[Preliminary, Do Not Cite]

The Marginal Returns to College Quality*

[Job Market Draft]

Steven T. Simpson†

October 1, 2012

Abstract

Estimating the returns to college quality is often complicated by lack of good data and/or reliable information about the selection model. To circumvent these complications, I use administrative data from Taiwan's highly competitive, centralized higher education market. In these data I observe the match outcomes for a universe of college applicants and a universe of college seats available, as well as the eventual earnings of these graduates. To facilitate this analysis, I employ a stacked regression discontinuity design that pools the estimates from over 1000 discontinuities. My results show that attending a marginally better program (college-department) does have a substantive impact on first-year earnings. I also use other proxies of college quality to show that attending these marginal better programs provides students access to more prestigious institutions and better financial aid in the form of public subsidy.

JEL Classification: I21, J24, J31

Keywords: college quality, returns to education, regression discontinuity.

*I would like to thanks Professor Jin-Tan Liu of National Taiwan University for giving me access to his data.
†Correspondence: Email: stsimpson@columbia.edu

Access to higher education has been expanding rapidly around the world (Becker et al., 2010; Goastellec, 2008). In one sense, this positive change could herald a more egalitarian society with decreasing inequality. In another, it means college quality has never mattered more to students who seek to distinguish themselves in a competitive, globalized economy.

How much college quality matters has been difficult to estimate precisely and convincingly. Using data from Taiwan and a regression discontinuity (RD) design, I provide quasi-experimental evidence that, on average, scoring just above the admissions threshold versus scoring just below are exposed to colleges of higher quality, with a 0.04 σ increase in average peer quality or about 2.5 points on their college entrance exam. That peer quality advantage corresponds to a 0.8 log-points (on a 2-digit log base) or 0.04 σ in log earnings. These results are economically substantive, statistically significant, and robust to specification testing.

I contribute to the 'college quality' literature in terms of data quantity and quality as well as research design. First, data quantity: mine is the first RD paper to use a national universe of microdata for college applications, college admissions, and earnings — all merged on national IDs — to identify causal estimates of the returns to college quality. Further, because my data catalogs student admissions down to the academic department level, I can estimate these returns at the college-major level, instead of just the college level, as in the other papers. Thus, my estimates can take into account the choice of college major as well as the choice of college, which had before been the purview of structural estimation strategies (Arcidiacono, 2004).

Second, data quality: having the universe of data for this admissions cohort means I also have the universe of discontinuities. I estimate these reduced-forms results using an RD framework with over 1000 discontinuities, generated by the Taiwanese governments' centralized student allocation system. Third, research design: the stacked RD I use allows me to approximate a local average treatment effect (LATE) at every point in the test score distribution. Further, this research design allows me to answer the question of whether (in this context) it is better to be a big fish in a little pond or a little fish in a big pond. The results suggest the latter.

Following the recent college quality literature (Black and Smith, 2006; Dale and Krueger, 2002, 2011), I measure college quality as a function of peer quality, proxied by the average college entrance exam score at that college. I also use other proxies such as college prestige/selectivity (Chevalier and Conlon, 2003) and financial aid in the form of public subsidy (Van der Klaauw, 2002; Winston, 1999). Other measures of college quality have been considered in the literature: such as endowments (Winston, 1999), faculty quality and faculty-student ratios (Saavedra, 2008; Winston, 1999)[1], or research productivity (Aghion et al., 2010). Each of these papers shows that peer quality is correlated with the other proxies of college quality, and thus my proxy is still a valid summary. Other mechanisms than peer quality may also be driving the reduced-form results I identify, such as peer group structure or tracking while in college (Carrell et al., 2011); yet the positive wage effects I identify fall in line with what Wiske-Dillon and Smith (2012) — a structural estimation — find. Among college students there appears to be a preference for being a little fish in a big pond, or as Dillon and Black refer to it "being underqualified" — that is, just making it into the higher quality school. Finally, reduced-form outcomes to school quality, such as aggregate productivity in the workforce or economic growth (Hanushek and Woessmann, 2012; Lucas, 1988) are beyond the scope of this paper.

My paper also contributes to the 'better schools' literature. Those studies focusing on better high schools (Clark, 2007; Jackson, 2010; Pop-Eleches and Urquiola, 2012) generally show that attending a higher quality high school leads to higher scores, though not more tries, on a college entrance exam. Unlike most papers in this literature, I can estimate wage returns in the labor market. Unlike any other paper in the 'college quality' or 'better schools' literatures, I can do this using all college-goers in a cohort.

The remainder of the paper is organized as follows: section 1 reviews the relevant literature and describes the educational system in Taiwan, while section 2 details and summarizes the administrative data sets used. Section 3 motivates the use of RD estimation strategies. Section 4 discusses results, and Section 5 concludes.

[1] Future drafts will include this aspect.

1 Literature and Context

One way to estimate the importance of college quality on future earnings is to perform a random survey of the relevant subpopulation of college-goers and ask where one went to college and how much that person income is. Then producing a summary table of these results would give an accurate description of the relationship between college quality, conditional on college attendance, and earnings. Moving beyond describing the relationship, one may want to estimate that correlation using some form of statistical analysis, like regression. However, as several researchers have noted in the past, there are individual characteristics — like sex, race, family background, to name a few — that could predict both the probability of attending a certain type and quality of college as well as the level of income that person achieves. And each of these characteristics would have been set prior to ever attending college. One could argue that these have allowed the achievement that is required to both get into a high-quality institution and get a high-paying job. Thus, the earnings benefit observed is better attributed to these factors and not college quality.

Another common way, one that tries to estimate the returns to college quality by controlling for all relevant characteristics. A serious concern is how many 'relevant' factors exist. Omitting one of these will again bias the results against finding the 'true' estimate. Lee and Lemieux (2010) point out that even an the extreme hypothetical case where a "kitchen sink" regression that included all relevant individual characteristics could perfectly predict admission into a high quality college or not, there is still no way to estimate the returns to college quality, independent of these covariates, because there is no independent variation left.

Thus, in recent years researchers have been looking beyond statistical fixes like "selection on observables" and matching, and more towards what works like research design in experimental settings. But because randomized experiments are expensive and often come up against ethical objections, a set of natural experimental methodologies have been designed to identify opportunities to approximate what works like a randomization but in everyday life, so that cross-section data can still be used. The regression discontinuity design is consider among these to be one of the most convincing research designs, if its stricter assumptions are met.

Dale and Krueger (2002) is one of the early papers in the natural experiments literature on school quality that sought to overcome selection problems between schools and students. Using a novel quality of the College and Beyond Survey that allows them to observe which college students applied to, were admitted at, and eventually attended, Dale and Krueger find that compared to those who were accepted but chose not to go, those who attended the more selective institution experienced a significantly higher wage in their first year of employment. However, Dale and Krueger show that 5 years later that earnings advantage fades, which they interpret to mean that college quality does not matter. Dale and Krueger (2011) return to this analysis using both more and better administrative earnings data further out in time and a more recent larger set of students. Their conclusion is the same: controlling for average SAT score—a measure of pre-existing student ability—absorbs all the effects of attending a selective college.

MacLeod and Urquiola (2011) argue against Dale and Krueger's interpretation. They believe that first-year earnings is exactly where we would expect college quality to play an important role. Employers use college quality to simplify the task of predicting a graduate's productivity before they have had a chance to observe it for themselves. The longer employers do observe the worker's productivity, the less they will need to rely on other signals of it, such as the average quality of the college the worker attended. This argument does not require that school quality work only as a signal of student ability. Indeed, MacLeod and Urquiola (2011)'s model of college reputation allows for both human capital and signaling effects. My analysis will not allow me to separate these two effects. In Lange et al. (2006)'s opinion, no study to date has been able to convincingly separate these two effects.

Hoekstra (2009) was a strong contribution to the college quality literature in terms of data and design. He uses application and admissions data at a flagship university in Texas, comparing the earnings outcomes for those who were marginally admitted and rejected. The argument towards randomization in the RD settting is that those who are just above and just below the admissions threshold are essentially the same in observable characteristics. Indeed, in many cases where admissions cutoffs are set based on a test scores, the random measurement error inherent in most tests is enough random variation to support this claim that around the cutoff students on either side are identical in expectation. Thus, those who are just were just below the threshold provide a good comparison group for those who are just above, and the differences in outcomes that occur later are much more plausibly argued to be due to, in this example, being admitted to the better college versus not.

The quality of Hoekstra's application and earnings data is what makes possible his RD design. Aside from the details admissions records, an important contribution of Hoekstra's paper is that he uses administrative earnings data from IRS tax records for 10 years after the students graduated; this is a significant improvement over less reliable records used in studies — like Dale and Krueger (2002) — that rely on self-reported earnings of survey respondents. Hoekstra finds that those who were admitted earned an economically significant higher income than those who were not. Such a rebuttal to the conclusions of Dale and Krueger (2002) could be held as more convincing, based on the higher standards of the RD research design compared to Dale and Krueger's sample strategy.

A critical weakness of Dale and Krueger (2002)'s sample design is that we are asked to assume that random reasons led students who were accepted but decided to reject admissions into the more selective institution. Hoekstra's RD design circumvents the need for this assumption. Instead of taking admitted students and matching them, he takes those that were in the admissions process marginal losers and marginal winners. In expectation, absent the arbitrariness of the flagship college's admissions criteria, we would expect them to have the same wage outcomes. A critical weakness of Hoekstra's design, though, is that he does not know what happened to those who were not admitted in his sample. Importantly, he does not know for sure whether they even went to college or not. Thus, the earnings advantage he finds in his study could be simply the earnings return to college attendance, instead of college quality. There are reasons to believe, though, that those who were competing at the margin for admissions to the flagship college in Texas would have been over the admission threshold at one of the non-flagship institutions, and very likely would have gone on to attend college. Dale and Krueger (2011) follows up and affirms their original work, by including even more and better administrative data on earnings, both farther out in time for the original cohort and with a new cohort. Still, in the end, Dale and Krueger (2002, 2011) are using a control strategy that includes many covariates (in the form of the average SAT score for every college that the student applied to) in an attempt to absorb all the influences of selection. By doing this, they many have unintentionally been absorbing both good and bad variation, and thus they null finding may not be conclusive.

Saavedra (2008) similarly uses an RD setup to estimate the returns to exceeding the admissions threshold at a top school in Colombia. He finds that doing so increases first-year earnings by about 35%. The most important contribution of this study is that it is able to estimate value-add of college quality on knowledge learned and on earnings, using scores on subject-specific standardized college *exit* exams commonly taken by college graduates at selective institutions. Rubinstein and Sekhri (2011) is the only other RD paper able to estimate the value-add of attending a more prestigious university. In most developing countries, public institutions are more prestigious than private ones, and thus are considered to produce better learning. Rubinstein and Sekhri show that there is no value-added to attending these more prestigious schools, as measured by exit exam scores. Unfortunately, their analysis does not include wages. The opposing views from the seminal papers above leave room for more work to add new and varied evidence to the discussion.

1.1 General context in Taiwan

Taiwan's market for higher education is centrally controlled by the Ministry of Education (MOE), as is every level of education there. One of the historically distinctive characteristics of that centralized control has been its constraint of supply — both publicly and privately provided, and again at every level. This control extends down to the number of students and teachers that are allowed to be in each and every school, even at the academic department level, in any given year.

Early research on the returns to education in Taiwan has focused on the 1968 expansion of compulsory schooling to include the 3 years of junior high school. With this reform, junior high school enrollment increased by 50% in the first year (Clark and Hsieh, 2000). Several studies have looked at the effect of this junior high school expansion on such varied outcomes as the labor market returns to education (Spohr, 2003), the decrease in wage inequality (Vere, 2005), the increase in female labor market participation (Tsai et al., 2009), the effect of maternal education on infant health (Chou et al., 2010), and the intergenerational transfer of human capital (Tsai et al., 2011).

Besides the historically limited supply, the other distinctive characteristic of Taiwan's higher education market is that admissions is determined by performance on a competitive examination and is centrally administered through a national, market-clearing student allocation system. Again, this centralized testing regime applies for admissions to both public and private institutions.[2] In the first week of July each year, any high school senior who hopes to attend college sits for the Joint College Entrance Exam (JCEE). The score on the college entrance exam T_i is a composite of subject-specific tests t_i. The number of subject tests a student must take ranges from 5 to 6 (max score for each is 100), and is predetermined by the MOE for each of the 4 educational tracks: social, science, medical, or agriculture.[3] (See Appendix Table A1.) Chinese and English are the only two subjects common to all the educational tracks. The scores on the JCEE T_i are a continuous measure, noted down to the hundredth of a decimal point with many data points occurring between any given test-point interval.

After observing her performance on the JCEE, the student must fill in an "aspirations card" that lists in ranked order her preferences over college departments.[4] In the notation that I use, I refrain from including an index for colleges because it simplifies notation. But more importantly, I do so because students cannot select a college only. The preference rank she lists must be for a specific academic department within a college. Thus, the discontinuities that I exploit are at the academic department level. Nevertheless, because of the research design I follow, I compare the quality of departments within a given college major; since for each college major a college typically has only one department, this means that the analysis reduces again down to a college quality comparison, taking into account college major. Because students are able to list preferences for academic departments in more than one college major, my results should be interpreted as a relevant estimate, based on one set of possible comparisons.

[2]Admissions into a 2-year junior college or technical college is based on a separate, yet equally centralized admissions system.

[3]According to the MOE's system of categorizing college majors and academic departments that teach them, there are 4 major educational tracks. Within these, there are 8 fields of study, which can be further broken down into 23 disciplines and 158 subfields or college majors. Academic departments can then be defined narrowly by how the majors they offer are categorized (`www.edu.tw/files/site_content/B0013/bcode.xls`).

[4]Over the years the maximum number of preferences a student could list has increased. Currently, because students must list specific academic departments and not just colleges, the maximum is 80.

A student is only eligible for admissions at departments that are within the educational tracks she has chosen to pursue. Thus, the first rule of admissions is that a student must take the set of subject tests required for her desired college major. For example, if a student chooses to take the required subject tests for the science track, but not the tests for medicine, she will not be considered by the matching algorithm for any department of medicine. (More regarding the JCEE test and the admissions process is discussed in Section 2.4). This aspirations card is returned to the College Entrance Exam Center. The College Entrance Exam Center is the independent government agency attached to the MOE tasked with the administration of all elements of the college admissions process, from JCEE development to managing the matching students to seats at university. (For ease of reading, hereafter I will use College Entrance Exam Center and MOE interchangably.)

1.2 Admissions and the student matching algorithm

The student matching algorithm used by the MOE falls into a class of matching algorithms known as the Deferred Acceptance algorithm (hereafter 'the algorithm'). Though the algorithm is used to prevent departments and students from choosing each other directly, it does allow for student preferences and department priorities to be taken into account. To use the language of Abdulkadiroğlu and Sönmez (2003), departments have a set of "priorities" over student qualities they would like to enroll, whereas students have preferences for specific departments. Department priorities are more abstract in that they are not for specific students, but rather for qualities that any applicant may have. The matching algorithm uses the departments' priorities and students' preferences to match students to seats in specific departments.

The main objective of the algorithm is to allocate unoccupied seats in a specific department to students who have the highest priority and a preference for being in that department. According to Abdulkadiroğlu and Sönmez (2003), the steps of the algorithm are as follows:

> "*Step 1*: Each student proposes to her first choice. Each [department] tentatively assigns its seats to its proposers one at a time following their priority order. Any remaining proposers are rejected.
>
> "In general, at

> "*Step k*: Each student who was rejected in the previous step proposes to her next choice. Each [department] considers the students it has been holding together with its new proposers and tentatively assigns its seats to these students one at a time following their priority order. Any remaining proposers are rejected" (p. 735).

This process terminates when all student preferences have been consider and all seats have been allocated.

The quote above suggests that departments are actively engaged in the priority ordering of students in each round. This is not the case in Taiwan. Departments notify the MOE of their priorities over student qualities before the admissions process begins and without ever seeing those who have applied to them. This can happen because in Taiwan's case, a student's priority order at a department is determined solely based on a student's JCEE scores. A department's priorities for specific student qualities are used by the algorithm to weight the exam performance of all students who have listed that department in their preference list; the weighting occurs in such a way that it influences the students' priority ordering for that department only.[5]

The outcome of the matching algorithm is final. Departments are required to admit all those that the algorithm assigns them; there are no wait-lists during the admissions process nor are there transfers between programs after admissions is over. The only way to be admitted into a department is to go through the MOE defined system. Students have no recourse if they do not like their match. Not liking their match does not mean that it was not their best possible match, given their performance on the JCEE and their priority ordering at departments listed in their preferences. If they so choose, they are allowed to retake the JCEE again the next year, but this is uncommon for admitted students.

According to Abdulkadiroğlu and Sönmez (2003) and many others, a Deferred Acceptance algorithm is a Pareto-efficient method for optimally matching one-to-one a finite number of students to a finite number of "objects" — in this case the objects are seats at college departments. Another advantage that this kind of matching algorithm has over other Pareto-efficient matching algorithms — such as the random serial dictatorship, which is often used for such real-life purposes as allocating students to dorm rooms — is that the Deferred Acceptance algorithm allows

[5]This weighted exam score T_{id} which provides the priority ordering to the matching algorithm corresponds to the running variable in the RD design.

for different schools (or college departments in the case of Taiwan) to have differing priorities over the same students. The same student may have different priority rankings at different schools, and the Deferred Acceptance algorithm — and thus my study — allows for these priorities to be accommodated.[6]

2 Data

This paper uses the universe of governmental administrative data on birth, education, and earnings outcomes for the 2000 cohort of college applicants in Taiwan. The micro-data are merged using national IDs. Table 1 provides the descriptive statistics. The sample of college admitted students appears to be evenly balanced in gender, and entering college at the expected age of around 18. Parents at the time of birth on average had less than a high school diploma, which is consistent with the historical expansion of public secondary education in Taiwan. Related to the parent's lower educational attainment, even though there has been worry that the market for college education has been expanding too quickly and thus starting salaries are less than before, it appears that the first-year earnings of college graduates is still high, compared to their parents' income. The average first-year earnings for newly graduated students are only slightly lower than the average of their parents' combined income four years prior at the time the student took the college entrance exam. Thus, a college degree does appear to still have value in Taiwan.

2.1 Admissions data

The main data set is the college exam and admissions data from the 2000 application season (for students admitted to college for the 2000-01 academic year), which contains all records of exams taken. These detailed records include scores on each of the subject tests t that the applicant i took. The admissions data also includes the department ID and name (including the college's name) where the applicant was admitted, if admitted at all. Approximately 111,150 applied to

[6] As Abdulkadiroğlu and Sönmez (2003) point out, a random lottery cannot take into account all the priorities that a department may have for students or priorities that are imposed upon them by an outside governing body. The example provides is of students being allocated to public schools, which often give higher priority to students who have siblings already in the school, who live within that school's catchment area, or who fall into certain mandated quotas.

college of which 67,898 or 61% were admitted.

The data does not supply the students' ranked preferences over all departments, only the name of the college-department where the student was admitted, if admitted at all. This realized outcome is only one of the several outcomes that we would like to observe in order to assess what could have happened. However, for the purposes of this paper, I make an assumption that students having been admitted into a particular academic department would have listed other academic departments which also offered the same college major. Making this assumption allows me to use information embedded in the examination and admissions records as well as knowledge of the admissions process to generate a set of other departments offering that college major where the student would have been eligible to apply. Admissions to college in Taiwan is limited by the educational track a student can apply within. As mentioned in Section 1, before taking the JCEE students must choose at least which educational track (social, science, medicine or agriculture) they want to pursue. This is because the MOE has predetermined which set of 5 to 6 subject-specific tests t are required in order to be eligible to apply to the departments in that specific educational track. Yet, even limiting the number of counterfactuals I consider to those departments where the student would have been eligible to apply based on the 5 or 6 subject tests she choose to take[7] would leave me with possibly hundreds of academic departments and dozens of college majors to consider for each student.

Pop-Eleches and Urquiola (2012) have a similarly unruly number of counterfactuals they could possibly consider: any prospective high school student in Romania is allowed to apply to any high school. Their simple solution is to limit the number of high schools (counterfactuals) they consider for each student to the ones in the village where the student lived. Following this guide to focus my analysis, I use the college major offered by the department where the student was admitted as a way to limit the number of counterfactuals I consider for that student. For example, for a student admitted into a department that teaches economics, I only consider for that student the potential outcomes at economics departments at other colleges. A limitation therefore of the data, and thus my study, is that a student's application strategy may be less oriented towards getting into a particular college major and more towards getting into a particular college. Therefore,

[7]Few students choose the grueling option of taking all 9 subject tests to make them eligible for any department.

my results should be interpreted as one of at least two possible and relevant counterfactuals to consider. However, because I eventually compare wage outcomes for these students, we expect that students graduating from the same major will have similar career paths and opportunities, and thus it is more preferable to compare within college major and across colleges rather than within colleges and across college majors.

2.2 Birth certificate data

For 100% of the individuals in the admissions data, I am able to match them to their birth certificate data, using the universe of all births in Taiwan for the birth cohorts 1978-1983. These data are collected by the Ministry of the Interior Affairs. From this data I get the student's date of birth (month and year), sex, and her parents' years of schooling (YOS) at time of child's birth.

2.3 Earnings data

The administrative data we use on earnings for the years 2004 to 2005; 2004 is the first year those enrolled at college in 2000 would be able to find a full-time job. The data is administered by Taiwan's Bureau of Labor Insurance and collected by the Ministry of Labor Affairs. It is an employer-employee matched data set, which includes the employee's national ID, monthly earnings for that year, and the employer's unique ID number.

As the name indicates, the Labor Insurance data (hereafter "the Labor data") is a record of healthcare insurance provided through employers. In 1995 Taiwan established universal healthcare coverage. All employed individuals were given access to this universal healthcare through their employer; and their monthly premia were paid half by the employer and half by the employee. The employee's share was based on a percentage of their monthly income. Thus, all employers were required by law to submit monthly earnings records for their employees in order for the premia to be removed from their earnings (Chou et al., 2003).

The data contain the complete universe of all private-sector workers in Taiwan, approximately 82% of the working population. Of those working adults not in the Labor data, about 10% are public sector employees, 3% of farmers/fishermen, and 5% of self-employed. The former three maintain a different form of public insurance that pre-dated the 1995 universal healthcare. All unemployed are not in this data, because their unemployment benefits cover their healthcare provision (Chou et al., 2003).

A significant limitation of the Labor data to which I have access is that only about 33% of those in my sample show up in the earnings data for 2004 and 2005. This could be for a multiple of reasons that are consistent with the average state of the labor market at this time. The main reason for the lower match rate, I think, is due to a gender-specific binding obligation. For males, those who graduated college in 2004 would have had to perform their 18-month national military service at this time[8]. For females, childbearing immediately following being newly married is a common expectation in Asia. Because for males graduating in June 2004 the 18-month military service would have ended in December of 2005, we would expect them to be looking for jobs starting January 2006. Thus, a desirable solution, though politically impossible, would be to get the Labor data for the 2006.

A second reason for the lower match rate is that, as mentioned above, about 10% of all working adults hold a position in the government. This means that greater than 10% of college graduates will work for the government: the market for government jobs is highly competitive, requiring applicants to pass a very strict test, but rewarding those who succeed with higher wages and better benefits. Therefore, I assume that a large share of the government bureaucracy is staffed by college graduates. Lesser reasons could be that access to graduate school grew noticeably about 3% a year, starting in 1999, and that the unemployment rate for college and university graduates rose to around 4% (Yang et al., 2010).

To check for evidence of whether this data limitation signals some element of selection, I regress an indicator for whether the individual has any wages on the model described further below. If we were to find a statistically significant (positive or negative) effect of having scored above the admissions threshold, then this would dampen our faith in the claims we might want to make about the outcomes of interest. If the effect were positive as well as statistically significant, then

[8] Reasons for not performing the service immediately (or at all) following graduation from college would be continued education at the post-graduate level, being deemed physically unfit for various reasons, being the only child (and thus would leave the parents without a child to assist them), having another family member currently servicing in the military, or fulfilling a bond with a government research position (`http://www.nca.gov.tw/ENGLISH/english.htm`).

we would be concerned that any wage impact we observe would really be due to selection into finding a job. If it were negative, then we might say that there were positive selection out of the private sector into the public sector. The residuals of that regression are displayed graphically in the sixth panel of Figure 3. It shows that there is no statistically significant effect, meaning that the selection into employment and possibly even sector of employment is not driving our wage results.

Even though I am sure than my wage results are not being driven by imbalance in the probability of public versus private employment, it does not diminish the fact that this smaller sample means that my reduced-form estimates are very noisy and imprecise. To overcome this limitation, I use the departmental average of first-year wages for the previous cohort in that department as a proxy for the kinds of wages I would expect for the 2000 cohort. Since the 1999 cohort had an extra year to appear in the Labor data, it increased the percentage of individuals who had wages from around one-third to over one-half. Using the 1998 data increases the match rate to over two-thirds, but fewer of the departments in existence in 2000 were operating in 1998. Therefore, for most of my results I prefer to use the department average for 1999. Though the results for the 1998 sample are generally slightly higher in magnitude, the 1999 results are more easily generalizable.

Because the earnings data is generated through the bureaucratic structures of the universal health-care system, it retains two key features of that system: it is measured at monthly intervals and somewhat roughly. The data I have access for is taken from the employee earnings records for the month of December for each year. There are advantages to this feature of the data. The first advantage of having monthly earnings from December is that we can capture earnings soon after graduates complete their learning in June. The second advantage enjoyed over monthly earnings data recorded in the typical survey is that we record all observations on the same month of the year, which protects against the noise induced by business-cycle seasonality in monthly earnings. A possible disadvantage is that if a college graduate were between jobs in either December 2004 or December 2005, the data would not capture her employment. The only way to remedy this would be to have the monthly earnings for every month, again politically impossible to obtain this data.

The Labor data is measured somewhat roughly. (See Figure 1.) The upper limit in our records is 42,000 NTD, which is sufficiently high for most earnings of new graduates. Any wage above 42,000 will be top-coded at the upper limit. And the lower limit is 11,000 NTD, which again is well below the average earnings. Again, any wage below 11,000 will be bottom-coded as the lower limit. Lastly, monthly earnings are listed in lumpy intervals, not in single-dollar units. These characteristics of the Labor data — the truncated distribution and the lumpy intervals — are artifacts of the fee schedule for healthcare with a floor, cap, and menu-pricing scheme.

I use first-year earnings because it is a sensitive measure of employers' reliance on college quality for a rough signal of students' actual (but yet to be directly observed) productivity. If employers use the average quality of a student's college cohort to predict productivity, the first-year earnings should follow a similar pattern as the variance in these signals (MacLeod and Urquiola, 2011). Further, if a college graduate's starting salary is predictive of long-term earnings growth, starting on a lower (higher) rung of this pay scale likely means that longer term earnings will also be lower (higher) (Oreopoulos et al., 2012).

2.4 Data preparation

Let T_{id} represent the score of student i being considered by the matching algorithm for admissions to department d. T_{id} does not necessarily equal the simple summation of raw scores on all the subject tests that a student may have taken. Instead, it is the combined realization of choices made by the student, the MOE, and the departments. Below I describe how each part of the following formula works to generate T_{id}:

$$T_{id} = \begin{cases} \sum_{i=1}^{N} \sum_{d=1}^{m} \sum_{j=1}^{J} t_i^j \rho_d^j w_d^j & \text{if} \quad t_i^j \geq \kappa_d^j \ \forall \ j \\ \oslash & \text{otherwise.} \end{cases}$$

The JCEE is a set J of 9 subject-specific tests created by the MOE , one t_i^j for each subject $j \in J$. Of the 9 subject tests only 5 to 6 are required by the MOE for admission into one of 4 educational tracks: social, science, medicine, or agriculture. The MOE determines which subject tests are required for admission to any given department, according to the educational track the department falls into. Departments have no control over whether a subject j is required or not. Thus, the dummy ρ_d^j in the formula above indicates (0=no/1=yes) whether or not a specific j is required for that specific department $d \in m \in M$, which is a part of a specific college major m according to the MOE's set of 158 college majors M. Notice that the formula above replicates counterfactuals only within college majors; for example, a student who is admitted into a linguistics department would have priority rankings replicated for her at other linguistics departments, but not a history department. I limit the counterfactuals in this way so that we can reasonably compare the earnings outcomes for those with the same college major.

Even though departments have no influence over required subject tests, the MOE does allow departments two ways to indirectly exercise their priorities for student quality: choosing a set of weights w_d^j and a set of subject-specific raw-score cutoffs κ_d^j to be applied to subject-specific test scores t_i^j from the JCEE. First, the MOE requires departments to choose weights w_d^j that increase the value of that required subject relative to other ones, allowing the department to emphasize that specific content knowledge over other content when summing them together. The menu of w consists of 1, 1.25, 1.5, 1.75, or 2. The default and most common weight is w=1, which leaves the raw score for that required subject score unchanged.

Second, the MOE requires departments to choose from a menu of raw-score cutoffs κ_d^j to use on t_i^j. Raw-score cutoffs let departments truncate the distribution of viable candidates at some point in the distribution of t_i^j. Departments are not allowed to set where the cutoffs occur. The same four κ^j are available for any of the required j. The default and most common raw-score cutoff is κ^j=0, which leaves the distribution of applicants unchanged. Specifically, the four κ^j are:

- A: only consider students with a t_i^j above the average of the upper 50th percentiles for subject j;
- B: only consider students with a t_i^j above the 50th percentile for subject j;
- C: only consider students with a t_i^j above the average of the lower 50th percentiles for subject j; or
- D: no raw score cut-off, which is analogous to κ^j=0.

Any student with a raw score t_i^j below the κ_d^j are automatically disqualified. In the data replication process I follow, if the condition $t_i^j \geq \kappa_d^j$ is not met, I approximate the algorithm's disqualification by giving i a missing value $\oslash$ for T_{id} for that department.

To reconstruct T_{id} as it was generated by the student matching algorithm, I use the MOE's records of department priorities for 2000, which detail how each department choose to use weights and raw-score cutoffs. The fact that the MOE allows each department to prioritize student quality along different margins of subject-specific knowledge means that there is a possibility that for each department on a student's preference list, she could have a different T_{id}. Still, even with the same score at any two departments, it is highly likely that the same student would have different priority orderings at those departments, if for no other reason than the two departments will likely not have the exact same mix of applicants. This corresponds to my description in Section 3 of how the matching algorithm generates for that student a unique priority ordering for each department in her preference list. I discuss below how this works in the context of multiple discontinuities being pooled together through a stacked RD framework.

Since the most common weights and raw-score cutoffs are the defaults, in many cases the total raw score and the total weighted score are the same. More importantly, even if a student's score were different for each of the departments she applied to, because the matching algorithm treats each department separately during the assignment of students, it only matters that within that specific counterfactual all the students are being sorted along the same priority ordering (running variable), which uses the same required subject tests, weights, and raw-score cutoffs.

2.5 Matching example

To illustrate how these three features (ρ^j,w^j, and κ^j) work in the student allocation process, take the hypothetical example with four students i_1, i_2, i_3, and i_4. Each is competing for the last seat in the same department. In this hypothetical example, the department's priorities for student quality lead it to value $t_i^{chinese}$ 1.5 times that of t_i^{math} and to want only students with a $t_i^{chinese}$ above $\kappa_d^{chinese(A)}$, that is, above the average of the top 50th percentiles in j=chinese.[9]

[9]This means that $w^{chinese}$=1.5, w^{math}=1, $\kappa^{chinese}$=A (or 54), and κ^{math}=D (or zero).

Let us assume for the purposes of this hypothetical example that $\kappa_d^{chinese(A)}$=54 and that these two subjects are the only ones required.

Looking at the matrix above, we can see that i_4 is automatically disqualified because he did not take the required math subject test. Next, we see that i_3 dominates in both raw and weighted total scores. However, because i_3's Chinese score is below the threshold $\kappa^{chinese(A)}$ for consideration, i_3 is also not admitted to d.

	$t_i^{chinese}$	t_i^{math}	Raw Total	Weighted Total	Enrolled?
i_1	64	66	131	162	yes
i_2	54	80	134	161	no
i_3	52	100	152	178$\Longrightarrow\oslash$	no
i_4	88	–	88	$\oslash$	no

Both i_1 and i_2 have $t_i^{chinese}$ above the raw-score threshold; and i_2 has the advantage in total raw score. However, because d's priority weights Chinese by 1.5, i_1 narrowly beats out i_2 for the seat. This is once again because i_1's total weighted score is higher, not because i_1's $t_1^{chinese}$ is highest. Had i_2 been able to score 1.5 more points either by increasing her $t_2^{chinese}$ by 1 point or t_2^{math} by 2, she would have bested i_1 for the seat.

3 Empirical Strategy

Students are given the chance to select which departments they would like to attend. Because the number of seats in these departments is finite and predetermined under the MOE's supervision, this capacity constraint creates a cutoff c_d in access to department d at the JCEE score for the last student admitted. Such a selection rule makes this context amenable to using an RD design to estimate the value of gaining access to the department with the higher average peer quality. Since I do not have the full list of student preferences, I approximate this using a fuzzy RD design and interpret the resulting counterfactual as an intention-to-treat (ITT).[10] Specifically, Abdulkadiroğlu and Sönmez (2003) make a helpful distinction in their analysis of student matching algorithms.

[10]See Imbens and Lemieux (2008) for a technical guide to recent RD techniques and Lee and Lemieux (2010) for survey of recent uses in the literature.

They show that a student may be eligible to be admitted to several schools according to the selection criteria created by each school; but this does not bind the student's choice. If a student is eligible for more than one school, she is given the one she prefers most, the one listed higher on her preference list. Thus, I make the distinction between being eligible for access and being admitted. A fuzzy RD takes into account that a student may be eligible for more than one, but may be admitted at only one.Below I motivate how this works with one discontinuity and then extend the framework to include many discontinuities.

3.1 Single discontinuity

Consider a hypothetical case in which there are two departments d_1 and d_2, ranked from lowest to highest by the average JCEE score for the department. One cutoff c_1 defines the access threshold to d_2, the department with the higher average peer quality. Students scoring above c_1 are eligible to attend d_1.

Let the following model (1) represent the first-stage admissions outcome, average peer quality $\overline{T_i^u}$, as a function of that student's achievement on the college entrance exam T_{id}:

$$\overline{T_i^u} = \beta 1(T_{id} \geq c_d) + b(T_{id} - c_d)^p + \gamma_c + X_i + \varepsilon_i \tag{1}$$

where $1(T_{id} \geq c_d)$ is an indicator function for whether a student's total weighted test score T_{id} is above the cutoff c_d for access to d. The function $b(T_{id} - c_d)^p$ is the summation of higher-order polynomials of order p, independently flexible above and below the cutscore. Within a narrow enough bandwidth around the cutoff (c_1 in this hypothetical case), a flexible specification of $b(T_{id} - c_d)^p$ fully controls for other factors driving access to d_2. The error term ε_i is assumed to be i.i.d.

A set of covariates X_i are included, but are not required for the coefficient of interest β to be causally identified. These include gender, age at time of testing, mother and father's years of schooling at time of birth, and family's income at time of testing. A main identifying assumption of RD is that individuals cannot manipulate their position on the running variable. The description of the rigors of the student matching algorithm above should help to motivate this assumption theoretically. Two practical ways to check for this type of manipulation are to check the density of the running variable and to test whether student characteristics are balanced across the threshold. First, Figure 2 shows two density plots. The one on the left shows the weighted JCEE scores in the original data set. These are the realized scores and are not recentered around any cutoffs. As one might expect from data taken from a large subpopulation, the results conform to a fairly normal distribution. The density plot in the right panel in Figure 2 represents the distribution of the replicated data around the recentered cutoff value of zero. Again, the data is reasonably close to a normal distribution. Second, even though it is fundamentally impossible to check whether all characteristics are balanced, I test for balance in the characteristics I have access to by inspecting whether there is overlap in their confidence intervals on either side of the discontinuity. Figure 3 displays the outcomes for the five covariates. The confidence intervals overlap, signifying that results are either statistically indistinguishable from zero. A sixth panel is provided in Figure 3 to test for whether we can predict whether someone ends up in my private sector wage data, given whether they are above or below the discontinuity. Again, there appears to be no differences between the characteristics on either side. The balance is the same for all five of the main covariates even when I condition the estimates of whether the sample had wages in my earnings data. From these results, we can infer that the covariates are balanced, and thus the positive effects we observe in the outcomes of interest are not due to substantive differences in student covariates around the cutoff.

We can interpret β as the improvement in average peer quality $\overline{T_i^u}$ that a student experiences by scoring above the cutoff for access to d_2. I measure average peer quality $\overline{T_i^u}$ in terms of the average UNWEIGHTED (or raw) test scores on the JCEE for each department, signified by the superscript u in T_i^u to differentiate it from the WEIGHTED test scores T_{id}. Using both the weighted and unweighted (raw) test scores each for their specific different purposes faithfully replicates how the test scores are used in the college admissions context in Taiwan.

I use the weighted scores T_{id} to construct the running variable because the algorithm uses the weighted score to sort students into priority ordering for seats, and thus the use of the weighted scores faithfully reconstructs the cutscore and the relative distance to the cutscore that each student would have experienced while the algorithm was allocating seats. However, the weights themselves have no substantive or economic meaning outside the allocation process, because it was the raw scores that students were given on the JCEE. Indeed, the weights are specific to the departments, and so it would be hard to compare the weighted average peer quality in one department to that in another. Since subject-specific test scores t_i are the JCEE's measure of student performance, I use their unweighted total sum to construct the average peer quality for each academic department. The outcome measure $\overline{T_i^u}$ has only a subscript for the individual i, because the only realization of average peer quality we are considering for each student is the one she actually experienced. Thus, since $\overline{T_i^u}$ represents the average peer quality for the department that the student actually got into, we can think of the i in $\overline{T_i^u}$ as being a function of the specific department she actually got into. Because the same students are being compared in multiple counterfactuals, I follow Pop-Eleches and Urquiola (2012) and cluster the standard errors at the individual level.

Taiwan's college admissions selection rule conditions access to any given department on the relative distance a student's test scores are above or below that department's ex-post realized cutoff. Such a rule is an as-good-as-random process. Thus, it allows my analysis the advantage of circumventing biases stemming from correlations a student's performance on the JCEE may have with other variables, which may also predict of access to higher peer quality. For example, though we may be able to look at one specific student's performance on the JCEE and predict some likely level of college quality she would be eligible for, we could not WITH PRECISION predict her relative distance (or priority ordering) to the as-of-yet unrealized and unobserved cutoff for any given department. The realized distance of T_{id} from c_d (in either JCEE scores or rank scores) works as if it were arbitrarily set because it is a function of the ex-post realization of c_d, which also is due to factors beyond the control of any student or department. Such factors would include the number of and the test scores for all others applying to that department, that department's priorities for specific student qualities, and the capacity constraint set by the MOE for that department.

Now consider the reduced-form model (2): outcome Y_i is again a function of a student's performance on the college entrance exam. As before, the term $b(T_{id} - c_d)^p$ flexibly controls for other factors leading to admission, while ν_i is a random error term:

$$Y_i = \phi 1(T_{id} \geq c_d) + b(T_{id} - c_d)^p + \gamma_c + X_i + \nu_i \tag{2}$$

The coefficient on the term $1(T_{id} \geq c_1)$ is now ϕ. In the case that the outcome is log wages, ϕ represents the increase in log-earnings (log-points) i receives from scoring just above c_1 compared to just below.

3.2 Multiple Discontinuities

My analysis exploits discontinuities between 1140 departments. I follow Pop-Eleches and Urquiola (2012) and use a stacked RD set up. To my knowledge, aside from being the first paper to use a universe of applications within a contained college market, this paper is also the first to use a stacked RD in a returns to college analysis.

Admissions at each department is a separate RD, a self-contained natural experiment with its own LATE estimate. By stacking all the RDs, we can aggregate over all the LATEs to provide one estimate of the impact. Besides simplifying the task of interpreting the results, stacking RDs has the desirable characteristics of increasing statistical power by pooling observations. Stacking entails re-centering the running variable T_{id} for each student i who applied to department d by the minimum threshold c_d for admission to that department: $T_{id} - c_d$. When stacking RDs, it is common to include a fixed effect γ_c for each cutoff c_d to take into account that they occur at different points in the distribution of T. As a robustness check, I also experiment with other fixed-effect specifications.

Sufficient density around the admissions cutoff score is a statistical necessity for regression discontinuity. When this is violated the estimated jump at the discontinuity, even when large, will be imprecisely estimated and thus less likely to show up as statistically different than zero. Since I take the lowest score in each department as the ex-post realization of the cut-score for access at that department, this assumption is dependent on there not being serious measurement error defining the cutoff. After inspecting the data, I find that about 20% of the departments have a conspicuous outlier defining the lowest score; that is, for those departments, the distance between the lowest total weighted score and the second-lowest is significantly different from the distances observed between other test scores in that department. These outliers appear to be coding errors,[11] and measurement error they produce can cause attenuation bias, forcing my estimates toward zero.

To get a sense of whether these large gaps are coding errors or characteristic of the gaps observed for that department, I calculate the exact distance between each of the lowest 8 test scores in each of the departments in my data set. I chose the number 8 because the smallest department with in my sample has an enrollment of 9. As a summary measure, I find that 90% of the data is associated with a department that has a gap of less than 3 points between the lowest and next-to-lowest test scores for the last two admitted students. Thus, I visually inspected the 8 test-scores distances for any department that had a test-score distance great than 3. My general goal for inspection was to differentiate between two types: The first are cases where dropping an outlier observation would allow me to use the rest of the otherwise tightly packed data from the department. The second are cases where the test scores for the department as a whole are widely dispersed, and thus I would need to drop the department as a whole.

Figure 4 shows the average for the 8 distances for each of the 3 types of departments: trimmed (red-solid line), dropped (blue-dotted line), and unchanged (black-dashed line). For those departments were I decide to trim the lowest one or two cases, the average gap for the first distance was around 40 test points and just under 10 for the second. This characterized 221 departments with me dropping 289 cases, for an average of 1.3 cases per department. In the instances where I chose to drop more than one of the lowest observations, it was because the conspicuous gap occurred between the second and third observations, instead of the first and second. My general rule was to bring gaps between the lowest and second-lowest observations down to less than 10 points and to drop less than 5% of the data for each department. In practice it was not hard to make this

[11] It seems reasonable to expect that the test scores in most department to be tightly packed. First, students are allowed to list preferences for multiple departments, which mechanically insures that all departments will be oversubscribed even without there more fewer slots at college than students wanting them. Second, conditional on actually listing that department as a preference, students are admitted solely on the basis of their test scores being above everyone else's who also wanted to get in. Lastly, the test scores are measured at a 100^{th} of a decimal point, which indicates that the system is constructed to distinguish students 0.01 points away from each other. However, it could also seem reasonable that for departments that teach a college major that is not well-assessed by the JCEE, for example, one that teaches Music or Art , that students will have ordered their preference ranking of departments not necessarily to get into a program with student of equally high performance of the JCEE. For those departments we may expect them to have more dispersal in the test scores, in general.

distinction. As can be seen in the black-dashed line (flat at the bottom), most the departments are characterized by test scores distances of less than 1 point. For those departments that I chose to drop as a whole, the dispersal in test scores was so significant across the 8 distances I observed that in order to get the gap between the lowest and second-lowest to under 10 points, I would have to trim the lowest 5 students or about about 15-20% of the observations in the department. Thus, on average 80% of the data for those departments would be more than 50 points away from the cut scores, and thus would not have produced precise estimates. Since trimming such a large percentage was undesirable and because leaving them in would still have kept them so far from the discontinuity, dropping them all together made more sense. This entailed loosing only 669 observations, or less than 1% of my data.

Lastly, because re-centering and stacking the data multiple times aligns the discontinuities for each department at the cutoff, it increases the probability that there will always be an observation at T=0 and, therefore, mechanically increases the density of re-centered test scores at the cutscore. This artificial peak has nothing to do with students being able to manipulate which side of the cutoff they land on. Nevertheless, to allay any doubts, I follow the strict recommendations of Barreca et al. (2011), and I run a 'donut-RD' by dropping observations at T=0. Pop-Eleches and Urquiola (2012) also drop data at zero in their running variable.

4 Results

To summarize my results, scoring above the cutoff provides applicants with access to a higher average peer quality, around 2 test points worth or about 0.04 σ in peer quality. This quality improvement translates into a wage differential of around 0.8 log-points (on a 2-digit log base) or 0.04 σ in log earnings. Scaling the log earnings improvement by the corresponding improvement in peer quality indicates that a 1 standard deviation improvement in peer quality corresponds to around a 1 standard deviation improvement in log earnings. All results are conditional on being admitted to college at all. They are robust across multiple specifications.

4.1 College Quality and First-Year Earnings

My preferred estimation strategy starts with a wide bandwidth — 200 test points on either side of the cutoff, roughly equivalent to the full sample (see Panel B, Figure 2) — on the running variable and works incrementally inward, cross-validating the results on a range of specifications (Imbens and Lemieux, 2008). The basic model fits a fourth-order polynomial in the running variable on the wider bandwidths. As the bandwidth narrows, I follow Angrist and Pischke (2009)'s recommendation to switch from a polynomial to a linear specification in order to improve the fit.

In the first two subsections below I show the results for the sample including all college-major to verify that my results are consistently positive, significant, and meaningful. But in third and following subsections I report estimates of heterogeneity for a selected sample (though still containing most) of the college majors for which the first-stage and reduce-form results are stronger.

4.1.1 First-stage

Table 2 provides the first-stage results. Panel A shows that scoring just above the cutoff gives students access to a department with higher average peer quality, the improvement ranging across bandwidths and specifications from 1.8 to 2.3 test points higher. Comparing these RD results to those from a simple OLS regression, we can see the RD results are several orders of magnitude larger than the highly significant 0.018 OLS results. Most likely, the reason for the substantial difference between OLS and RD results is due to the fact that the RD results are picking up the inherent non-linearity that is built into the education system in Taiwan.

These tabular results match well what we see graphically in Figure 5. The 4 panels in Figure 5 give the smoothed fits of average department peer quality around the discontinuity. The upper panels are at a 30-point bandwidth and use local polynomial smoothed fits, while the lower ones are at a 10-point bandwidth and use local linear smoothed fits. The left panels are unconditional fits, while the right panels are conditional, meaning that they are residuals from a regression on a linear trend in the test scores, the full set of covariates, and the cutoff fixed effects. The discontinuity of the residuals in the graph (right panels) is equivalent to the 2-point estimates in the Table 2, but are about one-third to one-half the size of the discontinuities in the unconditional fits in the left panel. Though not shown here, when covariates are not included, the jump at the discontinuity is very similar to the 2-point jump we see in the right panels and in Table 2. This would indicate that most of the differences between the unconditional and conditional is attributable to the cutoff fixed effects to absorb the influence of the cutoffs occurring at different points in the range of possible weighted test scores.

To relate this improvement in terms of percentages, I convert all scores in the sample to a percentage of the total possible scores. That is, for most departments the highest possible total score for the required subjects is 500, except for departments of medicine which is 600. To use the percentage-out-of-possible score, I divide all achieved scores by either 500 or 600, depending on which college major they were admitted into. The first stage results in Table 2, Panel B are qualitatively similar in direction and significance. The average improvement in average peer quality is around 0.4 percentage point improvement, which with a standard deviation of 1.0 percentage points, that is again equivalent to 0.04 σ.

The results looks quite similar when cross-validated across different bandwidths. The results hover around 2 test-points and are statistically significant at well over 0.01 level of significance. Interestingly, these results look quite similar when we look at the average quality of peers at the whole college (Panel C), not just at the department. In general, students who get into better departments are also getting into colleges that have also better peers as a whole. When comparing quartic fits at wider bandwidths (Columns 1 and 2) to the linear fits at smaller bandwidths (Columns 3 and 4) we see that, unlike with the department peer quality, the linearly fitted results on tighter bandwidths are slightly smaller. However, this could be a case of the constant term varying slightly across the specifications. When I compare the percentage change in total possible test scores (Panel D), there does not appear to be much difference between the different fits and bandwidths.

Compared to some of the other papers in the RD literature, this 0.04 σ improvement in peer quality appears noticeably smaller. For example, Saavedra (2008) finds a 0.5 σ change in peer quality for those admitted to a top institution in Colombia versus not who just missed the admissions cutoff. Importantly, these results use a single discontinuity for admission to one institution. If I compare my first-stage results to those in a paper that is more similar to my stacked RD framework (Pop-Eleches and Urquiola, 2012), I find that my results are equivalent to their results, which range around 0.10 σ.[12] Most other RD papers in the literature report one LATE based on one discontinuity in admission for the whole school, while my analysis — like Pop-Eleches and Urquiola (2012)'s — is reporting an average of all LATEs from all 1140 discontinuities in my sample.

4.1.2 Reduced-form

Table 3 and Figure 6 shows that the monetary returns to improvements in college quality. As discussed in Section 2.3, the match rate for earnings on the 2000 admitted cohort is only 33%, low enough that there is not sufficient density in data at the discontinuity off which to identify an effect. (See Panel C in Table 3.) Had I one more year of earnings data (that is, earnings data for 2006), I believe a significant share of those currently missing would then appear. However, getting this data is impossible at this point. Since I cannot go forward in time with my earnings data, I go backwards. For each department in the 2000 sample, I take the average earnings for the 1999 and 1998 cohort as a proxy of expected earnings for the 2000 cohort had I been able to observe all their first-year earnings. Panels A and B provide the coefficients from these reduced-form estimates for the 1999 and 1998 cohorts. The results are quite similar, highly statistically significant and consistently positive across the range of bandwidths: a 0.4 to 0.6 log-point increase in earnings at the discontinuity for scoring above the threshold is equivalent to around 0.02 to 0.04 σ increase.

Compared to the OLS regression results of 0.00002, these marginal returns are again orders of magnitude larger. Further, once these log wage returns are scaled by the corresponding college quality returns, I can get an impact measure of 0.5 σ increase in log-earnings for a 1.0 σ increase in college quality. Thus, we could expect that increasing the quality of peers in the department where the student was admitted by 55 points (or 10 percentage-point increase in possible points) on the college entrance exam, then we would expect a 8.5 log-point increase in earnings. Because

[12] Hoekstra uses the probability of admission to a single institution as his first stage, thus is not an adequate comparison group.

I use an averaged measures, it could be that my measure of standard deviations is too narrow. In this case, if I were to use the standard deviation of the individual log wages (0.33) as my actually measure of the variance in earning potential, then this would indicate that actually impact is a 0.25 σ increase in log-earnings, which is still a 8.25 log-point increase. This is similar to elsewhere in the literature. For example, Dale and Krueger (2011) find that attending a school with 100-points higher average SAT score is associated with a 6.8 log-point increase in earnings. With a sample standard deviation in SAT scores at around 122, this would suggest a corresponding 8.3 increase in log-earnings.

4.2 Heterogeneity in selected samples

Here I inspect the heterogeneity in returns *across* college majors. To do so, in the selected sample, I drop all observations for college majors within three fields of study: education, services, and liberal arts. I drop the education majors because many graduates will go work for the public school system in Taiwan and thus will be government employees, not showing up in my wages data. I drop the services majors, again, because many of these majors are oriented towards careers in the public sector, like the military, police, transportation, and environmental protection. Lastly, I drop all liberal arts majors — like dance or sculpture or literature — because both students' preference rankings for department and potential employers' desires for these graduates seem less correlated with performance on academic-type tests. This leaves a set of selected college majors associated with business & law, social sciences, natural sciences, and medicine.

Indeed, when I drop these 3 fields, the magnitudes of the reduced-form results improve. Looking in at Columns 1-4 in Panel A of Table 4, we see the effect on the full sample results of dropping each of the 3 fields in sequential order. Interestingly, dropping any and all of the 3 fields has almost little impact in the full sample. However, when we look at a more narrow bandwidth (Columns 5-8), we see that dropping each of the fields causes an increase in peer quality by about 6% for each field dropped. We can see in Figure 7 that when comparing the differences in impacts between the full set of college major (left panels) and the selected set (right panels) which drops all three fields simultaneously, the change in peer quality is mimimal.[13]

When we consider the differences in wage returns, the relative differences are starker; they more than double. This doubling of the impact is immediately obvious when comparing left (full set) and right (selected set) panels in Figure 8. When we observe the sequential change in Table 4, Panel B. Both at the 200 bandwidth and the 20 bandwidth, we see that dropping the education majors immediately doubles the returns. Given the point estimates and tight confidence intervals in Columns 9 and 10, we can argue that these estimates are statistically significantly different from each other. However, for the estimates in Columns 13 and 14 for the narrow bandwidth, we can say that the one point estimate is outside the confidence interval of the other; but because their confidence intervals overlap, we cannot reject that they share some point estimate in between them. After dropping the liberal arts and services fields, there is less improvement in the log-returns. This pattern in the difference between all-fields versus selected-fields follows what we might expect given that at least two of the three fields (education and services) have high rates of public-sector employment. Those who could not find employment in the public sector may be of lower quality and thus are relegated to working for lower paying jobs in the private sector.

4.2.1 Top-half vs. bottom-half in department quality

To further inspect the heterogeneity in returns to quality, I now turn to variation *within* college majors. If we divide our analysis into considering discontinuities for departments at the top-half of the quality distribution versus the bottom-half within each college major. Specifically, these regressions consider the quality improvements and log-returns to moving up the quality distribution but only within a departments relative position within the upper and lower halves. Therefore, if a department is at the lower end of the upper half, only the discontinuities for departments above it are being considered. But if it were on the upper end of the lower half, then it would be as if this were the highest-quality department for those regressions, and all discontinuities would be for those below it. Table 5A provides the first-stage and Table 5B the reduced-form results.

Focusing on the first-stage results in the Panels A and B of Table 5A for the top and bottom

[13] Though not shown here, the impact on peer quality at the department is also minimal.

halves of the quality distribution, we see that the sample sizes between the upper and lower halves are relatively balanced. The difference between upper and lower-quality departments (Columns 1 and 2 vs. 3 and 4) in departmental quality ranges from about a 20% to a 33% relative improvement in peer quality on average, depending on the bandwidth. When considering quality of the whole college, we see a qualitative similar picture, with slightly smaller relative increases. These are sizable relative improvements, but seem less meaningful if converted into sample its 0.01 σ. Yet, thinking in terms of scaling up by a small number, the resulting overall impact may still be considerable.

When we compare — upper versus lower halves — the wage differentials for going to a marginally better quality department, we see in Table 5A, Panels C (Columns 5 and 6) and D (Columns 13 and 14) that when using the full bandwidth sample, there is no difference. Unexpectedly, when we look at the narrow bandwidth sample, we find that actually the variation within the lower quality is greater than that in the upper-half of the distribution. But because the confidence intervals overlap, we cannot say that these estimates are significantly different from each other. However, when we look at Figure 10, the difference in the size of the jump at the discontinuity is noticeably larger for the bottom half (left) panel. Even if the difference in log wages is not statistically different in this division of the sample, the actual differences in peer quality between upper and lower halves, which enters a Wald estimate in the denominator, can scale up the overall impact estimate of peer quality on log earnings, which enters in the numerator.

Wages are not the only means by which monetary benefits may be transferred to students (and their families). Liu et al. (2006) show that national universities in Taiwan, like elsewhere in the Global South and East (Rubinstein and Sekhri, 2011), are publicly funded; thus, their tuition is considerably lower than private ones, in some cases as much as half. Using MOE records from 2000 for the tuition it allowed departments to charge at each institution, I estimate the differential benefit in reduced tuition for scoring just above the access threshold. The yearly tuition are transformed into logs; consequently, the coefficient of interest can be interpreted as the expected log-point *decrease* in log yearly tuition for scoring above the access threshold.

Table 5B, Panels A (Columns 1-2) and B (Columns 5-6) report the differential impact students could expect to experience between being in the upper half of the quality spectrum and the lower half. For either wide or narrow bandwidths, students in the upper half of the quality distribution can expect to consistently increase their tuition subsidy at twice the rate that students in the lower half can. Independent of any wage earnings benefit to scoring above the cutoff, this 0.06 σ reduction in tuition likely motivated parents and students to compete hard on admissions exams no matter which half of the distribution the child strove to be at. But considering that the benefits to be in the upper half are twice as high, this is a strong incentive.

4.2.2 Prestige effects: National vs Not-National Colleges

The obvious channel of these differential subsidies comes from departments being in either national (public) or non-national (private) college. This is by construction of the Taiwanese system, in which the only form of financial aid is subsidized tuition at public institutions only.[14] However, the evidence is also clear in Table 5A, Panels A and B. Student in the top half of the quality distribution are twice as likely to end up in a national university by scoring above the admissions threshold, compared to those in the bottom half. They are at least 9 times more likely to end up in one of the elite top-6 institutions (which are all national universities) by scoring above the threshold versus not.

Even though these relative differences are large, overall it is not clear whether a 3.1 percentage points increase in national college admittance signals that this is a rare or overly common event. In both cases the average impact would be small. Such an ambiguity motivates me now to divide the samples within college major explicitly along the lines of national college distinction. Thus, in each of my regressions the comparison of being above or below the cutoff is now restricted to being above or below the cutoff for admittance to the next higher-quality national college for the national subsample and next highest not-national for the not-national subsample. To be able to construct

[14]So far I have refrained from using the terms public versus private to distinguish national versus not-national. A main reason for this is that in most Western contexts the exact opposite connotations for prestige are being communicated. In both Western and non-Western contexts, public institutions are more highly subsidized. But in the West, this has meant that public colleges are for the average population of college-goers and private colleges are more elite and reserved for those privileged with greater financial resources or higher ability or both. Thus in the West, the wealth-transfer through tuition subsidy has been towards helping the less advantaged. The reverse is true in most non-Western cases, because on the meritocratic ideals of giving the most resources to those who, it is argued, use them the most ably.

such a comparison, it required that I work only with college majors that had departments for at least two national colleges and two not-national colleges. This restriction required that I drop 19 college majors, most of which contained 4 or fewer departments. Consequently 35 departments are left and none of the fields from the selected set of college majors dropped out completely. Even after dropping these 19 college majors, only about 3% of the observations in the selected sample was lost. This gives evidence that not only were these 19 college majors few in departments but also small in department size.

Using this new division of data, it becomes apparent that about half of the peer quality gains in Table 5B for the bottom half of the departments was due to lower quality national colleges being grouped with the predominantly not-national colleges at the bottom. Both at the department and college levels, this characterization of average peer quality is true. The difference in the potential to improve peer quality between national and not-national colleges is visually compelling in Figure 11. Within each college major, scoring above the threshold to the next level not-national college provides access to peer quality that is half the improvement in national colleges. The same can be said about log wages. Figure 12 shows that getting into the next better national college has twice the impact on log earnings as getting into the next better not-national college. Even with the noisier data for the national subsample (left panel Figure 12), Table 6 shows that the returns to getting into a better college are statistically different estimates for the national and not-national samples. Interestingly, now that the comparison is within national college distinction, going to a better national college actually increases log tuition, but going to a better not-national college is actually cheaper. It seems reasonable that 5 times higher returns to log earnings in the first year could repay the previous 4 years of marginally higher tuition.

If one looks primarily at the 200-bandwidth estimates, it appears from the corresponding magnitudes that improved peer quality, then, is driving the improvements in log wages. Two reasons lead me to hold out on this interpretation. First, when we focus on 20-bandwidth, we see that estimates on peer quality at the department go up for both the national and not-national subsamples, especially for the later by about 33%. But looking at log wages at the 20-bandwidth, it goes down in size from the full bandwidth by 50% in the not-national but only 25% in the national. Now the log wages improvement is nearly 4 times larger than the corresponding estimate in the not-national. One explanation be compositional changes due to how the data disperses over the running variable differently for the national and not subsamples. The relative change in sample sizes for going from 200 and 20 bandwidths diminishes the national sample size drastically more than the not-national. Yet, its the coefficient on not-national that changes more. Possibly when we decrease the bandwidths, some departments drop out more readily in the not-national subsample relative to national.

The second reason for my holding out of declaring this a peer quality story is that we are still comparing within national and not-national groups, not across them exclusively. In order to do this, I need only to set up a simpler stacked RD design in which all the student within each college major are centered around a single discontinuity which marks the cutoff for access to the lowest quality national college in that department. As long as one national college offers that college major, the data for that college major is included in the analysis. Since only one counterfactual per college major is being considered in this part of the analysis, I do not use the data from the replication, only the smaller original dataset. Because the sample size of the original data set, I prefer a linear specification on a slighter wider bandwidth, but show 10 points on either side of the 20-point bandwidth from before. Encouragingly, these results, shown in Table 7 and Figures 13 and 14, look remarkably like what we saw in Table 6 and Figures 11 and 12 for the replicated data. Getting into a national college is associated with 2 to 3 test-point improvement in average peer quality and about 3 log-point increase in monthly earnings.

5 Conclusion

Estimating the returns to a college education is a task common to both the general public and the research community alike. Families and governments around the world must decide how to allocate resources on behalf of their children and their citizens' education, with the hope that such educational investments will indeed payoff in the labor market and citizenry. As access to college has expanded around the world, it may not be enough to get just a college degree. To be competitive now, increasing numbers of students vie for seats at higher quality institutions.

These thoughts motivate my estimation of the returns to college quality. My paper's first contribution is to use data from Taiwan to overcome previous limitations to internal and external validity. As a context, Taiwan provides a highly centralized and very competitive market for students who are seeking admission to top schools, though not competitive for the universities itself. As a data source, Taiwan highly centralized government has been able to keep detailed and complete records of performance on college entrance exams, college admissions outcomes, and even earnings in labor market, thus allowing for researchers like Professor Jin-Tan Liu to merge them all together to great advantage. These factors and data provide a useful context in which college quality is finely sorted and its effects consistently measured.

This paper's second contribution is methodological. A stacked regression discontinuity design maps well onto the details of the student matching algorithm used in Taiwan, which sorts students by their performance on a central exam and matches their admissions preferences to seats in specific college departments. This research design allows me to efficiently summarize the impact of scoring above (below) an admission cutoff at any point in the entrance exam distribution, and then to relate that admissions outcome to the reduced-form impact in first-year earnings in the labor market.

Taken together, the analysis above presents results that are robust to a wide range of specifications and robustness checks. Despite the expansion of higher education, these positive and significant wage differentials indicate that there still is a college quality premium in the labor force. These positive wage returns to higher college quality would suggest that being the last admitted into a better college-department is (at least marginally) preferable to being the first admitted into the next-best program within the same major. In order for this to be generally true, it would suggest that employers in Taiwan are more sensitive to prestige signals than they are to other signals of college performance. In the estimates shown in Table 7 and Figures 13 and 14, there does appear to be evidence that the monetary returns to getting into a more elite national college is well above what could be explained by the change in peer quality that occurs when moving up from a lower quality (private) college to the next closest national college. Since this part of the analysis is on a selected sample of college majors, the higher returns estimated here must reflect both a returns to college quality as well as the differential returns to college majors.

The prestige effects may not necessarily represent or be dependent on greater current learning occurring in national colleges. Lee (2007) argues that there may be two different effects being identified in Asian versus Western education. Whereas in the US context students typically create their most important quality signals while in college (namely, GPAs, awards, recommendation letters, internships, as well as alma mater, to name a few), Asian students do most of their quality signaling prior to entering college, or in fact by the college they enroll into. Following the train of thought laid out by Lee, if most of Asian education is test-centric and the most important test is the college entrance exam, it could be that Asian students have less extrinsic incentive to put in effort after the most important signal has been sent. If this were true, we would expect the apex of their educational effort and output to occur in their senior year of high school.

If we believe that one of the main channels of benefit is exposure to better peers, this would also suggest that being the last (weakest) person admitted into a department is better than being the first (top) in the same department. In order for us to test this, we would want to estimate not only the change in levels (jump) at the discontinuity which is interpretable, but also the change in slope (kink). Even without a thorough-going analysis using a regression kink design, a cursory inspection of the slopes in the left panels of Figures 5 indicates that there is no significant kink in the slopes. Thus, I rely solely on the jump at the discontinuity to compare marginally similar students in different departments, instead of different students in the same department.

I suspect that the (jump) estimates I provide are lower bounds on the actual returns for two main reasons: First, I lack data for government workers who have a larger share of college graduates. Since the competition for these higher-paying jobs is intense and again exam-based, it would logically follow that college graduates scored well on the entrance exams will also perform well here. Second, a larger portion of those missing from my earnings data are males, which earn more than females, according to previously literature of gender wage gaps in Taiwan (Tsai et al., 2009; Vere, 2005). If I were to have the Labor data for 2006, I believe that a significant share of these males would be showing up in my earnings data. Unfortunately, as I mentioned above, accessing the Labor data for 2006 is politically impossible.

Increased first-year earnings is not the only benefit to higher college quality. Indeed, another monetary return is reduced tuition. The 0.06 σ reduction in tuition for attending a higher quality college is a non-negligible benefit. This is especially so when we consider the benefit lasted for four years of schooling. Again, the psychic benefits (to family) of attending a higher quality institution are important, especially when considering the reputation value of more prestigious ones. A full cost-benefit or welfare accounting would be useful, but are beyond the scope of this paper. Among other things, I would have to also account for the amount of money parents have spent on extra tutoring and/or post-high school cram schools as well as the foregone leisure students invested invested in studying in order to get into these higher quality schools. Given the intensity of the investments made by the family, a fair welfare analysis would also want to acknowledge the psychic benefits and costs. Since I do not have access to this information, I present these estimates of the savings on tuition to complement the wage differentials and as an acknowledgment of the overall welfare benefits.

The results in this paper may be specific to centralized systems like the one in Taiwan; nevertheless, such systems are common throughout the developing world where public higher education is rationed and thus very competitive. Moreover, these results provide a basis to further inspect the mechanisms that actually generate the positive relationship between college quality and earnings in such contexts.

References

Abdulkadiroğlu, Atila and Tayfun Sönmez, "School Choice: A Mechanism Design Approach," *American Economic Review*, Jun 2003, *93* (3), 729–747.

Aghion, Philippe, Mathias Dewatripont, Caroline Hoxby, Andreu Mas-Colell, and André Sapir, "The Governance and Performance of Universities: Evidence from Europe and the US," *Economic Policy*, 2010, *25* (61), 7–59.

Angrist, Joshua D. and Jorn-Steffen Pischke, *Mostly Harmless Econometrics: An Empiricist's Companion*, Princeton University Press, 2009.

Arcidiacono, Peter, "Ability Sorting and the Returns to College Major," *Journal of Econometrics*, Jul 2004, *121* (1-2), 343–375.

Barreca, Alan I., Melanie Guldi, Jason M. Lindo, and Glen R. Waddell, "Saving Babies? Revisiting the Effect of Very Low Birth Weight Classification," *Quarterly Journal of Economics*, Nov 2011, *126* (4), 2117–2123.

Becker, Gary S., William H. J. Hubbard, and Kevin M. Murphy, "Explaining the Worldwide Boom in Higher Education of Women," *Journal of Human Capital*, 2010, *4* (3), pp. 203–241.

Black, Dan A. and Jeffrey A. Smith, "Estimating the Returns to College Quality with Multiple Proxies for Quality," *Journal of Labor Economics*, 2006, *24* (3).

Carrell, Scott E., Bruce I. Sacerdote, and James E. West, "From Natural Variation to Optimal Policy? The Lucas Critique Meets Peer Effects," *National Bureau of Economic Research Working Paper Series*, 2011, *No. 16865.*

Chevalier, Arnaud and Gavan Conlon, "Does It Pay to Attend a Prestigious University?," *Institute for the Study of Labor (IZA) Working Paper*, 2003, *No. 848.*

Chou, Shin-Yi, Jin-Tan Liu, and James K. Hammitt, "National Health Insurance and Precautionary Saving: Evidence from Taiwan," *Journal of Public Economics*, Sep 2003, *87* (9-10), 1873–1894.

_ , _ , Michael Grossman, and Ted Joyce, "Parental Education and Child Health: Evidence from a Natural Experiment in Taiwan," *American Economic Journal: Applied Economics*, Jan 2010, *2* (1), 33–61.

Clark, Damon, "Selective Schools and Academic Achievement," *Institute for the Study of Labor (IZA) Working Paper*, 2007, *3182.*

Clark, Diana E. and Chang-Tai Hsieh, "Schooling and Labor Market Impact of the 1968 Nine-Year Education Program in Taiwan," Working Papers 215, Princeton University, Woodrow Wilson School of Public and International Affairs, Research Program in Development Studies. June 2000.

Dale, Stacy and Alan B. Krueger, "Estimating the Payoff to Attending a More Selective College: An Application of Selection on Observables and Unobservables," *Quarterly Journal of Economics*, Nov 2002, *117* (4), 1491–1527.

_ and _ , "Estimating the Return to College Selectivity over the Career Using Administrative Earnings Data," *National Bureau of Economic Research Working Paper Series*, Jun 2011, *17159.*

Goastellec, Gaële, "Changes in Access to Higher Education: From Worldwide Constraints to Common Patterns of Reform?," May 2008, *International Perspectives on Education and Society* (9), 1–26.

Hanushek, Eric A. and Ludger Woessmann, "Do Better Schools Lead to More Growth? Cognitive Skills, Economic Outcomes, and Causation," *Journal of Economic Growth*, Jul 2012, pp. 1–55.

Hoekstra, Mark, "The Effect of Attending the Flagship State University on Earnings: A Discontinuity-Based Approach," *Review of Economics and Statistics*, 2009, *91* (4), 717–724.

Imbens, Guido W. and Thomas Lemieux, "Regression Discontinuity Designs: A Guide to Practice," *Journal of Econometrics*, Feb 2008, *142* (2), 615–635.

Jackson, C. Kirabo, "Do Students Benefit from Attending Better Schools? Evidence from Rule-based Student Assignments in Trinidad and Tobago," *The Economic Journal*, Dec 2010, *120* (549), 1399–1429.

Lange, Fabian, Robert Topel, Eric Hanushek, and F. Welch, "Chapter 8 The Social Value of Education and Human Capital," in "Handbook of the Economics of Education," Vol. 1, Elsevier, 2006, pp. 459–509.

Lee, David S and Thomas Lemieux, "Regression Discontinuity Designs in Economics," *Journal of Economic Literature*, Jun 2010, *48* (2), 281–355.

Lee, Sanghoon, "The Timing of Signaling: To Study in High School or in College?," *International Economic Review*, 2007, *48* (3), 785–807.

Liu, Jin-Tan, Shin-Yi Chou, and Jin-Long Liu, "Asymmetries in progression in higher education in Taiwan: Parental education and income effects," *Economics of Education Review*, 2006, *25* (6), 647–658.

Lucas, Robert E., "On the Mechanics of Economic Development," *Journal of Monetary Economics*, July 1988, *22* (1), 3–42.

MacLeod, W. Bentley and Miguel Urquiola, "Anti-Lemons: School Reputation, Relative Diversity, and Educational Quality," 2011.

Oreopoulos, Philip, Till von Wachter, and Andrew Heisz, "The Short- and Long-Term Career Effects of Graduating in a Recession," *American Economic Journal: Applied Economics*, Jan 2012, *4* (1), 1–29.

Pop-Eleches, Cristian and Miguel Urquiola, "Going to a Better School: Effects and Behavioral Responses," *American Economic Review*, 2012.

Rubinstein, Yona and Sheetal Sekhri, "Do Public Colleges in Developing Countries Provide Better Education than Private ones? Evidence from General Education Sector in India," 2011, (0130).

Saavedra, Juan E., "The Returns to College Quality: A Regression Discontinuity Analysis," 2008, (Unpublished Manuscript).

Spohr, Chris A., "Formal Schooling and Workforce Participation in a Rapidly Developing Economy: Evidence from "Compulsory" Junior High School in Taiwan," *Journal of Development Economics*, Apr 2003, *70* (2), 291–327.

Tsai, Wehn-Jyuan, Jin-Tan Liu, Shin-Yi Chou, and Michael Grossman, "Intergeneration Transfer of Human Capital: Results from a Natural Experiment in Taiwan," *National Bureau of Economic Research Working Paper Series*, 2011, (No.16876).

_, _, _, and Robert Thornton, "Does Educational Expansion Encourage Female Workforce Participation? A Study of the 1968 Reform in Taiwan," *Economics of Education Review*, 2009, *28* (6), 750 – 758.

Van der Klaauw, Wilbert, "Estimating the Effect of Financial Aid Offers on College Enrollment: A Regression–Discontinuity Approach," *International Economic Review*, 2002, *43* (4), 1249–1287.

Vere, James P., "Education, Development, and Wage Inequality: The Case of Taiwan," *Economic Development and Cultural Change*, April 2005, *53* (3), 711–735.

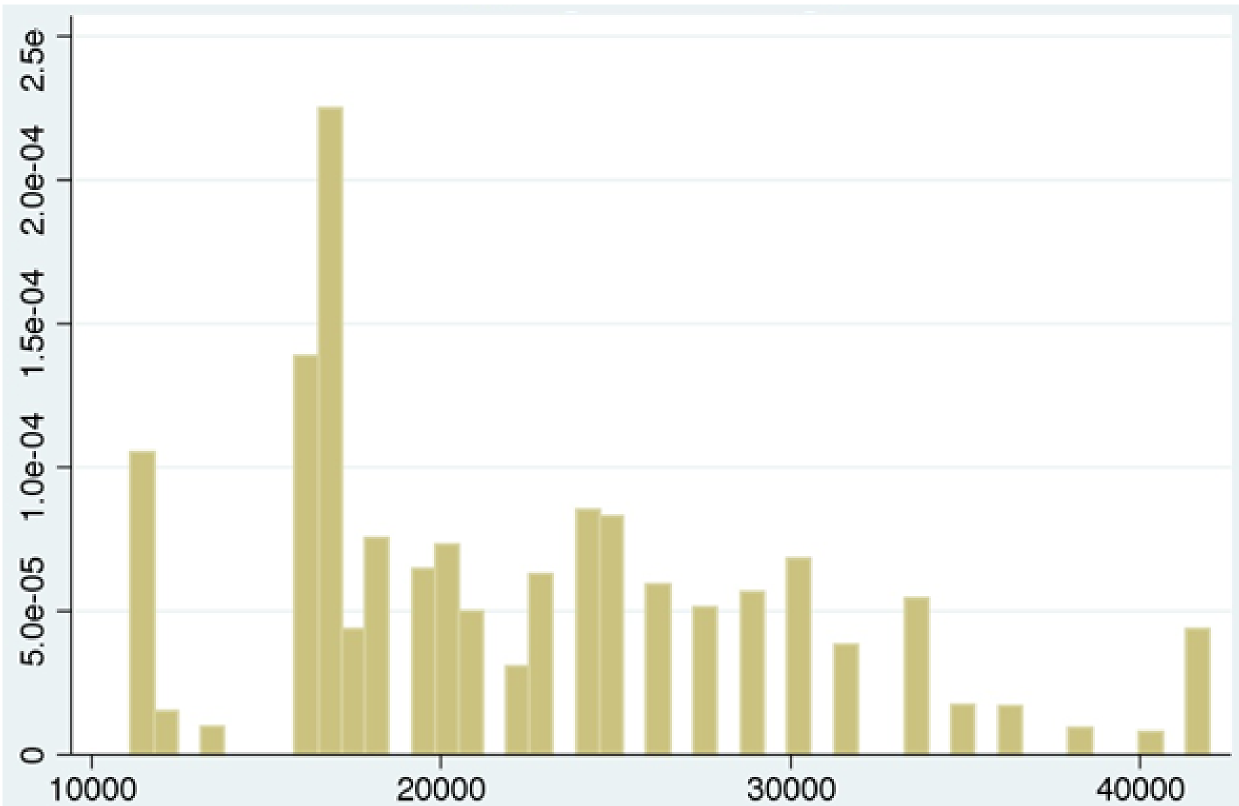
Winston, Gordon C., "Subsidies, Hierarchy and Peers: The Awkward Economics of Higher Education," *Journal of Economic Perspectives*, Jan 1999, *13* (1), 13–36.

Wiske-Dillon, Eleanor and Jeffrey Smith, "Determinants of Mismatch between Students and Colleges," 2012, (Unpublished Manuscript).

Yang, Chih-Hai, Chun-Hung A. Lin, and Chien-Ru Lin, "Dynamics of Rate of Returns for Postgraduate Education in Taiwan: The Impact of Higher Education Expansion," *Asia Pacific Education Review*, Nov 2010, *12* (3), 359–371.

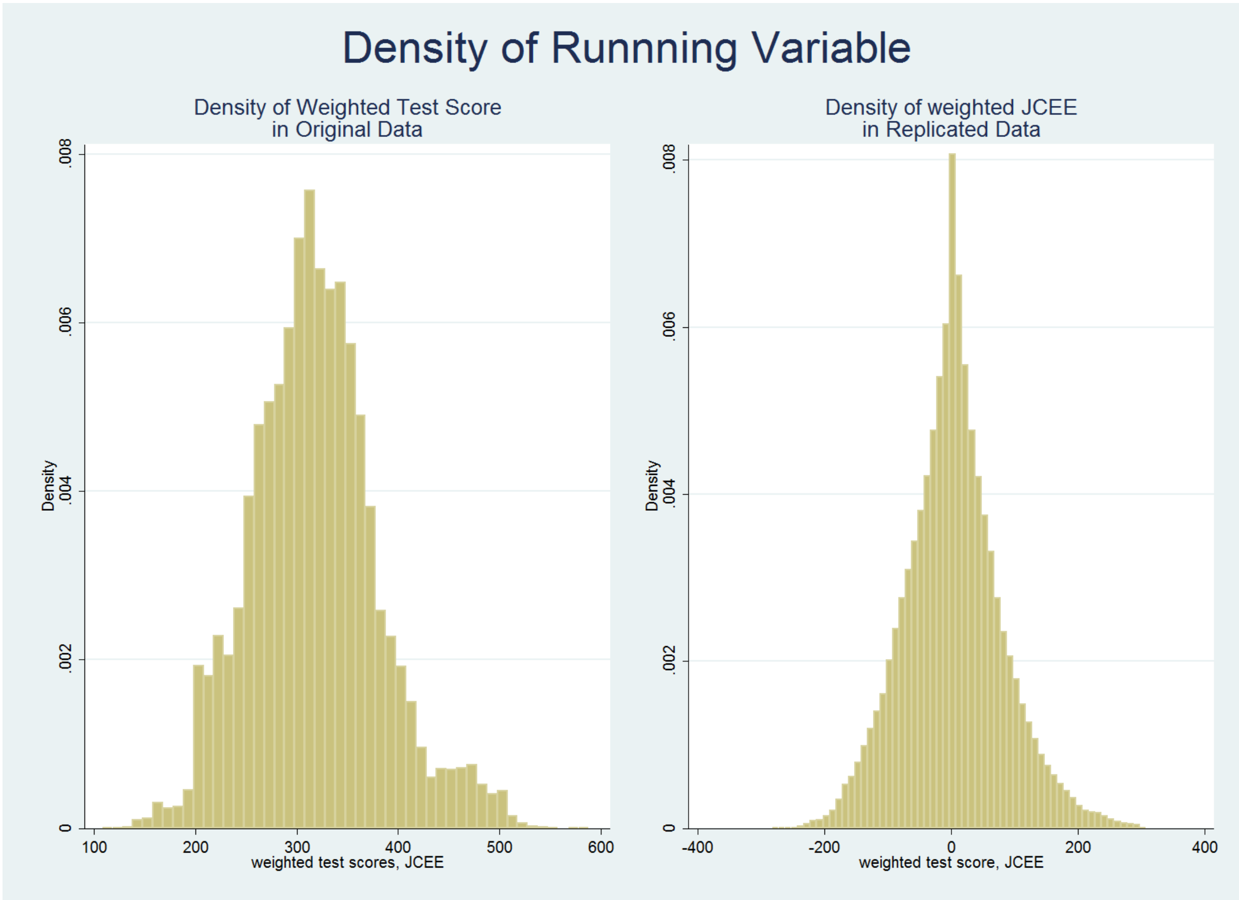
Figures

Figure 1: Histogram of wages



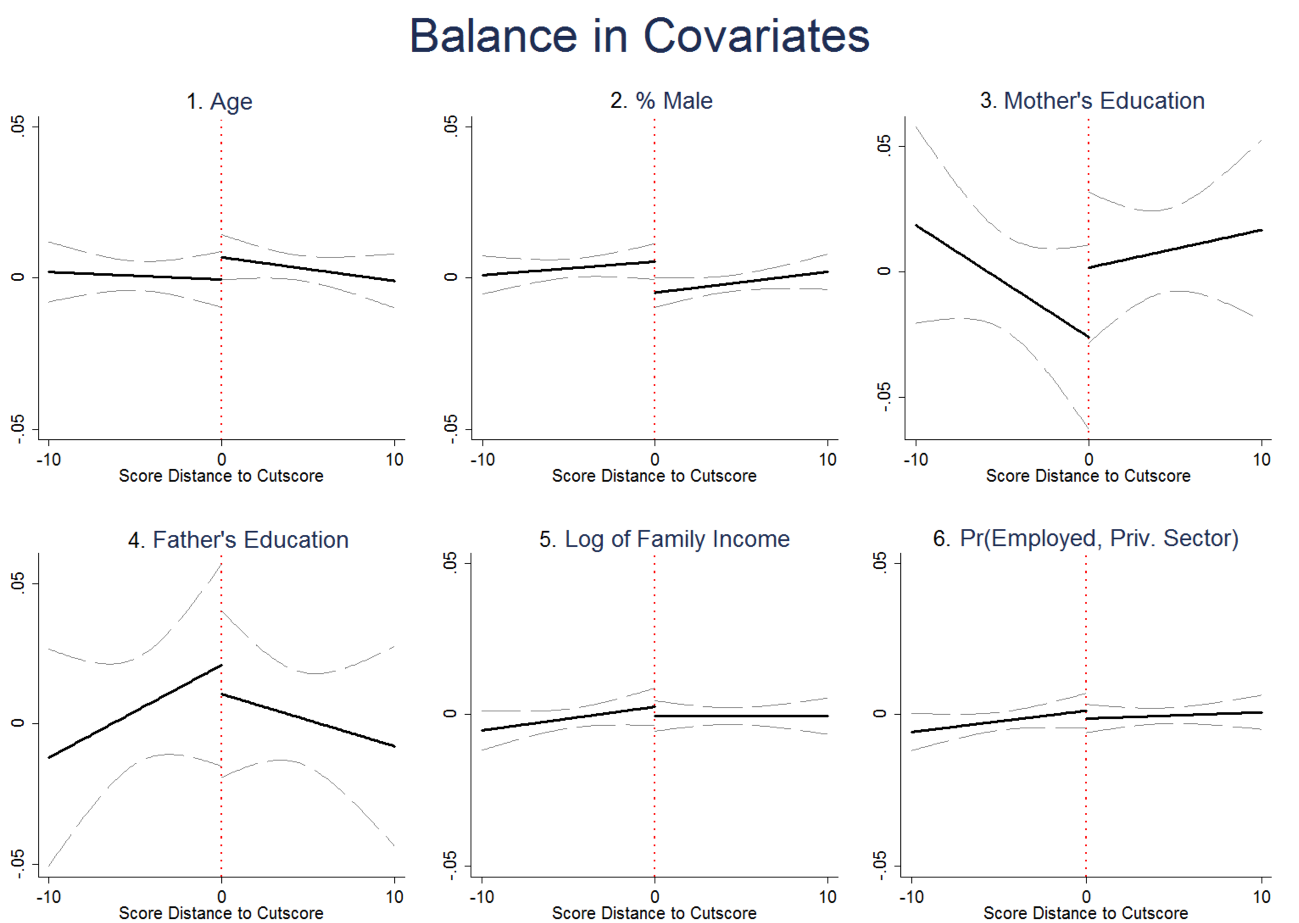
Notes: These represent the first year monthly earnings (in New Taiwan Dollars) for those entering college in 2000. The data were recorded in the month of December for 2004 and 2005. The data are from the Labor Insurance scheme that Taiwan follows in which an individual's premium is a function of his/her monthly salary. Thus, because of the intervals in the menu pricing, the wage records are also lumpy and truncated to a range between 11,000 NTD to 42,000 NTD.

Figure 2: Density plot of running variables in the original (left) and replicated (right) data sets



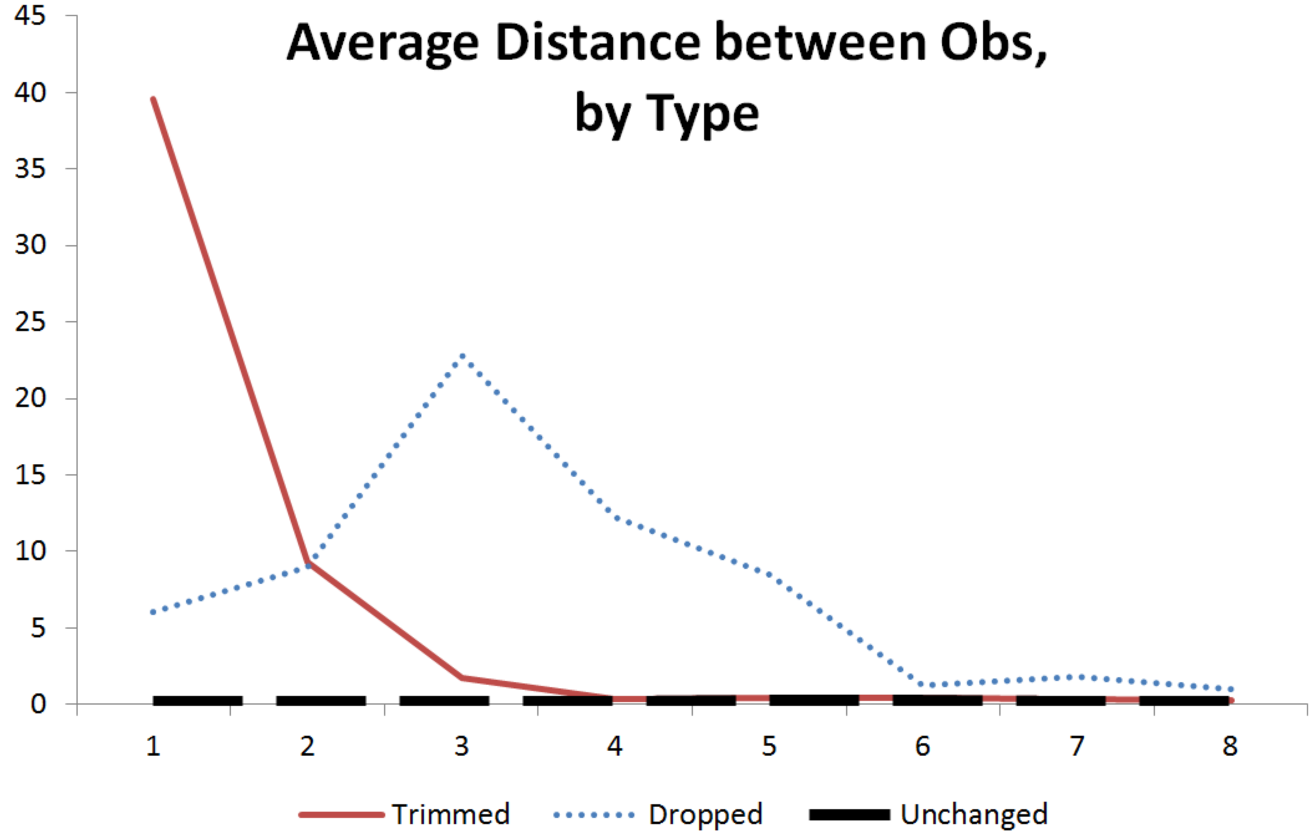
Notes: Density plot in the left panel represents the distribution of realized weighted JCEE scores in 10 test-point bins. Density plot in the right panel represents the distribution of simulated weighted JCEE scores in 10 test-point bins Both appear to be fairly approximating a normal distribution.

Figure 3: Balance in the Covariates



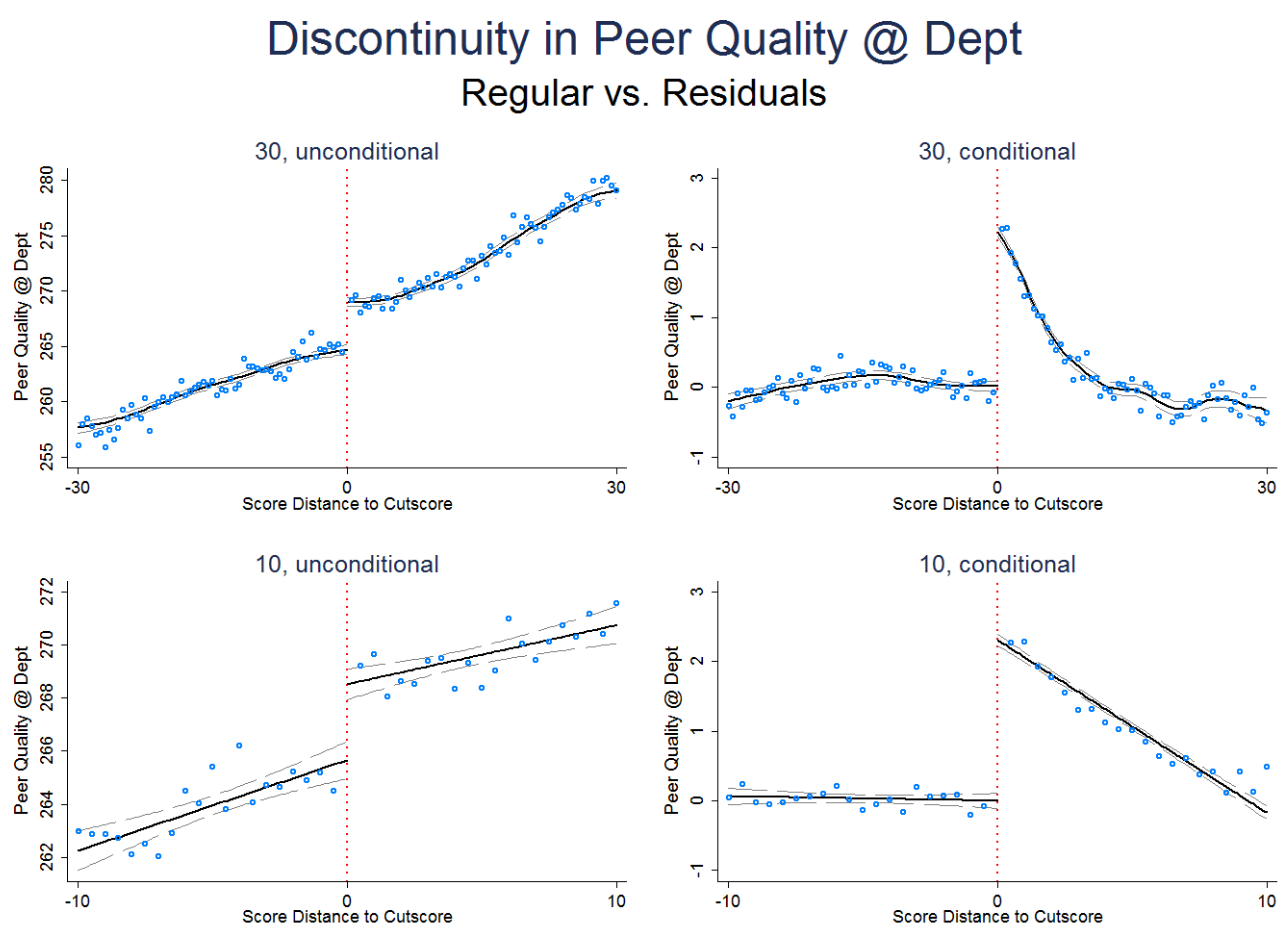
Notes: The plots above present the linearly smoothed fits of residuals from a regression of each of the covariates on to a model including a linear trend in re-centered test scores, the other covariates, and cutoff fixed effects.

Figure 4: Graphic motivation for cleaning the data



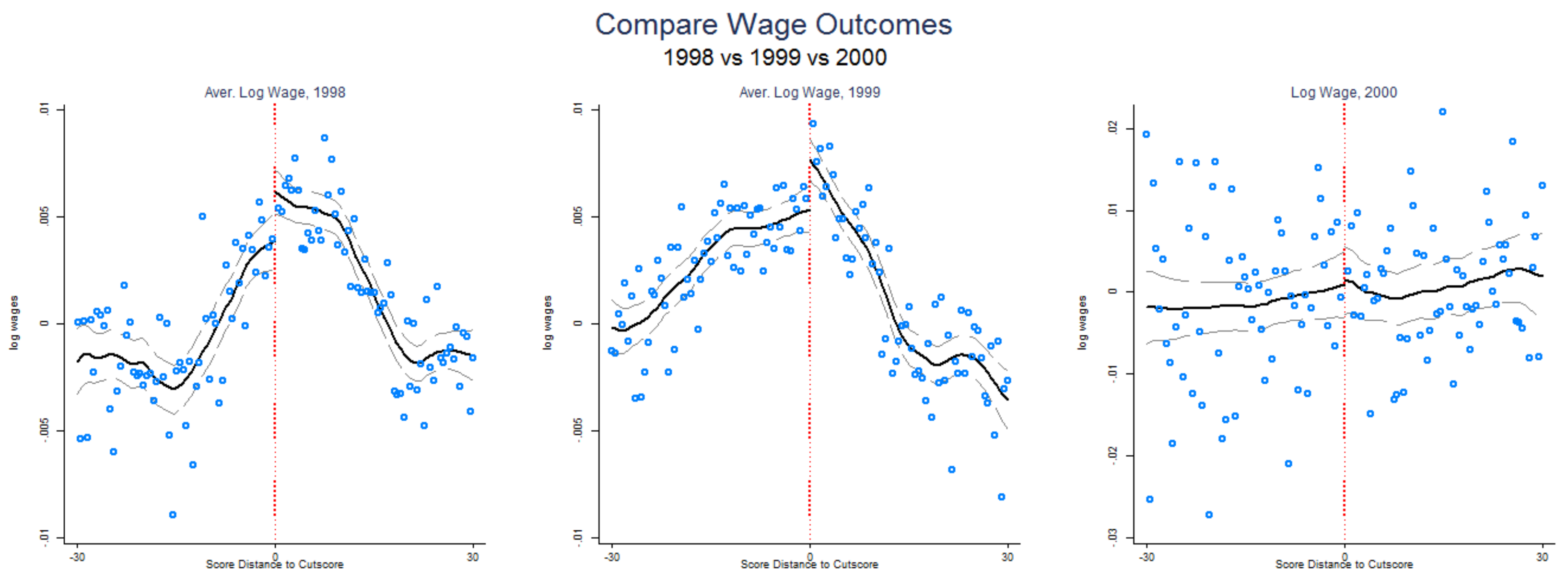
Notes: The figure above shows by type of department the average gap (or distance) between the lowest 9 data points in each department, with the 1 representing the distance between the lowest and next-to-lowest points. Only departments that had a gap of greater than 3 test points for the these 8 distances were included, because 90% of the data was shown to be associated with a department to had gaps less than 3 points. "Trimmed" (solid red line) represents the departments for which the gaps were mainly associated with extreme outliers at the very bottom, thus those observations for trimmed. "Dropped" (dotted blue line) represents departments that had gaps consistently throughout the first 8 distances, and thus the whole departments data was dropped. "Unchanged" (thick dashed black line) represents all other departments, most of which had gaps of less than 1 point.

Figure 5: First-stage results in college quality



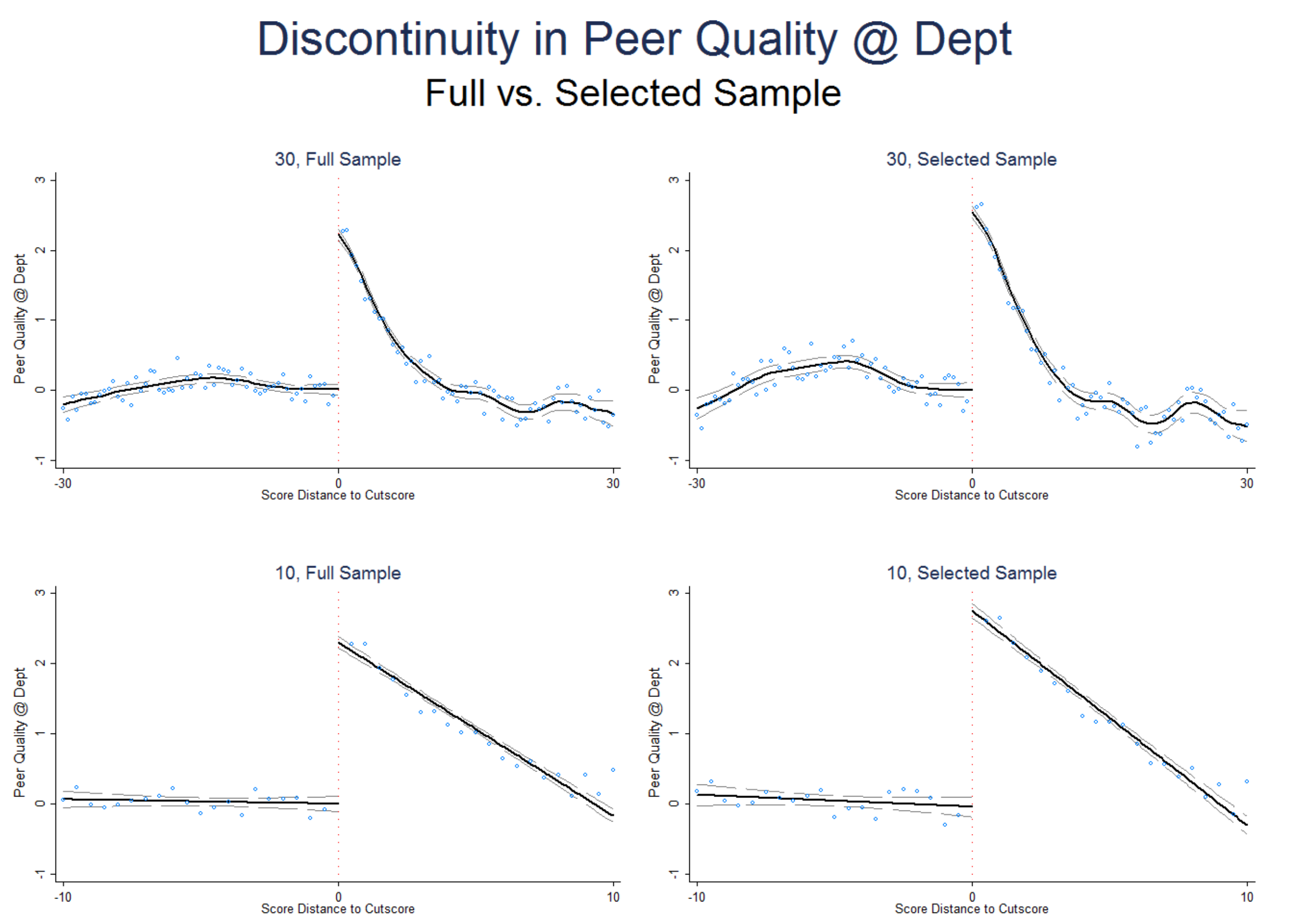
Notes: The 4 panels above give the smoothed fits of average department peer quality around the discontinuity. The upper panels are at a 30-point bandwidth and use local polynomial smoothed fits, while the lower ones are at a 10-point bandwidth and use local linear smoothed fits. The left panels are unconditional fits, while the right panels are conditional, meaning that they are residuals from a regression on a linear trend in the test scores, the full set of covariates, and the cutoff fixed effects.

Figure 6: Comparing measures of log wage outcomes.



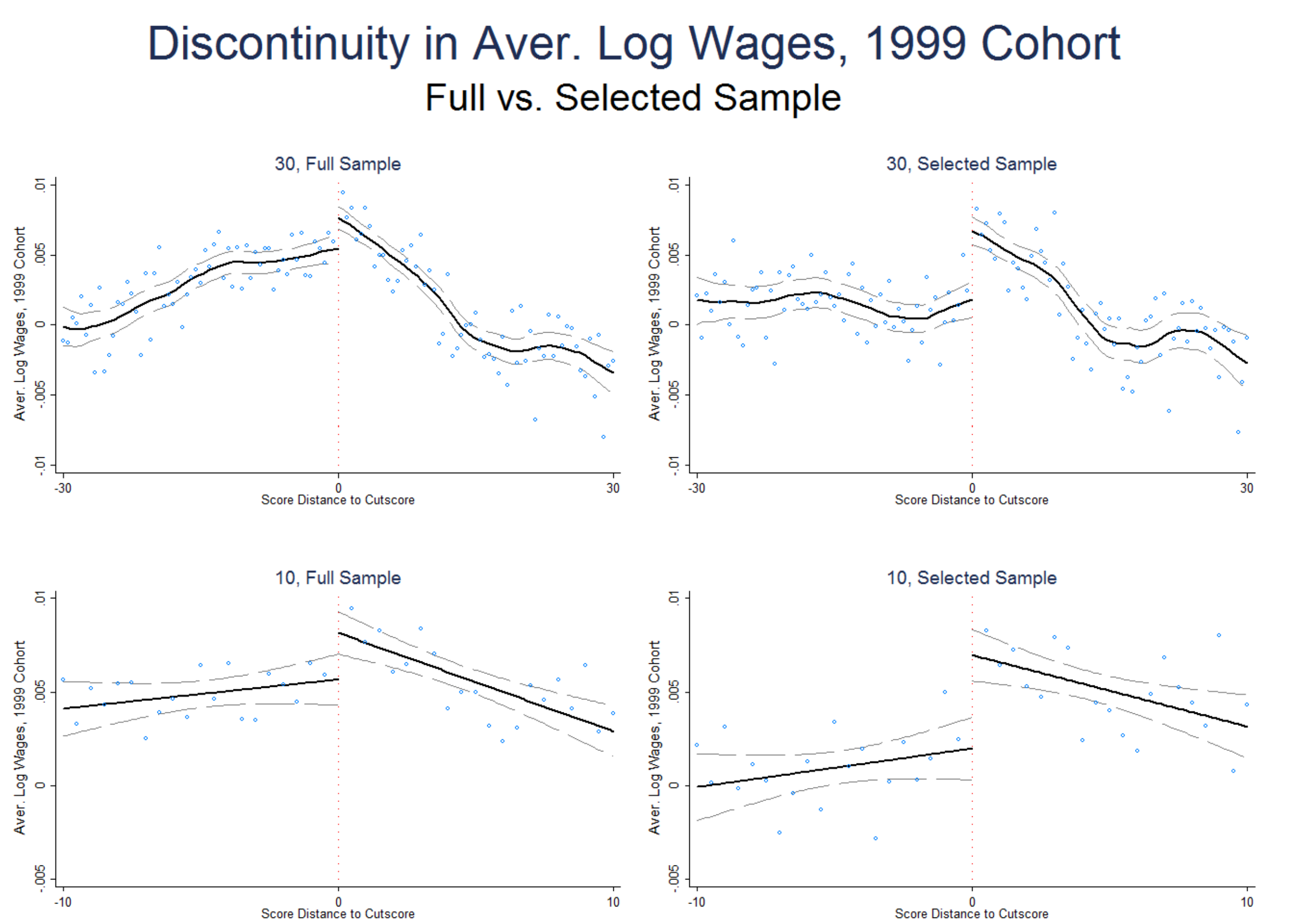
Notes: The 3 panels above give the smoothed fits of average log wages around the discontinuity. All use residuals from a regression on a linear trend in the test scores, the full set of covariates, and the cutoff fixed effects, and fit these with a local polynomial smooth and a 30-point bandwidth. Panel A gives the discontinuity using log of average wage for the 1998 cohort for each department. That is, individuals admitted into a particular department during the 2000 cohort are assigned the average log earning for those admitted in that same department during the 1998 cohort. Panel B gives the similar results for the 1999 cohort. Panel C uses the log of wages for that I observe for the 2000 cohort who could be matched.

Figure 7: First Stage: Full vs Selected Samples.



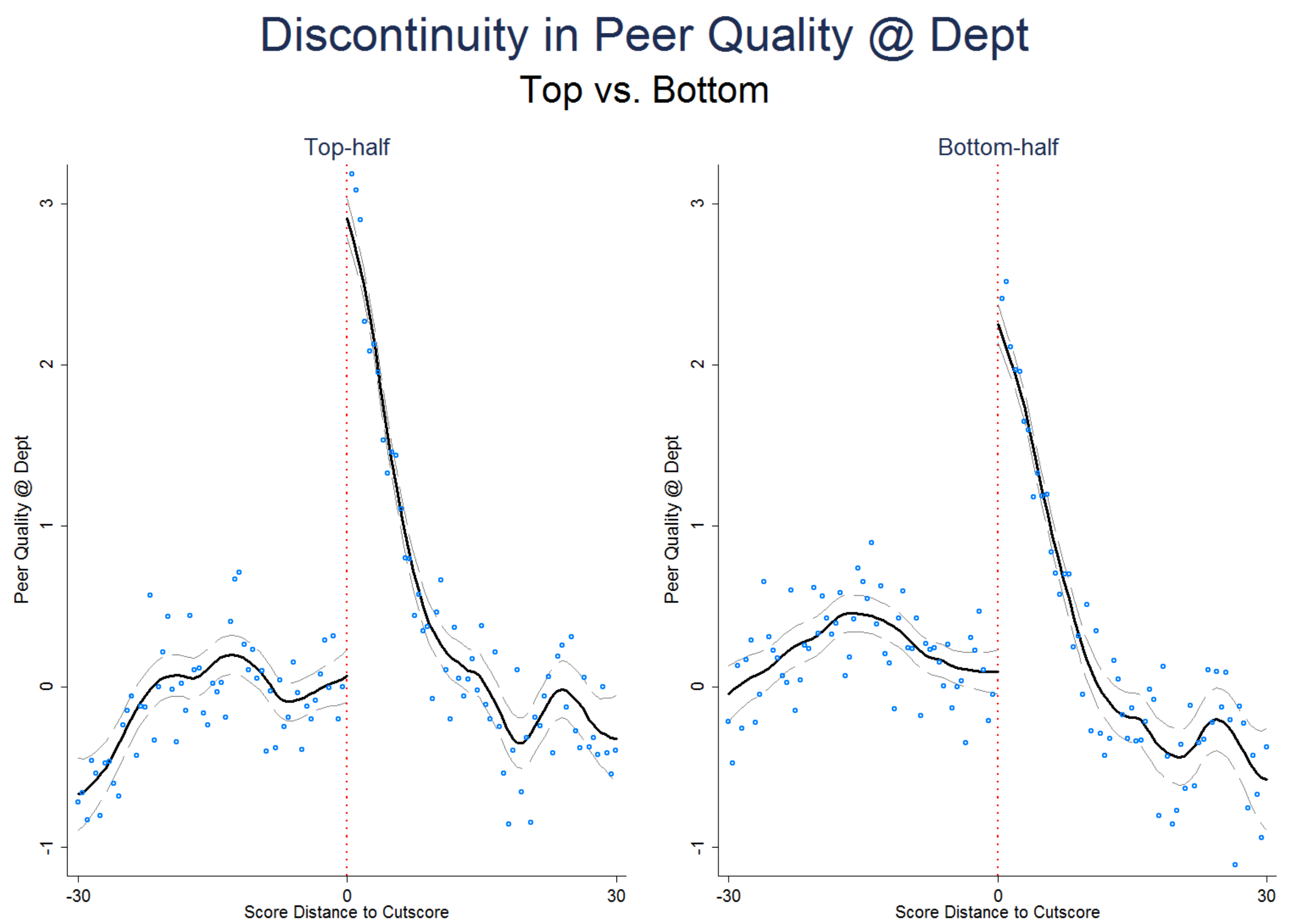
Notes: The 4 panels above give the smoothed fits of average department peer quality around the discontinuity. All use residuals from a regression on a linear trend in the test scores, the full set of covariates, and the cutoff fixed effects. The upper panels are at a 30-point bandwidth and use local polynomial smoothed fits, while the lower ones are at a 10-point bandwidth and use local linear smoothed fits. The left panels are fits for the full sample, while the right panels are on selected sample.

Figure 8: Reduced Form: Full vs Selected Samples.



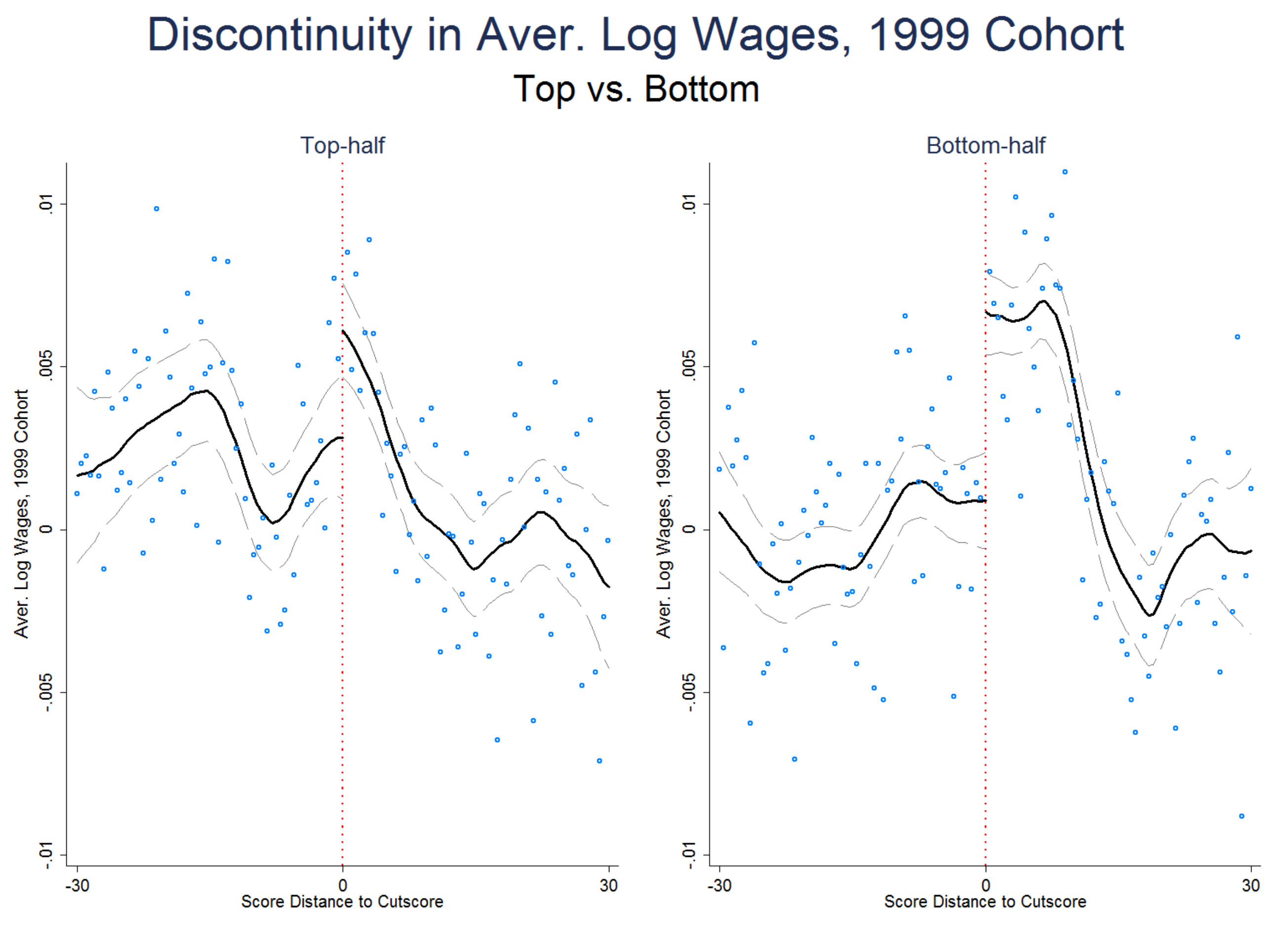
Notes: The 4 panels above give the smoothed fits of average log wages around the discontinuity. Average of log earnings for the 1999 cohort in each department is being used as a proxy of expected earnings for the 2000 cohort in that same department. All plots use residuals from a regression on a linear trend in the test scores, the full set of covariates, and the cutoff fixed effects. The upper panels are at a 30-point bandwidth and use local polynomial smoothed fits, while the lower ones are at a 10-point bandwidth and use local linear smoothed fits. The left panels are fits for the full sample, while the right panels are on selected sample.

Figure 9: First Stage: Top-half vs. Bottom-half of departments



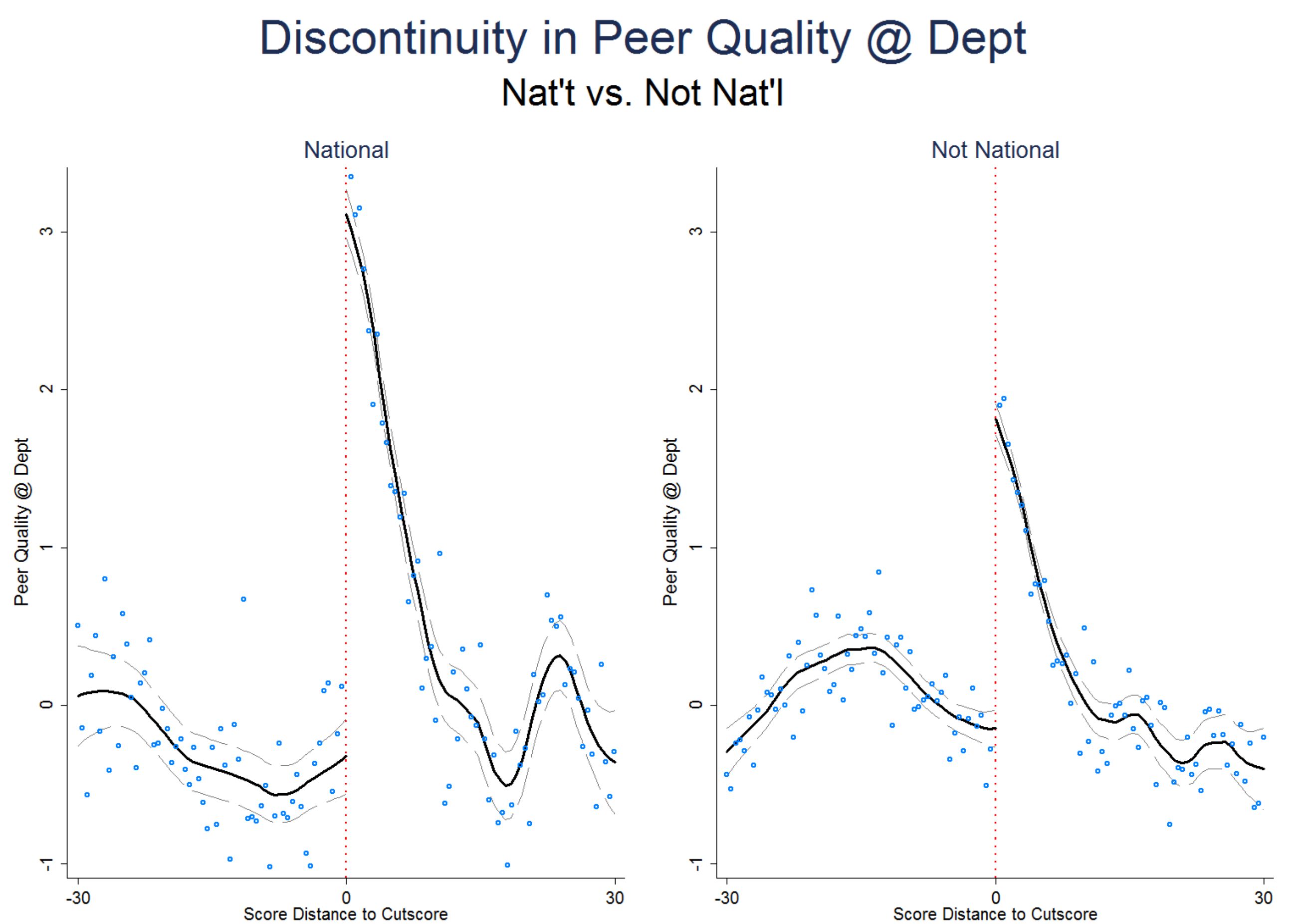
Notes: The 2 panels above give the smoothed fits of average department peer quality around the discontinuity. All use residuals from a regression on a linear trend in the test scores, the full set of covariates, and the cutoff fixed effects. The left panels are fits for the top-half of the departments (within each major), while the right panels are on the bottom-half of the departments.

Figure 10: Reduced-form Wages: Top-half vs. Bottom-half of departments



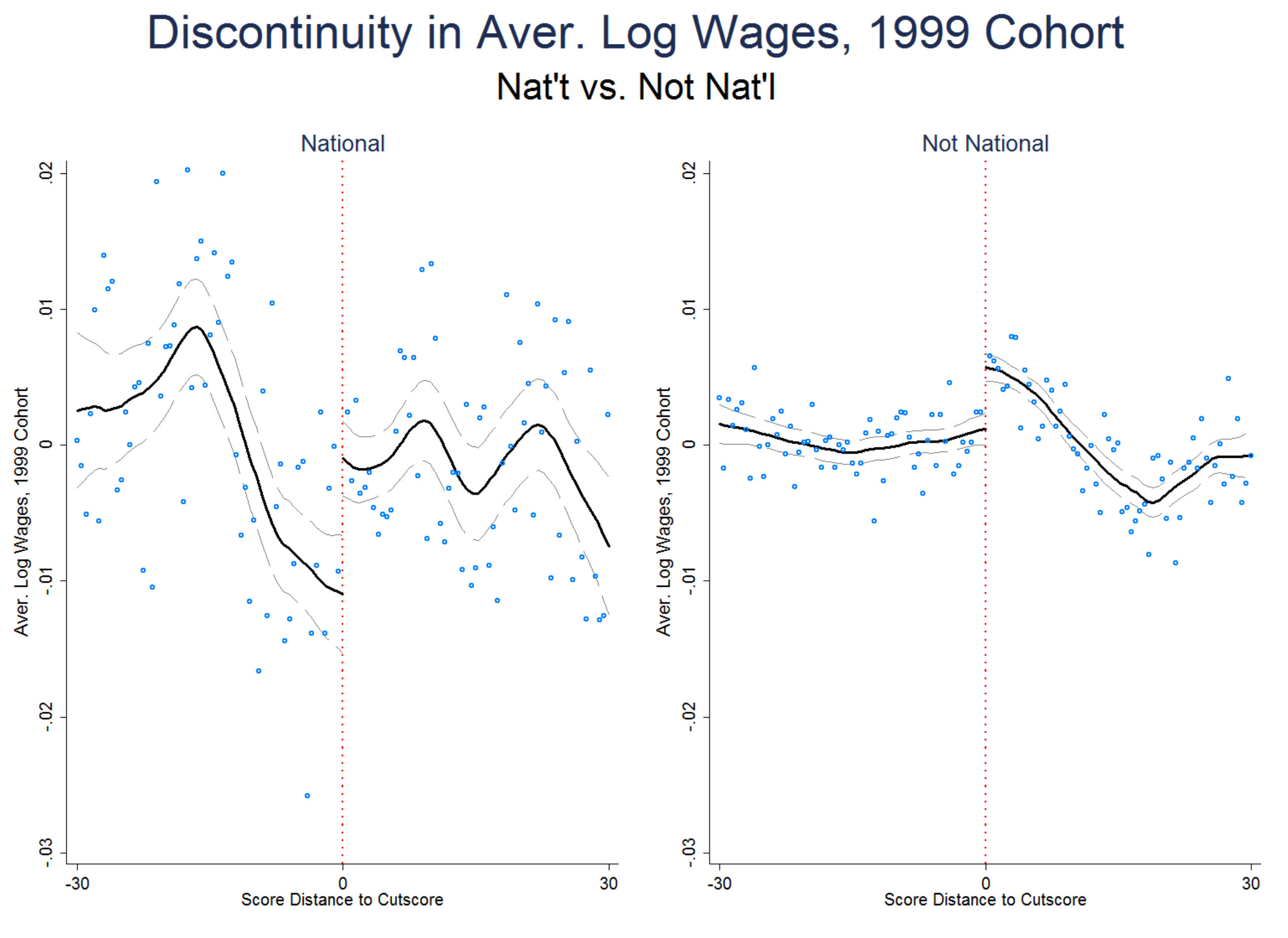
Notes: The 2 panels above give the smoothed fits of average log wages around the discontinuity. All use residuals from a regression on a linear trend in the test scores, the full set of covariates, and the cutoff fixed effects. The left panels are fits for the top-half of the departments (within each major), while the right panels are on the bottom-half of the departments.

Figure 11: First stage in quality: National vs. Not-national



Notes: The 2 panels above give the smoothed fits of average department peer quality around the discontinuity. All estimates are on the selected sample of college majors. All estimates using only college majors that have at least two department that is associated with a national college. All use residuals from a regression on a linear trend in the test scores, the full set of covariates, and the cutoff fixed effects. The left panels are fits for the departments (within each major) that are part of national college, while the right panels are the departments (within each major) that are not part of national college.

Figure 12: Reduced-form wages: National vs. Not-national

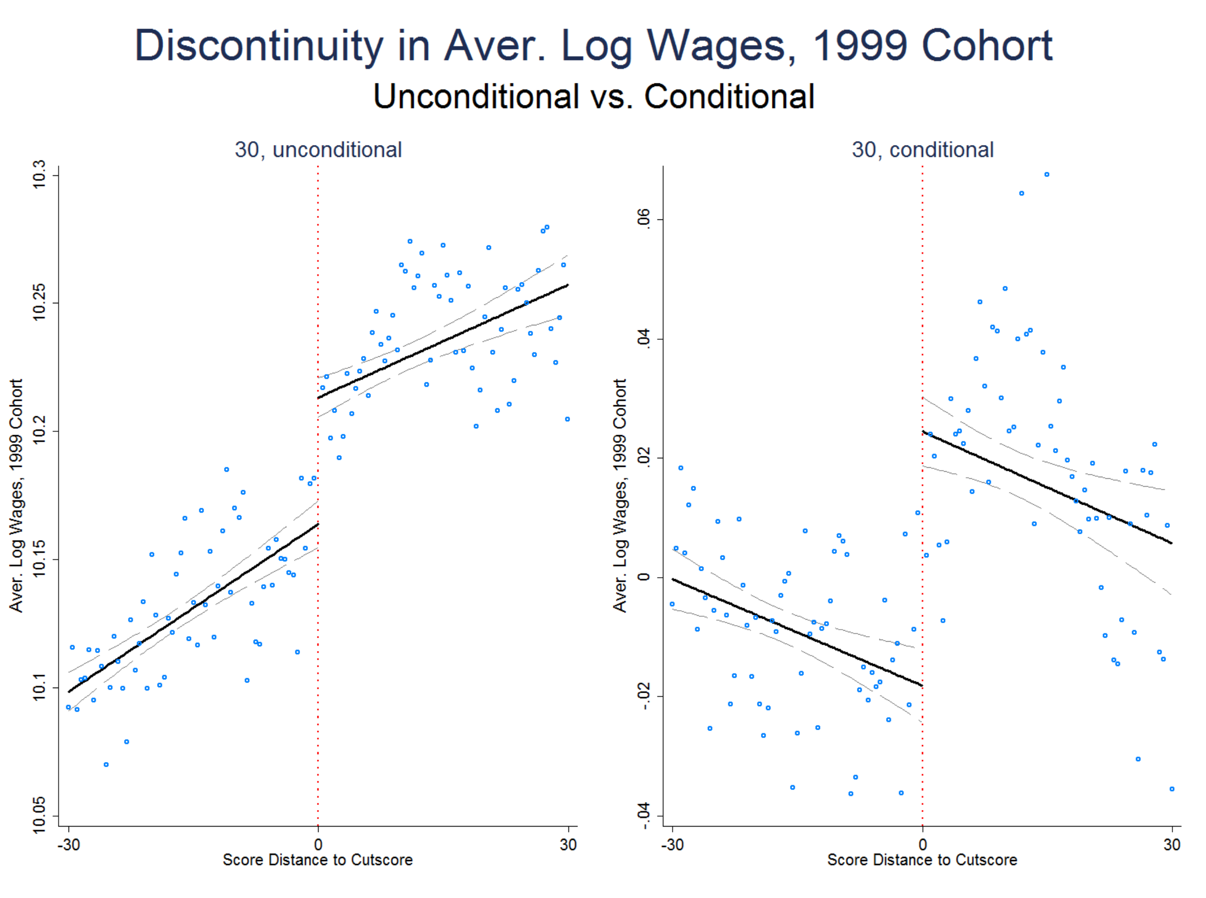


Notes: The 2 panels above give the smoothed fits of average log wages around the discontinuity. All estimates are on the selected sample of college majors. All estimates using only college majors that have at least two department that is associated with a national college. All use residuals from a regression on a linear trend in the test scores, the full set of covariates, and the cutoff fixed effects. The left panels are fits for the departments (within each major) that are part of national college, while the right panels are the departments (within each major) that are not part of national college.

Figure 13: First stage in quality: Original Data: National vs. Not-national

Notes: The 2 panels above give the smoothed fits of average department peer quality around the discontinuity. All estimates use the original (non-replicated/simulated data) recentering the cutoff for each college major at the lowest test score for the last person admitted into any of the national colleges. Thus the comparison is only for admission to a national university or not. Compared to the estimates above, which makes comparisons WITHIN the national vs. not-national labeling, these estimates are ACROSS the national vs. not-national labels. All estimates are on the selected sample of college majors. All estimates use only college majors that have at least one department that is associated with a national college. The left panels are unconditional fits, while the right panels are conditional, meaning that they are residuals from a regression on a linear trend in the test scores, the full set of covariates, and the cutoff fixed effects.

Figure 14: Reduced-form wages: Original Data: National vs. Not-national



Notes: The 2 panels above give the smoothed fits of average log wages around the discontinuity. All estimates use the original (non-replicated/simulated data) recentering the cutoff for each college major at the lowest test score for the last person admitted into any of the national colleges. Thus the comparison is only for admission to a national university or not. Compared to the estimates above, which makes comparisons WITHIN the national vs. not-national labeling, these estimates are ACROSS the national vs. not-national labels. All estimates are on the selected sample of college majors. All estimates use only college majors that have at least one department that is associated with a national college. The left panels are unconditional fits, while the right panels are conditional, meaning that they are residuals from a regression on a linear trend in the test scores, the full set of covariates, and the cutoff fixed effects.