Social Science Research 41 (2012) 287–305

Contents lists available at SciVerse ScienceDirect

Social Science Research

journal homepage: www.elsevier.com/locate/ssresearch

The effects of attending selective college tiers in China

Prashant Loyalka *, Yingquan Song, Jianguo Wei

China Institute for Educational Finance Research, Peking University, #5 Summer Palace Rd., Rm. 406, Education Building, Beijing 100871, China

ARTICLE INFO

Article history:
Received 24 October 2010
Revised 4 November 2011
Accepted 21 November 2011
Available online 8 December 2011

Keywords:
Higher education
International education
Quasiexperimental design
College selectivity
China

ABSTRACT

We estimate the effects of attending the first versus second-tier of higher education institutions on Chinese students' at-college and expected post-college outcomes using various quasi-experimental methods such as regression discontinuity, genetic matching, and regression discontinuity controlling for covariates. Overall we find that just attending the first versus second-tier makes little difference in terms of students' class ranking, net tuition, expected wages, or likelihood of applying for graduate school. The results do show, however, that just attending the first versus second tier makes it less likely that students will get their preferred major choice.

1. Introduction

In recent years, China's government has placed a great deal of emphasis on raising the quality of the country's most selective higher education institutions. It has allocated substantial funding to improve both the physical conditions and the teaching and research of the country's most selective universities and programs (Li, 2010). It has also granted larger per student subsidies to more selective universities compared to less selective universities (Bao, 2007). In addition, the government has controlled the growth of student enrollments in more selective institutions (Loyalka et al., 2009a), thus potentially allowing these institutions to maintain lower student-to-faculty ratios and a greater focus on instructional quality.

Another factor which affects quality in the higher education system is that China utilizes a college application and admissions process that systematically facilitates the entry of the highest ability students (in terms of college entrance exam scores) into the most selective higher education institutions (Loyalka et al., 2009a). This is accomplished by categorizing all of the nation's universities into various selectivity tiers: generally speaking, the most selective public 4-year universities and programs comprise the first tier, less selective public 4-year universities and programs comprise the second tier, still less selective 4-year private universities comprise the third tier and 3-year vocational institutions comprise the fourth tier. Each year provincial governments usually set official tier eligibility cutoff scores based on the overall results on an annual college entrance exam so that only higher scoring students are eligible to attend higher tier institutions. These and other elements of the college application and admissions process ultimately lead to a high concentration of the highest ability students in the first tier.

The government's classification of universities into a hierarchy of selective tiers, combined with greater governmental financial support for first-tier institutions and the heavy concentration of higher ability students in the first-tier, often leads students and families to conclude that attending a first-tier versus a second-tier institution impacts a student's college and post-college experiences. However little is actually known about the causal effects of attending first-tier versus second-tier

* Corresponding author. Fax: +86 10 62756183.
E-mail addresses: prashant@ciefr.pku.edu.cn (P. Loyalka), yqsong@ciefr.pku.edu.cn (Y. Song), jgwei@ciefr.pku.edu.cn (J. Wei).

0049-089X/$ - see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.ssresearch.2011.11.015

institutions on the outcomes of college students in China. Also little is known about whether attending a first-tier versus a second-tier institution really matters for the tens of thousands of students each year who are just at the cutoff of being eligible to attend a first-tier institution. Information about these issues is particularly important to students who score right below the cutoff, since they must then decide, for example, whether to repeat a full year of high school and take the college entrance exam again the next year in an attempt to score high enough to enter a more selective tier (a practice common in China). In response to the lack of knowledge about this issue, this paper measures the causal effects of attending a first-tier versus second-tier institution for students around the first-tier eligibility cutoff.

The academic literature posits a variety of reasons why attending more selective higher education institutions may affect student outcomes both during and after college. More selective institutions may provide better education through richer resources and improved instructional quality and thus increase an individual's human capital (Becker, 1964). Attending these types of institutions may further signal employers that a graduate has relatively higher abilities (Spence, 1973). Furthermore, selective institutions may offer students the opportunity to network with more able peers or faculty, providing advantages both in school and later in life (Nechyba, 2006; Ishida et al., 1997). Finally, students may participate to a different degree in various social activities or in civil society by attending more selective institutions (Feldman and Newcomb, 1969; Pascarella et al., 1988; Astin and Antonio, 2004).

Past empirical studies have mostly focused on the influence of university selectivity on wages and mostly in developed countries like the United States, Europe, and Japan (see Brand and Halaby (2006) for a review). Some of these studies have found significant effects of university selectivity on future wages (e.g. Brewer et al., 1999; Behrman et al., 1996; Ishida et al., 1997), while others have not (e.g. Black and Smith, 2004; Dale and Krueger, 2002). These divergent findings may in part be due to problems in dealing with selection bias when estimating selectivity effects (Black and Smith, 2004; Brand and Halaby, 2006). Studies have used different methods to try to correct for selection bias including covariate adjustments (e.g. Bowen and Bok, 1998), propensity-score matching (e.g. Brand and Halaby, 2006; Black and Smith, 2004), or selection corrections (e.g. Brewer et al., 1999; Loury and Garman, 1995). These methods, however, have relied on relatively strong assumptions like the ignorability of selection processes (e.g. propensity score matching, covariate adjustments) or strong modeling assumptions (e.g. selection corrections).

This paper builds on the previous empirical literature by utilizing different quasi-experimental methods to estimate the causal effects of attending first tier versus second tier higher education institutions in China on various student outcomes. One of these methods includes the regression discontinuity design (Thistlewaite and Campbell, 1960), which under ideal conditions provides causal estimates that have a high degree of internal validity (Shadish and Cook, 2009). Under ideal conditions—including the existence of strict tier eligibility cutoffs and access to data on all students who took the college entrance exam in a certain province and their future outcomes—the regression discontinuity design can estimate the average treatment on the treated effect for students around the eligibility cutoff (Battistin and Rettore, 2007).

In this paper, we further attempt to deal with a missing data problem that would ordinarily prohibit the use of regression discontinuity in estimating tier selectivity effects. In our particular example, because we utilize a random sample of all students from one province who take the college entrance exam and attend universities within that same province, we lack data on certain groups of students who took the exam (for example, students who attended universities outside the province). The nature of this missing data problem potentially differs on either side of the tier eligibility cutoff and thus leads to likely bias in the regression discontinuity (RD) estimates. We account for this missing data problem by running RD on various subsamples chosen according to the different types of college application choices that students make. These various subsamples are likely much more similar in observed and unobserved characteristics on either side of the tier eligibility cutoff line. After running various sensitivity tests for RD on these subsamples, we also compare the substantive conclusions from these RD estimates with those from other methods including genetic matching (Sekhon, 2006), and RD that controls for covariates.

In the end, the findings of the various quasi-experimental methods come to the same substantive conclusions. Specifically, this paper finds that just getting into and attending a first tier institution versus a second tier institution makes little difference in terms of class ranking, net tuition (college tuition minus financial aid), expected wages after graduation or the likelihood of preparing to apply for graduate school just after graduating college. The results do show however, that just getting into and attending a first tier versus a second tier institution makes it less likely that students will get their preferred major choice.

The organization of the rest of the paper is as follows. Section 2 discusses important issues that have been raised in the literature regarding the consistent estimation of college selectivity effects. Section 3 explains the classification of Chinese 4-year universities and majors into various tiers in greater detail, and also how the supply of university and major spots is determined in each province. It also briefly discusses the college application and admissions process that determines students' differential entry into these various tiers in order to give the reader the necessary background to understand the use of the RD design in the Chinese higher education context. Section 4 specifically discusses the application of the RD design and genetic matching to identify treatment effects on various student outcomes. Section 5 describes the data which was collected by the authors in November 2008—a simple random sample of approximately 4500 science-track students from Shaanxi province who were admitted into a university in Shaanxi in 2005. Section 6 presents the results from the main analyses and sensitivity analyses. Section 7 discusses the results and concludes.

2. Problems associated with measuring unbiased university selectivity effects

2.1. Defining and measuring university selectivity

Researchers have defined university selectivity or quality in different ways. Several studies measure "quality" through various characteristics such as an institution's mean student SAT scores (e.g. Dale and Krueger, 2002), student-to-faculty ratios (e.g. Behrman et al., 1996), or expenditures per student (e.g. Behrman et al., 1996). Studies that use the term "selectivity" often refer to college rankings either from the Barron's ratings (e.g. Brewer et al., 1999; Brand and Halaby, 2006), the Carnegie classification or possibly applicants-to-admissions ratios (Winston, 1999), occasionally dividing schools into private and public categories (Brand and Halaby, 2006). The distinction between university quality and university selectivity is not entirely clear in the literature, however, and we will use the term "selectivity" to refer to either selectivity or quality throughout the rest of this paper.

Beyond the basic definition, there is some discussion in the literature about how to best measure university selectivity (Brand and Halaby, 2006). Measuring university selectivity through different institutional characteristics reflects different channels through which the overall selectivity effect may operate (Zhang, 2005a). For example, high mean SAT scores may indicate positive peer effects while lower student-to-faculty ratios indicate better instructional quality. In fact, separate measures of university selectivity may result in a variety of conclusions about the significance of university selectivity effects, since each measure only reflects a single aspect of the selectivity of the institution (Zhang, 2005a). In light of this concern, Black and Smith (2004) create an index using the first principal component of average SAT scores of the entering class, average faculty salaries, and the average freshman retention rate. More recently, Black and Smith (2006) and Black et al. (2005) compare multiple proxies for college selectivity and address the possible presence of classical and non-classical measurement errors in these proxies; they find that by failing to address measurement error, previous studies likely underestimate the effect of university selectivity on earnings.

Therefore, having an easily-interpretable, error-free measure of university selectivity can be an important factor in identifying effects. As this paper measures the effect of going to a first-tier versus second-tier institution, we largely avoid having to deal with this issue because the classification of Chinese higher education institutions into first, second, third, and fourth tiers is ultimately determined by the government and is intimately related to the way in which students make college choices. More details about this classification are given in Section 3 below.

2.2. Dealing with selection bias

A second set of methodological concerns revolves around measuring the causal effect of university selectivity on student outcomes while avoiding selection bias (Black and Smith, 2004; Brand and Halaby, 2006). Behrman et al. (1996) describe in more detail how college choice is affected by individual, family, and school inputs at various stages in a student's life before college, and that this has implications for estimating the effects of college selectivity on students' future wages. It can however be quite difficult to obtain measures of all relevant pre-college home and school inputs that go in to educating students, especially measures that are error-free (Behrman et al., 1996).

Past studies use various methods to deal with selection bias. Several authors control for a large number of observable covariates (see Brand and Halaby, 2006). Dale and Krueger (2002) instead essentially match students accepted and rejected from the same sets of universities. Brand and Halaby (2006) and Black and Smith (2004) later utilize propensity score matching to try to identify causal effects. Besides assuming that students are selected into treatment and control groups on observable characteristics alone, both covariate adjustments and matching practically depend on finding the right balance in covariates between treatment and control groups (so that pre-treatment differences between both groups are removed), as well as in controlling for those covariates that are closely related to the actual selection mechanism as well as the outcomes of interest (Steiner et al., 2010).[1]

This paper utilizes quasi-experimental methods which, for the most part, rely on somewhat different assumptions to handle selection bias and identify college selectivity effects (see Section 4 for details). The next section first provides a more detailed background into China's university system in order to understand the application of these various identification strategies.

3. The classification of Chinese universities into tiers and the college application and admissions process

China's post-secondary education system expanded remarkably during the last 10 years, with gross enrollments in 4-year universities alone increasing from 2.7 million students in 1999 to approximately 10 million students by 2007 (National Bureau of Statistics of China, 2008). The expansion in enrollments has been further accompanied by an increase in the number of post-secondary institutions of varying types and quality including 4-year public universities, 4-year private universities,

[1] Other studies use selection corrections (e.g. Loury and Garman, 1995; Brewer et al., 1999), which often have relied on strong distributional assumptions as well as exclusion restrictions. Behrman et al. (1996), on the other hand, rely on twin data and a rich set of observable characteristics, and use a variance components model to assess the impact of college selectivity on female students' wages.

and 3-year vocational colleges. This section describes how universities and majors are classified into different tiers and explains the application and admissions process in some detail.[2] It thus provides the necessary context for understanding the identification strategies presented in Section 4.

3.1. The classification of 4-year universities into tiers

After consultations between central and local government agencies as well as with the universities themselves, most provinces classify all of the universities and majors across the country that its students are eligible to apply to into first, second, third, and fourth tiers. This is done separately for two types of tracks: a science and engineering track and a humanities track.

While the classification of universities and majors may thus differ somewhat across provinces and by track, in general, the first-tier consists of public higher education institutions that are under the control of central government ministries such as the Ministry of Education as well as a few of the most selective institutions under the jurisdiction of the provincial government. Second-tier institutions, on the other hand, are for the most part less-selective public institutions under the jurisdiction of provincial government agencies such as the provincial education bureau. Third-tier universities include private universities and "independent schools". The latter are lower-quality private subsidiaries of first or second-tier universities that charge significantly more tuition. Fourth-tier universities are 3-year vocational colleges, both public and private.

It is important to understand that higher education institutions are classified into tiers in order to facilitate the process by which students apply to and are selected into universities and majors. Before the application process begins, government agencies and universities work together to set quotas for how many students from each province will be allowed into each major within each university. These numbers are then aggregated to set by-province quotas for how many students will be allowed in each of the tiers, universities, and majors. High school graduates in each province then take annual province-wide college entrance examinations and subsequently fill out college and major choices on official forms provided by their provincial government. Based on the supply of spots available for its students in each tier and the distribution of student scores from the provincial college entrance examination, each provincial government then sets strict cutoff scores between the four tiers. Students are finally sorted into a single university and a single major (and of course, in a single tier) according to a complex system of rules that takes into account students' college entrance exam scores and the college and major choices they filled out.

The above description highlights several reasons why higher tier institutions are regarded as more selective than lower tier institutions. First, each provincial government authoritatively assigns universities and majors into one of the four selective tiers. Second, higher college entrance exam scores qualify students for higher tiers. Third, after filling out separate college and major choices within each tier, sorting rules give priority to the highest tier choices that each student qualifies for over his or her lower tier choices. Furthermore, financial constraints generally do not inhibit students from choosing higher tier institutions over lower ones, as they have to pay roughly the same or less to go to higher tier institutions than lower ones.[3] Therefore the classification of universities and majors into tiers is institutionalized and widely-accepted and understood by the general population in a province.[4]

3.2. The college entrance exam and college-major choices

In this sub-section, we further describe the college entrance exam, the way in which students fill out college and major choices, as well as the chronological order of the college application and admissions process.

After graduating from academic high school, most students take one of two distinct types of college entrance exams: the science track or the humanities track exam. Each type of exam is associated with a separate (but procedurally-equivalent) college application and admissions process. The science track exam (which is of relevance to this study) includes subject

[2] This description is based on policies from Shaanxi province in 2005—exact policies may differ slightly across provinces and across years (see Loyalka et al. (2009a) for a detailed description).

[3] An exception is that third-tier universities are much more expensive than fourth-tier colleges, but this study does not look at the causal differences between attending these tiers.

[4] One issue not emphasized in the above discussion is that sometimes more selective majors in a few provincial-level universities may be ranked in the first tier while other less selective majors in those universities are ranked in the second tier. Thus comparing students across the first tier eligibility cutoff line in the RD design (for example, see Section 4) in part involves comparing students from the same university who are in first-tier and second-tier departments respectively. Nonetheless, estimating this "tier selectivity effect" is still quite similar to estimating the college selectivity effect that is measured in other countries. This is because those majors from provincial-level universities that are counted as "first-tier institutions" generally are of higher quality and have greater resources than other departments in the university that are counted as "second-tier". Furthermore, because students are eventually admitted into only one university and major according to their college entrance exam performance and college and major choices (through the admissions process described in Section 3.2), average peer ability (as measured by exam scores) of first-tier majors is higher than that of second-tier majors (again by the design of the admissions system). Finally, in China, students are placed in one class of students in their major for the full 4 years of their study—students take almost every class together and are trained using the financial and teaching resources of the major department into which they were admitted. Because first-tier and second-tier majors also seem to differ in teaching quality, resources and peer effects (i.e. in the same way that first-tier and second-tier universities differ), students in the first and second-tiers may differ systematically in their schooling and post-schooling outcomes.

tests in math, Chinese language, foreign language, and a science comprehensive portion. Altogether, the exam is worth 750 points, with 300 points for the comprehensive portion and 150 for each other subject test.[5] Most provinces also have policies that grant 10 or 20 bonus points to students with particular background characteristics or notable achievements. For example, in Shaanxi province, students who are non-Han minorities may receive 10 bonus points, while students who have won math or science awards at the provincial, national, or international levels may also be eligible to receive extra points.[6] Finally, it is important to note that the rules and procedures surrounding bonus points are simply formulated, highly transparent and widely publicized. It is therefore highly unlikely that students or administrators would be able to precisely manipulate final entrance exam scores through the bonus point system.

After students take the province-wide college entrance exam in early June, each provincial government grades the exams and publishes a list of tier eligibility score cutoffs based on the overall score distribution and student quotas within each tier.[7] These tier eligibility cutoffs apply to all students who took the exam in that particular province. Although students do not yet know how they scored, and thus do not yet know the selectivity tier they may be eligible for, they are still required to fill in their college and major choices for the first and second tiers.[8] Shortly after this, they are sent their official scores and are then given the chance to fill in their college and major choices for the third and fourth tiers.

When making their university and major choices, students can pick several universities (with accompanying majors) in each tier. The first choice of university in each tier is of critical importance: if students qualify for a given tier but do not qualify for their first choice institution within that tier, universities that are listed as later choices may subtract a heavy penalty from students' scores during the selection process. There is also a fallback option within each tier wherein a student can choose to be admitted into any university and major in that respective tier that has available space if the colleges and majors they selected are already filled by higher-scoring students. It is thus important for students to make their college choices within each tier wisely, as "unwise" choices can prevent them from being admitted not only into their preferred university and major, but into any university within that tier at all.[9]

4. Regression discontinuity and genetic matching

In this section, we proceed to describe the regression discontinuity and genetic matching methods and discuss how these methods might be applied to identify the causal treatment effects of attending a first-tier versus second-tier institution.

4.1. The regression discontinuity design—background

The regression discontinuity design (RDD) is a quasi-experimental design in which the likelihood of receiving a certain treatment changes abruptly at a cutoff of a continuous function of one or more variables (Hahn et al., 2001). Under the appropriate conditions, the RDD can mimic a highly localized randomized experiment around the discontinuous value of the function. At the same time, these conditions mainly require that other unobservable factors continue to vary continuously around the cutoff value of the underlying functional rule (Jacob and Lefgren, 2004b). This type of situation often occurs as a result of an exogenous rule imposed by an outside institution (Imbens and Lemieux, 2007). Thus as in a program evaluation, "treated" units just to the right of the discontinuous cutoff value can be compared to "control" units just to the left of the cutoff (Lee and Card, 2008). When constant treatment effects exist, it is also possible to obtain the treatment effect for a random member of the concerned population (the average treatment effect or ATE), while with heterogeneous treatment effects the ATE is calculated only for units with values close to the discontinuity point (Shadish et al., 2002). Under the RDD, one can use parametric or nonparametric estimation techniques to try to obtain treatment effects; local linear regressions on either side of the cutoff are often used in practice (Lee and Lemieux, 2010).

A number of recent papers have used the regression discontinuity method to analyze educational issues such as estimating the effects of financial aid on career aspirations or college enrollments (Thistlewaite and Campbell, 1960; Van der Klaauw, 2002; Kane, 2003), evaluating the impact of remedial, childhood education or teaching training programs (Ludwig and Miller, 2007; Jacob and Lefgren, 2004a,b), finding the spillover effects of female education on infant health (McCrary and Royer, 2006), calculating the benefits of delayed primary school attendance (McEwan and Shapiro, 2007), or determining the effects of class size on student academic outcomes (Angrist and Lavy, 1999). Papers that offer an in-depth discussion of the practical and theoretical considerations underlying the RDD include Van der Klaauw (2002), Hahn et al. (2001), Imbens and Lemieux (2007), and Lee and Lemieux (2010).

[5] The entrance exam score is based on multiple choice, fill-in-the-blank, short-answer and essay questions. Some of the hand-written questions can earn partial credit according to the discretion of the grader. The nature and grading of the test, in addition to variation in the individual student's potential performance, increase the variability of a student's potential exam score. This all serves to increase the domain of interest around the tier cutoff score.

[6] Total bonus points are capped at 20 points, even if a student qualifies in multiple ways.

[7] The establishment of the cutoff values is therefore not based on any one individual's scores; thus the position of particular individuals' final entrance exam scores (plus bonuses) in relation to the cutoffs are not manipulated by administrators (or indeed any stakeholder). This inability of stakeholders to precisely manipulate the entrance exam score is important to the validity of our RDD results below.

[8] During this time, students can guess their total scores (and how their personal score fits into the distribution of scores of all students) based on answers published just after the exam.

[9] Similarly, a students' first choice of major is most important as their later choices (for each university) can be penalized. Students can also choose to accept any major if they fail to receive one of their listed major choices.

The basis of methods for causal inference such as the regression discontinuity design (as well as matching) is the Fisher–Neyman–Rubin model of causal inference. It defines the effect of a particular treatment 1 as the difference in the value of an individual's potential outcomes if she/he had received the treatment (Y_i^1) relative to the value of the individual's outcome if she/he had received another (mutually exclusive) treatment 0 (which is oftentimes regarded as a "control" condition Y_i^0). The effect of the treatment relative to the control condition for this individual is defined as $\beta_i = Y_i^1 - Y_i^0$. Because the individual could only have been exposed to only one condition (say treatment or control) at a given moment in time, in reality we only observe one of two *potential* outcomes, corresponding to either the treatment or control condition respectively. Thus on the whole, methods for causal inference try to address the "missing data problem" (the problem that we do not observe one of the two potential outcomes, known as the "counterfactual" outcome) by inferring the unobservable outcome from observable data. Unable to observe the treatment effect for a given individual, these methods generally attempt to estimate averages of $Y_i^1 - Y_i^0$ for certain populations.

In the RDD, treatment status (say $D = 1$ is treatment, $D = 0$ is control) is assigned on the basis of the value of an assignment variable (X). In the most straightforward case of RDD (the sharp RDD, discussed immediately below), we generally observe the potential outcome under the treatment condition for all individuals with a value of X greater than a specified cutoff c, as well as the potential outcome under the control condition for all individuals with a value of X less than c. We then model the relationship between the average outcome under the treatment condition and X on the right side of the cutoff ($E[Y_i^1|X]$) as well as the relationship between the average outcome under the control condition and X on the left side of the cutoff ($E[Y_i^0|X]$) (Lee and Lemieux, 2010). If the underlying functions $E[Y_i^1|X]$ and $E[Y_i^0|X]$ are continuous across c along X, then estimating $E[Y_i^1|X] - E[Y_i^0|X]$ as both functions approach c (in the limit) would estimate an average treatment effect at c: $E[Y_i^1 - Y_i^0|X = c]$ (Lee and Lemieux, 2010).

The RDD has two general forms: sharp and fuzzy. In sharp RDDs, the probability of receiving the treatment changes from zero to one at the cutoff value, whereas in fuzzy RDDs, the probability of receiving the treatment still jumps discontinuously, but by less than one, at the cutoff point (Hahn et al., 2001). The probability may jump by less than one at the cutoff value in the latter case if individuals do not comply with the treatment (or control) condition to which they are assigned. In the case of fuzzy RDD, we estimate the local average treatment effect (LATE) for those who comply with the treatment assignment (and who have values near the cutoff score) as:

$$E[\beta_i|X_i = c] = \frac{(E[Y_i^1 - Y_i^0|X = c)}{(E[D_i^1 - D_i^0|X = c)}$$

The fuzzy RDD usually only calculates the LATE for an unidentifiable group of compliers (Hahn et al., 2001). However, the fuzzy RDD in combination with strict eligibility cutoffs (in the sense that no one who scores below the cutoff can get the treatment) allows for the calculation of the ATE for treated participants around the localized cutoff value under weak regularity conditions (Battistin and Rettore, 2007). This holds regardless of how eligible individuals self-select into (or out of) the treatment (see Battistin and Rettore, 2007).

4.2. Advantages and disadvantages of using the regression discontinuity design

Using the RDD to estimate the effects of attending a first-tier versus second-tier higher education institution in the Chinese context has several advantages over the methods used in the previous university selectivity effects literature. First, the RDD can potentially estimate treatment effects for students at the tier eligibility cutoff with a high degree of internal validity (Shadish et al., 2002). As Lee and Lemieux (2010, p. 289) state "the RD design is a much closer cousin of randomized experiments than other competing methods" such as matching or instrumental variables, in terms of identifying causal effects. Furthermore, studies that use covariate adjustments or propensity score matching often face the question of whether or not to adjust for total years of schooling in their outcome equations (Black et al., 2005) or may have to account for the hierarchical nature of the relationship between student-level and university-level variables (Thomas, 2003; Zhang, 2005b). By contrast, using the RDD to estimate tier selectivity effects in China's higher education system generally avoids both of these types of problems.

Of course, the proposed RDD has its own limitations in estimating college selectivity effects. First of all, the design is restricted to measuring the difference in effects between the "least" selective institutions in the higher (first) tier with the "most" selective institutions in the lower (second) tier. This is because the distribution of students' entrance exam scores also differs between the institutions within each tier—that is, those first-tier institutions that tend to admit more students from just to the right of the first-tier eligibility cutoff line have lower average scores than other institutions in the same tier and likewise second-tier institutions that admit more students from just to the left of the first-tier eligibility cutoff line may well have higher average scores than other institutions in the same tier. In this study, universities that admit students just above and below the first-tier eligibility cutoff-line mostly consistent of provincial-level universities which receive roughly equivalent per-student appropriations (as compared to even more elite 985 or 211 universities which receive special 985

Table 1
Student traits within and outside the neighborhood of the 1st tier cutoff.

	>5 points below the cutoff	−5 to +5 from cutoff	>5 points above the cutoff
Age (in years)	22.62 (.96)	22.46 (.94)	22.20 (.89)
% From urban areas	.27 (.44)	.30 (.46)	.39 (.49)
% Female	.44 (.50)	.35 (.48)	.38 (.48)
% Repeat exam takers	.41 (.49)	.35 (.48)	.29 (.45)
No. of siblings	1.41 (.94)	1.29 (.91)	1.16 (.96)
% Went to key high school	.51 (.50)	.65 (.48)	.70 (.46)
No. of students	2226	408	1831

Notes: The above statistics are means following by standard deviations in parentheses.

and 211 project funds from the central government). However, this type of limitation of comparing institutions that only differ somewhat in terms of selectivity seems to exist in previous studies as well.[10]

Another concern emerges from the well-known limitation of RDDs under heterogeneous treatment effects—that is, while estimates are highly valid around the cutoff region, they have a low degree of external validity. Table 1 compares descriptive statistics on students in a close neighborhood of the cutoff for whom we will estimate treatment effects with students with scores above and below the maximum and minimum of that neighborhood respectively. As we can see from Table 1, the size of various measures usually correlated with socioeconomic status such as likelihood of being from an urban area are higher for students around the first tier cutoff than lower-scoring students and lower for students around the first tier cutoff than higher-scoring students. Students around the cutoff also fall in the middle on education-related measures such as the likelihood of retaking the college entrance exam. However as mentioned in the introduction, this study *is* actually concerned about the effect of attending a first-tier versus second-tier university for those individuals who score at the tier eligibility cutoff. This is because each year, tens of thousands of Chinese students find themselves just on either side of the tier cutoff and must make difficult decisions about whether or not to attend the university and major they eventually are admitted to.

It is not obvious a priori whether it is better for a student around the score eligibility cutoff to attend a first-tier versus second-tier institution. For example, students who do just get admitted into a first-tier university may rank lower than their peers or may not be able to study the first choice major they listed on their college and major choice form (see Section 3.2). On the other hand, these same students may benefit from positive peer effects and more financial aid (see Loyalka et al., 2009b). The RDD combined with the necessary data can thus answer an important question for these types of college students in China.

4.3. Application of the fuzzy RDD to measure college selectivity effects

This sub-section discusses the application of the fuzzy RDD to identify the LATE of going to a first-tier versus a second-tier higher education institution on a student's at-school and post-school outcomes. We first begin with a simple linear specification to illustrate the general use of the regression discontinuity design in the context of the current study[11]:

$$(7)\ Y = \beta_0 + \beta_1 D + \beta_2 S + X'\alpha + \varepsilon$$

where the dependent variable Y is an individual student outcome such as wages after graduation, D is the treatment indicator equal to 1 if the student is *admitted* into a first-tier institution and 0 otherwise, S is the college entrance exam score, X is a vector of student (pre-treatment) background variables whose inclusion can ideally improve the efficiency of the estimator, and ε is the error term which includes other unobservable variables that affect the outcome.β_0, β_1, β_0, and α are coefficients with β_1 being the treatment effect of being *admitted* into the first-tier on the student outcome. The fact that D is exogenous around the cutoff (i.e. it is based on a rule imposed by the provincial government) enables the identification of the treatment effect. It is also important to note that the treatment indicator D represents whether or not a student is *admitted* into the first-tier compared to whether or not a student actually *attended* a first-tier institution. This is because not all students that are admitted into higher education institutions in the first and second tiers decide to attend those institutions.

The fuzziness of the RDD in estimating university tier treatment effects comes mainly from the fact that a significant proportion of students who have scores above the first tier cutoff line are admitted into the second tier (as a result of the interaction of their college entrance exam score with their university and major choices). Furthermore, we found a small minority of students (24) who scored below the first tier cutoff line and yet was still admitted into the first tier. The actual RD esti-

[10] It is also important to note that the RDD measures the effect of going to a first-tier versus a second-tier higher education institution including both university *and* major and thus subsuming the effects of major assignment as part of the overall effect. Since one of the main goals of this paper is to discern whether it matters or not if a student who scores nears the tier eligibility cutoff attends a first-tier versus second-tier institution, the study *is* interested in examining the combination of both university and major effects (both aspects of tier-effects) on schooling or post-schooling outcomes such as wages.

[11] This specification is only for expository purposes. Unbiased estimates would be found if the potential outcome regression lines were both linear and had the same slope (strong assumptions). We use flexible local linear regression methods in our analyses instead (see Section 6).

mates (see Section 6.1) use local linear regression (with a triangular kernel) for the outcome and treatment equations on either side of the cutoff value (Nichols, 2007). We also use the automatic bandwidth selection procedure suggested by Imbens and Kalyanaraman (2009) for the continuous outcomes.[12] Bootstrapping is used for the standard errors (Nichols, 2007).[13]

A potential threat to the validity of the fuzzy regression discontinuity estimates comes from the fact that the dataset used in this paper lacks information on the entrance exam scores and outcomes of three potentially important subgroups of students: (1) students who were admitted to first and second tier institutions outside the province, (2) students who were admitted into second-tier military universities in the province, (3) students who were admitted into first or second tier universities but decided not to go. We attempt to overcome this missing data problem using various analytical strategies described in Section 6.1.

4.4. Matching—background and general application

As mentioned in Section 4.3, the dataset used in this paper lacks information on certain types of students and this threatens the validity of RDD estimates. In this subsection, we therefore introduce matching as a separate method to help estimate average treatment on the treated effects (ATT).[14] Matching is a data preprocessing method which takes observational data and tries to create samples of treatment and control observations that are as similar as possible in the distribution of pre-treatment characteristics (Z) (Ho et al., 2007b). Under the assumption that potential outcomes are independent of treatment assignment after conditioning on Z (known as the "ignorability" or "conditional independence" assumption), matching on Z (when done properly, in a sense which will be made clearer below) removes confounding factors that prevent an unbiased estimate of the treatment effect (Rosenbaum and Rubin, 1983).

However, it may not be possible to exactly match treated and control observations on a multi-dimensional covariate vector Z due to the "curse of dimensionality". In other words, there may not be enough observations available in the control group that exactly match the treatment group on the values of the covariates, especially if there are too many covariates (Rosenbaum and Rubin, 1983). Rosenbaum and Rubin (1983) instead show that if the ignorability assumption holds when conditioning on Z, then it also holds when conditioning on the "propensity score" which is the conditional probability of receiving the treatment given Z. Under the ignorability assumption and the additional assumption that the propensity scores in treatment and control groups overlap (these two assumptions together comprise the "strong ignorability" assumption), matching procedures that use correctly-specified propensity scores will minimize conditional bias in treatment effect estimates (Rosenbaum and Rubin, 1983).[15]

It is important to note that ignorability is a fairly strong and untestable assumption. In actuality, there may well be systematic differences between treatment and control groups even after matching on the available observable covariates (say a subset of Z). In particular, the failure to match on unobserved covariates, which are correlated with the treatment assignment and outcome variables, generally leads to biased estimates. Furthermore, failing to correctly model the propensity score can increase bias even if the ignorability assumption holds (Drake, 1993).

When the ignorability assumption is thought to hold, however, researchers can choose among a variety of matching methods to try and achieve "balance" between treatment and control samples (Ho et al., 2007b). Balance is about using the available pre-treatment covariates to obtain identical treatment and control groups on average; in practice this generally involves minimizing the difference in measures of the empirical distribution of pre-treatment covariates between the treatment and control groups. Unfortunately, there is no general consensus in the literature on exactly how matching should be done or how to measure whether a particular matching procedure is successful (Diamond and Sekhon, 2010).

In regards to propensity score matching specifically, Rosenbaum and Rubin (1984) discuss the importance of finding an estimated propensity score that achieves balance in the covariates.[16] They suggest an approach for how to estimate an appropriate propensity score. This involves iteratively estimating a propensity score and checking/improving balance until balance is maximized according to a particular loss function. In practice, however, few researchers commit to implementing this algorithm, as it is potentially cumbersome (Diamond and Sekhon, 2010).

In this paper, we utilize a recently developed matching method called genetic matching. This method uses a search algorithm to automate the iterative process of assessing and eventually maximizing overall covariate balance.[17] It is furthermore a generalization of multivariate matching and propensity score matching when the covariates to be matched on include an estimated propensity score (Sekhon, 2011). Genetic matching usually performs better than propensity score matching or multivariate matching alone in terms of reducing bias and mean-square error and under more general conditions (Sekhon, 2006).[18]

[12] We use bandwidths of varying size for each of the binary outcomes—the results are fairly consistent across different bandwidths.

[13] Bootstrapping was performed 500 times.

[14] The ATT is defined as: $E[Y_i^1 - Y_i^0|\mathrm{Ti} = 1] = E[Y_i|T_i = 1] - E(E[Y_i|T_i = 0, Z_i])$.

[15] The overlap condition is formally stated as $0 < \Pr(T|Z = z) < 1$ for all z.

[16] As Diamond and Sekhon (2010) state: "Since the propensity score is a balancing score, the estimate of the propensity score is consistent if and only if matching on this propensity score balances the observed covariates (p. 2)."

[17] Much of the discussion in this and the next paragraph is based on Diamond and Sekhon (2010).

[18] In particular, Genetic Matching reduces mean-squared error both when the "equal percent bias reduction" property holds and does not hold (see Sekhon (2006) for more details).

Table 2
Description of 2008 Shaanxi data: Shaanxi students admitted into Shaanxi first two tier HEIs.

	Science track		
	1st Tier	2nd Tier	Total
Answered questionnaire	1843	2622	4465
Declined to answer questionnaire	8	18	26
Not reachable by administrators	29	59	88
Not at the university[a]	45	229	274
Attended military university	0	413	413
Total	1925	3341	5266

[a] "Not at university" mostly refers to students who were admitted into the university (in Shaanxi) but never registered in the university. It also includes a very small number of students that dropped out or had other exceptional reasons for not being at the university.

Specifically, the search algorithm in genetic matching minimizes a generalized version of the Mahalanobis distance (MD) metric (MD itself measures the multivariate distance between subjects in different groups, for more details see Diamond and Sekhon, 2010). This generalized version of MD includes a diagonal matrix of weighting parameters W—one for each covariate to be matched on, including the estimated propensity score (however, adding the propensity score is optional – genetic matching does not in fact depend on knowing or estimating the propensity score). Each potential set of weighting parameters thus corresponds to a potential value of the generalized MD measure. Weights on each variable are then chosen iteratively by the algorithm so as to minimize the generalized MD metric. After weighting parameters are chosen in each iteration, the observed data is matched according to the generalized MD value and another specific matching procedure chosen by the researcher (for example, the researcher can choose between nearest-neighbor or optimal matching, can choose whether to match with or without replacement, etc.). The algorithm will produce many matched samples, evaluating the balance of covariates for each one according to the loss function, and ultimately identify the weights which correspond to minimum loss (and thus maximum balance) (Diamond and Sekhon, 2010).

As alluded to above, matching is a method that preprocesses the original dataset in the sense that it "prunes" observations from the original dataset to obtain treatment and control groups that are more similar in the distribution of pre-treatment covariates (Ho et al., 2007a). After the data is preprocessed in this way, a particular estimation procedure needs to be chosen to obtain the treatment effect estimate. In many studies, researchers simply difference the average outcomes across treatment and control groups. Several papers including Rubin (1973, 1979), and more recently Ho et al. (2007a) suggest using regression adjustments (e.g. linear regression) on the matched dataset to obtain results that are "doubly-robust". These estimation procedures are used in Section 6 below.

5. Data

The data used for the analysis is from a simple random sample of senior university students who took either the humanities or sciences college entrance exam in Shaanxi province in 2005 and were admitted into a first or second tier university in Shaanxi that same year.[19] Shaanxi province, located in the northwest China, is known for its relatively strong higher education system. Ten universities are regularly designated as first tier universities in the science-track, while another 15 public, 4-year institutions are regularly designated second tier universities. Zhou and Loyalka (2010) estimate that the per student instructional costs of Shaanxi's elite institutions (those for which they were able to get financial data) is approximately twice as much as that of the province's other 4-year public universities. Loyalka et al. (2009a) further show that there is extreme sorting by ability (as measured by college entrance exam scores) between university tiers in Shaanxi which has implications for tracking and peer effects.

Because of sample size requirements for the RD analyses, in this paper we focus only on the science track students (73% of students in the total sample). Table 2 shows that of the 5266 sampled students in the first two tiers, 413 students (7.8%) were admitted into military universities (in the second-tier only) and could thus not be surveyed. Out of the remaining 4853 sampled students, 4465 (92.0%) answered a short survey questionnaire, 274 (5.6%) never ended up attending the university that admitted them or dropped out before their senior year, 26 (0.5%) declined to take the survey, and 88 students (1.8%) were officially registered at the universities but could not be reached by college administrators.[20] At the same time, the response rate to all relevant survey questions was over 99%.

[19] Our survey was thus conducted at the end of 2008, at the end of the second semester of the students' fourth and final year and included students from both the science and humanities tracks. A population frame of local students who took the "pure" science-track or "pure" humanities track college entrance exam and were admitted into non-military universities in the first three tiers in 2005 was obtained from the provincial government. We did not survey every student at each university, as the costs for surveying such a large sample were prohibitive.

[20] Furthermore, according to 2006 enrollment data, approximately 36% of Shaanxi students who attended four-year universities went outside the province (Loyalka et al., 2009a).

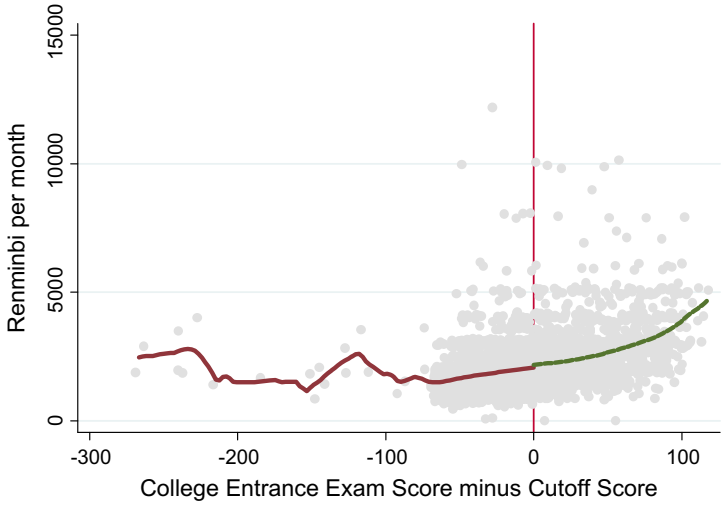

Fig. 1. Expected salary for first year after graduation.

The data contain information about college entrance exam scores which are a key component for the analyses in this paper. Administrative college entrance exam scores were obtained for approximately 85% of the students (both tracks combined). Self-reported scores were used whenever administrative scores were not available.[21]

Covariates that we later control for our in our various analyses in the next section below include age (in months), gender, an indicator for urban or rural residential status, college entrance exam score (plus bonuses), and dummy variables for whether or not the student took the exam in previous years and whether or not a student chose to go to a non-military Shaanxi university. We also sometimes control for student major choice (a categorical variable with 11 major categories, parallel to the definition used by China's Ministry of Education).

The causal analyses in the next section focus on five outcomes: (a) whether a student receives his or her first choice of major (yes or no), (b) whether a student receives any type of financial aid (yes or no), (c) whether a student ranks in the top-third of his or her class (yes or no), (d) expected monthly wages (in Chinese yuan) in the first year after graduation and (e) whether a student has plans to go to or prepare for graduate school immediately after graduation (yes or no). The outcomes were measured at the end of the first semester of students' senior year in college (about three and a half years after students learned about their assignment) and may illustrate some of the tradeoffs in attending first versus second tier universities. Arguably the most important outcome to examine is expected monthly salary—students who just get into the first tier versus the second tier on the whole enter a more selective institution and are surrounded by peers with better exam scores (a proxy for cognitive ability and thus may expect or actually earn higher wages upon graduation.[22] At the same time, these students may be assigned a major of lower selectivity (outcome "a" above) despite controlling for their major choice) which could potentially depress their expected and actual wages.

6. Results

The treatment effect estimates are based on three different types of analyses: RD (Section 6.1), matching (Section 6.2), and "RD plus covariates" (Section 6.3). It is important to note that although all of these different analyses estimate treatment effects for students near the first-tier eligibility cutoff, they estimate somewhat different types of treatment effects for somewhat different samples of students. Notably, the RD and associated "RD plus covariates" analyses estimate local average treatment effects (in the fuzzy cases, for an unidentifiable group of compliers) while the matching analyses estimate average treatment on the treated effects. The RD-related analyses also depend upon a different set of identifying assumptions (see Section 4.1) and a larger range of observations around the cutoff compared to the matching analyses. Nonetheless, as the

[21] We also compared the self-reported college entrance exam scores of students with administrative scores and found that approximately 75% of students reported correctly.

[22] We note the concern that students may answer "yes" to the question about plans to go to graduate school (outcome e) and that this may affect the question about expected salary (outcome d). To at least partially address this question, we asked students to write down their expected salary after graduation in the case that they did not go to graduate school. We also asked students about their minimum acceptable salary in the hypothetical case that they did have to work after graduation. Results from causal analyses that used expected monthly salary or minimum acceptable salary as the outcome variable were substantively the same.

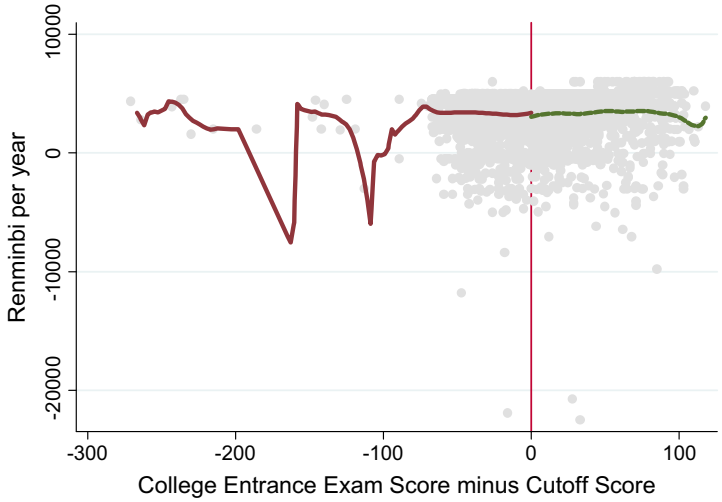

Fig. 2. Net tuition paid.

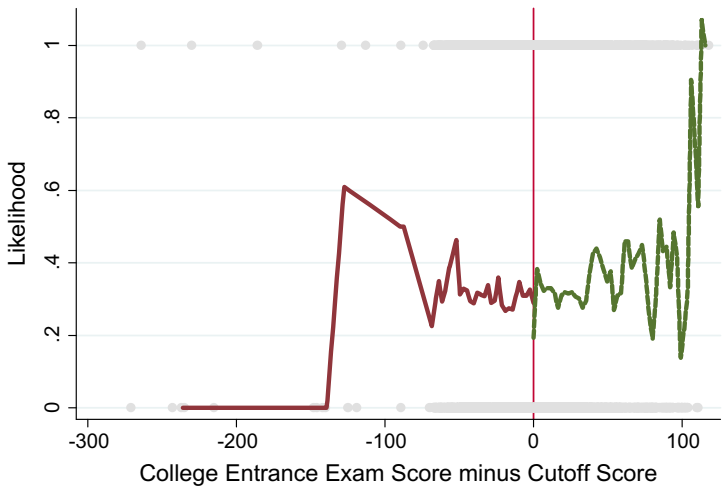

Fig. 3. Likelihood of ranking in top-third of class.

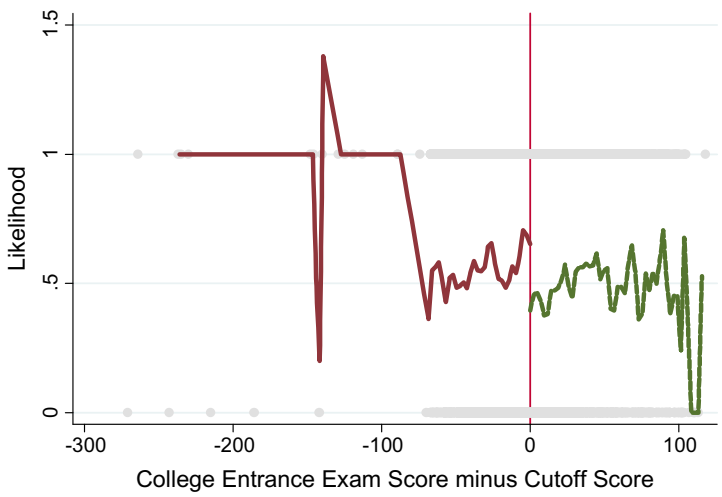

Fig. 4. Likelihood of getting first major choice.

results show below, the different analyses generally lead to the same substantive conclusions—that attending the first-tier results in a lower chance of getting one's first choice of major.

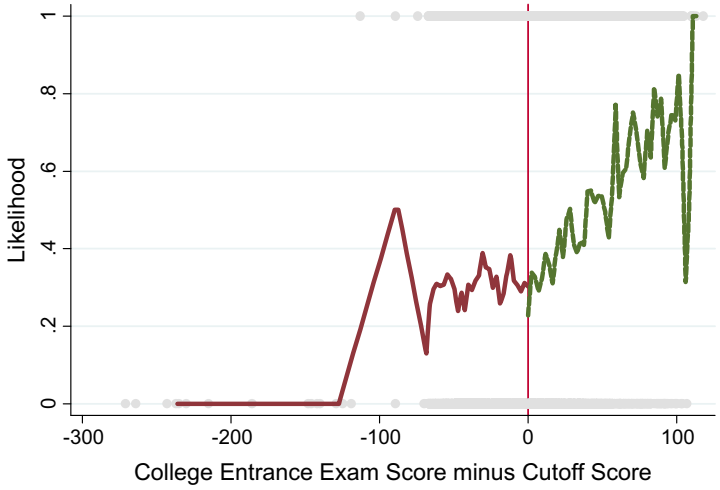

Fig. 5. Likelihood of going to graduate school.

6.1. Regression discontinuity

We first present standard RD plots (Figs. 1–5)—scatterplots of each of the five outcomes of interest against the assignment variable (college entrance exam score), overlaid with local linear regression lines estimated using the entire sample of data on each side of the designated cutoff. One can see from these figures that with the exception of receiving one's first major choice (Fig. 4), there is little evidence of discontinuities at the cutoff line for the various outcomes. Table 3 further presents fuzzy RD estimates for students in the science track.[23] The first two columns show treatment effect estimates using the entire sample of students (RD analysis 1). The only significant result is that attending a first-tier institution reduces the probability that a student will get his or her first major choice.

We try to account for the missing data on students that went to college outside the province and students that went to military universities by first narrowing the sample to students who listed non-military universities in Shaanxi as their first choices in the first and second tier.[24] By narrowing the sample in this way, we hope to compare students who are more similar on either side of the cutoff—in particular students who had a much smaller probability of leaving the sample. The only significant treatment effect estimate at the 10% level in this limited sample is whether or not a student gets his or her first major choice (Table 3, RD analysis 2).[25] Going to the first-tier reduces the probability that a student will get his or her first major choice.

We also further reduce the sample of students to those who chose Shaanxi non-military universities as their first choices *and* whose scores were also within 30 points on either side of the first-tier eligibility cutoff. This is done because students with relatively high scores on the left of the cutoff would be more likely to get their first choice of university (which in this case is a Shaanxi non-military university). Furthermore, students with relatively low scores on the right of the cutoff may arguably have been less likely to select universities outside the province for their latter choices in each tier (i.e. after their first choice). The treatment effect estimates for these students again show that going to the first-tier reduces the probability that a student will get his or her first major choice (Table 3, RD analysis 3).

Finally, we reduce the sample from RD analysis 3 further to students that had taken the college entrance exam in years previous to 2005. We reason that these students would be more likely to attend university in 2005 after being admitted into university (rather than repeating the college entrance exam and applying for university again the next year). None of the results from this analysis are significant (Table 3, RD analysis 4).

Sensitivity tests were run after each of the above analyses to test the validity of the RD analyses. In particular, we tested for discontinuities in other pre-treatment covariates along the discontinuity in the running variable (see Imbens and Lemieux, 2007). These covariates include age and number of siblings, and dummy indicators for gender, rural versus urban residence, whether or not the student went to a key high school and whether the student had previously taken the college entrance exam. Discontinuities in these covariates would invalidate the assumptions of RDD (see Section 4). However, we were not able to reject the null hypotheses that the covariates were continuous around the cutoff for each of the four RD analyses (results not shown). In particular, for RD analysis 1, we could not reject the null hypotheses that (1) the proportion of students who selected Shaanxi non-military universities for their first choice was continuous around the first tier eligibil-

[23] Because less than 1% of the students in the sample are in the non-eligible treated group, complete data would allow a fairly close estimate of the ATT (see Section 4.3).

[24] The data set unfortunately only has information on the first choices of students in each tier. Later choices within each tier may also determine whether or not students stay in the sample.

[25] Moderately smaller bandwidths result in significant estimates at the 5% level.

Table 3
Regression discontinuity estimates.

Analysis Description	Science track students			
	RD analysis 1 RD using all the science students data	RD analysis 2 RD with data limited to those who chose Shaanxi non-military HEIs as first choice	RD analysis 3 Same as RD2 but also limited to + and −30 around the cutoff	RD analysis 4 Same as RD3 but also excluding those who took exam before
Expected salary (monthly)	66.59 (329.53)	189.09 (288.49)	189.08 (281.31)	−123.03 (549.06)
Top third of class (yes/no)	−0.18 (0.2)	−0.22 (0.27)	−0.23 (0.27)	−0.04 (0.56)
Net tuition	−698 (548.48)	−562.37 (786.39)	−562.37 (781.05)	−1535.65 (2209.1)
Received 1st choice major (yes/no)	**−0.50 (0.24)	*−0.51 (0.31)	*−0.52 (0.3)	−0.01 (0.59)
Graduate school (yes/no)	0 (0.21)	−0.13 (0.28)	−0.12 (0.25)	0.17 (0.44)
McCrary density Test (log diff in height + *SD*)	**0.23 (0.06)	−0.03 (0.07)	0.21 (0.14)	*0.57 (0.31)
No. of observations	4465	3097	1612	555

Notes: Bootstrapped SEs (500 times).
* Significant at the 10% level.
** Significant at the 5% level.

ity cutoff; (2) the proportion of students who had taken the entrance exam previously was continuous around first tier eligibility cutoff.

Finally, we also tested for discontinuities in the density of the running variable using McCrary's DCdensity test (McCrary, 2008). For the entire sample of students (RD analysis 1), the density test rejects the null hypothesis that the density of the running variable (the college entrance exam score) is continuous around the cutoff at the 5% level.[26] Fig. 6 shows this graphically (by indicating the underlying density of the assignment variable is discontinuous at the cutoff); the darker line on either side of the cutoff represents the estimate of the empirical density while the thin lines around the darker line indicate the confidence intervals for the estimate of the empirical density. The density tests for the samples of students in RD analyses 2 and 3, however, do not reject the null hypothesis that the density of the exam score is continuous around the cutoff (see Fig. 7).[27] The density test also rejects the same null hypothesis for RD analysis 4 at the 10% level (see Fig. 8).

We note that we cannot reject the null hypothesis of a continuous density of the assignment variable across the cutoff when we limit the sample to those first-time college entrance takers who chose a non-military university in the province as their first choices (i.e. analysis 2). In regards to analysis 4, we note that the null hypothesis is rejected at the 10% level only. Furthermore, when we do not limit the range around the cutoff to observations within 30 points on either side, we cannot reject the null hypothesis of a continuous density of the assignment variable across the cutoff. Nonetheless, we use caution in interpreting the results from analysis 4.

It is important to note that the results of the sensitivity tests are not sufficient to validate the RD estimates from RD analyses 2 and 3. The tests (especially for RD analysis 4) may also suffer from inadequate power.[28] Students just on either side of the tier eligibility cutoff may still be different on observable or unobservable variables (such as students' college and major choices after the first choice in each tier), even after limiting the sample by the means used above. We therefore further conduct analyses using matching and matching plus RD below.

6.2. Matching

We perform both propensity score matching and genetic matching using the "Matching" package in R (Sekhon, 2011). We first limit the sample to treatment and control observations within 5 points on either side of the first-tier eligibility cutoff (again separately for each track).[29] We also retain all treatment observations in both types of matching in order to later estimate the ATT for students near the cutoff line.

[26] The discontinuity estimate for the density of the entrance exam score at the cutoff is .23 and the standard deviation is .06.
[27] The density test graph for RD analysis 3 is not shown—it's similar to RD analysis 2.
[28] This is especially true when expected wages (rather than actual wages) is used as an outcome variable. Rounding error (i.e. measurement error) may inflate the SEs substantially in this case.
[29] Because our RDD estimates are to the cutoff, we used 5 points on either side of the cutoff to obtain roughly comparable matching estimates. We also surmised that student scores may vary plus or minus 5 points on any given day given the nature of the college entrance exam (which has short answer and essay questions) and natural variation from one hypothetical test day to another. We have 408 observations within plus or minus 5 observations. Finally, we had also ran estimates on matched data from plus or minus 10 points on either side of the cutoff and found similar results.

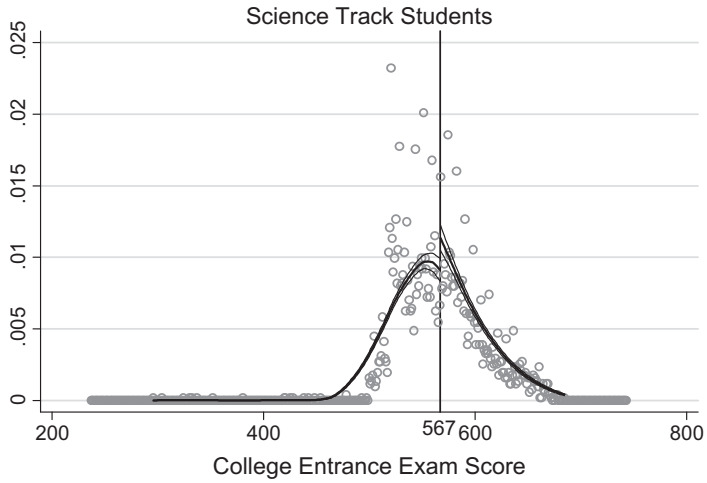

Fig. 6. Density test of the exam score. *Note*: McCrary (2008) density plot of the assignment variable (exam score). The null hypothesis, that the exam score is continuous at the cutoff, is rejected at the 5% level (log difference in height = .25, standard error = .06).

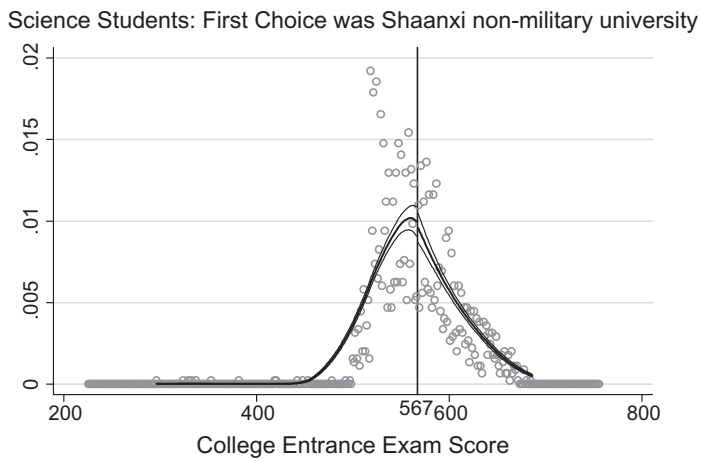

Fig. 7. Density test of the exam score. *Note*: McCrary (2008) density plot of the assignment variable (exam score). The null hypothesis, that the exam score is continuous at the cutoff, is not rejected even at the 10% level (log difference in height = .02, standard error = .07).

Before implementing the matching procedures, we fix the estimand of interest (in our case the ATT), the ratio of treatment to control observations (in our case, 1–2 matching with replacement[30]), and the loss criteria. In regards to the loss criteria, we follow Sekhon's (2006) default setting, where loss is the minimum *p*-value observed across a series of balance tests performed on distributions of matched baseline covariates.[31] The tests are *t*-tests for the standardized difference in means between as well as bootstrapped nonparametric Kolmogorov–Smirnov (KS) tests.[32] Although the *p*-values are not true probabilities, they are useful measures of balance (Sekhon, 2006). Importantly, Sekhon (2006) uses experimental and simulated datasets to compare the results of his default genetic matching procedure (which maximizes the smallest *p*-value of standardized balance metrics for observed covariates and higher order terms and interactions) with that of various types of matching including propensity score matching, other types of multivariate matching, and even his genetic matching procedure which tries to find balance by reducing standardized means. He shows that the genetic matching procedure using the default loss criteria dominates the other methods in terms of reducing bias and MSE when assumptions required for equal percent bias reduction (EPBR) hold and, even more so, when they do not. Furthermore, he shows that the genetic matching estimates come very close to retrieving benchmark (experimental or simulation dataset) estimates.

[30] Ties are handled deterministically and "Abadie-Imbens" standard errors are estimated (see Sekhon (2011) for more details).

[31] It is important to note that for genetic matching, Diamond and Sekhon (2010) state that their "algorithm uses *p*-values so that results from different tests can be compared on the same scale. [That is], as the sample size is fixed within the optimization, the general concern that *p*-values depend on sample size does not apply (p. 7)."

[32] The *p*-values of bootstrapped tests are consistent even when the covariate distributions are not fully continuous.

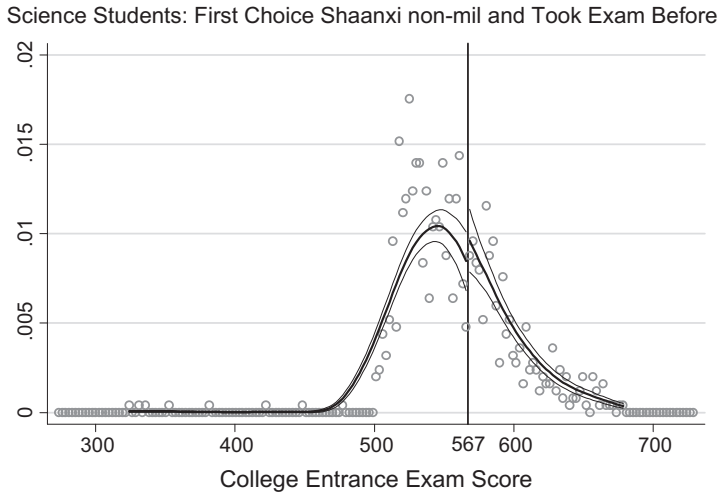

Fig. 8. Density test of the exam score. *Note:* McCrary (2008) density plot of the assignment variable (exam score). The null hypothesis, that the exam score is continuous at the cutoff, is rejected at the 10% level (log difference in height = .57, standard error = .31).

In regards to propensity score matching, we include a number of variables and interactions representing individual, household, and high school student characteristics in a logit equation to estimate the propensity score.[33] The variables that were chosen were similar in nature to the ones used in Black and Smith (2004) and Brand and Halaby (2006). We tried multiple specifications for the propensity score until finally we arrived at one that achieved good balance (in the propensity score matching). The final propensity score was constructed using the following variables: age, college entrance exam score ("CEE score"), female (Y/N), urban or rural residence ("urban"), whether the student was retaking the college entrance exam for a second time ("retook"), whether the student applied for a non-military school in the province as their first major choice ("choice_Shaanxi"), whether the student applied for an engineering major ("engineering"), age and score squared, as well as interactions between CEE score and all of the other (above-listed) variables. We obtained much better balance on most covariates and interactions than prior to matching, but fell somewhat short of obtaining adequate balance on CEE score. That is, the *p*-value for the balance metric for this covariate was 0.07, while Sekhon (2011) notes that the *p*-values should generally be above 10.

To try and improve balance, we then included both covariates and the propensity score in the more general genetic matching procedure. In particular, we matched on the propensity score, age, CEE score, retook, choice_Shaanxi, engineering. We again checked for balance in the distribution of all of the above covariates and all interactions between the continuous and discrete variables. This time we found that the minimum *p*-value was .14 (for age interacted with female) (see Table 4). We then ran (regression) estimates, controlling for age, CEE score, female, urban, retook, choice_Shaanxi, and engineering, on this matched dataset.

The matching (ATT) estimates are shown in Table 5 (columns 1 and 2). The propensity score and genetic matching estimates agree substantively with the general trend of the RD estimates—the only significant result is that going to a first-tier institution reduces the likelihood that a student will get his or her first major choice.

6.3. RD plus covariate adjustment

This subsection uses regression discontinuity with covariate adjustments to estimate tier selectivity effects. According to Imbens and Lemieux (2007) controlling for pre-treatment covariates (which according to the RD identification strategy should be continuous across the cutoff—something that we tested for in Section 6.1), does not fundamentally change the identification strategy of RD. Including the covariates may possibly have little effect on the RD estimates as the covariates are independent of treatment assignment conditional on the assignment variable being close to the cutoff. However, including the covariates may eliminate some bias because in practice our analysis does include values of the assignment variable (college entrance exam score) that are not close to the cutoff (Imbens and Lemieux, 2007; Frölich and Melly, 2008). Including the covariates may also increase the precision of the estimates somewhat (Imbens and Lemieux, 2007; Frölich and Melly, 2008).

Given these considerations, we run RD plus covariates for each of the five outcome variables of interest and also for each of the four subsamples mentioned in Section 6.1 (i.e. RD analyses 1–4). Our pre-treatment covariates include age, urban ver-

[33] As we noted in Section 4.4, unbiased estimates from propensity score matching depend on the assumption of ignorability as well as correctly specifying the propensity score. There is however some evidence from Steiner et al. (2010), that having information on covariates that are strongly related to the treatment assignment and outcome of interest may result in matching estimates being quite close to experimental estimates. In our situation, we find that we have covariates that are strongly correlated with the treatment assignment as well as to each of the outcome variables. By at least controlling for students' basic demographic characteristics, socioeconomic/educational background, schooling choices, entrance exam performance (the key item in the college applications and admissions process), and whether students had taken the entrance exam previously, we covered a range of pertinent covariates. We posit thus that the bias of our estimates has been reduced substantially after matching.

Table 4
Balance in covariates before and after genetic matching.

	Before matching	After matching
T-test p-values		
Female (Y/N)	0.08	0.38
Urban (or rural)	0.13	0.80
Retook the CEE (Y/N)	0.79	0.71
Chose Shaanxi (Y/N)	0.07	0.69
Chose engineering	0.12	0.71
KS bootstrap p-values		
CEE score	<2.22e−16	0.20
CEE score ∗ female	<2.22e−16	0.19
CEE score ∗ urban	0.01	0.71
CEE score ∗ retook	<2.22e−16	0.50
CEE score ∗ chose_Shaanxi	<2.22e−16	0.14
CEE score ∗ engineering	<2.22e−16	0.54
Age	0.63	0.23
Age ∗ female	0.09	0.16
Age ∗ urban	0.29	0.24
Age ∗ retook	0.71	0.38
Age ∗ chose_Shaanxi	0.31	0.48
Age ∗ chose_engineering	0.41	0.76

Table 5
Propensity score matching and genetic matching estimates.

Method	Science students	
	PS matching	Genetic matching
Expected salary (monthly)	−206.60 (219.55)	160.82 (163.20)
Top third of class (yes/no)	.04 (.09)	.01 (.07)
Net tuition	195.25 (303.13)	115.58 (242.08)
First major choice (Y/N)	−.29** (.08)	−.25** (.07)
Graduate school (Y/N)	.02 (.08)	.08 (.06)
Original # of observations		408
Original # of treated observations		133
Matched # of observations		133
Matched # of observations (unweighted)		267

** Significant at the 5% level.

sus rural, gender, "retook", "chose Shaanxi", and whether the student chose engineering as a first major choice. From Table 6, we can see that there are no statistically significant effects of attending the first tier on expected income, net tuition, class rank, or plans to go to graduate school. In contrast to the other outcomes, we do find statistically significant and negative effects on receiving one's first college major choice for the first three subsamples when a bandwidth somewhat larger than the one estimated by cross-validation is used. Overall, our findings generally correspond with the previous findings from both RD and matching separately that attending the first tier increases the likelihood that a student must forego his or her first major choice.

We also use RD plus covariates to try and decompose "between direct and indirect effects of the treatment effect (see Frölich, 2007)."[34] Specifically, we reran the RD analysis on five of our outcomes (expected monthly income in the first year after graduation, minimum acceptable income in the first year after graduation, net tuition fee, class rank, and preparations to go to grad school) with major assignment dummies as covariates. Major assignment dummies are based on an aggregation of major categories which in turn are based on the majors coding system established by the Ministry of Education. Altogether we have four categories of major assignment: a combination of economics, law, and management; a combination of education, literature, and history; a combination of science, agriculture, and medicine; and finally engineering. We thus add three dummy variables (the fourth category, engineering, is omitted) in the RD plus covariates analysis. In the end, none of the analyses showed a significant effect of the treatment on the outcome. We thus conclude that the college tier effects, even after controlling for major, have no significant impact on expected salary, net tuition fees paid, class rank, or preparations to go to grad school.

[34] Frölich (2007) state that "in the case where X is a post-treatment variable, and a change in treatment status D may have an effect on Y via X as well as a direct effect on Y" one can disentangle "the direct from the indirect effect... (under certain conditions)...by controlling for X" in the regression discontinuity analysis.

Table 6
Regression discontinuity controlling for pre-treatment covariates estimates.

	Science track students			
Method Description	RD analysis 1 RD using all the science students data	RD analysis 2 RD with data limited to those who chose Shaanxi non-military HEIs as first choice	RD analysis 3 Same as RD2 but also limited to + and −30 around the cutoff	RD analysis 4 Same as RD3 but also excluding those who took exam before
Expected salary (monthly)	200.74 (146.37)	46.19 (150.82)	46.51 (184.30)	−183.88 (412.61)
Top third of class (yes/no)	−.08 (.29)	−.11 (.49)	−.07 (.37)	−.16 (3.78)
Net tuition	−399.23 (342.34)	−532.03 (417.83)	−561.86 (410.57)	−887.13 (974.66)
Received 1st choice major (yes/no)	−.46[ii] (.34)	−.27[ii] (.49)	−.23[ii] (.42)	.12 (3.66)
Graduate school (yes/no)	−.10 (.36)	−.36 (.47)	−.28 (.43)	.39 (1.45)

Notes: (1) *Significant at the 10% level, **significant at the 5% level.
(2) Bootstrapped SEs (500 times).
[ii] significant at the 5% or 10% level when bandwidth is increased by 200% (magnitude of estimate is similar).

7. Discussion and conclusion

Despite the fact that the estimators from the RD, genetic matching and RD plus covariate designs above are all slightly different, their estimates lead to the same basic conclusions.[35] Namely that attending a first tier institution reduces the chance that a student will get his or her first major choice. In addition, going to the first tier seems to have no effect on a student's immediate plans to go to graduate school, class ranking, net tuition or expected monthly wages. We discuss some of the implications associated with these main results below.

This study altogether evaluates the impact of a certain type of admissions policy—namely the imposition of strict tier eligibility cutoffs—on students' at-college and post-college experiences. This policy may well be of concern to families who place great emphasis upon sending their children to higher-tier institutions. Currently, students who just miss the cutoff to attend a first-tier institution may spend an entire year and substantial funds to prepare to retake the college entrance exam again the next year—indeed, high schools in China charge substantial tuition fees and more so for students who repeat their senior year. Students and their parents who are not informed of the tradeoffs of attending a less selective first tier institution versus a more selective second tier institution may be willing to devote this time and expense to repeat the college entrance exam. According to our findings, however, there are no tangible economic benefits to such an expenditure of time and money, as students who just miss the first tier cutoff enjoy the same at-college and post-college outcomes as those who just make it into the first tier.

Therefore it seems that attending the first versus second tier matters little for students around the eligibility cutoff, unless they strongly prefer a certain major over another. Put another way, the most consistent result of this study is that students around the eligibility tier cutoff often have to sacrifice their first choice of major if they wish to attend a first-tier institution. This is in addition to the fact that these students may already be somewhat conservative in listing their first-tier major choices (as they know they have to compete with other students in the first tier for spots in various majors) and less conservative when listing their second-tier major choices. It could also be that the opaque rules associated with China's high stakes college admissions process, not to mention the lack of counseling and information about how to make college/major choices, prevent students from revealing their ideal preferences for colleges/majors, especially around the tier cutoff.[36] The policy of imposing a tier eligibility cutoff thus adds to potential inefficiencies associated with not allowing students to choose and/or change their majors once they get to college.

At first glance, the results of this study would also seem to imply that there may not be a large difference in the quality of higher education institutions that currently accept students on either side of the first tier eligibility cutoff (i.e. the most selective second tier universities and the least selective first tier universities). While this may be the case, this interpretation should be treated with some caution. First, the estimated effects of going to the first-tier versus the second-tier include university *and* major effects rather than just university effects alone. The "null effect" findings of this study may be the result of counteracting effects from university and major selectivity; for example, going to a more selective first tier university may have a positive effect on students' expected wages ceteris paribus, but this effect may be offset by a potential negative effect on wages of attending their second choice major (as opposed to attending their first choice major). At the same time, we did

[35] This may not be surprising as the different analyses focus on similar samples and are exposed to similar threats to internal validity.
[36] See Abdulkadiroglu et al. (2009) for more information about alternative "strategy-proof" college admissions mechanisms, in which students are not required to hide their true preferences.

control for major assignment through an RD plus covariates analysis (see Section 6.3), but did not find significant effects of attending a first tier institution on expected salary, net tuition, class ranking, or planning on going to graduate school.

One plausible concern is that our analyses examine the effects of college selectivity on student expected wages in the first year of graduation. This outcome variable may not have a great deal of variation compared to wages several years after graduation. Wages in China may also not fully represent a graduate's productivity since many students choose to work in the public sector which has other benefits not reflected in wages. At the same time, we are less concerned with the idea that expected wages may not reflect actual earnings since the main task of Chinese college students in their last year of undergraduate study is to find a suitable job and as a number of authors (e.g. Webbink and Hartog, 2004) find that students in general are quite apt at estimating actual wages.

This paper is one of the first to look at the repercussions of attending different types of tiers within China's higher education system. Research about the effects of a university education in China mainly focuses on estimating the returns to having a university education in general (Fleisher and Li, 2004; Heckman and Li, 2003). The only exception we find is a concurrent study by Fan et al. (2010) which also uses a regression discontinuity design to find positive wage differences between graduates who attended Chinese 4-year versus 3-year colleges. However, Fan et al. (2010) rely heavily on the accuracy of self-reported college entrance exam scores for a sample of individuals who may have graduated from college as much as 20 years prior; potential measurement error in these scores would bias the causal estimates.

Admittedly, there are several limitations to the current study. Perhaps most notably, our sample of universities on either side of the tier eligibility cutoff did not include a significant number of aspiring world-class (985 and 211) universities that have received substantial funding from the Chinese government for over a decade. Future research could attempt to estimate the causal effects of attending these most selective universities on various student outcomes. In this case, however, researchers would have to do so without the benefit of the strictly-defined eligibility cutoff lines and rules (between university tiers in China) discussed in this paper. Another limitation is that our sample was limited to local students from a single province—longitudinal data on high school graduates from one or multiple provinces would provide more diverse and reliable causal estimates. Future research could further use methods similar to the ones discussed in this paper to examine the impacts of attending second-tier versus third-tier (expensive, private) institutions as well as third-tier versus fourth-tier (vocational) institutions, but would have to further account for the fact that the rate of repeating the college entrance exam is much higher around the cutoffs for these tiers.

Further exploration into tier selectivity effects is clearly needed. Information about these types of effects, combined with information on instructional costs at different types of institutions, could enable policymakers to better determine the allocation of student quotas, resources, fees, and financial aid within and across tiers (Black et al., 2005). The knowledge of tier selectivity effects and the sorting of students from different backgrounds across different tiers can also help policymakers better understand the potential impact of the higher education system on the labor market and social stratification. Finally, tier effects measured just around an eligibility cutoff can expose the direct consequences (on student at-college and post-college experiences) of policy-mandated tier eligibility cutoffs themselves.

References

Abdulkadiroglu, A., Pathak, P., Roth, A. (2009). Strategy-proofness Versus Efficiency in Matching with Indifferences: Redesigning the New York City High School Match. National Bureau of Economic Research Cambridge, Mass., USA.

Angrist, J.D., Lavy, V., 1999. Using Maimonides' rule to estimate the effect of class size on scholastic achievement. Quarterly Journal of Economics 114, 533–575.

Astin, H.S., Antonio, A.L., 2004. The impact of college on character development. New Directions for Institutional Research 122, 55–64.

Bao, V., 2007. Research on Mechanisms for Raising Funds for Higher Education After Expansion. CIEFR Working Paper, 2007, 7-8A (in Chinese).

Battistin, E., Rettore, E., 2007. Ineligibles and eligible non-participants as a double comparison group in regression-discontinuity designs. Journal of Econometrics 142 (2), 715–730.

Becker, G., 1964. Human Capital. National Bureau of Economic Research, New York.

Behrman, J., Rosenzweig, M., Taubman, P., 1996. College choice and wages: estimates using data on female twins. Review of Economics and Statistics 78 (4), 672–685.

Black, D.A., Smith, J.A., 2004. How robust is the evidence on the effects of college quality? Evidence from matching. Journal of Econometrics 121 (1-2), 99–124.

Black, D.A., Smith, J.A., 2006. Estimating the returns to college quality with multiple proxies for quality. Journal of Labor Economics 24 (3), 701–728.

Black, D.A., Daniel, K., Smith, J.A., 2005. University quality and wages in the United States. German Economic Review 6 (3), 415–443.

Bowen, W., Bok, D., 1998. The Shape of the River: Long-term Consequences of Considering Race in College and University Admissions. Princeton University Press, New Jersey.

Brand, J.E., Halaby, C.N., 2006. Regression and matching estimates of the effects of elite college attendance on education and career achievement. Social Science Research 35 (3), 749–770.

Brewer, D., Eide, E., Ehrenberg, R., 1999. Does it pay to attend an elite private college? Cross cohort evidence on the effects of college type on earnings. Journal of Human Resources 34 (1), 104–123.

Dale, S.B., Krueger, A., 2002. Estimating the payoff to attending a more selective college: an application of selection on observables and unobservables. Quarterly Journal of Economics 117 (4), 1491–1528.

Diamond, A., Sekhon, J.S., 2010. Genetic Matching for Estimating Causal Effects: A General Multivariate Matching Method for Achieving Balance in Observational Studies. Technical Report WP2006-35. Mimeo UC Berkeley.

Drake, C., 1993. Effects of misspecification of the propensity score on estimators of treatment effect. Biometrics 49 (4), 1231–1236.

Fan, E., Meng, X., Wei, Z.C., Zhao, G.C., 2010. Rates of Return to University Education: the Regression Discontinuity Design. CEPR Discussion Papers 634. Centre for Economic Policy Research, Research School of Economics, Australian National University.

Feldman, K.A., Newcomb, T.M., 1969. The Impact of College on Students, vol. 1. Jossey-Bass, San Francisco.

Fleisher, B.M., Li, H. Li, S., Wang, X., 2004. Sorting, Selection, and Transformation of the Return to College Education in China. IZA Discussion Papers 1446, Institute for the Study of Labor (IZA).

Frölich, M., 2007. Regression Discontinuity Design with Covariates. IZA Discussion Papers 3024, Institute for the Study of Labor (IZA).
Frölich, M., Melly, B., 2008. Quantile Treatment Effects in the Regression Discontinuity Design. IZA Discussion Papers 3638. Institute for the Study of Labor (IZA).
Hahn, J., Todd, P., Van der Klaauw, W., 2001. Identification and estimation of treatment effects with a regression-discontinuity design. Econometrica 69 (1), 201–209.
Heckman, J.J., Li, X., 2003. Selection Bias, Comparative Advantage and Heterogeneous Returns to Education: Evidence from China in 2000. NBER Working Paper W9877, National Bureau of Economic Research, Inc.
Ho, D., Imai, K., King, G., Stuart, E., 2007a. Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. Political Analysis 15 (3), 199–236, <http://gking.harvard.edu/files/abs/matchp-abs.shtml>.
Ho, D., Imai, K., King, G., Stuart, E., 2007b. Matchit: nonparametric preprocessing for parametric causal inference. Journal of Statistical Software <http://gking.harvard.edu/matchit>.
Imbens, G., Kalyanaraman, K., 2009. Optimal Bandwidth Choice for the Regression Discontinuity Estimator. NBER Working Paper No. 14726, National Bureau of Economic Research, Inc.
Imbens, G., Lemieux, T., 2007. Regression Discontinuity Designs: A Guide to Practice. Technical Working Paper 337, NBER Technical Working Paper Series.
Ishida, H., Spilerman, S., Su, K., 1997. Educational credentials and promotion chances in Japanese and American organizations. American Sociological Review 62 (6), 866–882.
Jacob, B.A., Lefgren, L., 2004a. The impact of teacher training on student achievement: quasi-experimental evidence from school reform efforts in Chicago. Journal of Human Resources 39 (1), 50–79.
Jacob, B.A., Lefgren, L., 2004b. Remedial education and student achievement: a regression-discontinuity analysis. Review of Economics and Statistics 86, 226–244.
Kane, T.J., 2003. A Quasi-experimental Estimate of the Impact of Financial aid on College-going. NBER Working Paper Series, Working Paper 9703, <http://www.nber.org/papers/w9703>.
Lee, D.S., Card, D., 2008. Regression discontinuity inference with specification error. Journal of Econometrics 142 (2), 655–674.
Lee, D.S., Lemieux, T., 2010. Regression discontinuity designs in economics. Journal of Economic Literature 48 (2), 281–355.
Li, H.Z., 2010. Higher education in China–complement or competition to US universities? In: Clotfelter, C.T. (Ed.), American Universities in a Global Marketplace. University of Chicago Press, Chicago.
Loury, L.D., Garman, D., 1995. College selectivity and earnings. Journal of Labor Economics 13 (2), 289–308.
Loyalka, P.K., Song, Y.Q., Wei, J.G., 2009a. Institutional Mechanisms and Sorting Across China's Higher Education System. CIEFR Working Paper 9001.
Loyalka, P.K., Song, Y.Q., Wei, J.G., 2009b. Information, College Choice and Financial Aid: Evidence from a Randomized Control Trial in China. CIEFR Working Paper 9004.
Ludwig, J., Miller, D., 2007. Does head start improve children's life chances? Evidence from a regression discontinuity design. Quarterly Journal of Economics 122 (1), 159–208.
McCrary, J., 2008. Manipulation of the running variable in the regression discontinuity design: a density test. Journal of Econometrics 142 (2), 698–714.
McCrary, J., Royer, H., 2006. The Effect of Female Education on Fertility and Infant Health: Evidence from School Entry Policies Using Exact Date of Birth. NBER Working Paper No. W12329, National Bureau of Economic Research, Inc.
McEwan, P., Shapiro, J., 2007. The Benefits of Delayed Primary School Enrollment: Discontinuity Estimates Using Exact Birth Dates. Wellesley College and LSE Working Paper.
National Bureau of Statistics of China, 2008. China Statistical Yearbook. China Statistics Press, Beijing.
Nechyba, T., 2006. Income and peer quality sorting in public and private schools. In: Hanushek, E., Welch, F. (Eds.), Handbook of the Economics of Education 2. Elsevier/North Holland, pp. 1327–1368.
Nichols, A., 2007. Causal Inference with Observational Data: Regression Discontinuity and Related Methods in Stata. North American Stata Users' Group Meetings 2007 2, Stata Users Group.
Pascarella, E.T., Ethington, C.A., Smart, J.C., 1988. The influence of college on humanitarian/civic involvement. The Journal of Higher Education 59 (4), 412–437.
Rosenbaum, P.R., Rubin, D.B., 1983. The central role of the propensity score in observational studies for causal effects. Biometrika 70 (1), 41–55.
Rosenbaum, P.R., Rubin, D.B., 1984. Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. The American Statistician 39 (1), 33–38.
Rubin, D.B., 1973. Matching to remove bias in observational studies. Biometrics 29 (1), 159–183.
Sekhon, J.S., 2006. Alternative Balance Metrics for Bias Reduction in Matching Methods for Causal Inference. Working paper.
Sekhon, J.S., 2011. Multivariate and propensity score matching software with automated balance optimization: the matching package for R. Journal of Statistical Software 42 (7), 1–52.
Shadish, W.R., Cook, T.D., 2009. The renaissance of experiments. Annual Review of Psychology 60, 607–629.
Shadish, W.R., Cook, T.D., Campbell, D.T., 2002. Experimental and Quasi-experimental Designs for Generalized Causal Inference. Houghton-Mifflin, Boston.
Spence, A.M., 1973. Job market signaling. Quarterly Journal of Economics 87 (3), 355–374.
Steiner, P.M., Cook, T.D., Shadish, W.R., Clark, M.H., 2010. The importance of covariate selection in controlling for selection bias in observational studies. Psychological Methods 15 (3), 250–267.
Thistlewaite, D., Campbell, D., 1960. Regression-discontinuity analysis: an alternative to the ex-post facto experiment. Journal of Educational Psychology 51 (6), 309–317.
Thomas, S., 2003. Longer-term economic effects of college selectivity and control. Research in Higher Education 44 (3), 263–299.
Van der Klaauw, W., 2002. Estimating the effect of financial aid offers on college enrollment: a regression-discontinuity approach. International Economic Review 43 (4), 1249–1287.
Webbink, D., Hartog, J., 2004. Can students predict starting salaries? Yes! Economics of Education Review 23, 103–113.
Winston, G.C., 1999. Subsidies, hierarchy and peers: the awkward economics of higher education. The Journal of Economic Perspectives 13 (1), 13–36.
Zhang, L., 2005a. Advance to graduate education: the effect of college quality and undergraduate majors. The Review of Higher Education 28 (3), 313–338.
Zhang, L., 2005b. Do measures of college quality matter? The effect of college quality on graduates' earnings. The Review of Higher Education 28 (4), 571–596.
Zhou, S. and Loyalka, P., 2010. Measuring Instructional Costs in Chinese Universities, Mimeo.