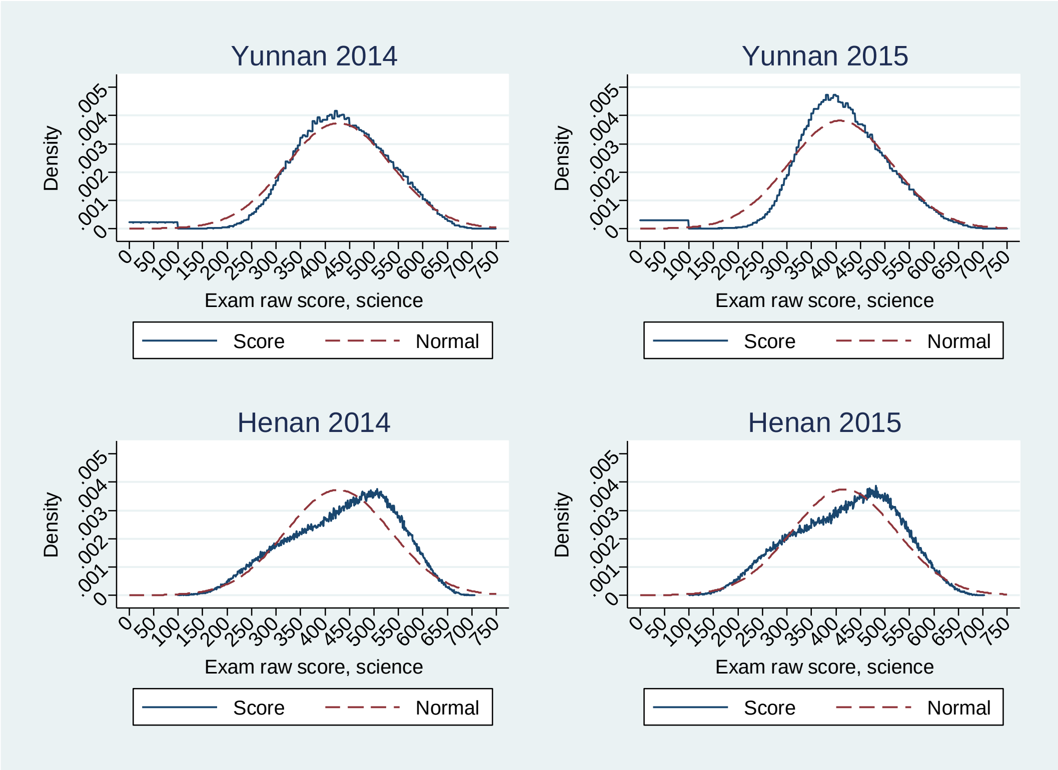
Notes: The y-axis is the fraction of exam takers, arts and science combined, that are admitted through general admissions from 2013 through 2015. The data on admissions for each university for the period 2013-2015 are obtained from gaokao.chsi.com.cn. The numbers of exam takers are obtained from the websites of provincial education examination authorities.

Figure 2: Distribution of places by distance



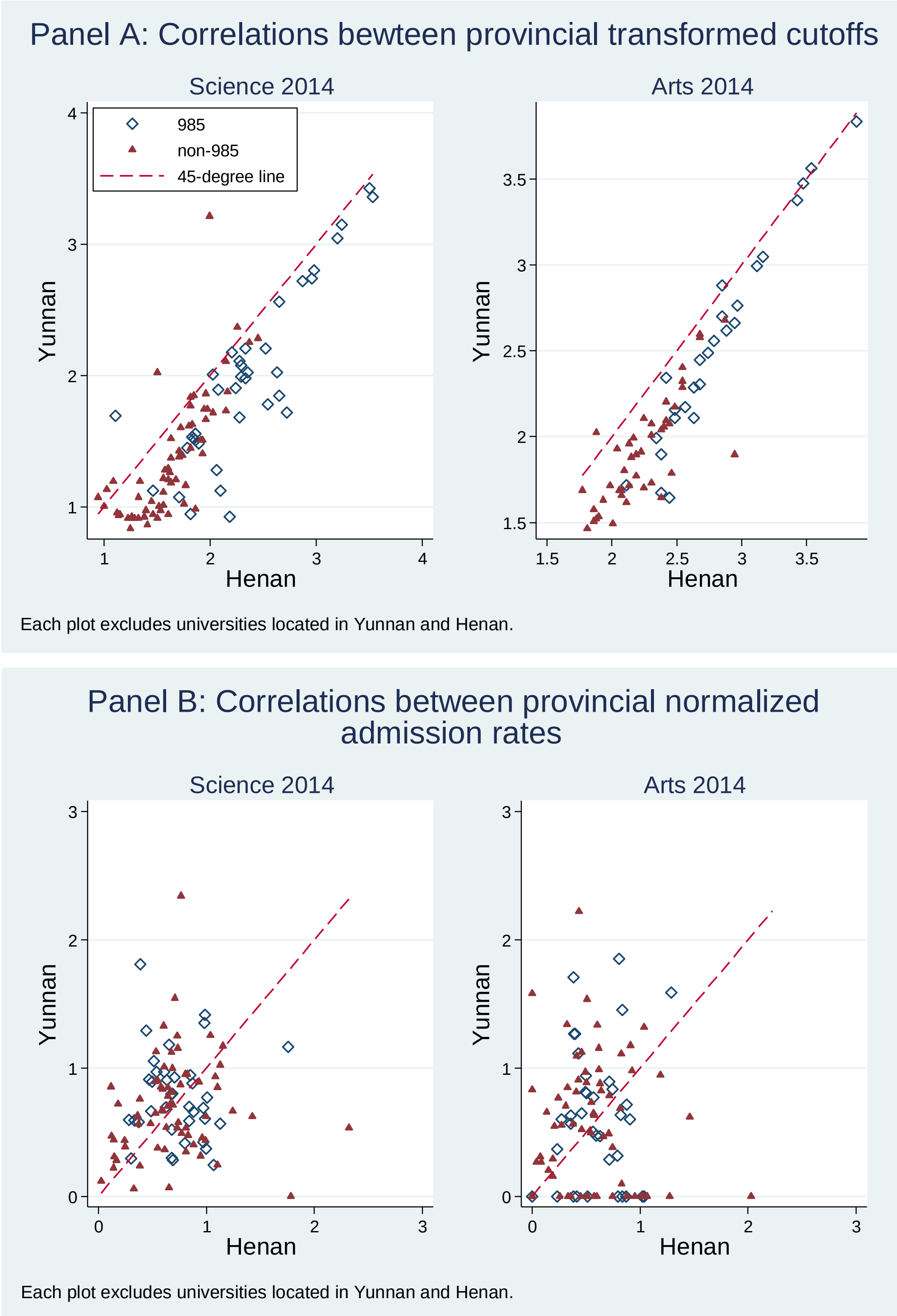
Notes: Figure 2 plots the average normalized admission rate for each box of university-province pairs. The normalized admission rate of a province to a university is the fraction of places of a university going to a province divided by the share of the exam takers of the province. The order of the boxes is arranged according to the distance between the university and the province of candidates. The first bar at the left end of the x-axis represents the average normalized admission rate for the provinces where the universities are located. Each of the remaining four boxes contains the same number of university-province pairs (around 700).

Figure 3: Distributions of raw exam scores



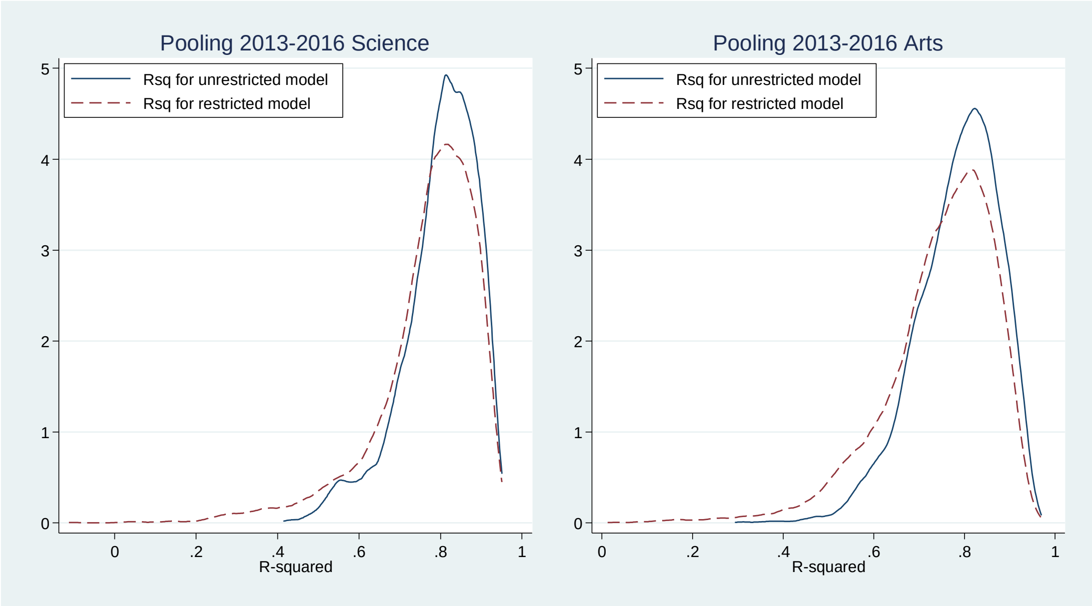
Note: The dashed curve in each chart represents the density plot of a normal distribution with the same mean and variance as the raw exam score distribution.

Figure 4: Correlations between provincial transformed cutoffs and normalized admission rates



Note: In Panel B, the normalized admission rate of a province to a university is the fraction of places of a university going to a province divided by the share of the exam takers from the province.

Figure 5: Density plots of the R-squared of the unrestricted and restricted models



Notes: For province pair (*i*, *k*), the unrestricted model is: $c_{i,j}^{*} = \phi_{ik}^{0} + \phi_{ik}^{1} c_{k,j}^{*} + error_{i,k,j}$.

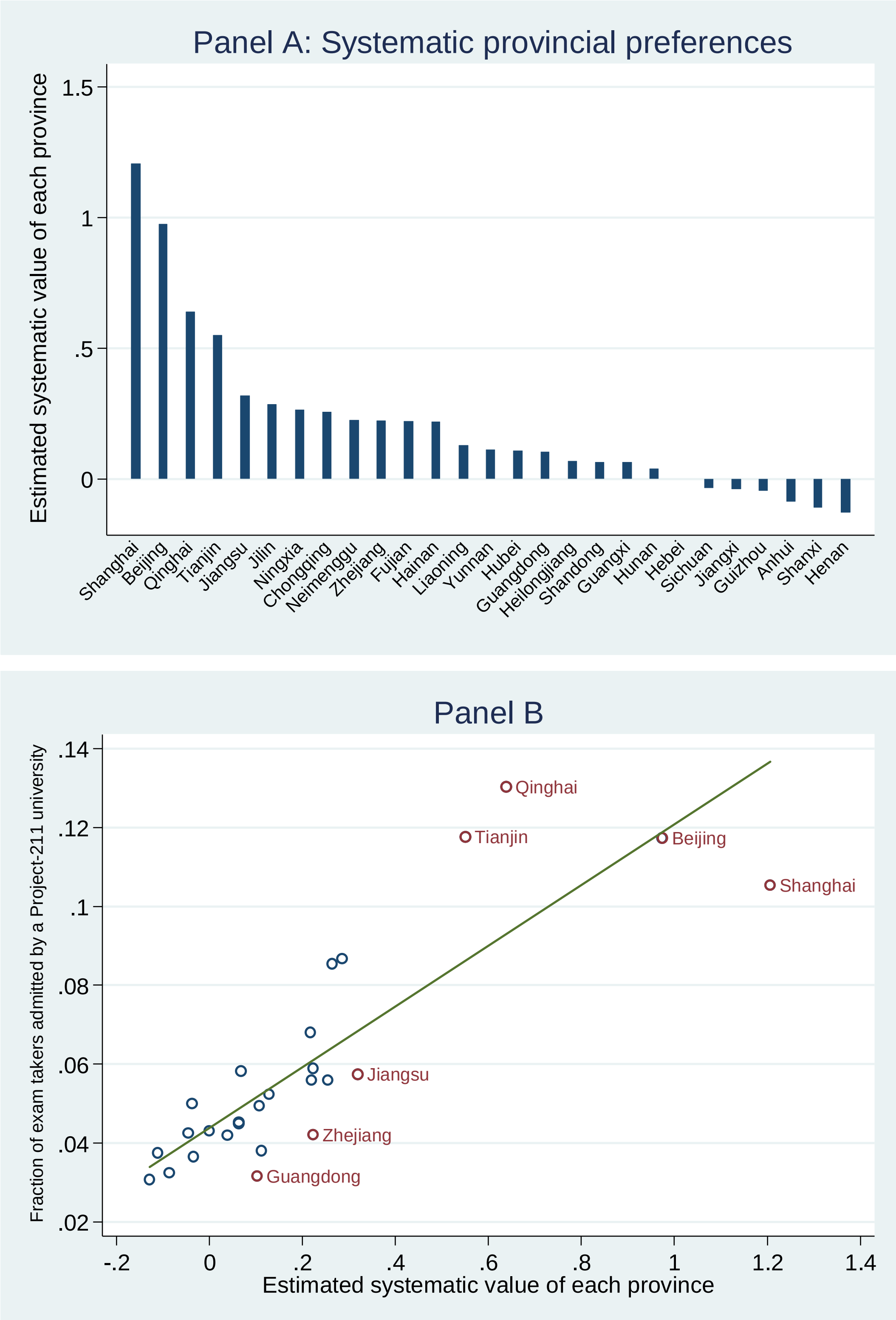
The R-squared of the unrestricted model is calculated using the following formula:

$$R_{ur\_ik}^{2} = 1 - \sum_{j} (c_{i,j}^{*} - \hat{\phi}_{ik}^{0} - \hat{\phi}_{ik}^{1} * c_{k,j}^{*})^{2} / \sum_{j} (c_{i,j}^{*} - \overline{c}_{i}^{*})^{2}.$$

For province pair (*i*, *k*), the restricted model is: $c_{i,j}^{*} - c_{k,j}^{*} = \varphi_{ik}^{0} + error_{i.k,j}$. The R-squared of the restricted model is calculated using the following formula:

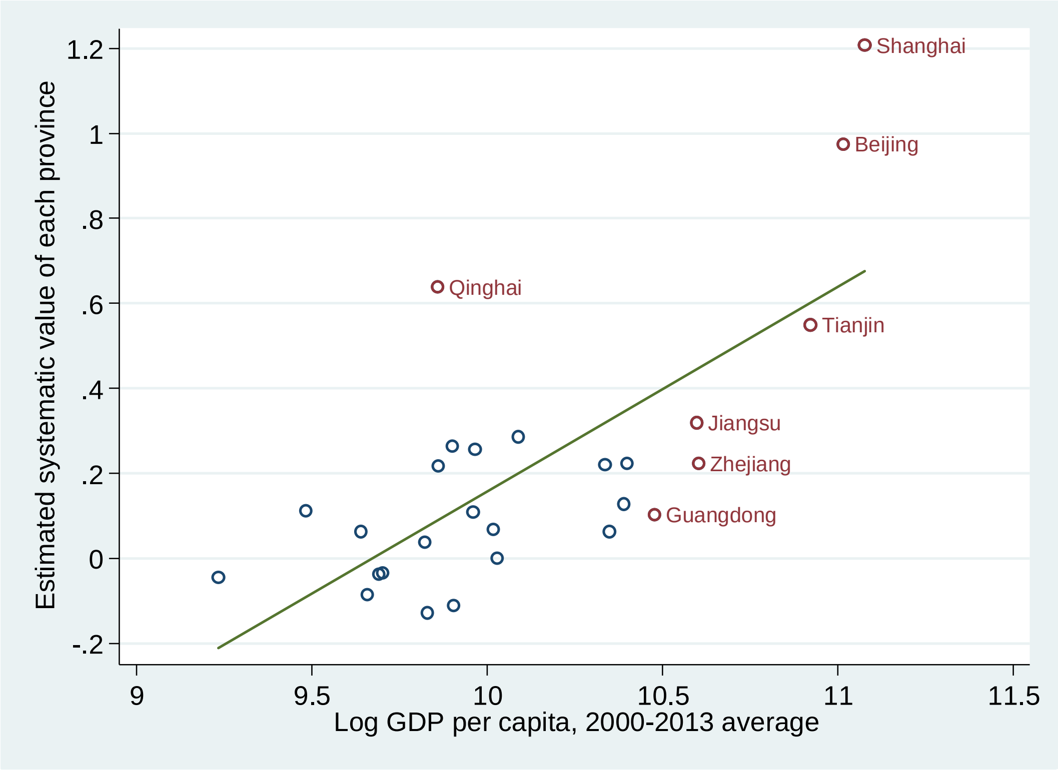
$$R_{R\_ik}^{2} = 1 - \sum_{j} (c_{i,j}^{*} - c_{k,j}^{*} - \hat{\varphi}_{ik}^{0})^{2} / \sum_{j} (c_{i,j}^{*} - \overline{c}_{i}^{*})^{2}.$$

Figure 6: Systematic provincial preferences



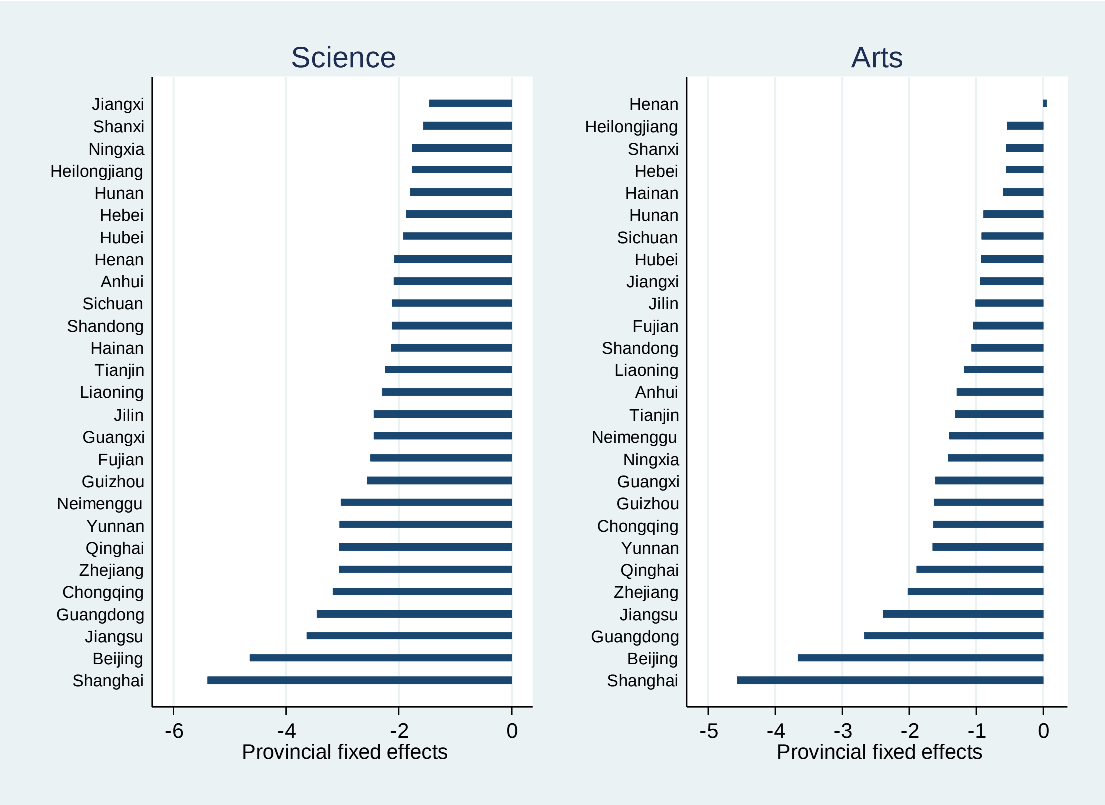
Notes: Panel A of Figure 6 displays the estimated systematic value of each province ($\hat{a}$) averaged over years and streams, weighted by the number of exam takers. Panel B of Figure 6 shows the relationship between $\hat{a}$ and the fraction of exam takers in a province admitted by a Project-211 university. In Panel B, the OLS fit is superimposed.

Figure 7: Correlation between estimated systematic value and provincial per-capita GDP



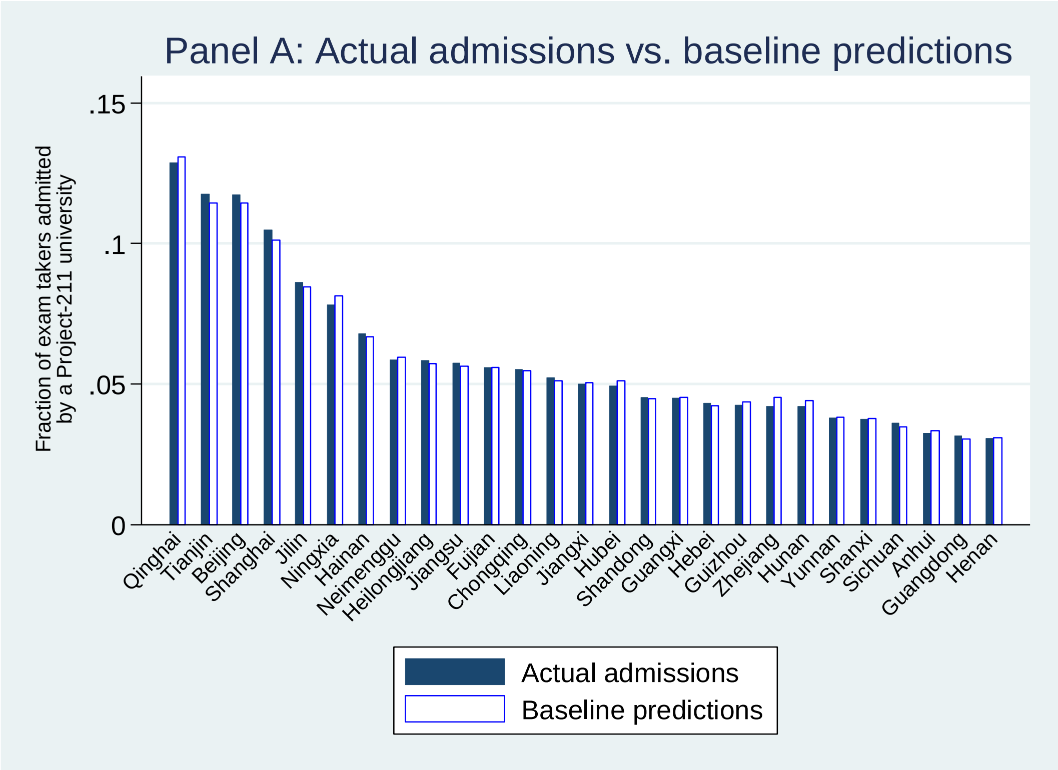
Notes: The y-axis is the estimated systematic value of each province ($\hat{a}$) averaged over years and streams, weighted by the number of exam takers. Data on per-capita GDP of each province is obtained from the *National Bureau of Statistics*. The OLS fit is superimposed.

Figure 8: Provincial fixed effects



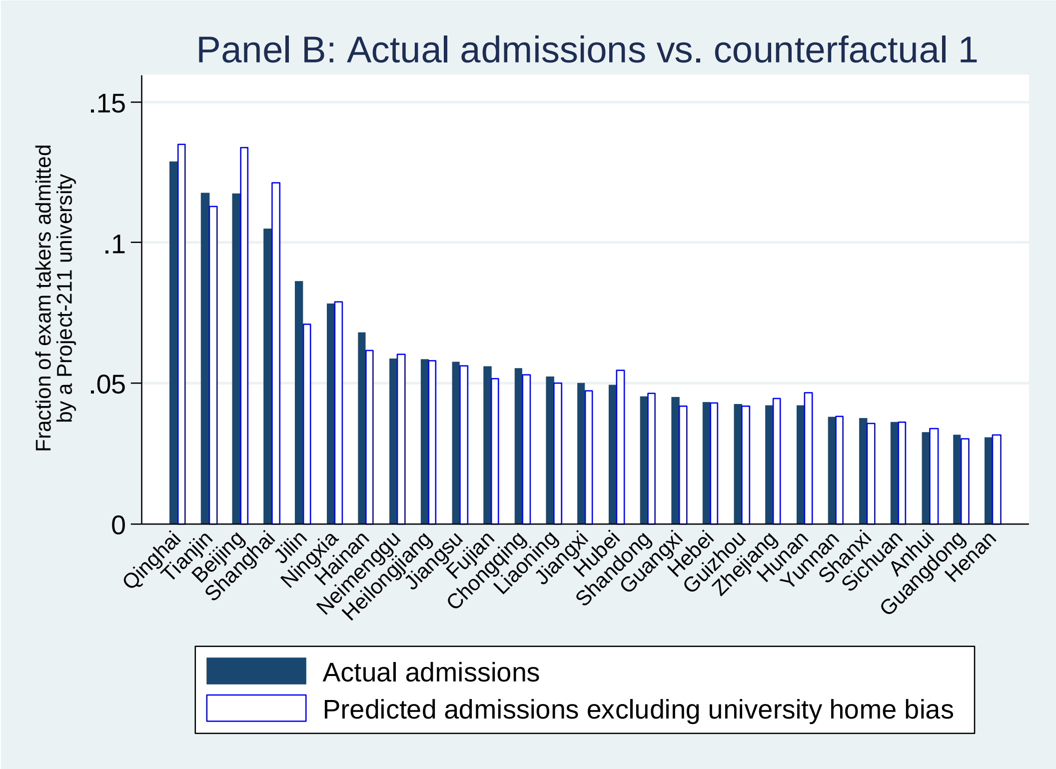
Note: Figure 8 plots the provincial fixed effects estimated by regression (13).

Figure 9: Actual and predicted geographic distribution of places



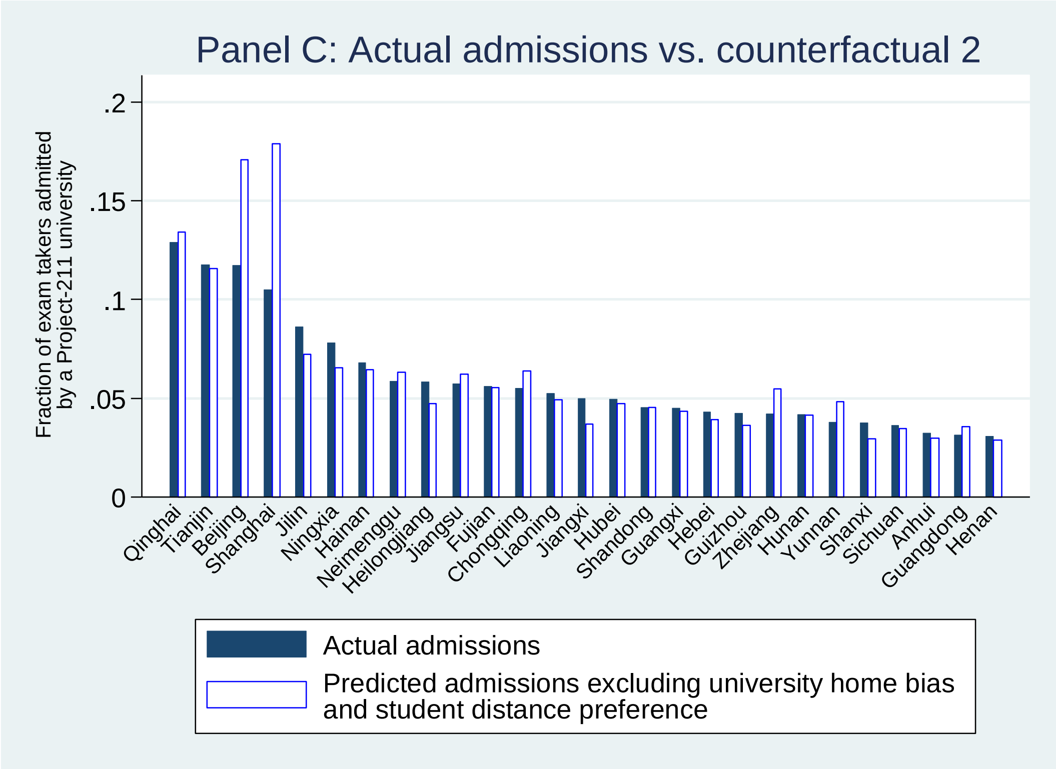
Notes: The solid bars represent the actual distribution of university places. The hollow bars represent the distribution in the benchmark case. For illustration purposes, we average the predictions from 2013 through 2015.

Figure 9: Actual and predicted geographic distribution of places (cont'd)



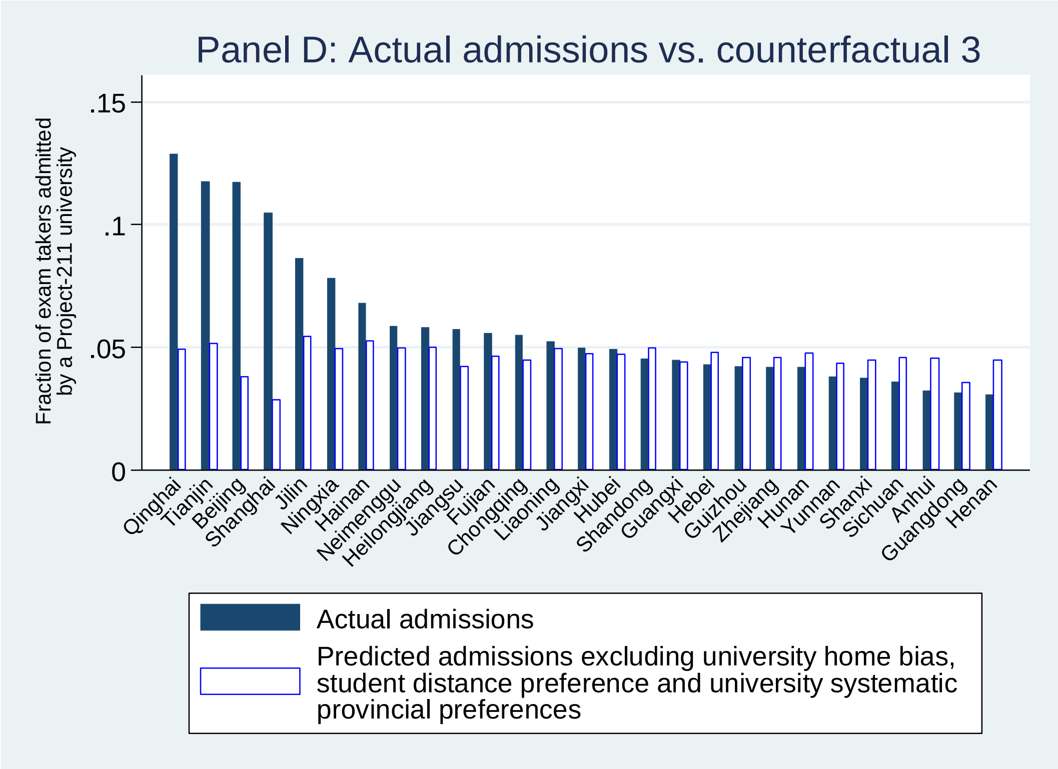
Notes: The solid bars represent the actual distribution of university places. The hollow bars represent the simulated distribution of university places of the counterfactual experiment that eliminates universities' home bias.

Figure 9: Actual and predicted geographic distribution of places (cont'd)



Notes: The solid bars represent the actual distribution of university places. The hollow bars represent the simulated distribution of university places of the counterfactual experiment that eliminates universities' home bias as well as students' preference for attending universities close to home.

Figure 9: Actual and predicted geographic distribution of places (cont'd)



Notes: The solid bars represent the actual distribution of university places. The hollow bars represent the simulated distribution of university places of the counterfactual experiment that eliminates universities' home bias, students' preference for attending universities close to home, and universities' systematic preference.

Table 1: Geographic distribution of universities in sample

Economic-Geographical division	Province	Number of 211 universities	Number of 985 universities
Northeast region	Heilongjiang	4	1
	Jilin	3	1
	Liaoning	4	2
East region	Beijing	24	8
	Fujian	2	1
	Guangdong	4	2
	Hainan	1	0
	Hebei	1	0
	Jiangsu	11	2
	Shandong	3	2
	Shanghai	9	4
	Tianjin	3	2
	Zhejiang	1	1
Central region	Anhui	3	1
	Henan	1	0
	Hubei	7	2
	Hunan	3	2
	Jiangxi	1	0
	Shanxi	1	0
West region	Chongqing	2	1
	Gansu	1	1
	Guangxi	1	0
	Guizhou	1	0
	Neimenggu	1	0
	Ningxia	1	0
	Qinghai	1	0
	Shaanxi	7	2
	Sichuan	5	2
	Yunnan	1	0
Total		107	37

Note: Table 1 shows the locations of the universities in our sample.

Table 2: Summary statistics

	Admissions			Cutoff percentile		
	Mean	S.D.	Obs.	Mean	S.D.	Obs.
	(1)	(2)	(3)	(4)	(5)	(6)
Science						
2013	99	249	2,889	0.91	0.09	2,781
2014	92	221	2,889	0.92	0.09	2,816
2015	93	227	2,889	0.92	0.09	2,787
2016	NA	NA	NA	0.92	0.09	2,808
Total	95	233	8,667	0.91	0.09	11,192
Arts						
2013	21	77	2,889	0.96	0.06	2,231
2014	20	72	2,889	0.97	0.05	2,267
2015	20	72	2,889	0.97	0.04	2,230
2016	NA	NA	NA	0.97	0.05	2,280
Total	21	74	8,667	0.97	0.05	9,008

Note: Table 2 reports the means and standard deviations of the number of admissions and cutoff percentile for each university in each province.

Table 3: Correlations between university systematic provincial preferences and provincial characteristics

Dependent variable: Estimated systematic value of province					
	(1)	(2)	(3)	(4)	(5)
Log GDP per capita	0.480***		-0.108		
	(0.094)		(0.098)		
Log pre-college educational spending per student		0.633***	0.725***	0.654***	0.695***
		(0.056)	(0.100)	(0.052)	(0.061)
Ethnic minority share				0.390**	0.336*
				(0.164)	(0.167)
Percentage of cohort admitted to a grammar high school					-0.492
					(0.382)
Constant	-4.646***	-5.473***	-5.211***	-5.709***	-5.836***
	(0.955)	(0.501)	(0.552)	(0.471)	(0.475)
Observations	27	27	27	27	27
R-squared	0.509	0.837	0.845	0.868	0.877
Adjusted R-squared	0.489	0.831	0.832	0.857	0.861

Notes: *** p<0.01, ** p<0.05, * p<0.1. Standard errors are in parentheses. The dependent variable is the estimated systematic value of each province ($\hat{a}$) averaged over years from 2013 through 2016 and streams, weighted by the number of exam takers. All value variables are converted to 2012 constant prices using the provincial-level CPI data obtained from the National Bureau of Statistics. The GDP per capita of each province is averaged over the period from 2000 through 2013. Per-student pre-college educational spending is calculated by the total government spending on pre-college education (including primary schools, middle schools and high schools) divided by the total number of students in primary, middle and high schools. Each province’s per-student educational spending is averaged over the period from 2004 through 2013. The ethnic minority share of each province is the percentage of the province's total residents who do not belong to the Han ethnic group according to the 2000 Chinese Census. The percentage of a cohort admitted to a grammar high school is calculated by the ratio of the high-school admissions to the number of primary-school graduates for the same cohort, and it is averaged between the cohorts who were newly enrolled in high schools during the period from 2010 through 2013.

Table 4: Determinants of university home bias

Dependent variable: University home bias		
	Science	Arts
	(1)	(2)
Dummy: Project 985	0.117**	0.024
	(0.050)	(0.046)
Dummy: Provincial universities	0.236***	0.154***
	(0.058)	(0.053)
Dummy: Beijing universities	-0.140**	-0.150***
	(0.053)	(0.049)
Dummy: Shanghai universities	-0.228***	-0.241***
	(0.077)	(0.070)
Constant	-0.057	-0.026
	(0.040)	(0.036)
Observations	99	98
R-squared	0.303	0.281

Notes: *** p<0.01, ** p<0.05, * p<0.1. Standard errors are in parentheses. For both streams, we are unable to estimate the home bias for eight universities that are located in either Shaanxi or Gansu as we do not have score distribution data for these two provinces. Additionally, for the arts stream, we are unable to estimate the home bias of the University of Science and Technology of China as it does not recruit from the arts stream.

Table 5: Determinants of student preference

Dependent variable: Popularity of university in each province				
	Science		Arts	
	(1)	(2)	(3)	(4)
Log (distance between university and province of candidates+1)	-0.424***	-0.445***	-0.391***	-0.358***
	(0.027)	(0.032)	(0.032)	(0.038)
Dummy: Project 985		1.816***		1.606***
		(0.085)		(0.080)
Log (average monthly salary of graduates)		8.316***		5.173***
		(0.275)		(0.199)
Log (population size of the city of university)		0.133***		0.190***
		(0.026)		(0.039)
Log (GDP per capita of the city of university)		0.031		0.772***
		(0.063)		(0.078)
Log (hedonic house price of the city of university)		-0.777***		-0.271**
		(0.082)		(0.099)
Fixed effects of the province of candidates	Y	Y	Y	Y
Year fixed effects	Y	Y	Y	Y
University fixed effects	Y		Y	
Observations	8,384	8,384	6,728	6,728
R-squared	0.868	0.715	0.852	0.568
Adjusted R-squared	0.866	0.714	0.849	0.566

Notes: Standard errors (in parentheses) are clustered at the province of candidates. The average monthly salary of graduates during the first five years after entering the labor force is calculated by IPIN.com based on data on a sample of 40 million young workers who had recently graduated from college. The hedonic quality-controlled home price index of the city of university is as of June 2013 and it is obtained from the Hang Lung Center for Real Estate at Tsinghua University. The distance from a university to a candidate's province is measured by the line distance between the city where the university is located and the capital city of the candidate's province.

Table 6: Actual and predicted normalized admission rates

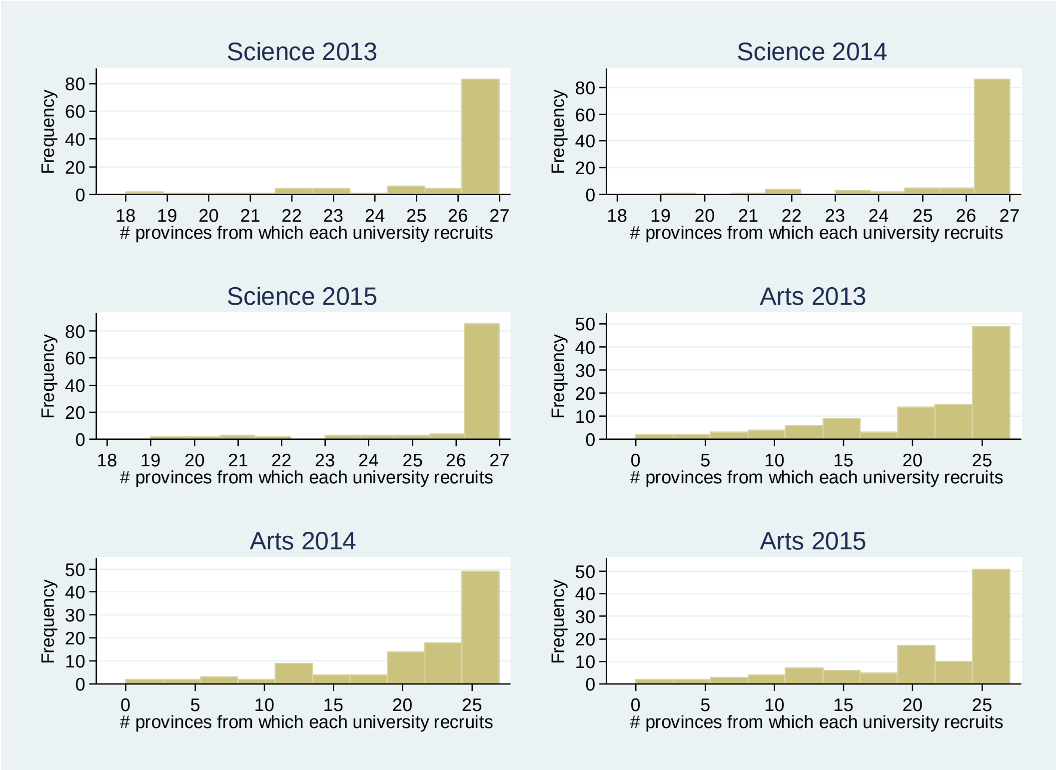
	Home province of university
Actual admissions	14.54
Baseline predictions	13.42
Counterfactual 1	14.73
Counterfactual 2	3.9
Counterfactual 3	1.75

Notes: The normalized admission rate of a province to a university is the fraction of places of a university going to a province divided by the share of the exam takers of the province. In counterfactual 1, we exclude the effect of university home bias. In counterfactual 2, we also eliminate the students' preferences for shorter distance based on the estimates in columns 1 and 3 of Table 5. In counterfactual 3, we eliminate the systematic preferences of the universities as well.

For Online Publication

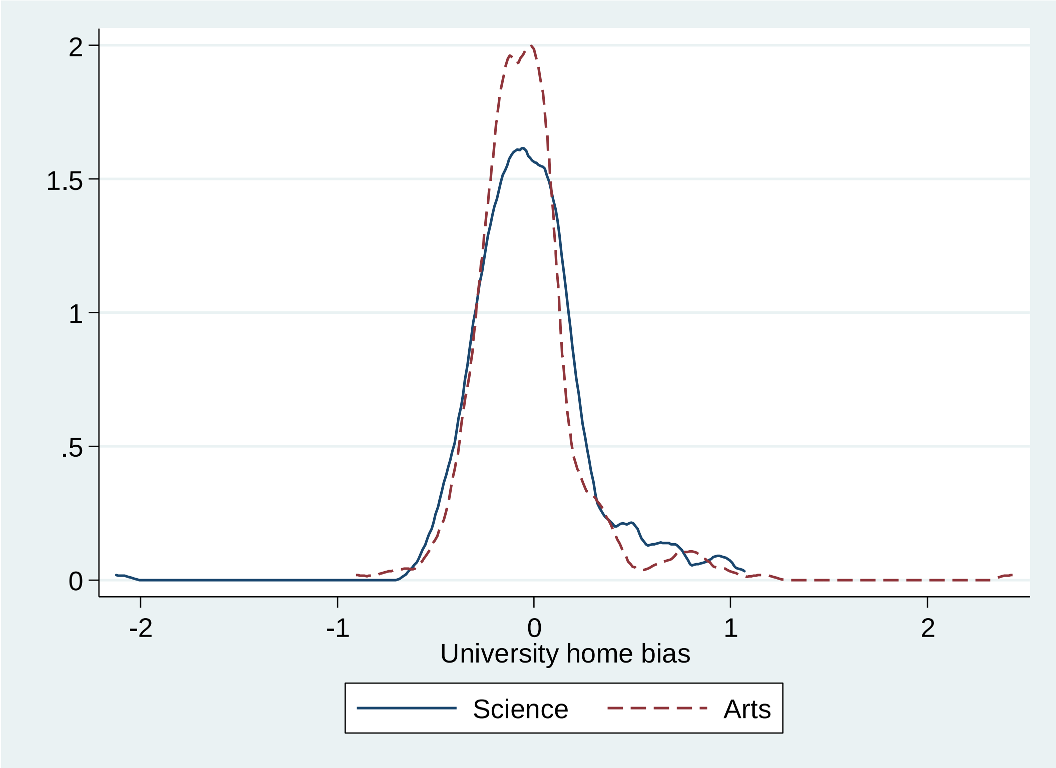
Appendix A: Tables and Figures

Figure A1: Number of provinces from which university recruits



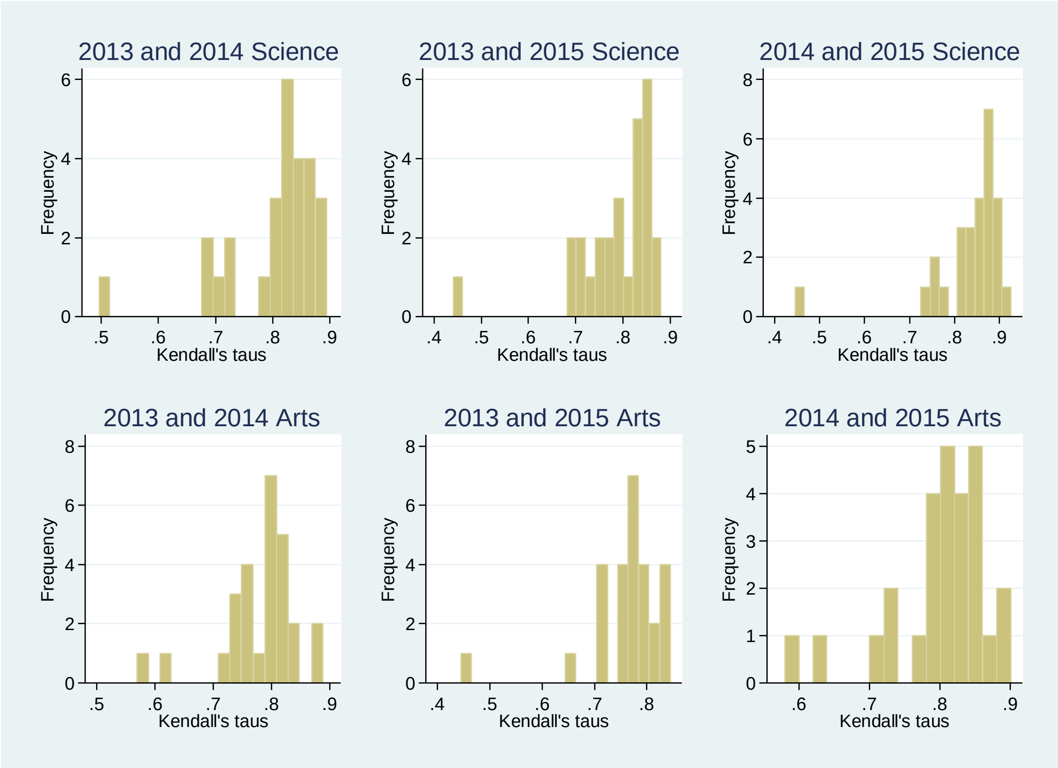
Note: The x-axis represents the number of provinces from which each university recruits.

Figure A2: Density plots of university home bias



Notes: We use the Epanechnikov kernel function for the density plots. The smoothing parameter is set at 0.05. To plot the density graphs, we pool the estimates of university home bias from all years. For the science stream, the cutoff scores of Southwest University in Chongqing in 2016 and Ningxia University in Ningxia from 2013 through 2016 are missing, therefore we are unable to estimate the home bias of these universities in the corresponding year. For the arts stream, we are not able to estimate the home bias for three universities in 2013 (University of Science and Technology of China, Beijing University of Chemical Technology, and Donghua University), three universities in 2014 (University of Science and Technology of China, Beijing University of Chemical Technology, and Northeast Forestry University), three universities in 2015 (University of Science and Technology of China, Donghua University and Qinghai University), and one university in 2016 (University of Science and Technology of China). The main reason for the missing is that these universities do not recruit students from the arts stream. In addition, the arts stream cutoff score for Qinghai University in Qinghai Province is missing for 2015. Furthermore, for both streams, we are unable to estimate the home bias for eight universities that are located in either Shaanxi Province or in Gansu Province as we do not have score distribution data for these two provinces.

Figure A3: Rank correlation coefficients of student preference in each province across years



Note: The x-axis is the rank correlation coefficient of student preference in each province across the specified years.

Table A1: General admissions, special admissions and planned quotas

Year	Number of general admissions (gaokao.chsi.com.cn)	Number of special admissions (gaokao.chsi.com.cn)	Number of admissions (gaokao.chsi.com.cn)	Number of planned quotas (Ministry of Education)	Share of general admissions in planned quotas
	(1)	(2)	(3)	(4)	(5)
2013	381354	433563	52209	527947	72.23%
2014	354744	NA	NA	518798	68.38%
2015	357558	421556	63998	519566	68.82%

Notes: Table A1 reports the total number of admissions from the 107 universities in all 31 provinces. The data in columns 1 and 2 are obtained from gaokao.chsi.com.cn. The data in column 4 are obtained from the Ministry of Education. The number in column 3 is the sum of those from columns 1 and 2. The number in column 5 is the number in column 1 divided by that in column 4.

Table A2: Numbers of exam takers and admissions

	Number of exam takers in the 27 provinces			Number of general admissions from the 107 universities in the 27 provinces			Number of admissions in the 27 provinces (degree programs only)	Number of admissions in all 31 provinces (degree programs only)
	Science	Arts	Total	Science	Arts	Total	Total	Total
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
2013	4,303,928	2,997,191	7,301,120	287,364	62,088	349,452	3,486,538	3,814,331
2014	4,395,310	3,011,071	7,406,381	265,999	58,276	324,275	3,500,605	3,834,152
2015	4,335,850	2,900,745	7,236,595	268,947	58,898	327,845	3,569,987	3,894,184
2016	4,321,161	2,866,503	7,187,665	NA	NA	NA	3,726,576	4,054,007

Notes: Columns 1-3 report the numbers of exam takers, which are obtained from the websites of provincial education examination authorities. Columns 4-6 report the numbers of general admissions from the 107 universities in the 27 provinces, which are obtained from gaokao.chsi.com.cn. Column 8 reports the numbers of university admissions (including only the undergraduate-degree programs), which are obtained from the Education Statistics Yearbooks of China. From various news articles, we obtain the numbers of university admissions (including only the undergraduate-degree programs) in Gansu, Shaanxi, Xinjiang and Tibet. We use these numbers and the numbers in column 8 to calculate the numbers of university admissions in the 27 provinces in our sample and report them in column 7.

Table A3: Universities in sample

University	Province	City	Project-985	Central university
Anhui University	Anhui	Hefei	0	0
Hefei University of Technology	Anhui	Hefei	0	1
University of Science and Technology of China	Anhui	Hefei	1	1
Peking University	Beijing	Beijing	1	1
Beijing University of Technology	Beijing	Beijing	0	0
Beihang University	Beijing	Beijing	1	1
Beijing University of Chemical Technology	Beijing	Beijing	0	1
Beijing Jiaotong University	Beijing	Beijing	0	1
University of Science and Technology Beijing	Beijing	Beijing	0	1
Beijing Institute of Technology	Beijing	Beijing	1	1
Beijing Forestry University	Beijing	Beijing	0	1
Beijing Normal University	Beijing	Beijing	1	1
Beijing Sport University	Beijing	Beijing	0	1
Beijing Foreign Studies University	Beijing	Beijing	0	1
Beijing University of Posts and Telecommunications	Beijing	Beijing	0	1
University of International Business and Economics	Beijing	Beijing	0	1
North China Electric Power University, Beijing	Beijing	Beijing	0	1
Tsinghua University	Beijing	Beijing	1	1
Communication University of China	Beijing	Beijing	0	1
China University of Geosciences, Beijing	Beijing	Beijing	0	1
China University of Mining & Technology, Beijing	Beijing	Beijing	0	1
China Agricultural University	Beijing	Beijing	1	1
Renmin University of China	Beijing	Beijing	1	1
China University of Petroleum, Beijing	Beijing	Beijing	0	1
China University of Political Science and Law	Beijing	Beijing	0	1
Central University of Finance and Economics	Beijing	Beijing	0	1
Minzu University of China	Beijing	Beijing	1	1
Southwest University	Chongqing	Chongqing	0	1
Chongqing University	Chongqing	Chongqing	1	1
Fuzhou University	Fujian	Fuzhou	0	0
Xiamen University	Fujian	Xiamen	1	1
Lanzhou University	Gansu	Lanzhou	1	1
South China University of Technology	Guangdong	Guangzhou	1	1
South China Normal University	Guangdong	Guangzhou	0	0
Sun Yat-sen University	Guangdong	Guangzhou	1	1
Jinan University	Guangdong	Guangzhou	0	1

Note: Table A3 presents the list of the 107 universities in our sample.

Table A3: Universities in sample (cont'd)

University	Province	City	Project-985	Central university
Guangxi University	Guangxi	Nanning	0	0
Guizhou University	Guizhou	Guiyang	0	0
Hainan University	Hainan	Haikou	0	0
North China Electric Power University, Baoding	Hebei	Baoding	0	1
Northeast Forestry University	Heilongjiang	Harbin	0	1
Northeast Agricultural University	Heilongjiang	Harbin	0	0
Harbin Engineering University	Heilongjiang	Harbin	0	1
Harbin Institute of Technology	Heilongjiang	Harbin	1	1
Zhengzhou University	Henan	Zhengzhou	0	0
Huazhong University of Science and Technology	Hubei	Wuhan	1	1
Huazhong Agricultural University	Hubei	Wuhan	0	1
Central China Normal University	Hubei	Wuhan	0	1
Wuhan University	Hubei	Wuhan	1	1
Wuhan University of Technology	Hubei	Wuhan	0	1
China University of Geosciences, Wuhan	Hubei	Wuhan	0	1
Zhongnan University of Economics and Law	Hubei	Wuhan	0	1
Hunan University	Hunan	Changsha	1	1
Hunan Normal University	Hunan	Changsha	0	0
Central South University	Hunan	Changsha	1	1
Southeast University	Jiangsu	Nanjing	1	1
Hohai University	Jiangsu	Nanjing	0	1
Nanjing University	Jiangsu	Nanjing	1	1
Nanjing University of Aeronautics and Astronautics	Jiangsu	Nanjing	0	1
Nanjing University of Science and Technology	Jiangsu	Nanjing	0	1
Nanjing Agricultural University	Jiangsu	Nanjing	0	1
Nanjing Normal University	Jiangsu	Nanjing	0	0
China Pharmaceutical University	Jiangsu	Nanjing	0	1
Soochow University	Jiangsu	Suzhou	0	0
Jiangnan University	Jiangsu	Wuxi	0	1
China University of Mining and Technology	Jiangsu	Xuzhou	0	1
Nanchang University	Jiangxi	Nanchang	0	0
Northeast Normal University	Jilin	Changchun	0	1
Jilin University	Jilin	Changchun	1	1
Yanbian University	Jilin	Yanji	0	0
Dalian Maritime University	Liaoning	Dalian	0	1
Dalian University of Technology	Liaoning	Dalian	1	1

Note: Table A3 presents the list of the 107 universities in our sample.

Table A3: Universities in sample (cont'd)

University	Province	City	Project-985	Central university
Northeastern University	Liaoning	Shenyang	1	1
Liaoning University	Liaoning	Shenyang	0	0
Inner Mongolia University	Neimenggu	Hohhot	0	0
Ningxia University	Ningxia	Yinchuan	0	0
Qinghai university	Qinghai	Xining	0	0
Chang'an University	Shaanxi	Xi'an	0	1
Shaanxi Normal University	Shaanxi	Xi'an	0	1
Xidian University	Shaanxi	Xi'an	0	1
Xi'an Jiaotong University	Shaanxi	Xi'an	1	1
Northwest University	Shaanxi	Xi'an	0	0
Northwestern Polytechnical University	Shaanxi	Xi'an	0	1
Northwest A&F University	Shaanxi	Xianyang	1	1
Shandong University	Shandong	Ji'nan	1	1
Ocean University of China	Shandong	Qingdao	1	1
China University of Petroleum, East China	Shandong	Qingdao	0	1
Donghua University	Shanghai	Shanghai	0	1
Fudan University	Shanghai	Shanghai	1	1
East China University of Science and Technology	Shanghai	Shanghai	0	1
East China Normal University	Shanghai	Shanghai	1	1
Shanghai University of Finance and Economics	Shanghai	Shanghai	0	1
Shanghai University	Shanghai	Shanghai	0	0
Shanghai Jiao Tong University	Shanghai	Shanghai	1	1
Shanghai International Studies University	Shanghai	Shanghai	0	1
Tongji University	Shanghai	Shanghai	1	1
Taiyuan University of Technology	Shanxi	Taiyuan	0	0
University of Electronic Science and Technology of China	Sichuan	Chengdu	1	1
Sichuan University	Sichuan	Chengdu	1	1
Southwestern University of Finance and Economics	Sichuan	Chengdu	0	1
Southwest Jiaotong University	Sichuan	Chengdu	0	1
Sichuan Agricultural University	Sichuan	Ya'an	0	0
Hebei University of Technology	Tianjin	Tianjin	0	0
Nankai University	Tianjin	Tianjin	1	1
Tianjin University	Tianjin	Tianjin	1	1
Yunnan University	Yunnan	Kunming	0	0
Zhejiang University	Zhejiang	Hangzhou	1	1

Note: Table A3 presents the list of the 107 universities in our sample.

Table A4: Correlation of cutoff percentiles across years

Science (balanced panel of 2,722 university-province pairs)				
	2013	2014	2015	2016
2013	1			
2014	0.8258	1		
2015	0.7698	0.8781	1	
2016	0.7487	0.8377	0.8505	1

Arts (balanced panel of 2,113 university-province pairs)				
	2013	2014	2015	2016
2013	1			
2014	0.8085	1		
2015	0.8627	0.841	1	
2016	0.8686	0.8008	0.889	1

Note: Table A4 reports the correlation coefficients of the cutoff percentiles of university-province pairs across years.

Table A5: Estimates of systematic provincial preferences

Panel A: Science

Province	Average	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE
		2013		2014		2015		2016	
Shanghai	1.206	1.253	0.030	1.077	0.029	0.999	0.027	1.050	0.027
Beijing	0.974	1.068	0.031	0.925	0.030	0.777	0.029	0.751	0.028
Qinghai	0.638	0.738	0.029	0.577	0.028	0.670	0.027	0.547	0.026
Tianjin	0.550	0.538	0.029	0.478	0.028	0.436	0.026	0.385	0.026
Jiangsu	0.319	0.304	0.029	0.207	0.028	0.235	0.027	0.194	0.026
Jilin	0.286	0.196	0.029	0.162	0.027	0.229	0.026	0.243	0.025
Ningxia	0.264	0.226	0.029	0.169	0.028	0.182	0.027	0.117	0.026
Chongqing	0.255	0.437	0.029	0.217	0.027	0.159	0.026	0.164	0.026
Neimenggu	0.223	0.227	0.029	0.192	0.027	0.200	0.026	0.191	0.025
Zhejiang	0.223	0.148	0.028	0.099	0.027	0.203	0.026	0.185	0.025
Fujian	0.220	0.180	0.028	0.103	0.027	0.195	0.026	0.210	0.025
Hainan	0.217	0.213	0.029	0.161	0.027	0.121	0.026	0.081	0.026
Liaoning	0.128	0.121	0.029	0.041	0.027	-0.013	0.026	0.071	0.026
Yunnan	0.112	0.181	0.028	0.208	0.027	0.027	0.026	-0.074	0.025
Hubei	0.108	0.092	0.029	0.064	0.028	0.100	0.026	0.065	0.026
Guangdong	0.103	0.107	0.029	0.045	0.028	0.073	0.026	-0.015	0.026
Heilongjiang	0.068	0.112	0.029	0.016	0.028	0.039	0.026	0.000	0.026
Shandong	0.063	0.051	0.029	-0.012	0.027	-0.004	0.026	-0.064	0.025
Guangxi	0.063	0.192	0.029	0.041	0.027	0.025	0.026	-0.051	0.025
Hunan	0.038	0.063	0.029	-0.002	0.027	-0.024	0.026	-0.023	0.025
Hebei	0	0	NA	0	NA	0	NA	0	NA
Sichuan	-0.034	0.011	0.029	-0.082	0.028	-0.077	0.026	-0.102	0.026
Jiangxi	-0.038	-0.019	0.028	-0.049	0.027	-0.064	0.026	-0.187	0.025
Guizhou	-0.045	0.096	0.029	-0.065	0.027	-0.149	0.026	-0.208	0.025
Anhui	-0.086	-0.083	0.029	-0.161	0.027	-0.173	0.026	-0.168	0.025
Shanxi	-0.111	-0.094	0.029	-0.120	0.027	-0.085	0.026	-0.137	0.025
Henan	-0.128	-0.120	0.029	-0.099	0.027	-0.117	0.026	-0.170	0.025

Note: The numbers in the first column represent the estimates of the systematic values of each province averaged over years and over streams, weighted by the number of exam takers.

Table A5: Estimates of systematic provincial preferences (cont'd)

Panel B: Arts

Province	Average	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE
		2013		2014		2015		2016	
Shanghai	1.206	1.609	0.030	1.281	0.029	1.294	0.029	1.376	0.029
Beijing	0.974	1.299	0.029	1.205	0.027	1.049	0.027	1.068	0.027
Qinghai	0.638	0.694	0.031	0.682	0.029	0.668	0.028	0.567	0.027
Tianjin	0.550	0.880	0.027	0.839	0.025	0.714	0.025	0.711	0.025
Jiangsu	0.319	0.577	0.027	0.488	0.025	0.437	0.025	0.409	0.025
Jilin	0.286	0.485	0.027	0.491	0.025	0.433	0.025	0.511	0.025
Ningxia	0.264	0.485	0.030	0.468	0.028	0.385	0.028	0.386	0.027
Chongqing	0.255	0.325	0.027	0.264	0.025	0.257	0.025	0.263	0.024
Neimenggu	0.223	0.312	0.028	0.274	0.026	0.246	0.025	0.209	0.025
Zhejiang	0.223	0.408	0.025	0.310	0.024	0.313	0.024	0.325	0.024
Fujian	0.220	0.268	0.027	0.263	0.025	0.353	0.025	0.402	0.025
Hainan	0.217	0.423	0.030	0.326	0.028	0.293	0.027	0.324	0.026
Liaoning	0.128	0.392	0.026	0.271	0.025	0.192	0.024	0.275	0.024
Yunnan	0.112	0.230	0.028	0.225	0.026	0.127	0.026	0.087	0.025
Hubei	0.108	0.147	0.027	0.177	0.025	0.149	0.025	0.103	0.025
Guangdong	0.103	0.281	0.028	0.159	0.026	0.129	0.025	0.089	0.026
Heilongjiang	0.068	0.205	0.027	0.096	0.025	0.045	0.025	0.068	0.025
Shandong	0.063	0.303	0.025	0.198	0.024	0.171	0.023	0.137	0.023
Guangxi	0.063	0.178	0.028	0.067	0.026	0.056	0.026	0.045	0.025
Hunan	0.038	0.099	0.026	0.085	0.024	0.101	0.024	0.068	0.024
Hebei	0	0	NA	0	NA	0	NA	0	NA
Sichuan	-0.034	0.044	0.026	0.011	0.025	-0.042	0.024	-0.004	0.024
Jiangxi	-0.038	0.113	0.027	0.035	0.025	0.032	0.025	-0.032	0.024
Guizhou	-0.045	0.110	0.028	0.065	0.026	-0.043	0.026	0.000	0.025
Anhui	-0.086	0.002	0.026	0.006	0.024	0.048	0.024	-0.008	0.024
Shanxi	-0.111	-0.097	0.026	-0.110	0.024	-0.137	0.024	-0.106	0.024
Henan	-0.128	-0.076	0.026	-0.084	0.024	-0.174	0.024	-0.182	0.023

Note: The numbers in the first column represent the estimates of the systematic values of each province averaged over years and over streams, weighted by the number of exam takers.

Table A6: Correlations of university preference parameters across years

Panel A: Systematic provincial preferences

Correlation of *a* across years, science (balanced panel of 27 provinces)				
	2013	2014	2015	2016
2013	1			
2014	0.9865	1		
2015	0.9626	0.9747	1	
2016	0.9526	0.9617	0.9864	1

Correlation of *a* across years, arts (balanced panel of 27 provinces)				
	2013	2014	2015	2016
2013	1			
2014	0.9877	1		
2015	0.9815	0.9888	1	
2016	0.9814	0.9828	0.9917	1

Note: Table A6, Panel A presents the correlations of universities' systematic provincial preferences across years.

Table A6: Correlations of university preference parameters across years (cont'd)

Panel B: Home-bias parameters

Correlation of ρ across years, science (balanced panel of 97 universities)				
	2013	2014	2015	2016
2013	1			
2014	0.6928	1		
2015	0.5432	0.8077	1	
2016	0.2878	0.7441	0.7025	1

Correlation of ρ across years, arts (balanced panel of 94 universities)				
	2013	2014	2015	2016
2013	1			
2014	0.3528	1		
2015	0.5449	0.5484	1	
2016	0.4979	0.5146	0.6601	1

Notes: Table A6, Panel B presents the correlations of universities' home bias across years. For the science stream, the cutoff scores of Southwest University in Chongqing in 2016 and Ningxia University in Ningxia from 2013 through 2016 are missing, therefore we are unable to estimate the home bias of these universities in the corresponding year. For the arts stream, we are not able to estimate the home bias for three universities in 2013 (University of Science and Technology of China, Beijing University of Chemical Technology, and Donghua University), three universities in 2014 (University of Science and Technology of China, Beijing University of Chemical Technology, and Northeast Forestry University), three universities in 2015 (University of Science and Technology of China, Donghua University and Qinghai University), and one university in 2016 (University of Science and Technology of China). The main reason for the missing is that these universities do not recruit students from the arts stream. In addition, the arts stream cutoff score for Qinghai University in Qinghai Province is missing for 2015. Furthermore, for both streams, we are unable to estimate the home bias for eight universities that are located in either Shaanxi Province or in Gansu Province as we do not have score distribution data for these two provinces.

Table A7: Actual and predicted Project-211 fraction

Panel A: 2013

Province	Actual admissions	Baseline predictions	Counterfactual 1	Counterfactual 2	Counterfactual 3
Anhui	3.40%	3.57%	3.54%	3.00%	4.67%
Beijing	11.80%	11.26%	13.17%	18.87%	3.76%
Chongqing	6.07%	6.18%	6.40%	7.84%	4.69%
Fujian	5.08%	5.14%	4.61%	5.22%	4.72%
Guangdong	3.36%	3.20%	3.24%	3.78%	3.79%
Guangxi	4.80%	4.86%	4.32%	4.86%	4.46%
Guizhou	5.15%	5.33%	5.12%	4.39%	4.81%
Hainan	7.48%	7.50%	6.86%	7.28%	5.92%
Hebei	4.06%	3.92%	3.94%	3.65%	4.92%
Heilongjiang	5.88%	5.71%	5.98%	5.09%	5.24%
Henan	2.96%	3.03%	3.05%	2.74%	4.70%
Hubei	4.78%	4.85%	5.31%	4.57%	4.97%
Hunan	4.52%	4.59%	4.85%	4.21%	4.92%
Jiangsu	5.93%	5.85%	6.02%	6.68%	4.47%
Jiangxi	5.80%	5.90%	5.44%	3.89%	5.17%
Jilin	8.75%	8.44%	6.45%	7.02%	5.77%
Liaoning	5.31%	5.21%	4.82%	5.29%	5.01%
Neimenggu	6.39%	6.55%	6.51%	6.44%	5.33%
Ningxia	9.81%	9.96%	9.51%	6.93%	5.38%
Qinghai	14.53%	14.34%	15.50%	14.71%	5.37%
Shandong	4.78%	4.66%	4.83%	4.80%	5.27%
Shanghai	10.58%	10.33%	13.01%	20.07%	2.80%
Shanxi	3.84%	3.92%	3.87%	2.88%	4.76%
Sichuan	3.98%	3.82%	4.07%	3.81%	4.93%
Tianjin	12.59%	12.32%	12.07%	12.34%	5.49%
Yunnan	4.28%	4.37%	4.28%	4.97%	4.55%
Zhejiang	4.22%	4.54%	4.43%	5.45%	4.93%

Note: Table A7 presents the actual and simulated distributions of university places across provinces.

Table A7: Actual and predicted Project-211 fraction (cont'd)

Panel B: 2014

Province	Actual admissions	Baseline predictions	Counterfactual 1	Counterfactual 2	Counterfactual 3
Anhui	3.14%	3.18%	3.23%	2.86%	4.47%
Beijing	11.18%	10.92%	12.82%	16.97%	3.58%
Chongqing	5.07%	4.98%	4.71%	5.70%	4.26%
Fujian	5.16%	5.15%	4.68%	4.99%	4.48%
Guangdong	2.89%	2.83%	2.85%	3.43%	3.55%
Guangxi	4.05%	4.00%	3.61%	3.89%	4.19%
Guizhou	4.05%	4.07%	3.97%	3.54%	4.52%
Hainan	6.69%	6.47%	5.81%	6.02%	4.80%
Hebei	4.24%	4.20%	4.24%	3.89%	4.60%
Heilongjiang	5.17%	5.17%	5.11%	4.30%	4.74%
Henan	3.18%	3.14%	3.20%	2.99%	4.35%
Hubei	4.82%	4.96%	5.20%	4.54%	4.49%
Hunan	4.00%	4.41%	4.73%	4.08%	4.65%
Jiangsu	5.56%	5.44%	5.44%	5.85%	4.12%
Jiangxi	4.68%	4.70%	4.39%	3.53%	4.46%
Jilin	8.41%	8.43%	7.60%	6.93%	5.26%
Liaoning	5.15%	5.09%	5.14%	4.81%	4.78%
Neimenggu	5.40%	5.44%	5.51%	6.01%	4.68%
Ningxia	7.56%	7.92%	7.81%	6.34%	4.72%
Qinghai	12.83%	12.71%	13.60%	12.32%	4.86%
Shandong	4.49%	4.45%	4.60%	4.33%	4.82%
Shanghai	10.86%	10.34%	12.09%	16.93%	2.82%
Shanxi	3.73%	3.75%	3.53%	2.86%	4.38%
Sichuan	3.39%	3.25%	3.34%	3.22%	4.35%
Tianjin	11.75%	11.27%	11.05%	11.54%	4.96%
Yunnan	3.59%	3.59%	3.65%	5.28%	4.08%
Zhejiang	3.99%	4.33%	4.25%	5.00%	4.44%

Note: Table A7 presents the actual and simulated distributions of university places across provinces.

Table A7: Actual and predicted Project-211 fraction (cont'd)

Panel C: 2015

Province	Actual admissions	Baseline predictions	Counterfactual 1	Counterfactual 2	Counterfactual 3
Anhui	3.19%	3.25%	3.39%	3.02%	4.59%
Beijing	12.23%	12.11%	14.10%	15.35%	4.07%
Chongqing	5.39%	5.28%	4.80%	5.64%	4.52%
Fujian	6.52%	6.45%	6.20%	6.40%	4.72%
Guangdong	3.20%	3.09%	2.99%	3.50%	3.42%
Guangxi	4.63%	4.69%	4.62%	4.28%	4.61%
Guizhou	3.52%	3.71%	3.45%	2.99%	4.49%
Hainan	6.21%	6.05%	5.83%	6.00%	5.06%
Hebei	4.64%	4.54%	4.72%	4.22%	4.86%
Heilongjiang	6.45%	6.27%	6.28%	4.77%	5.06%
Henan	3.06%	3.10%	3.25%	2.92%	4.44%
Hubei	5.22%	5.50%	5.85%	5.08%	4.68%
Hunan	4.05%	4.20%	4.42%	4.15%	4.77%
Jiangsu	5.72%	5.63%	5.40%	6.10%	4.10%
Jiangxi	4.50%	4.51%	4.36%	3.64%	4.61%
Jilin	8.72%	8.52%	7.20%	7.73%	5.30%
Liaoning	5.25%	5.06%	5.01%	4.62%	5.07%
Neimenggu	5.82%	5.89%	6.06%	6.45%	4.92%
Ningxia	6.09%	6.52%	6.31%	6.37%	4.77%
Qinghai	11.30%	12.18%	11.38%	13.23%	4.56%
Shandong	4.33%	4.29%	4.51%	4.50%	4.88%
Shanghai	10.04%	9.71%	11.27%	16.59%	2.99%
Shanxi	3.70%	3.66%	3.29%	3.07%	4.36%
Sichuan	3.46%	3.32%	3.39%	3.37%	4.48%
Tianjin	10.94%	10.74%	10.74%	10.82%	5.03%
Yunnan	3.54%	3.51%	3.55%	4.23%	4.42%
Zhejiang	4.38%	4.68%	4.68%	6.01%	4.45%

Note: Table A7 presents the actual and simulated distributions of university places across provinces.

Appendix B: Details of National College Entrance Exam

Province	2013	2014	2015	2016	Total exam score	Textbook version
Anhui				I	750	PEP Version A、BNP
Beijing					750	PEP Version B
Chongqing				II	750	PEP Version A
Fujian				I	750	PEP Version A
Guangdong				I	750	PEP Version A
Guangxi	II	II	II	III	750	PEP Version A
Guizhou	II	II	II	III	750	PEP Version A
Hainan					900	PEP Version A
Hebei	I	I	I	I	750	PEP Version A
Heilongjiang	I	I	II	II	750	PEP Version A
Henan	I	I	I	I	750	PEP Version A、BNP
Hubei				I	750	PEP Version A
Hunan				I	750	PEP Version A
Jiangsu					485	JEP
Jiangxi			I	I	750	BNP
Jilin	I	I	II	II	750	PEP Version A
Liaoning			II	II	750	PEP Version A
Neimenggu	I	I	II	II	750	PEP Version A\B
Ningxia	I	I	II	II	750	PEP Version A
Qinghai	II	II	II	II	750	PEP Version A
Shandong					750	PEP Version A\B
Shanghai					630	SEP
Shanxi	I	I	I	I	750	PEP Version A
Sichuan				III	750	PEP Version A
Tianjin					750	PEP Version A
Yunnan	I	I	II	III	750	PEP Version A
Zhejiang					810	PEP Version A

Notes: This table presents the versions of exam papers and textbooks used in each province. *I* indicates a province that used the National-I version of the exam papers. *II* indicates a province that used the National-II version of the exam papers. *III* indicates a province that used the National-III version of the exam papers. The others used their own exam papers. *PEP* means the People's Education Press, *BNP* means the Beijing Normal University Press, *JEP* means the Jiangsu Education Press, and *SEP* means the Shanghai Education Press.

Appendix C: Data Appendix

Provincial test score distribution

Provincial test score distributions are mainly collected from the websites of the Provincial Education Examinations Authorities (PEEA). Other sources include admissions guide books published by PEEA (Fujian and Shanghai) and the following admissions consulting websites: gaokao.com (Gaokaowang); gaokao.eol.cn (Zhongguo Jiaoyu Zaixian); gxeduw.edu (Gaokao Xinxiwang); gaokao.chsi.com.cn (Yangguang Gaokao). In total, we obtained 216 score distribution tables from 27 provinces from 2013 through 2016.

Among the 216 tables, 174 provide for each score the fraction of the exam takers who received the score. Another 36 tables provide the fraction of exam takers within an interval. For these tables, we interpolate linearly. Some of the tables are truncated from below, but in most cases, the cutoff scores of all the universities are above the truncation score. Of the remaining six tables, two provide the score associated with each percentile rank, and four provide the percentile rank that corresponds to the cutoff score of each university. Most distribution tables include the number of exam takers in each stream. When they do not, we use the number of exam takers in each stream reported by the PEEA or provincial education bureaus. When we have only the total number of exam takers but not the breakdown into the science and arts streams, we use the average breakdowns in years where the data are available.

Admission data

Admission data from 2013 through 2015 are obtained from gaokao.chsi.com.cn, a website affiliated with the Ministry of Education. Some universities' websites also provide data on admission quotas and the number of admitted students. We compare the admission data from gaokao.chsi.com.cn with those from universities' websites and find that they largely agree.

Cutoff-score data

Cutoff scores are mainly obtained from two sources: universities' websites, and two publications that contain the cutoff scores of Chinese universities: Quanguo

Zhongdiandaxue Luqufenshuxian 2017 (2015 and 2016 cutoff scores) and Quanguo Gaoxiao Luqufenshuxian Tongji 2015 (2013 and 2014 cutoff scores). For the science stream, 9,901 university-province-year observations are from the universities' websites, and 1,297 university-province-year observations are from the two publications. For the arts stream, 8,003 school-province-year observations are from the universities' websites, and 1,007 school-province-year observations are obtained from the two publications.

There are 103 cases of missing cutoff scores. These are the cases where our admission data show that students were admitted from a certain province by a university, but the corresponding cutoff scores are not available from either the university's website or the two publications. In another eight cases, the cutoff scores are below the truncation score of the test-score distribution table. These school-province-year observations are discarded and not used in either the university-side or student-side estimation.