The Impact of Letter Grades on Student Course Selection and Major Choice: Evidence from a Regression-Discontinuity Design

Joyce Main[*] and Ben Ost[†]

Abstract

This research examines the effect of undergraduate course letter grades on future course selection and major choice. Using a Regression-Discontinuity design, we exploit the fact that the probability of earning a particular letter grade jumps discontinuously around letter grade cutoffs. This variation in letter grades allows us to isolate the impact of letter grades on major choice and course selection. We collect original numerical scores for 65 introductory courses across 6 fields and merge this with administrative data including student-level characteristics and transcripts. Since grading cutoffs exist throughout the distribution of scores, we are able to estimate local treatment effects at a variety of localities to examine the distribution of treatment effects. Contrary to the findings of the previous literature, we find no evidence that students respond to their letter grades in terms of course or major choices.

---

[*]Visiting Assistant Professor, School of Engineering Education, Purdue University
[†]Assistant Professor of Economics, University of Illinois at Chicago

1 Introduction

The extent to which students respond to their letter grades is crucial to understanding student major choice and course taking behavior. These decisions are of particular concern to policy makers given the considerable effort that has been devoted to improving major persistence, especially in the sciences. A common concern, first explicated in Sabot and Wakeman-Linn (1991), is that differential grading standards across the disciplines distort course taking behavior. In particular, students may decide to avoid a science major due to the generally lower grades given in the sciences. Since technological advancements and production are reliant on individuals with scientific backgrounds, a relative decrease in the number of students choosing to major in these fields has the potential to impair social welfare..

Many studies have investigated whether students strongly respond to their letter grades. This literature has overwhelmingly found that students respond to their letter grades such that students with higher letter grades in introductory courses are much more likely to major in that subject. The response of students to their letter grades is generally argued as efficient in that it promotes students sorting towards their comparative advantage; however, grading imbalances across fields distort this sorting process. Simulations from this literature suggest that equating letter grades across the university would have the impact of encouraging more students to pursue science. While research on this topic has spanned many years, institutions and disciplines, a fundamental obstacle to identifying the impact of letter grades on major choice is the possibility of unobserved factors which influence both major probabilities and introductory course letter grades. In particular, if students with more interest in a subject work harder in that subject, we would expect to see students with the highest performance also to be the most likely to major. Though several studies have controlled for overall performance to identify a student's comparative advantage, this approach does not address concerns that students may put *relatively* more effort into their intended major's introductory courses.

Our study overcomes this obstacle by implementing a Regression Discontinuity (RD) design to identify the causal impact of letter grades on major choice and course performance. We supple-

ment administrative records with a refined measure of course performance, collected directly from course instructors. This data allows us to observe not only the letter grade a student receives, but the exact numerical score he/she earned in the course. By comparing the major and course choices made by students with similar numerical scores, but different letter grades, we identify the causal impact of the letter grades. To implement this analysis, we collect original numerical scores from 65 introductory courses across 6 fields at a large selective research institution (LSRU). We combine these data with each student's full transcript, demographic information and major choices.

To examine this issue, we take two distinct approaches. In our first approach, we reproduce the typical analysis of the literature, as if we did not know exact numerical scores. We find evidence of a clear relationship between letter grades and major choices, which matches that found in the previous literature. When we add a numerical score control to this regression, however, we find that the entire correlation is explained by a linear function of numerical score. Once controlling for numerical score, none of the letter grades indicators are statistically significant predictors of major choice. Furthermore, we are unable to reject the hypothesis that the letter grades jointly do not contribute to the model.

In our second approach, we use the exact numerical score to implement an RD design, testing whether students are more likely to major and take more course work in fields in which they earn higher letter grades. We find no evidence that students respond to their letter grades based on the RD specification. Since cutoffs exist throughout the entire distribution, we are able to estimate a variety of local treatment effects and find no evidence that students respond to their letter grades whether at the top or bottom of the overall distribution. While we are unable to examine students who are not on a grade margin, the students who are of most interest to policy makers are precisely the students who are marginal and thus the RD research design is well suited to this application.

As in any RD design, our major concern is the possibility that students manipulate their scores in order to fall just above a grade cutoff. In order to evaluate the likelihood of manipulation in this context, it is important to distinguish between manipulation of numeric scores and manipulation of letter grades. An example of manipulation of numeric scores would be artificially raising one's

numeric score from an 89 to a 90 whereas an example of manipulation of letter grades would be artificially raising one's letter grade from a B to an A. Ultimately, only letter grades are of consequence, so *a priori,* we expect that students and professors would be more likely to manipulate letter grades than numeric scores. Our data bear this out. We show that students are frequently granted higher letter grades than their numerical score dictates, which is clear evidence of letter grade manipulation. However, we find no evidence of manipulation of the numerical scores themselves: the histograms of numeric scores around each letter grade cutoff show no evidence of scores heaping just above grade cutoffs.

The key assumption made for the RD design is that the *numeric scores* are not manipulated to fall just above or below a cutoff—manipulation of letter grades will not bias estimates. If students are granted higher letter grades than their numerical score dictates, this simply converts the strict RD design to a fuzzy RD design. Importantly, even if the students who argue for higher grades are unobservably different than students who do not argue, the fuzzy RD design will yield consistent estimates. Essentially, our estimates compare the major choices made by students who earn an 89 to the major choices made by students who earn a 90. If there is letter-grade manipulation, some students with an 89 will be given an A, however, the overall group of students who earn an 89 will still have a lower average letter grade than the group of students who earn a 90. As long as *numeric scores* are not manipulated, the group of students who earn an 89 will be otherwise similar to the group of students who earn a 90 and thus, comparing major choices made in the two groups yields the causal impact of letter grades on major choices. In order to ensure that the underlying course scores are not manipulated, we obtain the original spreadsheets used by professors in calculating numerical scores and confirm with the professors that these spreadsheets were not altered, even when a student successfully petitioned for a higher letter grade.

The outline of the text is as follows. We begin by reviewing the literature in Section 2 paying particular attention to the magnitudes found in previous research. Section 3 describes our data, Section 4 presents our first regression approach and Section 5 presents our RD approach. We provide a discussion of implications and how our work relates to previous research in Section 6

4

and conclude in Section 7.

2 Literature Review

A large literature has examined the determinants of major choice with particular emphasis on examining persistence in the sciences. Given the breadth of topics covered in this literature, we focus here on describing the literature that examines the role of grades in determining major and course choices. Using data from Williams College, Sabot and Wakeman-Linn (1991) estimate how students respond to letter grades and examine how differential grade inflation across disciplines might distort major choice decisions. The authors find that controlling for performance in other subjects, receiving an A instead of a B in an introductory course increases the likelihood of taking a second course by approximately 10-20 percent for economics and English. Using a simulation, Sabot and Wakeman-Linn show that if economics graded as leniently as English at Williams College, enrollment in higher level economics courses would rise by 11.9 percent.

This basic point has been made repeatedly since that time and has been shown in a wide variety of disciplines and institutions. Christopher et al. (1994) examines the determinants of majoring and persisting in natural science and engineering at four highly selective institutions and similarly finds that letter grades are strongly correlated with declaring and remaining in these science majors. Similarly, Ost (2010) finds that students with a one point higher physical science GPA are 11 percentage points more likely to major in physical sciences and students with a one point higher life science GPA are 11 percentage points more likely to major in life science. Using data from a liberal arts college, Rask (2010) also finds that letter grades are important in predicting persistence in STEM fields such that a one letter grade change increases the probability of persisting by approximately five percentage points. Given that STEM departments grade more strictly than most departments in his study, Rask simulates the effect of equating grading standards across departments and concludes that this would increase STEM persistence by 2-4 percent.

In addition to discouraging persistence in STEM fields, student response to letter grades may

explain racial or gender imbalances in certain majors. Rask and Tiefenthaler (2008) find that economics students are sensitive to their grades in introductory courses and in particular, women appear more sensitive to these grades than men. Rask and Tiefenthaler posit that this sensitivity differential explains part of the gender imbalance in economics in higher level courses since women with equal performance to men leave economics at a higher rate. Owen (2010) confirms this finding for economics and finds that changing from a B to an A increases the probability of majoring by 15 to 20 percentage points among women while having no statistically significant impact for men.

While the literature examining the impact of introductory grades on course and major choice is well developed, the majority of the above studies rely on regression frameworks for identification. Several underlying behaviors are consistent with a strong correlation between letter grades and major choices and the regression framework is unable to distinguish between these underlying behaviors. First, it is possible that low letter grades in an introductory course cause students to leave a subject – either because they care about maintaining a high GPA or because they learn that their comparative advantage lies elsewhere. These two potential behavioral stories are intuitive and have been the primary interpretation of the literature. However, the relationship between major choice and introductory grades could also plausibly be generated by student response to underlying factors. In particular, students may choose to work hardest in the subject in which they intend to major, and as a result, they may earn their highest letter grades in their major fields. The policy implication of this phenomenon is very different. If students respond to their letter grades, then equating average letter grades across departments has the potential to increase enrollments in initially low grading departments. If, on the other hand, students simply work hardest in their intended major, equating grading standards across departments will not have any direct impact on enrollment or major choice behavior.

The only study of which we are aware that is able to rule out an underlying factor and plausibly identify a causal impact of grades is Owen (2010). In her paper, Owen examines the impact of letter grades on major choice in economics using a similar RD methodology to the one used in our paper. She finds evidence of a strong impact of letter grades on major choices among women in

economics and given her identification strategy these are interpreted causally. Given that Owen (2010) is the only paper that has estimated the causal impact of letter grades on major choices, we consider the replication of her analysis to be a contribution. This is particularly true because like many studies in this field, Owen (2010) focuses on a single institution and discipline and thus the results may not generalize to other settings.[1]

We extend Owen (2010) by considering a different institution and 6 disciplines. Also, in an attempt to improve the precision of the estimates, we have collected more than ten times the number of observations as was used in the Owen. As a result, instead of using 30-60 observations on either side of the threshold, we are able to use nearly 1,000 students on either side of the threshold. The large amount of data facilitates breaking out the data more finely than previously possible and exploring interactions between grade responsiveness and factors such as financial aid status, gender, discipline and overall GPA. In Main and Ost (2011), we attempt to replicate the exact analysis in Owen (2010). We are unable to replicate her findings, despite studying a similar institution and restricting our sample to just economics students. Our large sample provides sufficient precision such that we are able to rule out the effect sizes found by Owen for our sample. We discuss potential reasons for this difference in the discussion section.

3 Data

The data used in this paper come from three distinct sources that are merged together. First, we collected grading spreadsheets from instructors at LSRU who teach large introductory courses. In collecting these data, when possible, we obtained the original spreadsheets that professors used to record grades throughout the semester. In total we collected data from 65 course offerings across 6 disciplines. Due to confidentiality agreements made with specific instructors, we are unable to disclose the exact disciplines for certain subjects, and thus categorize courses as "Physical Sci-

---

[1]Owen performs secondary analyses using a small liberal arts school, but the small sample at the second school prevents her from using a regression discontinuity design.

ence", "Life Science" or "Economics".[2] Two key pieces of information come from the grading spreadsheets. First, the spreadsheets include each student's final numerical score in a given course. Second, we carefully went through each spreadsheet and coded instances in which the professor indicated that he/she had altered a students numerical grade. The first key variable that records numerical scores is of central importance to our entire analysis while the second is useful in assessing the extent to which grade manipulation might impact our results. Importantly, the data collected from instructors do not represent the universe of students at LSRU because it is restricted to only students who enrolled in one of the 65 course-offerings. In total, the spreadsheet data includes 20,774 students-course observations representing 9,565 students over an 11 year period (2000-2010).

Second, the registrar at LSRU provided the entire transcript for each student in the study population for the entire duration of their enrollment at LSRU. This data includes unique course identifiers and letter grades received for every course completed in addition to information on a students declared major(s).[3] From the transcript, we calculate cumulative GPA, semester GPA and categorize course taking behavior. Using a unique student identifier, this data is merged to admissions data from LSRU. The admissions data include basic demographic variables, financial aid information and SAT/ACT scores for each student. In addition, the admissions data include information on students' intended majors, which they list on their application for admission. The match rate between the three sources of data is very high for the years 2005-2010, but because LSRU changed administrative systems during the timeframe, we are unable to match all admissions variables prior to 2005. The 2000-2010 data has 20,334 student-course observations matched to transcripts and where possible, we use all of these observations. For some analyses, noted in the text, this sample is reduced as a result of missing admissions data in early years. The sample that focuses on 2005-2010 timeframe includes slightly over 13,000 observations.

---

[2]Data was also collected for another social science discipline, but this is excluded from the main analyses because less than 1 percent of enrolled students intend to major in this subject. In practice, all results presented are robust to the inclusion of this subject, but estimates become less precise.

[3]If a student enrolls in a class but drops the course within the first several weeks, this course will not appear on the transcript or in our data. If a student drops the course after the designated drop period, we observe that student-course combination in our data.

The final merged dataset thus includes a complete course history for each student and two related measures of performance for the collected introductory courses. The first measure of performance is the exact numerical score the student received in the course (for example a 91/100). The second measure of performance is the letter grade from the student transcript, ranging from an F to an A+. These letter grades are converted to the LSRU GPA scale ranging from 0 to 4.3 where a B+ is a 3.3 rather than a $3.\bar{3}$ and an A- is a 3.7 rather than as $3.\bar{6}$. Throughout the remainder of the paper, we refer to these performance measures as numerical score and letter grade respectively.

Because different courses use different scales, the numeric scores are standardized to a 0 to 4.3 scale which is analogous to the 0 to 4.3 GPA scale but is measured continuously. This standardization makes across-course comparisons possible and also facilitates comparisons to the previous literature. In practice, this standardization is performed by mapping course grading cutoffs to the GPA scale and then mapping each student's score according to the distance from the cutoff. More exactly, we use the following formula, where γ_1 and γ_2 are the grade cutoffs in the original distribution, y is the student's percentage score in the course and α_1 and α_2 are the grade cutoffs being mapped to on the 0 to 4.3 scale.

$$Standardized\ Score = (y - \gamma_1)\frac{\alpha_2 - \alpha_1}{\gamma_2 - \gamma_1} + (\gamma_2 - \gamma_1) \tag{1}$$

For example, if a course initially grades on a 100 point scale where 97 or above is an A+ and 93 or above is an A, we map 97 to a 4.3 and map 93 to a 4.0. A student who received a 95 would be mapped to a 4.15 and a student who received a 96.4 would be mapped to a 4.255. While the GPA scale ranges from 0 to 4.3, the continuous version allows for some grades to exceed 4.3 since anyone who earns a numerical score above the A+ cutoff will be mapped to above a 4.3.

The first three columns of Table 2 show descriptive statistics for our data split by course discipline. Of the 2,072 students we observe taking introductory economics, 43 percent are female, 2 percent are black and 7 percent are hispanic. These demographic characteristics are fairly similar in engineering and the physical sciences but are dramatically different in the life sciences, where

the gender imbalance is reversed and there is higher representation of black students. SAT scores (or ACT equivalents) are highest among students taking engineering and physical science courses and lowest among students taking life science courses; however, this pattern is not reflected in cumulative college GPA.

The most substantive difference between the three course categories is the intentions of students taking these courses. Nearly 70 percent of students taking engineering or physical science introductory courses intend to major in the course discipline. This is in stark contrast to the less than 5 percent of students taking economics who intend to major. The primary cause of this difference is the fact that students majoring in engineering are required to apply to the engineering school and list engineering as their intended major whereas there is no such requirement for economics majors (who enroll in the liberal arts portion of LSRU). Another potential reason for this difference is that introductory economics requires less technical background than do introductory engineering courses and thus students may be more likely to enroll in introductory economics purely out of topical interest. Of students who enroll in economics 17.3 percent choose to major in economics. The analogous figure is 60 percent for engineering and 53 percent for life sciences. This does not imply that engineers and life science courses have higher major persistence but simply reflect the fact that economics is a popular course among all students.

The last three columns of Table 2 restrict the attention to only students who eventually major in the course subject. These students are fairly similar to the other students in their classes with the notable exception that students who eventually major perform better in their introductory courses than students who do not major. Importantly, the demographic characteristics are similar between the students taking introductory courses and those majoring in the subject, suggesting that for this recent timeframe, persistence rates are similar for men and women. Compared to the average student taking an introductory course, a larger fraction of students who eventually major intended to major in that subject.

3.1   Data Issue: Imputing Grading Cutoffs

While our data is improved over previous research, one important limitation is that we do not exactly know the grading cutoffs used for the majority of the sample courses. Since knowing the grading cutoffs is crucial to our entire analysis, we put in considerable efforts to ensure that grading cutoffs are imputed accurately. Unlike many imputation procedures, it is not simply adequate to obtain an unbiased estimate of the cutoffs – we require that our imputation procedure perfectly and exactly obtains grading cutoffs. We are fairly confident that the imputation procedure that we use meets this high standard. The imputation procedure involves a quantitative imputation followed by manually inspecting each course to ensure that the imputation is not driven by students with manipulated letter grades. The quantitative procedure chooses the cutoff for grade X according to the highest numerical grade received by an individual with a letter grade below X. In order to explain the complete imputation procedure it is useful to consider an example. Table 1 shows 10 scores from students around the B+/A- cutoff in a hypothetical course. Because each course has hundreds of students, the density around any given cutoff is quite high and the example below is representative of the typical course in terms of density.

Table 1: Hypothetical Course

Student ID	Numeric Grade	Letter Grade
1	89.544	B+
2	89.662	B+
3	89.781	A-
4	89.824	B+
5	89.932	B+
6	90.031	A-
7	90.125	A-
8	90.132	A-
9	90.209	A-
10	90.311	A-

In the above example, the algorithm identifies student 5 as having the highest numerical grade of any student with below an A- letter grade. The imputed cutoff is then calculated as the average of that student with the next highest students score. In this case, averaging student 5, and student

6 yields and estimated cutoff of 89.9815. This imputation procedure is relatively simple, but performs exceptionally well. For the sample of courses for which we know the exact cutoffs, the imputation is typically within 0.02 points of the correct cutoff and always within 0.1 points of the correct cutoff. Once the grade cutoffs are imputed following the above procedure, we manually inspected each course to make sure that cutoffs appear appropriate and are not driven by students whose numeric grades were manipulated.

Note that in this example, student 3 received an A- but falls below imputed cutoff point. This situation is common in our data and we attribute this phenomenon to either persistent students who argue for higher grades or professors who take into account motivation or performance trends in assigning letter grades. Importantly, we observe the original distribution of numerical scores, prior to the manipulation that results in student 3 receiving an A- and thus, this type of grade manipulation will not bias our estimates.

4 Regression Model

Before examining the evidence from the Regression Discontinuity model, we first consider how models used in the literature are altered when we include a control for numerical score. While a variety of models have been used to estimate the impact of grades on major choice, the key features of every model examines how course letter grades relate to major choice, conditional on general academic performance in other courses (Ost, 2010; Owen, 2010; Rask, 2010; Rask and Tiefenthaler, 2008; Sabot and Wakeman-Linn, 1991). We follow the literature in our approach and estimate the following model as a baseline.

$$Y_{it} = X_i\beta + Z_{it}\alpha + \sum_{j=D-}^{A+} \delta_j J_{it} + \gamma_j + \epsilon_{it} \tag{2}$$

Y_{it} is one of two measures of major choice. The first measure is an indicator for whether the student eventually majors in the relevant subject while the second measure is a count of the total

number of credit hours taken in the relevant subject over the following three semesters.[4] X_i is a vector of time invariant characteristics including demographics, SAT score or ACT equivalent, and an indicator for whether the student listed the field as his/her intended major on the LSRU application. The vector Z_{it} includes cumulative GPA in time t, GPA in time t, and credit hours taken in time t. γ_j is a course fixed effect intended to capture important determinants of major choice such as professor or peer quality (Carrell et al., 2010; Ost, 2010).

The key variables of interest are the coefficients on the dummy variables denoted by δ_j. Equation 2 is estimated as a linear probability model, but using a probit to predict major choice or a count model to predict subject credit hours yields similar results.

Equation 2 is the model typically estimated in the literature and the results for our sample are given in columns (1) of Table 3. Just as in the most papers in the literature, the results presented in the first and fourth columns of Table 3 paint a clear picture of the relationship between letter grades and major choice. Controlling for performance in other classes, students with better letter grades are more likely to major in the field and the magnitude of this difference is large. A student who receives an A- in an introductory course is 5 percentage points more likely to major in the subject than a student who receives a B+. Moving from an A- to a B- lowers the probability of majoring by nearly 9 percentage points and moving from an A- to a C- lowers the probability of majoring by over 17 percentage points. While these effect sizes are large, they are consistent with the rest of the literature which find that, controlling for overall GPA, an increase of one point on a four point scale in one's introductory class is associated with a 15-20 percentage point increase in the probability of majoring in the subject.

Column (4) of Table 3 shows the results from the same model when predicting the number of credit hours taken in the field in the following three semesters. This variable is intended to capture more nuanced variation in subject interest, but naturally, credit hours taken is correlated with eventual major choice. The results for credit hours are less consistent than for major choice

---

[4]Looking at course behavior over the following three semesters is motivated by a desire to smooth idiosyncratic course taking behavior driven by the availability of certain courses in only the spring or fall; however, all results presented in the paper are similar when looking at course taking behavior only in the semester immediately following the introductory course.

and better letter grades are not monotonically associated with more credit hours. Lower letter grades in introductory courses are still generally associated with taking fewer subsequent credit hours and the impact is statistically significant when considering large letter grade changes. For example, students who receive a B- take 1.285 more credit hours than students who receive a C-.

While the relationship between letter grades and major choice is strong, whether this should be interpreted causally is unclear. It is possible that higher letter grades cause students to major in a subject, or it is plausible that students with the most interest or talent for a subject will both perform well in their introductory course and subsequently choose to major. To distinguish between these two explanations, we add numerical score as an additional control that is intended to proxy for a student's natural talent or interest in a subject. Specifically, we estimate

$$Y_{it} = X_i\beta + Z_{it}\alpha + \sum_{j=D-}^{A+} \delta_j J_{it} + \gamma_j + \omega S_{it} + \epsilon_{it} \tag{3}$$

where S_{it} is a students numerical score for class t and all other variables are defined as in equation 2. If students actually respond to the letter grades that they receive, then one would expect the dummy variables to remain significant after the inclusion of the numerical score. Column (2) of Table 3 shows that the inclusion of the numerical score eliminates the correlation between letter grades and major choices. The relationship between letter grades and major choice is no longer monotonic, the coefficients are reduced by an order of magnitude and there are no statistically significant differences between a B+ and other letter grades.

An alternative test of the importance of letter grades is given by the incremental F-test comparing a model with numeric score *and* letter grade dummies to a model with just numeric score . Specifically, we first estimate

$$Y_{it} = X_i\beta + Z_{it}\alpha + \gamma_j + \omega S_{it} + \epsilon_{it} \tag{4}$$

where all variables are defined as in equation 3. We use the incremental F-test to examine whether the model given by equation 3 that includes the letter grade dummies contributes any explanatory

power compared to the model given by equation 4 that excludes the letter grade dummies. This test is shown in the bottom panel of Table 3 and shows that adding letter grade dummies does not improve the model, when numeric score is already controlled for. Similarly, when using future credit hours as the outcome, the incremental F-test shows that adding letter grade dummies does not improve the model, when numeric score is already included.

In summary, we are able to replicate the findings of literature using a similar model, but these findings are not robust to the inclusion of the numerical score variable that we collected.

4.1 Analysis by Gender

Several papers have noted that women may be more sensitive to grade feedback than men (Crocker and Major, 1989; Owen, 2010; Rask and Tiefenthaler, 2008; Seymour, 1995). In order to investigate this possibility we re-estimate equations 2 through 4 on only the female students in our sample. Table 4 shows that results are fairly similar when focusing only on women. The relationship between letter grades and major persistence remains strong, though it is no longer entirely monotonic. Column 2 shows that once we control for numerical score, the relationship between letter grades and major choices is dramatically reduced in magnitude and is no longer statistically significant. As with the entire sample, female students with lower letter grades tend to take fewer credit hours in a subject that they perform poorly in; however, this relationship is not robust to the inclusion of the numeric score control. Once controlling for numeric score the dummy variable for earning an "A" is negative and marginally significant and the F-test rejects the hypothesis that the letter-grade dummy variables do not improve the model at the 10% level. However, the overall relationship is highly non-monotonic and does not show broad evidence in support of the notion that earning a higher letter grade increases the number of credit hours taken in the field. That said, given that the initial relationship between future credit hours and grades is relatively weak among women, we find these results to be inconclusive regarding whether letter grades matter in determining course choice among women.

5 Evidence from Regression Discontinuity Design

Based on the regression analysis, we conclude that the relationship between letter grades and major choice is likely driven by an underlying continuous process. To test this further, we use a regression discontinuity (RD) design to test for a structural break around each grade cutoff.

5.1 Heaping and Sorting

Given that RD estimates rely on comparability between students on either side of the threshold, a threat to identification occurs if students sort around the cutoff in a systematic and unobserved fashion. In the case of sorting around a grade cutoff, one might be especially concerned, because grade cutoffs are sometimes known ahead of time and students have a strong incentive to put in just enough effort for their numerical score to fall above a cutoff, or a student might argue with his or her professor to receive a higher grade even when the numerical score falls just below the cutoff (**?**). Furthermore, even if students are unable to successfully petition for higher grades, it is plausible that professors will artificially raise certain students' numerical score based on student interest and motivation, student improvement during the semester or extenuating circumstances. Whether driven by students or professors, this type of grade manipulation will likely generate a very specific heaping pattern in the histogram or numerical scores – a pattern that can be tested for directly. If many students who should have received scores just below the cutoff receive scores just above the cutoff, this will result in a hump in the histogram just above the cutoff and a valley in the histogram just below the cutoff. If no such pattern is evident in the histogram then this provides compelling evidence that students are not systematically sorting around the cutoff.

Importantly, if a student receives a higher letter grade than their numerical score justifies, this by itself does not violate the RD assumption in any way. The assumption is not that every student with a score below the cutoff receives the lower grade, but rather that the scores themselves are not manipulated in order to fall just above or below the cutoff. At LSRU, professors maintain their own personal records in addition to reporting official grades to the university. As long as professors

do not manipulate their own personal records, manipulation of the official grade will not invalidate the RD research design in this application. To determine the likelihood of professors manipulating their personal records, we spoke with each professor who provided us the data to directly discuss this issue. Our conversations suggest that the professors in our sample never change the numerical scores in their own records, but sometimes will change official letter grades based on student petitions or their own judgement. In any case, if professors do manipulate the raw numerical scores, this has the potential to bias estimates and the direction of this bias is likely in favor of finding a larger impact of grades on major choices. Under the plausible assumption that those most likely to major in a subject are also most likely to have their numeric scores artificially raised, the RD estimates will confound inherent interest or motivation with letter grades and overstate the impact of grades.[5] If grading thresholds are set endogenously to the score distribution, this will not bias estimates so long as the threshold is set independently of a student's unobserved motivation or subject interest. For example, if a professor sets grading cutoffs by looking for "natural breaks" in the distribution, this is will generate a valley on either side of the threshold, but it is unlikely to result in students being unobservably different on either side of that threshold. Regardless, if endogenous grading scales are used, this will be evident in the histograms, particularly if professors look for "natural breaks" to determine cutoffs.

Figure 1 shows the histogram of numerical scores centered around the B-/B cutoff, which is the modal score. Since sorting and manipulation might be masked by the standardization process, the only modification made in the histogram is subtracting the cutoff, which cannot alter the basic shape of the histogram. Figure 1 shows that this histogram of letter grades follow a bell shape, increasing up until B/B- and then decreasing. In order to look more precisely at heaping, Figures 2(a) through 2(i) show a zoomed in version of Figure 1, with the histogram of numeric scores centered around each cutoff. Scores are reported on the original 0 to 100 scale, but are standardized so that the cutoff is at zero in each figure. The histogram is shown with a bin size of 0.2 percentage

---

[5]We find it highly unlikely that the numerical grades in our sample have been manipulated both because the professors assured us that they were not and also because the professors have no incentive to manipulate their own records. The only grade that has any bearing is the official grade submitted to the university so we would expect that pressure to modify grades would be focused solely on this consequential variable.

points, but the patterns are not sensitive to displaying other similarly small bin sizes. As shown in Figure 1 the histogram steadily increases for lower grades, peaks in the B range and then steadily decreases in the A range. Broadly, these histograms show no clear evidence of sorting around cutoffs, given that the histograms tend to move smoothly on either side of these cutoffs. The two histograms that are closest to exhibiting a heaping pattern are Figure 2(a) around the D+/C- cutoff, Figure 2(g) around the B+/A- cutoff, and Figure 2(i) around the A/A+ cutoff. In these three figures, relative to the overall histogram trend, there appears to be slight heaping to the left of the cutoff. This is somewhat surprising given that if heaping were to occur, we would expect that students would be pushed just over the threshold, not artificially kept just under the threshold. Based on the mass of evidence from these histograms, combined with direct correspondence with professors about manipulation, we conclude that there is no evidence of manipulation of the raw numerical scores.

5.2 First Stage

The RD design requires that the latent variable (numerical scores) impacts the treatment (letter grades) in a discontinuous fashion. To examine whether this assumption holds, we examine whether there is a discontinuous jump in the probability of receiving grade X around the numeric threshold for X. For example, Figure 3(a) plots the fraction of students receiving a letter grade above D- versus the student's standardized numerical score. It is clear from Figures 3(a) through 3(i) that there is a large discontinuous increase in the probability of receiving a grade as one's test score crosses the necessary threshold. These figures also show that while a large discontinuity exists, numerical scores do not perfectly dictate letter grades. As the numeric score approaches the cutoff, more students are bumped up to the higher grade such that just below the cutoff nearly 20% of students receive a higher letter grade than their numerical score dictates. Regardless, there remains a large discontinuity at the cutoff since nearly all students who receive a numerical score above the cutoff are given the higher letter grade. The fact that numerical scores do not perfectly dictate letter grades transforms our empirical approach from a strict RD to a fuzzy RD, but the

intuition and implementation of the design is largely the same.

An alternative presentation of the same basic result is shown in Figure 4. This figure plots average letter grades (converted to a 0 to 4.3 scale) against average numeric score (standardized to the same scale). Each dot in this figure represents a bin of students who have a given numeric score. If numeric scores were perfectly predictive of letter grades, one would expect to see a perfect step graph where the letter grade jumps discontinuously at each cutoff and the average letter grade in between each cutoff is constant. Figure 4 shows a pattern that is close to a stepwise pattern, but exhibits a very slight slope, particularly as numeric scores approach each cutoff. The discontinuities are very clear and are particularly stark for grades above a D+.

5.3 Second Stage

Given that letter grade assignment jumps discontinuously around grade cutoffs, if letter grades impact major choices, we expect that the fraction of students majoring in a subject will jump discontinuously around the grade cutoffs as well. As a first step, we simply plot the fraction of students majoring in the course subject against these students' numeric scores in the introductory course. Figure 5 shows the relationship between major choice and numeric scores. On this figure, the points that land on a vertical line correspond to students who just barely earned a numerical score at or above the grade cutoff. If the proportion of students majoring in a subject jumps discontinuously at each line, this would therefore be evidence that letter grades are impacting major choices. Instead, Figure 5 shows little evidence of discontinuous jumps at grading cutoffs. Only the 3.0 (B) cutoff shows a potential jump relative to trend, and the increased probability at 3.0 is not persistent as numeric scores rise above 3.0. Also, the discontinuity at 3.0 is of similar magnitude to other jumps that occur far away from grade cutoffs (for example near 2.5). On the whole, visual inspection of the relationship between numeric grades and major choice shows little evidence of discontinuous jumps which is striking when one compares this to Figure 4 which shows clear discontinuous jumps at every grade cutoff. The combination of Figures 4 and 5 paint a picture which is consistent with the regression results previously presented – introductory performance is

correlated with major choices, but the letter grades themselves do not appear to impact major choice.

The results are fairly similar when considering course choices in the three semesters following the introductory course. Figure 7 shows no evidence of a consistent jump in the number of subsequent credit hours taken as the numeric score crosses letter grade cutoffs.

To empirically estimate the magnitude of any potential discontinuities, we use local linear regression.

5.3.1 RD: Local Linear Regression

To estimate a local linear regression at each cutoff, we restrict the sample to within 0.25 points of each threshold and use a rectangular kernel; however, results shown are robust across a number of bandwidth choices and are not sensitive to the choice of kernel. Specifically we estimate:

$$Y_{it} = X_i\beta + Z_{it}\alpha + \gamma_j + \omega C_{it} + \delta A_{it} + \xi(C_{it})(A_{it}) + \epsilon_{it} \qquad for \; |C_{it}| < 0.25 \tag{5}$$

The variable, C_{it} is student i's standardized numeric score for course j with the relevant grade cutoff subtracted. The variable A_{it} is an indicator defined as $\mathbf{1}(C_{it} >= 0)$ and the interaction of C_{it} and A_{it} is included to allow the slope to vary on either side of the cutoff. The parameter of interest is δ, which is the estimated discontinuity. The linear model is fit to only points within 0.25 points of the cutoff, which ensures that no figure includes more than one potential discontinuity. Figures 7(a) through 7(i) show how major choices change around each cutoff. Each figure plots major choice conditional on covariates against numeric scores and also includes a note of the estimated discontinuity ($\hat{\delta}$) along with a standard error taken from estimating equation 5.[6] The lines on either side of the cutoff are graphed based on the coefficient estimates from equation 5 ($\hat{\omega}$ and $\hat{\xi}$), rather than from fitting a line to the conditional major choice variable.

Estimating equation 5 on the nine letter grade cutoffs yields no statistically significant estimates. Of the nine estimates, five are negative and four are positive, and none of the figures show

---

[6]The conditional major choice variable is the residual from a regression of major choice on covariates.

visual evidence of a discontinuity. Furthermore, the point estimates are uniformly small and an order or magnitude less than earlier findings (Owen, 2010). While these results are generally robust across specification choices, some combinations of kernels and thresholds yield statistically significant discontinuities for certain thresholds; however, the statistically significant estimates are quite sensitive to specification and so we do not consider them to be strong evidence of a discontinuity. In results shown in Main and Ost (2011), we similarly find no evidence of a discontinuity when focusing just on women in economics as was done in Owen (2010).

Similar to our results for predicting major choice, we find little evidence that letter grades influence credit hours taken. Figures 7(a) through 7(i) show how conditional credit hours change around each cutoff. The discontinuity estimates noted on these figures are taken from estimating equation 5 using subject credit hours taken in the following three semesters as the dependent variable. As can be seen in these figures, four of the estimates are negative, five of the estimates are positive and none of the nine estimates are statistically significant. The estimated discontinuities shown should be interpreted as the causal impact of earning a score slightly below the threshold or the "intent to treat" (ITT). In order to obtain an estimate of the "Treatment on the Treated" (TOT) it is necessary to scale up these estimates by a factor of approximately 5/4. This accounts for the fact that the first stage discontinuity is only 0.8 since 20% of students just below the threshold receive the lower grade. Regardless of whether one considers the ITT or the TOT however, the effect magnitudes are small, statistically insignificant and inconsistent across cutoffs.

Given that there is no visual evidence of any discontinuities in Figures 5 or 7 and none of the local linear regressions yield statistically significant estimates, we conclude that the regression discontinuity design provides no evidence that major or course choices are influenced by letter grades.

6 Discussion

In their influential 1991 paper, Sabot and Wakeman-Linn develop a model of course choice in which students derive utility from learning, from their grades and from discounted future benefits. In their model, while students' human capital benefits from learning through their coursework, good grades themselves improve satisfaction. This notion of a direct benefit to higher grades has informed future research and is supported by theoretic intrinsic and extrinsic factors. In addition to contributing to a "warm glow of achievement", many extrinsic rewards such as graduate scholarships, academic honors and parental approval are direct functions of letter grades. The result from this model implies that students will pursue subjects in which they are best suited to learn, but this optimal behavior can potentially be distorted by the direct incentive of letter grades if different fields have different grading functions. If student behavior is indeed distorted by letter grade considerations, then one might expect that two students with roughly the same level of learning, but different letter grades, would have a different probability of majoring in a field. We find little evidence that this is the case. Taken as a whole, we believe that the results from the regression discontinuity design combined with the regression analysis do not provide support for the notion that letter grades causally impact major or course choices.

This finding has a number of implications. First, it suggests that if students learn about their ability through their relative performance in their coursework, this learning is not informed by the ultimate letter grade earned in the course. Second, this weakens the confidence with which we can predict the implications of policies being considered at several institutions to equalize grades across disciplines. Simulations of the impact of letter grades assume a causal impact of letter grades on major choices and our findings provide some reason to be skeptical.

Owen (2010) finds very different results from our paper in that she finds a very large positive impact of grades on major choices for women in economics. In Main and Ost (2011). we show that even when we exactly follow her methodology and restrict our sample to mimic Owen (2010), we find no evidence of a grading impact. There are several possible reasons that our results differ, but none are entirely satisfactory explanations. First, while Owen (2010) and our paper both examine

highly selective research universities, these universities may have different institutional factors that impede or facilitate choosing an economics major. Given that neither Owen nor we are permitted to reveal the institution used, a direct comparison of these institutional factors is not possible. That being said, by comparing our descriptive statistics, it is clear that these two institutions are slightly different in terms of who takes introductory economics. In our sample approximately 17 percent of students in introductory economics proceed to major in the field whereas in Owen's sample, only 12 percent major in the field. The average grades in the two samples are comparable and in both Owen's sample and our own, 44 percent of the students in introductory economics are women. Given that both institutions are selective research universities in the Northeastern United States, it is possible that the impact of grades on major choices is institution- or sample-specific, and therefore, further replications at other universities are necessary to characterize the effect.

A second potential reason for the difference in results is that the institution Owen analyzes gives grades without plusses and minuses, thereby making sharper discontinuities. While this can potentially explain the difference in results for the regression discontinuity estimates, it is not a convincing explanation for why we find such different results in a simple regression setting. We find similarly large "effects" of letter grades when not controlling for numerical scores but in our sample, controlling for numeric score eliminates these effects whereas in Owen's sample, the effect of letter grades is robust to controlling for numerical score. Furthermore, the sum of our effects across all 9 grading thresholds is substantially smaller than the point estimate Owen finds for just the B/A threshold, suggesting that the difference in grading scales at the institution is unlikely to fully explain the difference in our results. Furthermore, Owen extends her analysis to a liberal arts college that uses a plus/minus grading system and she finds large effects, directly contradicting the notion that the grading scale alone explains our divergent findings.

In both Owen's and our study the true effect that grades have on average major choices is potentially understated because the samples are necessarily restricted to students who choose to enroll in an introductory course. While these students are the appropriate sample when considering the determinants of major attrition, they are not representative of students in general. In particular, one

might expect that given that science and economics courses have a reputation of giving relatively low grades, only the students who are least responsive to course grades would elect to enroll in such a course. Although we find no evidence that these students respond to their letter grades by changing their course of study, it is very possible that certain students avoid enrolling in the first place due to a fear of low grades. Students at LSRU are likely well informed regarding average grades across disciplines since median grade reports are made public to the student body. If the knowledge of median grades results in only the least grade-sensitive students enrolling in low grading departments, this might explain why we find no effect of letter grades on major choices for our sample. Bar et al. (2009) finds evidence that students responded to the introduction of public median letter grades at Cornell and to the extent that LSRU students in our time frame are similarly responsive, the entire impact of grades on major choices may occur through the initial decision of whether to enroll in the introductory courses.

The regression discontinuity design aims to obtain the causal impact of letter grades by comparing two students of similar ability and motivation who received different letter grades. An interesting alternative exploration is to isolate the unobserved portion by comparing students who earn identical numeric scores but earn different letter grades. As we argue, a student who earns a score just below a grade cutoff but receives the lower grade is likely unobservably different than a student who earns a score just below a grade cutoff and receives the higher grade. This latter group of students had their letter grades artificially raised and we view this as suggesting that the student either demonstrated promise or argued forcefully for the higher grade, either of which we expect to be correlated with a higher likelihood of major persistence. To test this theory, we add an indicator for whether the students grade was artificially raised to the model given by equation 4. Columns (1) and (3) of Table 5, however, show no evidence that students who are given a higher grade than their numerical score dictates are more likely to major or take more credit hours. Columns (2) and (4) similarly show that this relationship does not appear even when allowing for the effect to differ across the distribution of numerical scores. This result has two possible interpretations and we cannot distinguish between the two. First, it is possible that grade adjustments are made primarily for

24

students with extenuating circumstances that are uncorrelated with interest in the major. Second, it is possible that students do not petition for higher grades differently in subjects in which they plan to major compared to other subjects. In other words, if certain students petition for higher grades in all their courses regardless of their majoring plans, this would result in no correlation between having one's grade raised and majoring in the field.

7   Conclusion

This paper examines the causal impact of letter grades on major and course choices. Contrary to the broader literature, we find no evidence that letter grades themselves raise the probability of persisting in a major. As in past research, we document a strong correlation between letter grades and major choices, but we find that this correlation is explained by a continuous underlying process, namely course performance. In other words, we find that students are more likely to major in a subject when they earn high scores in the introductory course, but students who just barely receive an "A" are no more likely to major than students who just barely miss the "A".

There exists a large grading gap across the disciplines such that students interested in science face a trade-off between taking coursework in their preferred field, and maximizing their GPA. If students strongly respond to these GPA incentives, this might discourage prospective scientists from pursuing that major. Previous research has found that students strongly respond to these GPA incentives and thus based on this literature, policy makers have deduced that rigorous grading practices in the sciences may discourage some prospective majors. Using an RD design, we find no evidence that students respond to their letter grades, casting doubt as to whether policies aimed at equalizing grades across the disciplines will indeed have the effect predicted by the previous literature.

One important question is whether major attrition in the sciences due to grading standards is problematic. Science majors are theoretically beneficial to society because they produce positive externalities and improve our global competitiveness; however, it is possible that the students most

likely to produce these positive externalities are also likely to have performed well in their courses. If poor grades cause students to leave the sciences, then relatively harsh grading standards might in some ways be beneficial as a screening device. For this reason, one might hope that students on the border of A/A+ do not respond to their letter grades, but students who perform at the bottom leave the major as a result of their low grades. Our results show no evidence of students responding to letter grades at either end of the performance distribution.

We find no evidence of a causal impact of letter grades on major choice contrary to previous literature including Owen (2010) which also employs RD methodology. It is therefore unclear whether letter grades directly impact student major choices.

References

Bar, Talia, Vrinda Kadiyali, and Asaf Zussman. (2009). Grade information and grade inflation: The cornell experiment. *Journal of Economic Perspectives*, *23*(3), 93–108.

Carrell, S.E., M.E. Page, and J.E. West. (2010). Sex and science: How professor gender perpetuates the gender gap*. *Quarterly Journal of Economics*, *125*(3), 1101–1144.

Christopher, Strenta A., R. Elliott, R. Adair, M. Matier, and J. Scott. (1994). Choosing and leaving science in highly selective institutions. *Research in Higher Education*, *35*(5), 513–547.

Crocker, J, and B Major. (1989). Social stigma and self-esteem: The self-protective properties of stigma. *Psychological review*, *96*(4), 608–630.

Main, Joyce, and Ben Ost. (2011). Comment: Grades, gender, and encouragement: A regression discontinuity analysis. *Unpublished manuscript*.

Ost, B. (2010). The role of peers and grades in determining major persistence in the sciences. *Economics of Education Review*.

Owen, Ann. (2010). Grades, gender, and encouragement: A regression discontinuity analysis. *The Journal of Economic Education*, *41*(3), 217–234.

Rask, Kevin. (2010). Attrition in stem fields at a liberal arts college: The importance of grades and pre-collegiate preferences. *Economics of Education Review*.

Rask, Kevin, and Jill Tiefenthaler. (2008). The role of grade sensitivity in explaining the gender imbalance in undergraduate economics. *Economics of Education Review*, *27*(6), 676–687.

Sabot, Richard, and John Wakeman-Linn. (1991). Grade inflation and course choice. *Journal of Economic Perspectives*, *5*(1), 159–70.

Seymour, E.. (1995). The loss of women from science, mathematics, and engineering undergraduate majors: An explanatory account. *Science Education*, *79*(4), 437–473.
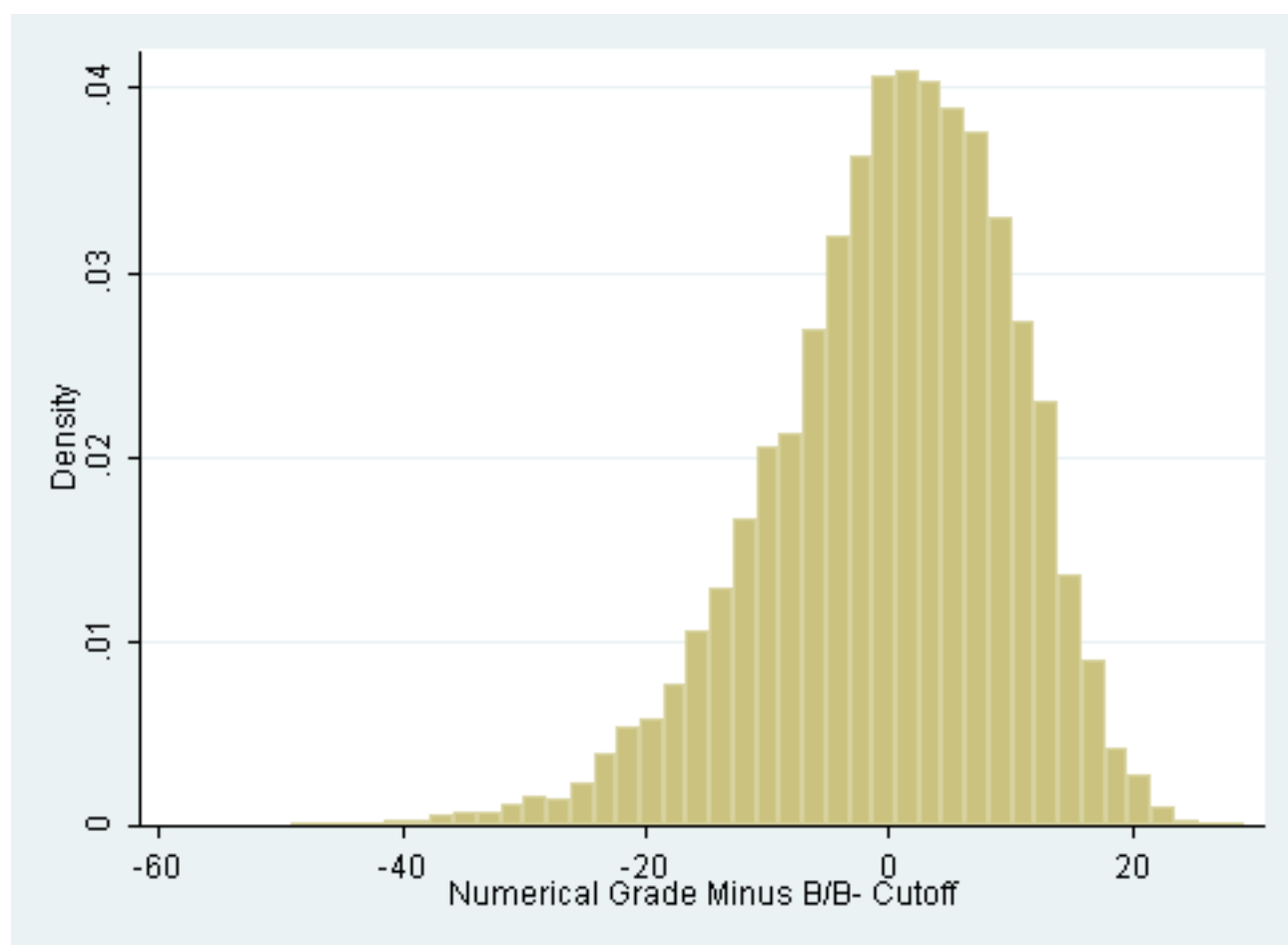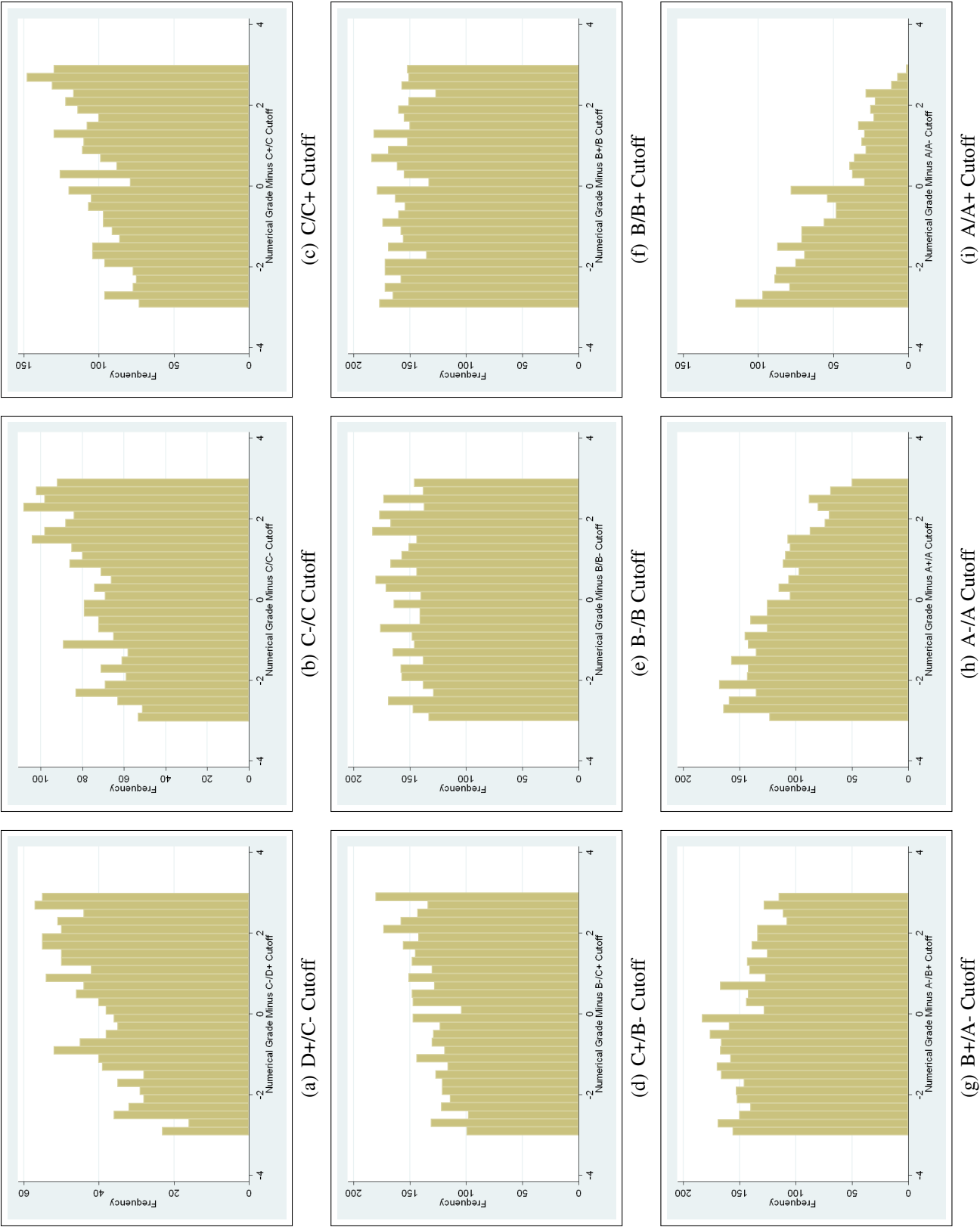
Figure 1: Histogram Normalized to B/B- Cutoff
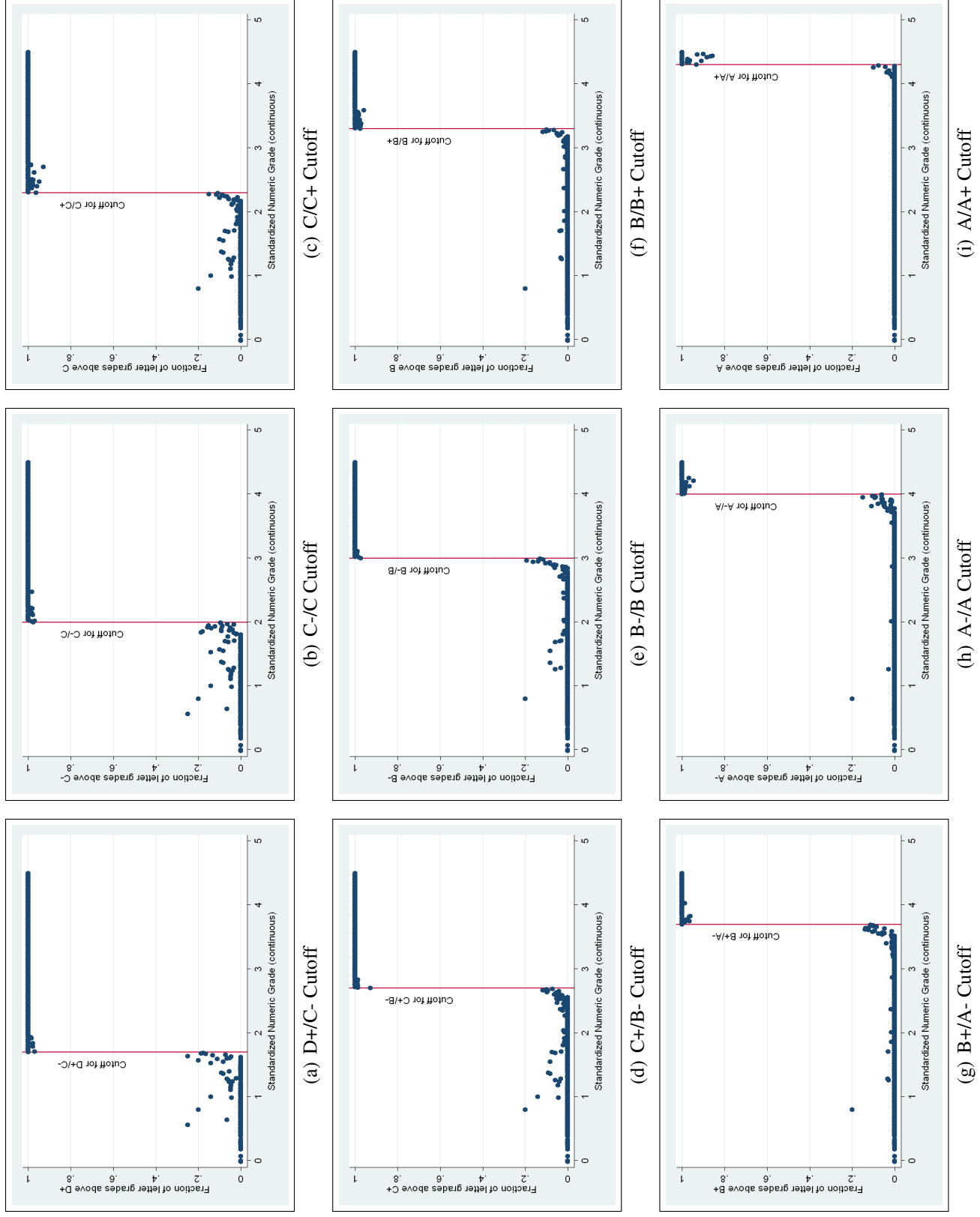
Figure 2: Histograms Around Each Cutoff

Figure 3: Discontinuity in Probability of Receiving a Given Grade Around Each Cutoff
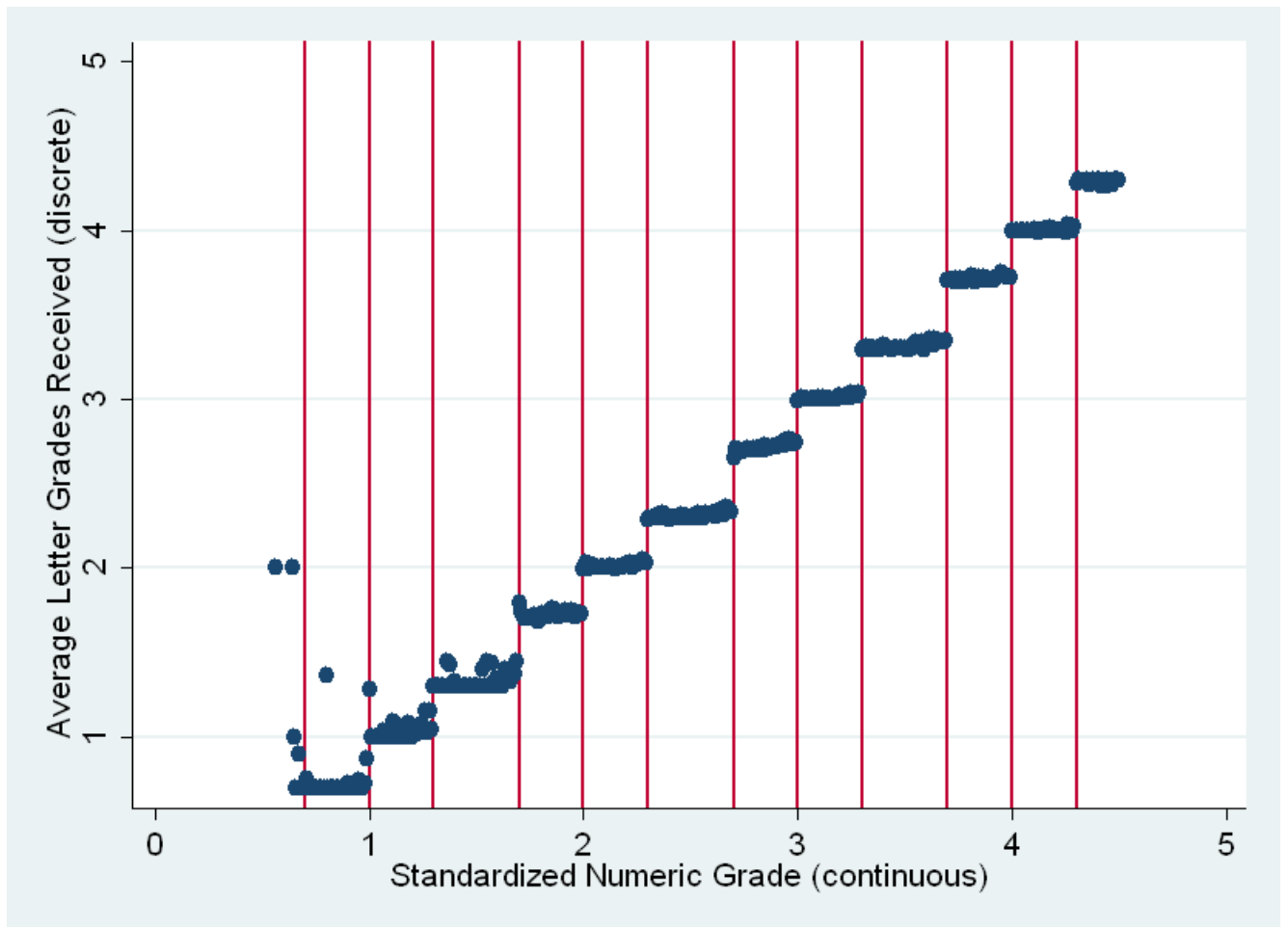
Figure 4: Average Letter Grades vs Numerical Score

Notes: Each point represents the average letter grade given to students in a given numerical score bin. The vertical lines show each cutoff value where 0.7, 1.0, 1.3, 1.7, 2.0, 2.3, 2.7 3.0, 3.3, 3.7 4.0 and 4.3 are the cutoffs for D-, D, D+, C-, C, C+, B-, B, B+,A-,A and A+ respectively.
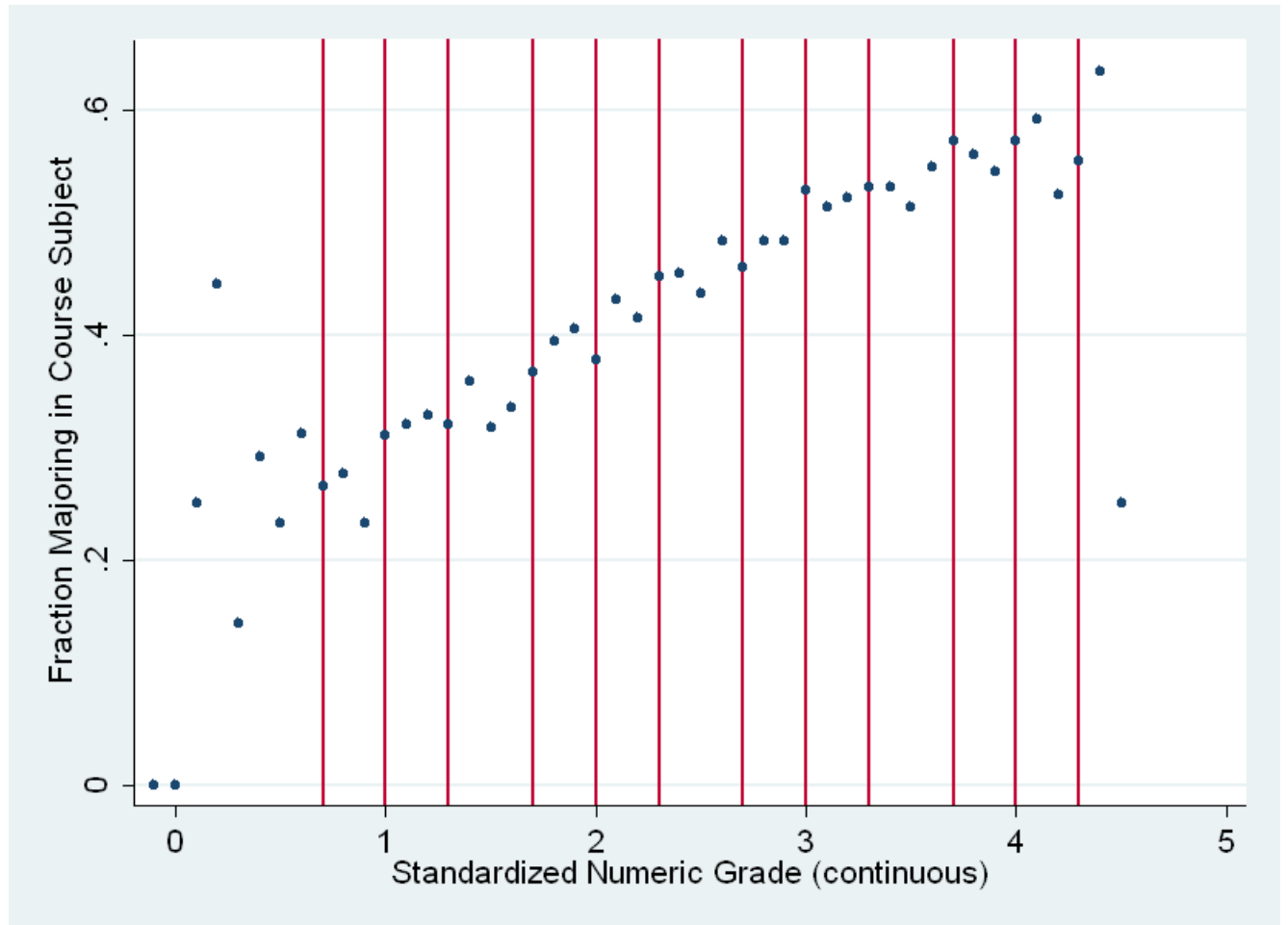
Figure 5: Fraction Majoring in Subject vs Numerical Score

Notes: Each point represents the fraction of students who major in the subject in a given numerical score bin. The vertical lines show each letter grade cutoff value where the cutoffs are the same as in Figure 4. Since the bin size is constant and the density is lowest at very high or very low scores, the variance is much larger at the extremes do to small sample sizes. The outliers at a numerical score of 4.5 and 0.4 represent very few students and thus these points should be interpreted with caution.
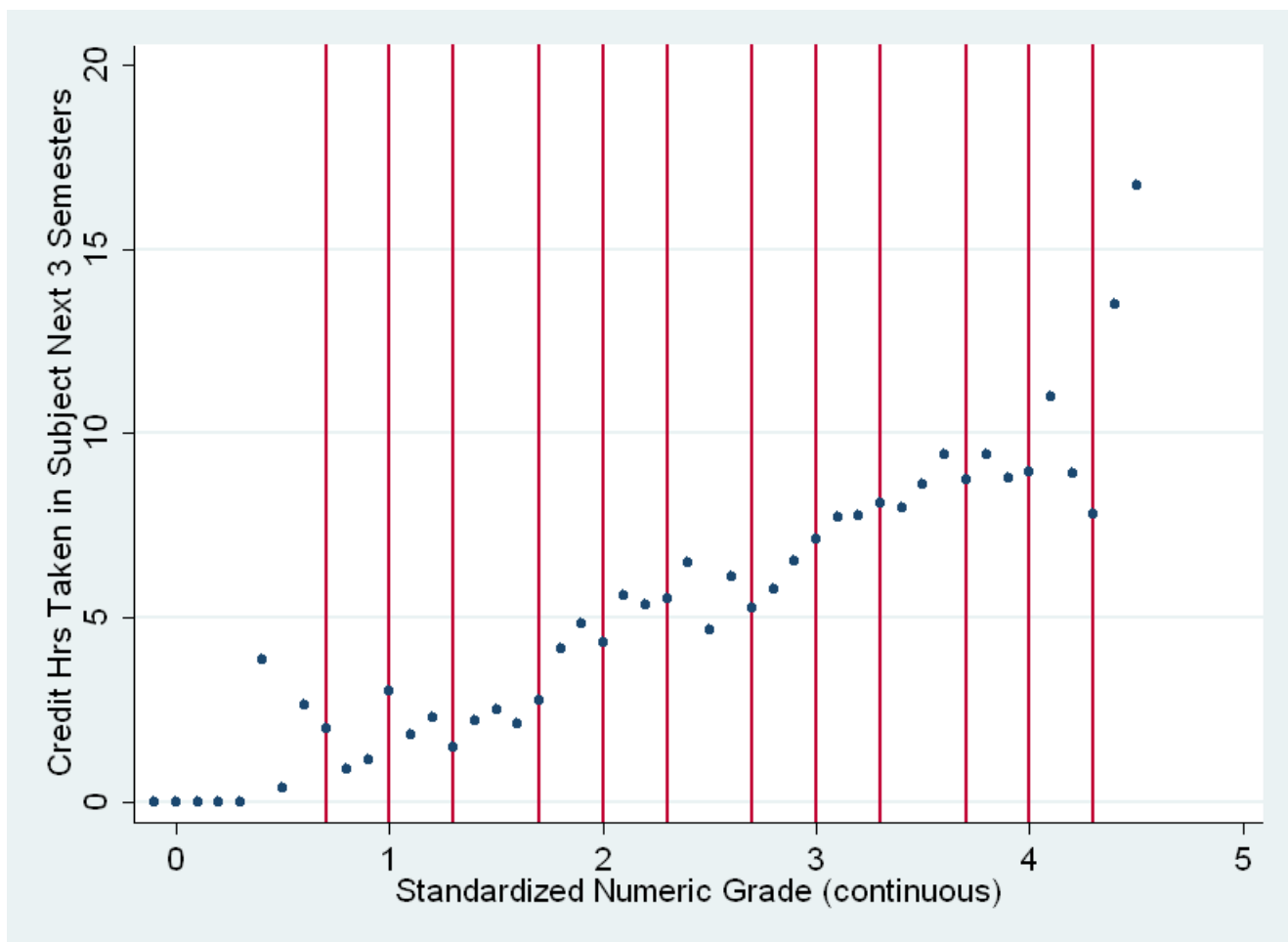
Figure 6: Credit Hours in Subject vs Numerical Score

Notes: Each point represents the fraction of students who major in the subject in a given numerical score bin. The vertical lines show each letter grade cutoff value where the cutoffs are the same as in Figure 4. Since the bin size is constant and the density is lowest at very high or very low scores, the variance is much larger at the extremes do to small sample sizes.

Figure 7: Local Linear Regression Predicting Major Choice Around Each Cutoff

Figure 8: Local Linear Regression Predicting Credit Hours Around Each Cutoff

Tables

Table 2: Descriptive Statistics

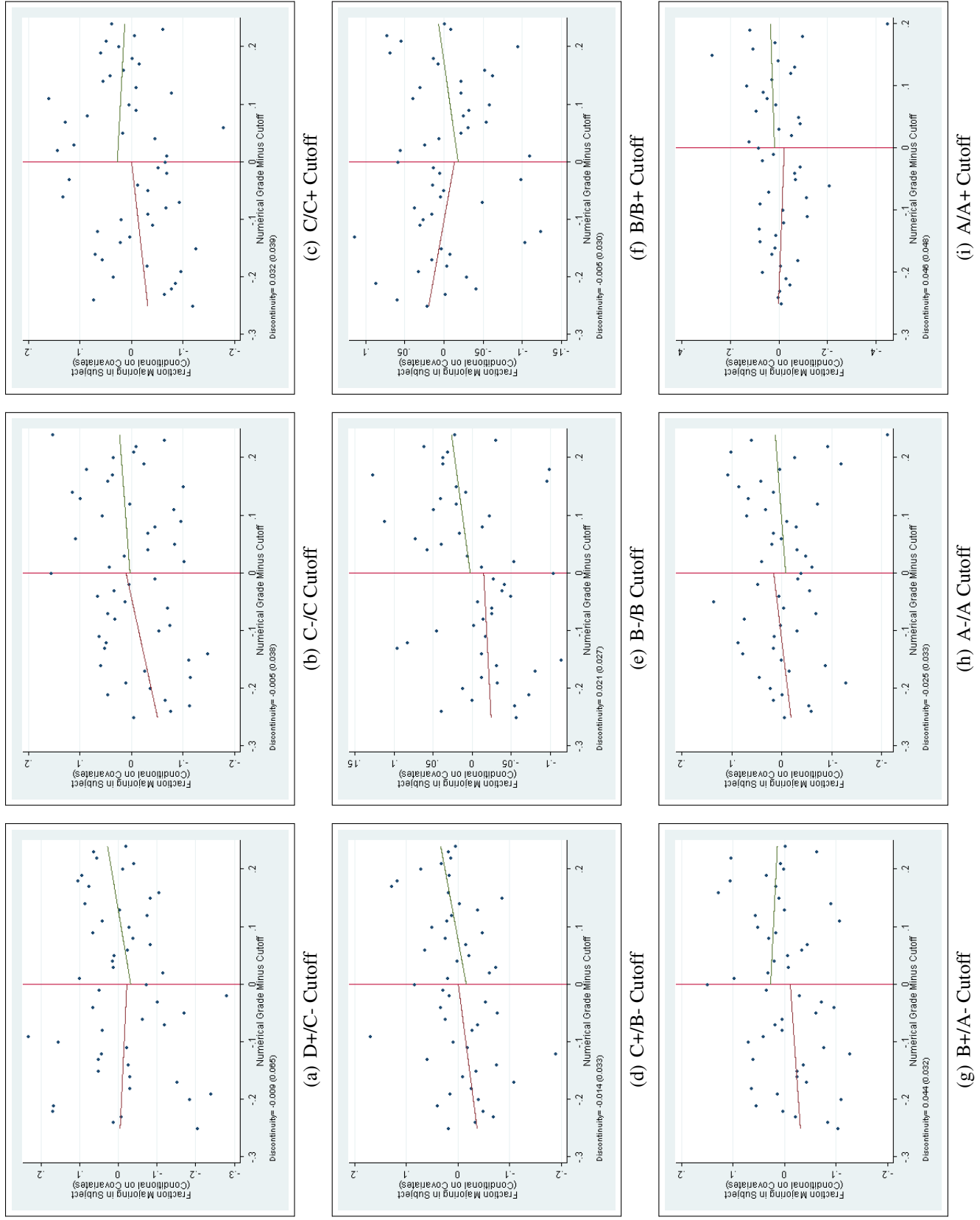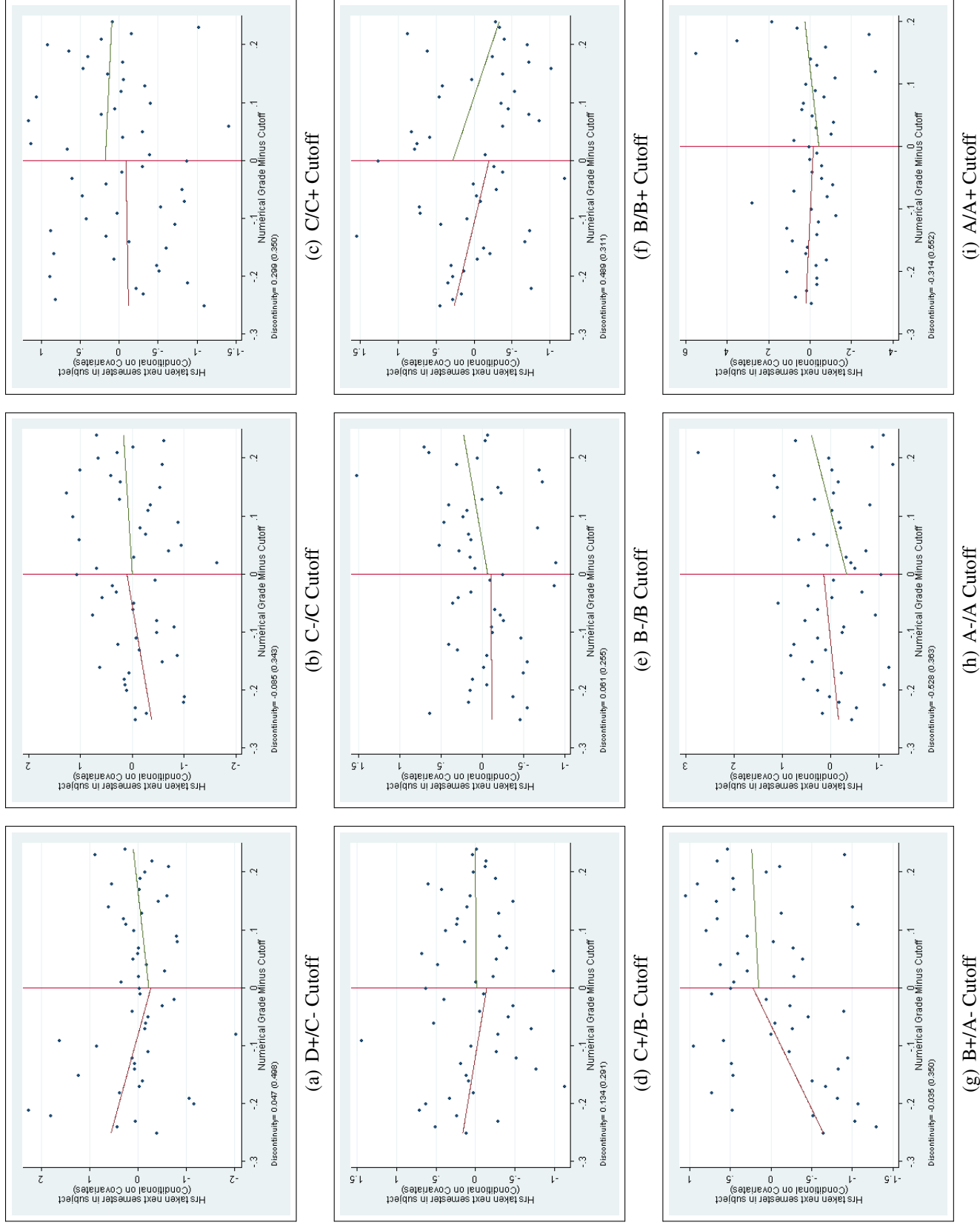| Discipline	Students Enrolled in Introductory Courses by Field			Restricted to Students who Eventually Major in the Introductory Course Subject		
	Economics	Engineering and Physical Science	Life Science	Economics	Engineering and Physical Science	Life Science
Major in Subject	0.17	0.59	0.53	1.00	1.00	1.00
Numeric Score (4.3 scale)	3.27	3.14	2.81	3.42	3.22	2.96
Intend to Major in Subject	0.03	0.69	0.36	0.16	0.84	0.61
Cumulative GPA	3.25	3.17	3.12	3.35	3.21	3.17
SAT or ACT equiv.	1393.12	1426.72	1386.17	1395.10	1429.27	1396.67
Female	0.43	0.41	0.60	0.44	0.40	0.57
Black	0.02	0.01	0.05	0.02	0.01	0.05
Hispanic	0.07	0.07	0.07	0.07	0.07	0.06
Observations	2,072	4,643	6,959	361	2,752	3,655

Table 3: Relationship between Letter Grades and Major and Course Choice

Dependent Variable:	Major in Subject			Credit Hrs in Subject During Following 3 Semesters		
	(1)	(2)	(3)	(4)	(5)	(6)
A plus	0.067*	0.008		-0.020	-1.028	
	(0.027)	(0.033)		(0.742)	(0.863)	
A	0.051**	0.009		-0.084	-0.802	
	(0.016)	(0.021)		(0.452)	(0.552)	
A minus	0.048**	0.025		0.724	0.337	
	(0.015)	(0.017)		(0.409)	(0.446)	
B	-0.011	0.013		-0.391	0.021	
	(0.013)	(0.015)		(0.342)	(0.392)	
B minus	-0.049***	-0.005		-0.442	0.308	
	(0.015)	(0.021)		(0.360)	(0.503)	
C plus	-0.081***	-0.014		-0.550	0.592	
	(0.016)	(0.027)		(0.391)	(0.653)	
C	-0.106***	-0.015		-1.045*	0.521	
	(0.018)	(0.035)		(0.443)	(0.842)	
C minus	-0.127***	-0.017		-1.285*	0.604	
	(0.020)	(0.042)		(0.501)	(1.006)	
Below C minus	-0.218***	-0.058		-2.215***	0.525	
	(0.023)	(0.058)		(0.532)	(1.340)	
Numerical score		0.072**	0.091***		1.226*	0.859***
		(0.024)	(0.008)		(0.567)	(0.191)
N	13,674	13,674	13,674	13,046	13,046	13,046

Incremental F-tests of whether dummy variables jointly contribute to model fit

Incremental F-test:	Column (3) → (2)	Column (6)→ (5)
	F statistic: 0.93 (p=0.498)	F statistic: 1.15 (p=0.324)

* Significant at 10%; ** significant at 5%; *** significant at 1%. Standard errors clustered at the student level reported in parentheses.

Note: The outcome is major choice in columns (1)-(3) and credit hours in column (4)-(6). All regressions also control for demographics, cumulative and current college GPA, credit hours taken contemporaneously with the courses analyzed, SAT score or ACT equivalent, an indicator for whether the student listed the major as their "intended major" on their application to LERU, and a course fixed effect. The omitted group for the letter grade dummies is B plus.

Table 4: Relationship between Letter Grades and Major and Course Choice (Only Females)

Dependent Variable:	Major in Subject			Credit Hrs in Subject During Following 3 Semesters		
	(1)	(2)	(3)	(4)	(5)	(6)
A plus	0.117**	0.058		1.176	0.371	
	(0.044)	(0.050)		(1.128)	(1.228)	
A	0.027	-0.015		-0.882	-1.450*	
	(0.024)	(0.030)		(0.569)	(0.649)	
A minus	0.063**	0.040		-0.102	-0.413	
	(0.022)	(0.024)		(0.534)	(0.561)	
B	-0.019	0.006		-0.745	-0.417	
	(0.019)	(0.021)		(0.426)	(0.476)	
B minus	-0.071***	-0.027		-1.332**	-0.732	
	(0.020)	(0.027)		(0.458)	(0.607)	
C plus	-0.079***	-0.012		-1.093*	-0.179	
	(0.022)	(0.035)		(0.475)	(0.739)	
C	-0.106***	-0.013		-1.231*	0.022	
	(0.024)	(0.044)		(0.533)	(0.925)	
C minus	-0.136***	-0.025		-2.571***	-1.060	
	(0.028)	(0.053)		(0.629)	(1.131)	
Below C minus	-0.216***	-0.054		-2.732***	-0.547	
	(0.031)	(0.072)		(0.626)	(1.463)	
Numerical score		0.073*	0.093***		0.984	0.936***
		(0.030)	(0.011)		(0.616)	(0.228)
N	6,953	6,953	6,953	6,696	6,696	6,696

Incremental F-tests of whether dummy variables jointly contribute to model fit

Incremental F-test:	Column (3) → (2)	Column (6)→ (5)
	F statistic: 1.27 (p=0.249)	F statistic: 1.75 (p=0.072)

* Significant at 10%; ** significant at 5%; *** significant at 1%. Standard errors clustered at the student level reported in parentheses.

Note: The outcome is major choice in columns (1)-(3) and credit hours in column (4)-(6). All regressions also control for demographics, cumulative and current college GPA, credit hours taken contemporaneously with the courses analyzed, SAT score or ACT equivalent, an indicator for whether the student listed the major as their "intended major" on their application to LERU, and a course fixed effect. The omitted group for the letter grade dummies is B plus. The entire table is restricted to women.

Table 5: Relationship Between Unobservables and Major and Course Choice

Dependent Variable:	Major in Subject		Credit Hrs in Subject During Following 3 Semesters	
	(1)	(2)	(3)	(4)
Numerical Score	0.091***	0.092***	0.856***	0.869***
	(0.008)	(0.008)	(0.192)	(0.193)
Letter Grade Was Raised	0.002	0.027	-0.130	0.966
	(0.022)	(0.077)	(0.404)	(1.617)
Letter Grade Was Raised x Numerical Score		-0.009		-0.395
		(0.027)		(0.589)
N	13,674	13,674	13,046	13,046

* Significant at 10%; ** significant at 5%; *** significant at 1%. Standard errors clustered at the student level reported in parentheses.