Grades, gender, and encouragement: A regression discontinuity analysis

Owen, Ann L.

Hamilton College

October 2008

Grades, gender, and encouragement: A regression discontinuity analysis

Ann L. Owen*
aowen@hamilton.edu
Hamilton College

October 2008

Abstract

This study employs a regression discontinuity design in order to provide direct evidence on the effects of grades earned in economics principles classes on the decision to major in economics and finds a differential effect for male and female students. Specifically, for female students, receiving an "A" for a final grade in the first economics class is associated with a meaningful increase in the probability of majoring in economics, even after controlling for the numerical grade earned in the class. This suggests that, for female students, the feedback that is embedded in the course letter grade has an encouragement effect on their decision to study economics further. It finds no evidence of a similar effect for male students.

*198 College Hill Road, Clinton, NY 13323, phone: 315-859-4419, fax: 315-859-4477.
I am grateful to Jennifer Borton, Emily Conover, Wayne Grove, Elizabeth Jensen, David Toomey, Julio Videras and Stephen Wu for helpful comments. I am especially indebted to Rachel Friedberg and David Weil for their assistance and advice. Shirin Rashid provided valuable research assistance.

Grades, gender, and encouragement: A regression discontinuity analysis

October 2008

Abstract

This study employs a regression discontinuity design in order to provide direct evidence on the effects of grades earned in economics principles classes on the decision to major in economics and finds a differential effect for male and female students. Specifically, for female students, receiving an "A" for a final grade in the first economics class is associated with a meaningful increase in the probability of majoring in economics, even after controlling for the numerical grade earned in the class. This suggests that, for female students, the feedback that is embedded in the course letter grade has an encouragement effect on their decision to study economics further. It finds no evidence of a similar effect for male students.

1 Introduction

Many studies have focused on the role that performance in economics classes plays in the decision to become an economics major and sensibly conclude that students who perform better in economics principles classes are more likely to major in economics. (See, for example, Horvath, Beaudin and Wright, 1992; Dynan and Rouse, 1997; or Chizmar, 2000). However, it is difficult to infer causation from this correlation: Does the better performance cause an interest in economics or does the interest in economics cause better performance? This study employs a regression discontinuity design in order to provide more direct evidence on the effects of grades earned in principles classes on the decision to major in economics and finds a differential effect for male and female students. Specifically, it finds that, for female students, receiving an "A" for a final grade in the first economics class is associated with a meaningful increase in the probability of majoring in economics, even after controlling for the numerical grade earned in the class. This suggests that, for female students, the feedback that is embedded in the course letter grade has an encouragement effect and positively influences the probability of studying economics further. It finds no evidence of a similar effect for male students.

To reach these conclusions, data from over 1,300 introductory economics students in five different semesters were used. First, a preliminary probit analysis predicting the probability of becoming an economics major shows that the receipt of an A in the class, even after including the raw score in the course (i.e., the numerical course grade) is significant in determining the probability of majoring in economics for female students. The nature of the data, however, is ideal for a regression discontinuity design which allows for more robust inferences about causality. Specifically, if the numerical course grade is a good measure of the student's ability to do well in economics classes, it should be positively correlated with the probability of majoring in economics. If grades cause students to major in economics, then we should see a discontinuous jump in the probability of majoring in economics at the threshold numerical score for receiving an "A" in the course. Because students who are close to either side of the threshold are very similar to each other, the presence of a discontinuity provides more convincing evidence that the treatment of receiving an A actually caused the increase in probability. This study documents such a jump for female students, but not for male students, leading to the conclusion that grades may affect female students in a different way than they affect male students.

Because it suggests that the experience in the first economics class has a greater effect on female students than male students, this finding is more broadly related to the findings in Jensen and Owen (2001, 2000). They report that female students are less likely to take introductory economics thinking that they will be an economics major, and female students are more likely to change their minds about future economics courses after they take introductory economics. Interestingly, our results suggest that in addition to their overall performance in the course, the instructor feedback in the form of a letter grade has an additional encouraging effect for women. We find no evidence that the letter grade, after controlling for course performance, has the same impact on male students. This finding may also be related to the findings of some psychologists who have argued that female students are more sensitive to instructor feedback (Beyer and Langenfeld, 2000).[1]

Positive feedback may also be particularly important for female students in economics if they did not expect to do well in the class. Steele (1997) argues that women in quantitative fields face a stereotype threat that makes it more difficult for them to identify with these academic domains. In such a case, receiving an A might help to overcome the stereotype, while receiving a lower grade may simply reinforce it. In fact, Ballard and Johnson (2005) demonstrate that negative stereotyping exists among female students in introductory economics classes by documenting that female students have lower expectations for their performance, even after controlling for family background, academic experience, and mathematics experience. In a corroborating finding, Nowell and Alston (2007) survey students in economics and quantitative courses and find that male students exhibit more overconfidence in their abilities than female students.

While the study described below is not able to provide evidence for the reason that grades affect the decision of female students, it does provide convincing evidence of a causal role. Regardless of the reason for this effect, this is an important result for instructors of economics because it documents a causal link between grades and student choices. These conclusions are demonstrated in the following three sections: The next section describes the data and methods, the third section presents results, and some concluding thoughts are provided in Section 4.

---

[1] In a more dated study, Dweck, Davidson, Nelson and Enna (1978) provide evidence that female students are more likely to attribute negative feedback to their own low ability than male students.

2 Data and Methods

To explore the effects of grades received in introductory economics classes on the probability of majoring in economics, approximately 1,800 students who took introductory economics in five different semesters at a highly-selective research university located in the Northeastern United States over the time period 2003 to 2006 were studied. Grade data from the instructors were matched up with registrar data on the subsequently declared major. Specifically, course numerical grades, course letter grades, student gender, and the major that each student later declared were used in the analysis. The one-semester introductory course which the students completed is the first class required for the economics major at this institution and is the prerequisite for intermediate microeconomics and macroeconomics.[2] The course is typically team taught with one male and one female instructor; however, in one semester in the sample, the two instructors were both male, and in one semester, there was only one female instructor.

Students who failed the course or who were taking this course as either a junior or a senior were removed from the sample as these students would not be in a position to declare an economics major after completing the course. Several students who had not yet declared a major were also removed from the sample. Finally, because the study focuses on the effect of letter grades awarded, students taking the course on a pass/fail basis (did not receive a letter grade) were also excluded. The resulting sample was 1,355. Descriptive statistics for key variables in the analysis are in Table 1.

Different sections of the course had slightly different distributions for the numerical course scores. In order to make them comparable, the numerical course scores

[2] As is the case at most institutions, students can place out of this course with satisfactory performance on AP exams, the international baccalaureate (IB) HL exam in economics, or the British A-level exam.

were standardized so that each section's numerical grades had a mean of zero and a standard deviation of one. Even so, the threshold standardized course score for receiving a letter grade of "A" in the class was not identical across sections so, prior to pooling all the data, the threshold for receiving an "A" was subtracted from each student's standardized score. In other words, in the transformed variable, any student with a score equal to zero or greater received an A for the course. A histogram of this variable for female and male students is shown in Figures 1A and 1B.

At this institution, students who pass the class receive a letter grade of either an A, a B, or a C. There are no pluses or minuses awarded and students who earn less than a C receive no credit for the course.[3] As can be seen in Table 1, 35 percent of the students received an A, while 39 percent of the students received a B. Twelve percent of the students declared economics as a major subsequent to completing the class, while female students represented slightly less than half the sample, at 44 percent.

Using these data, this paper employs two different methods to examine the effect of grades on the decision to major in economics. Both are based on the idea that the numerical score earned in the course is a measure of the student's performance that would reflect the student's ability and interest in economics. Theoretically, the numerical score should be positively related to majoring in economics. However, inferring causality from a positive coefficient in a probit model predicting whether or not the student becomes an economics major is problematic. The coefficient estimates may be biased if students intending to major in economics have systematically different performances. One source of bias would be if students of overall higher (lower) ability

[3] The absence of pluses or minuses makes the distinction between letter grades sharper and may make the magnitude of the effect larger.

were interested in economics. In this case, their numerical scores in the course would be systematically higher (lower) than the other students. A similar issue arises if students interested in economics work harder in economics classes because it is the subject in which they intend to major.

The first method used to isolate the effect of the grade received is to enter a dummy variable indicating the letter grade received in the class in a probit estimation predicting the probability of majoring in economics. In these estimations, the course grade is also entered (in the standardized form discussed above) to control for the student's interest and ability in economics. In other words, any residual effect of the letter grade on the probability of majoring in economics can be interpreted as a "treatment" effect of receiving that letter grade. Thus, by controlling for the raw score in the course, this method allows us to isolate the effect of receiving that particular grade. By interacting the letter grade received with a dummy for gender, we are able to determine if the treatment of receiving a particular letter grade varies among male and female students.

Of course, the dummy variable in the probit regression may simply be picking up a non-linearity in the relationship. Although we also allow the numerical score to enter in a non-linear fashion in some of the probit estimations, an empirical technique that better matches up with the idea that there is a "treatment" effect of receiving a letter grade is regression discontinuity analysis. For example, if there is an effect of receiving an A in introductory economics, then there should be a discontinuous jump in the probability of majoring in economics at the threshold grade for receiving an A (in our case at zero), even though students who are close to either side of the threshold are arguably very similar. Therefore, if the only difference between students who are close to either side of the threshold is that one set received an A and the other did not, then we can attribute a causal role to receiving the A in determining the probability of becoming an economics major. Regression discontinuity analysis allows us to estimate the size of the jump in probability and test for its statistical significance. Essentially, the procedure entails estimating local regressions near the hypothesized threshold, using these regressions to predict the probability of majoring in economics, and then using a Wald test to determine if these differences in probabilities are statistically significant. Checks for robustness of the conclusions include 1) ensuring that there are not other thresholds in the data where they should not be, and 2) varying the definition of "local" regressions to include different bandwidths around the threshold.[4]

Using regression discontinuity requires a few assumptions. First, there must be a discontinuous jump in the treatment at the threshold. This assumption is easily satisfied in this case because the treatment changes discretely from a B to an A or a C to a B at a threshold score in the class. Second, individuals must not be able to manipulate their treatment. In this case, this would amount to students being able to either select their own grades, regardless of their scores, or manipulate the instructor to award grades that are inconsistent with their scores. Both of these situations are implausible in this context so this criterion is met as well.

Essentially, in order for the discontinuity analysis to provide valid causal evidence, we want students who are just above and just below the threshold to be

---

[4] Interested readers can find additional advice for implementing a regression discontinuity analysis in Nichols (2007a) or Imbens and Lemieux (2008).

identical in relevant characteristics.[5] Therefore, one might also be concerned if instructor grading methods resulted in this assumption being violated. Specifically, if instructors determined the threshold by looking for "breaks" in the distribution, there would be a cause for concern. However, in our data, the large number of students in each section minimizes the potential for meaningful breaks in the data and an examination of the raw grades in each section supports this conclusion: The average difference between the cutoff numerical grade for an A and the next lowest grade is .17 (out of a 100 point scale). Especially given the imprecision with which a grade reflects any student's effort, ability or knowledge, this is not a meaningful difference and we can be confident that students just above and just below the threshold performed similarly.

A final assumption in this study is that a linear probability model is employed in the estimation of the local regressions. When we use a linear probability model in place of a probit in estimations that mirror those presented in our preliminary analysis, we find similar results, giving us confidence that the use of the linear probability model is not affecting our conclusions.

3 Results

3.1 Main sample results

Prior to discussing the regression discontinuity analysis, a preliminary probit analysis is presented in Table 2. In each column, the probability of declaring economics as a major is predicted using a variety of control variables. In the first column, only the standardized grade is used as an explanatory variable (less the threshold for receiving an A to make this numerical score comparable across sections). As expected, this variable is

[5] Unfortunately, we do not have access to additional student characteristics other than those used in the analysis to verify this claim directly.

positively and significantly related to majoring in economics. In the second column, a dummy variable for the receipt of an A is added, but now both measures of course performance, though positively related, are statistically insignificant. The third column in Table 2 shows a separate effect of receiving an A for male and female students. As expected, female students are less likely to major in economics, but the interaction of being female and earning an A has a positive and statistically significant effect on the probability of majoring in economics, whereas there is no evidence of a similar effect for male students. The results in column 4 of Table 2 finds similar effects of earning a B. (The omitted category is receiving a C in the course.) Although the coefficient on the interaction of receiving a B and being female is smaller than that for receiving an A and being female, it remains positive and statistically significant. Again, there is no evidence for similar effects among male students.

The marginal effects of receiving an A for female students associated with the estimations reported in Table 2 are notable. For example, the marginal effect of a female student receiving an A in the estimation reported in column 4 is .24. This more than compensates for the negative marginal effect of .15 of being female.

Columns 5 and 6 of Table 2 add dummy variables for different sections of the class. As mentioned earlier, one dimension in which the sections differ is in the gender of the instructors. In addition, some sections were taught in the Fall semester, while others were taught in the Spring semester. There may be seasonal effects as well, with students more interested in economics taking their first class in their first semester in the Fall. While there is not enough variation in the data to control for each of these effects separately, section dummy variables should allow us to capture these effects together. In fact, the section that has a positive, significant effect on the probability of majoring in economics was taught by one female instructor in the fall semester. The one section that has a negative interaction effect with female students was taught by two male instructors in the spring semester.[6] Nonetheless, the conclusion regarding the effect of receiving a better letter grade for female students remain intact, even after controlling for section characteristics.

Finally, the results in Column 7 are for a specification that allows for a non-linearity in the relationship between the standardized grade less the threshold for receiving an A and the probability of majoring in economics. Although there is no evidence for a non-linearity in this relationship, the conclusions drawn earlier are robust to allowing for this possibility. Although not reported in the table, we also found that the interaction of being female and the higher order terms for the numerical score variable are insignificant, suggesting that the effect of being female has an intercept, but not a slope effect.

The preliminary analysis in Table 2 suggests that a regression discontinuity analysis may prove fruitful if male and female students are allowed to be affected differently by grades in the analysis. As a first step, Figures 2 and 3 present a graph of the probability of majoring in economics against the standardized course grade less the standardized threshold for receiving an A. To calculate these probabilities, the data were grouped into "bins" that were .2 standard deviations wide. Then, the probability of majoring in economics for male students was calculated as the number of male students majoring in economics divided by the total number of male students in that bin. A

[6] Rask and Bailey (2002) find evidence that undergraduates are more likely to declare majors in fields in which they have been taught be instructors of the same gender or race.

similar calculation was performed to obtain the probability for female students. A vertical line is drawn on each graph at 0, the threshold for receiving an A in the class. Two points are clear from comparing Figures 2 and 3. First, the probability of becoming an economics major for male students is higher. Second, there is no easily discernible relationship between the grades that male students received and the probability of becoming an economics major, but there does seem to be a clear pattern for female students, with higher probabilities associated with better grades. Visual inspection also suggests that there may be a discontinuous jump at the threshold for receiving an A in the probability of majoring in economics for female students. Although this analysis is admittedly ad hoc, this point will be investigated more rigorously in the analysis described below.

As discussed above, to employ a regression discontinuity analysis, local linear or "kernel" regressions are estimated on either side of the hypothesized threshold.[7] Essentially, regressions are estimated on small samples near the threshold, with the bandwidth determining how small the sample is. Figures 4 and 5 present the smoothed values of the results of the local linear regressions in which the probability of majoring in economics is predicted by the standardized grade less the A threshold. At the hypothesized discontinuity, the graph is not connected; however, in order to determine if the two predicted probabilities at either side of the threshold are in fact different from each other, it is necessary to perform a Wald test. Before reporting these results, it is necessary to comment on the fact that the graph in Figure 4 actually depicts a decline in the probability of majoring in economics for female students who were near, but below the threshold for an A. This same pattern is not evidenced in Figure 5, which examines

[7] The regression discontinuity routine by Nichols (2007b) was used in this analysis.

the probability of male students majoring in economics. Although these graphs are drawn for particular bandwidths (.10), this qualitative feature of both graphs is present in alternative bandwidth selections. The decreased probability of majoring in economics for female students who earned a high B, but not quite enough for an A is consistent with a discouragement/encouragement effect of letter grades for female students.

The results of the Wald tests indicating whether or not the differences in predicted probabilities are significant appear in Table 3. Table 3A presents the differences in the probabilities on either side of the A threshold for male and female students. Each estimate is made three times, for differing bandwidths. The bandwidth reported in the first row of Table 3A is .05. Because the range of the standardized grade runs from -4.6 to 1.3, this bandwidth is only a small fraction of the entire range. However, as depicted in Figure 1, there are many more observations near the threshold; there are approximately 30 observations in this bandwidth on either side of the threshold. The smaller bandwidth of .025 (row 2 of Table 3A) contains about half that number of observations on either side of the threshold and the bandwidth of .10 (row 3 of Table 3A) contains about twice as many. As bandwidth decreases, we are estimating the regressions with data points that are closer to the hypothesized discontinuity; however, the accuracy of the estimation may suffer from a small number of data points. Nonetheless, for all three bandwidths, there is a positive and significant jump in the probability of becoming an economics major at the threshold for receiving an A in the class for female students. The estimates of the magnitude of this increase in the probability are consistent in all three estimations, ranging from 15 to 18 percentage points. In contrast, the corresponding analysis for male students yields more widely varying estimates and reveals no significant discontinuity in the probability of majoring in economics at the threshold for receiving an A.

Table 3B displays the evidence for a discontinuity at the B threshold. In order to perform this analysis, it was necessary to transform the data by subtracting the threshold grade for a B rather than an A from the standardized numerical score, but the same steps in the analysis described above were performed. However, unlike the probit results, the results in Table 3B provide no evidence for a discontinuity at the B threshold for either male or female students.

Finally, in order to provide a convincing analysis, it is necessary to demonstrate that the evidence found for a discontinuity is not spurious. One way to do this is to show that breakpoints do not exist at other locations where they should not. Figure 3 suggests a few places in which we might look for additional breakpoints. Visual examination of this figure suggests that looking for a breakpoint at -1, -0.8, -0.6 or 0.6 may be worthwhile. Table 3C presents the results from this exercise, showing that there is no evidence of a breakpoint at 3 of these 4 points. When the standardized grade less the threshold for the A is equal to -0.8, there is some weak evidence of a discontinuity. Two of the three differences are significant at the 10 percent level; however, it is difficult to put much confidence in these results. First, unlike the results for the discontinuity at the A threshold, the two significant differences in probability are of very different magnitudes (.09 vs. .28). Second, the two significant differences are for the smallest and the largest bandwidths. The intermediate bandwidth, which contains an overlapping sample, does not produce significant results, calling the validity of these results into question. That said, one reason why a discontinuity may exist at -0.8 is that it is close to the threshold for receiving a B in many of the sections. Therefore, the weak evidence for a discontinuity here is not inconsistent with the idea that letter grades do affect female students.

3.1 Results from alternative sample

The main results presented above are based on a sample from an institution that does not allow students to earn half letter grades (i.e., pluses or minuses). Although the nature of this grading system provides a unique opportunity to implement regression discontinuity tests and make a more confident assertion about causality, one may wonder if these results can be generalized. At many institutions, pluses and minuses are routinely attached to letter grades and it is interesting to investigate whether the same effects exist when the distinction between A's and B's is blurred by the existence of pluses and minuses. In this section, we briefly discuss the extension of our results to such a sample.

The data contained in the alternative sample is obtained from 19 different sections of introductory microeconomics taught at a selective residential liberal arts college in the Northeast during the 2000 to 2007 time frame. The data come from four different instructors, all male. Except for one section, all sections of introductory microeconomics were taught in the Fall semester. At this institution, students who pass the class can earn letter grades ranging from D- to A+. Students in this sample have a different experience than those in the research university described earlier. Notably, class sizes are significantly smaller, with the average number of students in each section being 29 (vs. 374 at the research university). Unfortunately, this implies a considerably smaller sample of 384 students after juniors, seniors, and students who had not yet declared a major are removed from the sample. The fact that there are now 11 possible breakpoints rather than two, makes this a less than ideal sample for regression discontinuity tests. In addition, the smaller sample, combined with a smaller percentage of females in the sample, and a smaller percentage of students earning A's precludes us from bootstrapping standard errors in the regression discontinuity tests. Therefore, we can only present the preliminary probit results as suggestive evidence of the overall phenomena—grades have an encouragement effect for female students.

The descriptive statistics for this alternative sample in Table 4 suggest three interesting differences between this sample and that of the research university. First, the fraction of students who take introductory economics and then later major in economics is relatively large. This is likely related to the fact that this college has no distribution requirements; only one other small program requires its students to take introductory economics as part of that program's requirements. Because few students other than economics majors are required to take introductory economics at this institution, students in these classes are probably more likely to be in the classes because of an interest in economics. Second, and possibly related to this first observation, is that there are fewer female students in these classes than at the research university. Finally, the fraction of students who earn a grade in the A range (A, A-, A+) is considerably smaller, indicating different grading standards at the two different institutions. On the one hand, this may create a larger impact of receiving an A if earning an A at this institution carries more meaning. On the other hand, the ability to earn a B+ may decrease the distinction associated with receiving a grade in the A range.

The data manipulation described earlier is repeated on this sample and Table 5 presents the results from a probit analysis showing corroborating results.[8] Results in Column 3 of Table 5 show that receiving an A and being female is associated with a higher probability of subsequently declaring economics as a major, but no evidence of this effect is found for male students who receive A's. These effects exist even after controlling for the student performance in the class via the standardized course grade. Unlike the earlier results, there is no evidence of an effect of receiving a grade in the B range; but, as before, including this effect does not make the effect of receiving an A for female students insignificant (Column 4 of Table 5). Columns 5 and 6 show that the positive and significant effect of receiving an A for female students is robust to the inclusion of dummy variables for instructor. The dummy variables for the four male instructors, (not shown here), were statistically insignificant. Finally, the results in column 7 indicate that the conclusions drawn from the earlier results are robust to allowing for nonlinearity in the effect of the numerical course grade.

Although the nature of this sample does not allow us to perform the regression discontinuity tests, the corroborating evidence presented here does suggest that the effect identified in the larger sample may be generalized to other contexts. This evidence supports the idea of an encouragement effect of grades for female students in introductory economics.

[8] Because there are many more smaller sections in this sample, rather than including section dummy variables, the estimations in columns 5 through 7 in Table 5 includes instructor dummy variables. Although not shown, the instructor dummy variables are insignificant. Because one instructor had few female students, none of whom majored in economics, it is not possible to estimate instructor and female student interactions in the probit specification.

4 Conclusion

This study has used two different methods to demonstrate the effect of grades in introductory economics on the probability of majoring in economics. Both methods suggest that female students respond to the grades given and have higher probabilities of majoring in economics when they receive an A in introductory economics. Although this study can document the treatment effect of receiving an A, it cannot discern the reasons why female students react differently. One idea that has been proposed in the psychology literature that may be relevant here is that female students may be more likely to interpret the grades received as a reflection of their ability. In addition, these effects may be accentuated by stereotype threat: a low grade for a female student may simply act to confirm the stereotype that she should not be good at economics, and a high grade may be necessary to overcome that expectation.

Regardless of the reason, these results have important implications for instructors. In particular, tough grading standards, even applied fairly, may disproportionately discourage female students from studying economics. Instructors who have such grading policies should be sensitive to these effects and look for other means to encourage female students.

5 References

Ballard, Charles and Marianne Johnson, 2005, "Gender, Expectations, and Grades in Introductory Economics at a US University," *Feminist Economics* 11(1): 95-122.

Beyer, Sylvia and Kelly Langenfeld, 2000, "Gender Differences in the Recall of Performance Feedback," ERIC working paper #ED444079.

Chizmar, John F., 2000, "A Discrete-Time Hazard Analysis of the Role of Gender in Persistence in the Economics Major," *Journal of Economic Education* Spring: 107-118.

Dweck, Carol S., William Davidson, Sharon Nelson, and Bradley Enna, 1978, "Sex Differences in Learned Helplessness: II. The Contingencies of Evaluative Feedback in the Classroom and III. An experimental Analysis," *Developmental Psychology*, 14(3): 268-276.

Dynan, Karen E. and Cecilia E. Rouse, 1997, "The Underrepresentation of Women in Economics: A Study of Undergraduate Economics Students," *Journal of Economic Education* 28(4): 350-368.

Horvath, Jane, Barbara Q. Beaudin, and Sheila P. Wright, 1992, "Persisting in the Introductory Economics Course: An Exploration of Gender Differences," *Journal of Economic Education*, Spring: 101-108.

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

Jensen, Elizabeth J. and Ann L. Owen, 2001, "Pedagogy, Gender, and Interest in Economics," *Journal of Economic Education*, Fall: 323-343.

Jensen, Elizabeth J. and Ann L. Owen, 2000, "Why are Women Such Reluctant Economists? Evidence from Liberal Arts Colleges," *American Economic Review* 90(2): 466-470.

Nichols, Austin, 2007a, "Causal Inference with Observational Data," *Stata Journal*, StataCorp LP, vol. 7(4), pages 507-541.

Nichols, Austin, 2007b. rd: Stata module for regression discontinuity estimation. http://ideas.repec.org/c/boc/bocode/s456888.html.

Nowell, Clifford and Richard M. Alston, 2007, "I Thought I got an A! Overconfidence Across the Economics Curriculum," *Journal of Economic Education*, Spring: 131-142.

Rask, Kevin N. and Elizabeth M. Bailey, 2002, "Are Faculty Role Models: Evidence from Major Choice in an Undergraduate Institution," *Journal of Economic Education*, 33(2): 99-124.

Steele, Claude, 1997, "A Threat in the Air: How Stereotypes Shape Intellectual Identity and Performance," *American Psychologist* 52(6): 613-629.

Table 1: Descriptive Statistics

Variable	Mean	Std. Dev.	Definition
Female	0.44	0.50	=1 if female
Economics Major	0.12	0.33	=1 if economics major
A	0.35	0.48	=1 if A in course
B	0.39	0.49	=1 if B in course
Standardized Grade – Threshold for A	-0.45	0.87	Numerical course score standardized to a mean of 0 and standard deviation of 1 by section minus the standardized threshold for obtaining an A in that section.

Figure 1A

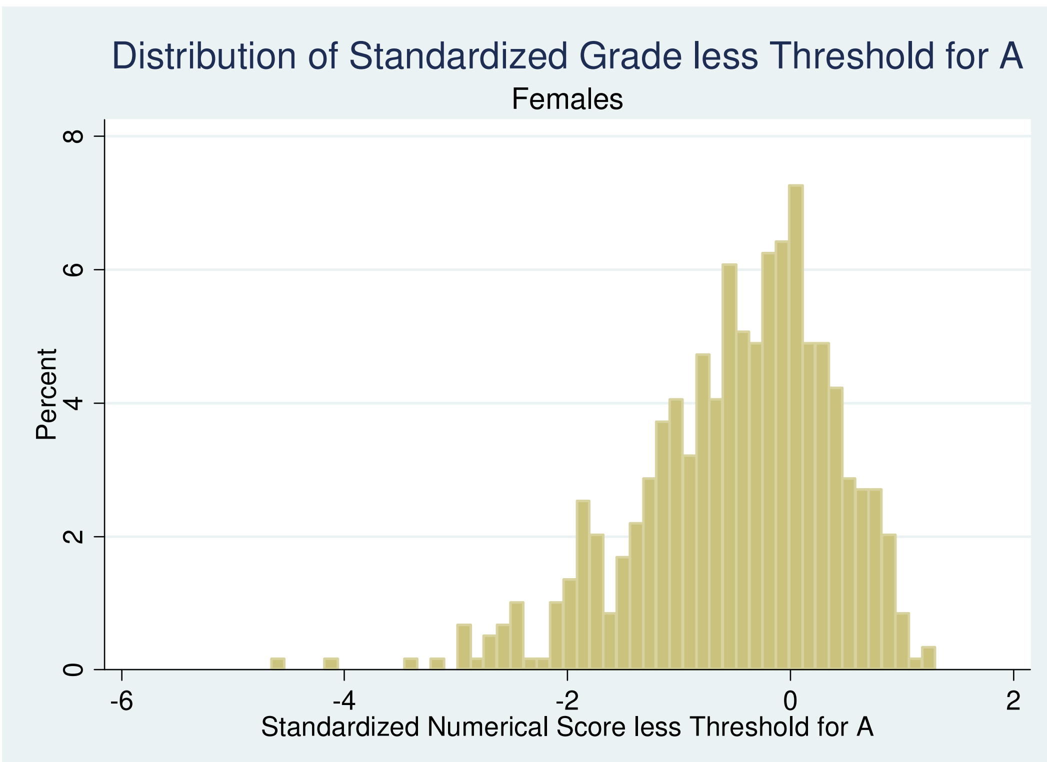
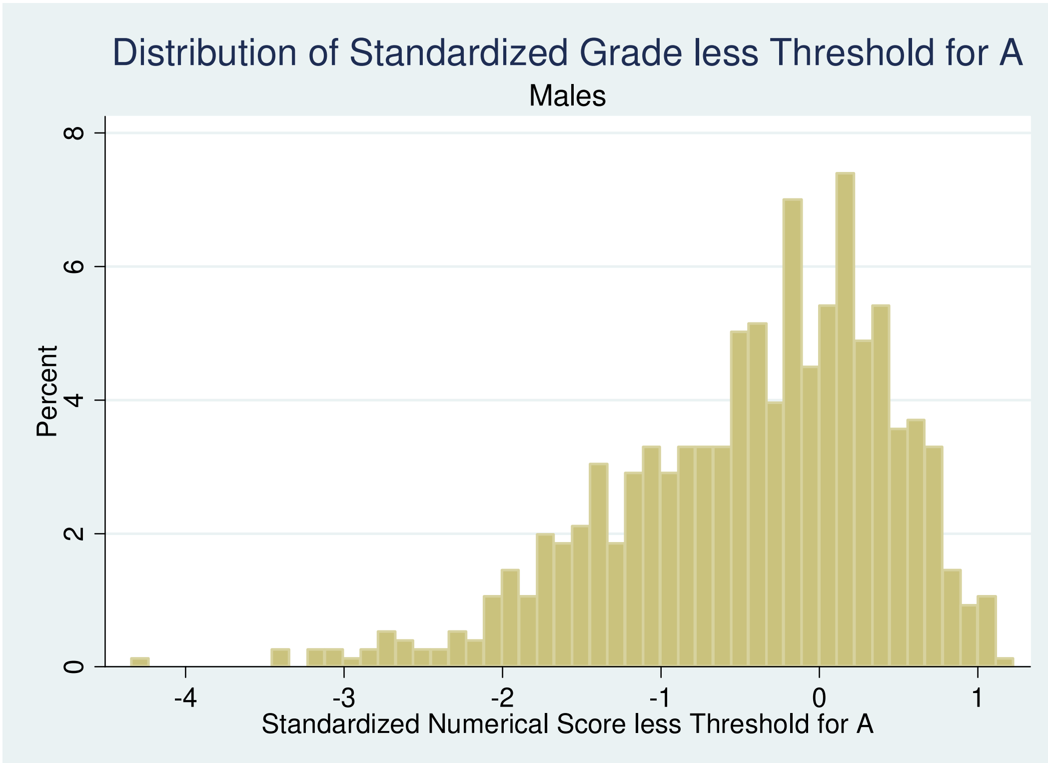


Figure 1B



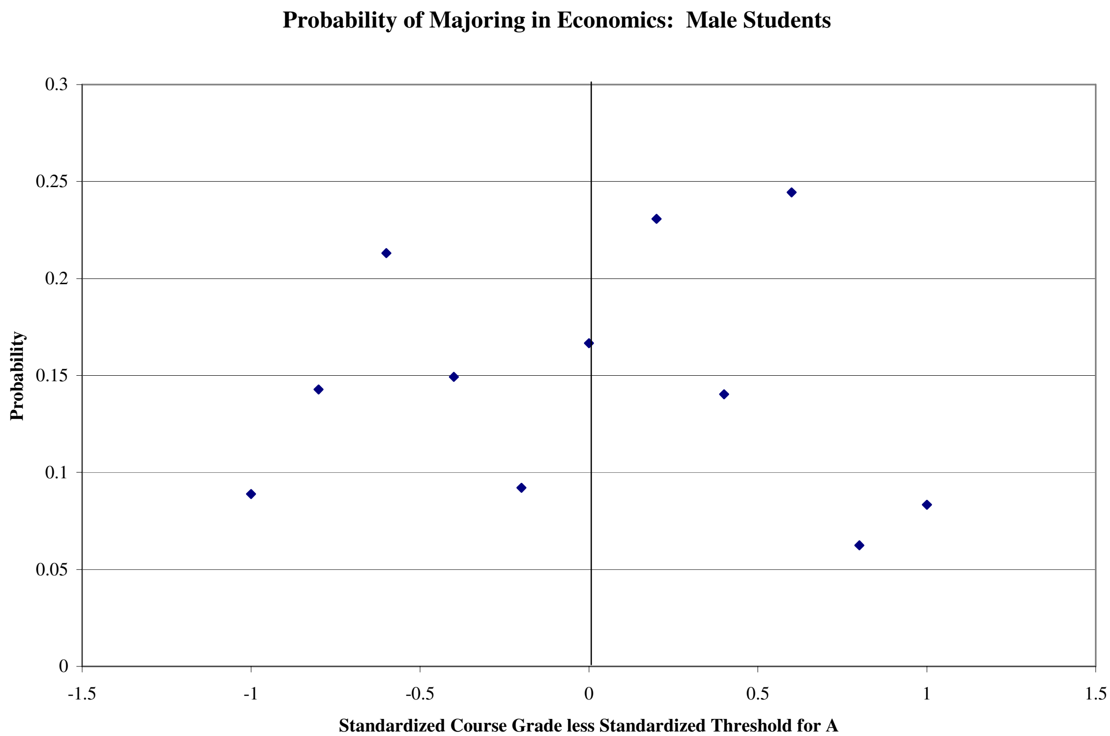
Distribution of Standardized Grade less Threshold for A
Males
Percent
0
2
4
6
8
-4
-3
-2
-1
0
1
Standardized Numerical Score less Threshold for A

Table 2: Preliminary Analysis: Probit Results, Probability of Majoring in Economics

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Standardized grade – threshold for A	0.189	0.102	0.108	0.017	0.011	0.020	-0.114
	(3.47)**	(1.29)	(1.34)	(0.14)	(0.09)	(0.16)	(0.64)
A		0.197					
		(1.48)					
Female*A			0.373	0.893	0.861	0.847	1.027
			(2.10)**	(2.64)**	(2.46)**	(2.38)**	(2.59)**
Male*A			0.073	0.199	0.252	0.234	0.424
			(0.48)	(0.73)	(0.90)	(0.84)	(1.28)
Female			-0.422	-0.815	-0.751	-0.586	-0.590
			(3.45)**	(3.39)**	(3.06)**	(1.98)**	(1.98)**
Female*B				0.572	0.559	0.550	0.632
				(2.06)**	(1.96)**	(1.89)*	(1.98)**
Male*B				0.018	0.037	0.044	0.138
				(0.09)	(0.18)	(0.21)	(0.57)
Section1					0.504	0.564	0.548
					(3.81)**	(3.36)**	(3.26)**
Section2					-0.153	-0.185	-0.201
					(1.05)	(0.97)	(1.05)
Section3					-0.031	0.077	0.063
					(0.22)	(0.43)	(0.35)
Section4					-0.270	-0.069	-0.078
					(1.71)*	(0.35)	(0.40)
Section1*Female						-0.151	-0.140
						(0.55)	(0.51)
Section2*Female						0.069	0.080
						(0.23)	(0.27)
Section3*Female						-0.302	-0.284
						(1.01)	(0.95)
Section4*Female						-0.618	-0.596
						(1.75)*	(1.68)*
(Standardized grade – Threshold for A)^2							0.081
							(0.55)
(Standardized grade – Threshold for A)^3							0.050
							(0.90)
Constant	-1.083	-1.192	-1.022	-1.112	-1.193	-1.249	-1.401
	(22.82)**	(13.50)**	(10.23)**	(5.14)**	(4.92)**	(4.91)**	(4.53)**
Observations	1349	1349	1349	1349	1349	1349	1349

Absolute value of z statistics in parentheses
* significant at 10%; ** significant at 5%

Figure 2



Probability of Majoring in Economics: Male Students
0.3
0.25
0.2
0.15
0.1
0.05
0
Probability
-1.5
-1
-0.5
0
0.5
1
1.5
Standardized Course Grade less Standardized Threshold for A

Figure 3

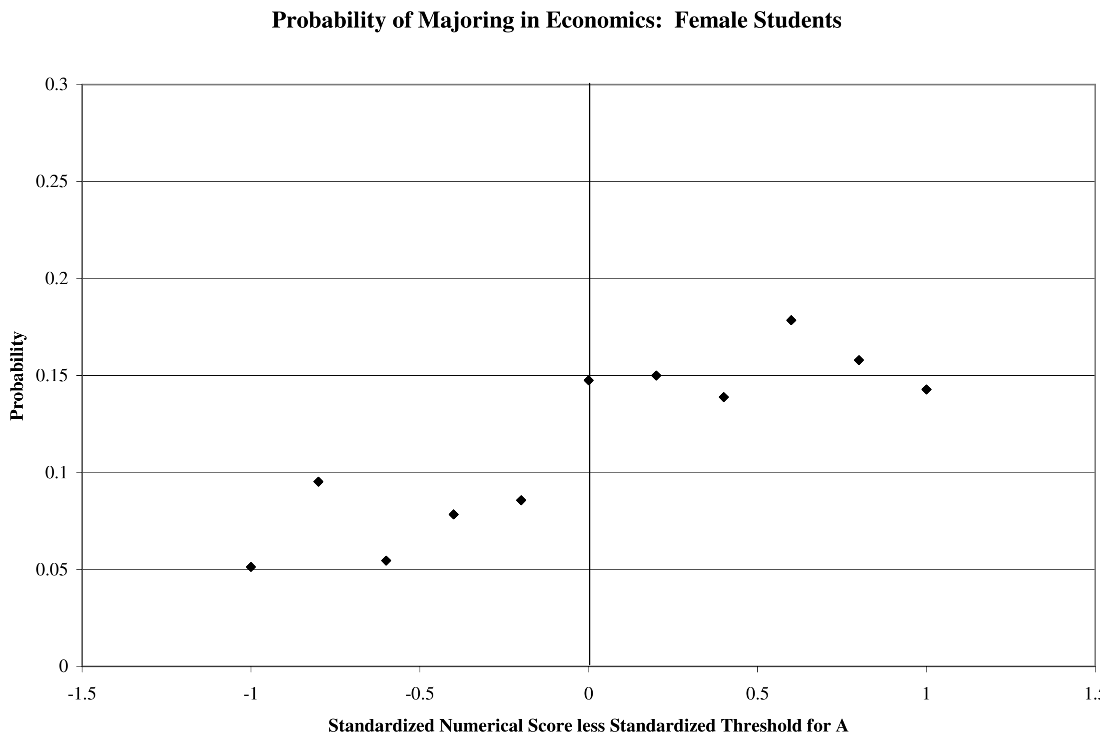
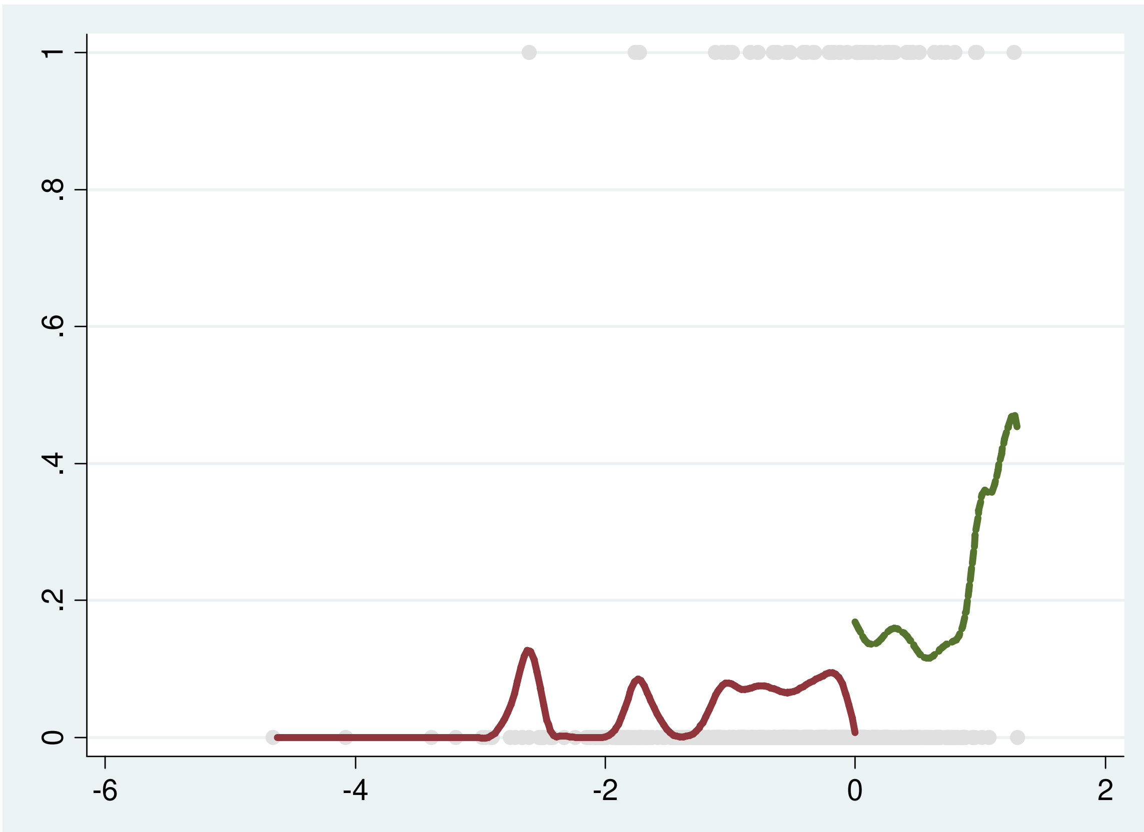
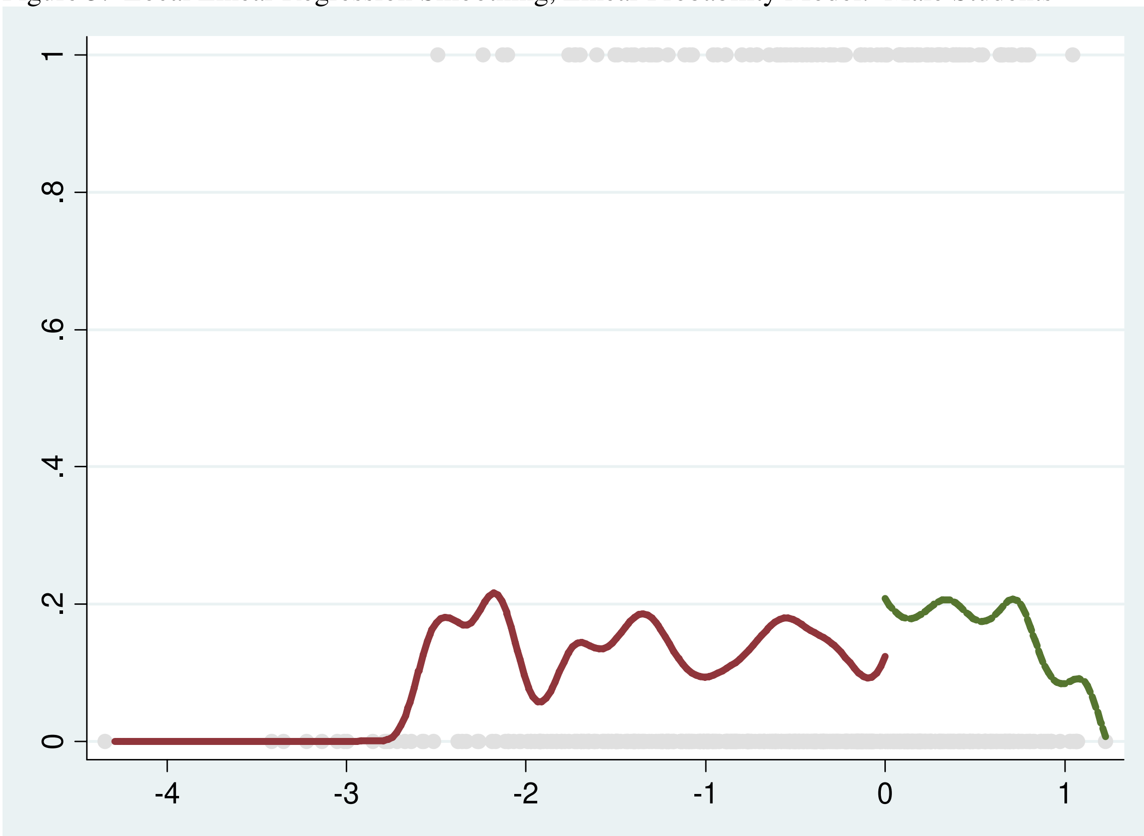

Probability of Majoring in Economics: Female Students
0.3
0.25
0.2
0.15
0.1
0.05
0
Probability
-1.5
-1
-0.5
0
0.5
1
1.5
Standardized Numerical Score less Standardized Threshold for A

Figure 4: Local Linear Regression Smoothing, Linear Probability Model: Female Students



Estimation method: Gaussian kernel, bandwidth of .10

Figure 5: Local Linear Regression Smoothing, Linear Probability Model: Male Students



Estimation method: Gaussian kernel, bandwidth of .10

Table 3: Regression Discontinuity Results

3A: Estimated impact on probability of majoring in economics of receiving A

Bandwidth	Female Students	Male Students
.05	.181** (2.20)	.051 (0.36)
.025	.148* (1.65)	.170 (0.77)
.10	.162** (2.30)	.085 (0.96)

**significant at 5%, *significant at 10%, z-statistics in parentheses, calculated via bootstrap method

3B: Estimated impact on probability of majoring in economics of receiving B

Bandwidth	Female Students	Male Students
.05	.001 (0.01)	.091 (0.62)
.025	-.041 (0.27)	.124 (0.64)
.10	.027 (0.46)	.028 (0.29)

***significant at 5%, *significant at 10%, z-statistics in parentheses calculated via bootstrap method

3C: Results for impact of other possible breakpoints, Female Students

Bandwidth	Female Students
Breakpoint=0.6	
.05	.131 (1.25)
.025	.148 (1.26)
.10	.074 (0.73)
Breakpoint=-0.6	
.05	-.101 (1.24)
.025	-.097 (0.91)
.10	-.057 (1.00)
Breakpoint=-0.8	
.05	.124 (1.46)
.025	.278* (1.91)
.10	.087* (1.71)
Breakpoint=-1.0	
.05	-.036 (0.35)
.025	.076 (0.46)
.10	-.081 (.068)

**significant at 5%, *significant at 10%, z-statistics in parentheses calculated via bootstrap method

Table 4: Descriptive Statistics, Alternative Sample

Variable	**Mean**	**Std. Dev.**	**Definition**
Female	0.34	0.47	=1 if female
Economics Major	0.36	0.48	=1 if economics major
A	0.18	0.39	=1 if A, A- or A+ in course
B	0.26	0.44	=1 if B or B+ in course
Standardized Grade – Threshold for A	-0.75	1.04	Numerical course score standardized to a mean of 0 and standard deviation of 1 by section minus the threshold for obtaining an A in that section

Table 5: Probit Analysis Predicting Economics Major, Alternative Sample

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Standardized Grade – Threshold for A	0.343	0.293	0.328	0.368	0.306	0.354	-0.578
	(4.85)**	(2.94)**	(3.24)**	(2.49)*	(2.96)**	(2.28)**	(1.67)*
A		0.161					
		(0.71)					
Female*A			0.601	0.697	0.649	0.737	1.595
			(2.00)**	(1.78)*	(2.11)**	(1.81)*	(2.90)**
Male*A			-0.165	-0.325	-0.112	-0.295	0.580
			(0.62)	(0.85)	(0.41)	(0.75)	(1.07)
Female			-0.587	-0.840	-0.562	-0.837	-0.892
			(3.51)**	(3.83)**	(3.33)**	(3.77)**	(3.95)**
Female*B				0.394		0.419	0.583
				(1.24)		(1.29)	(1.58)
Male*B				-0.265		-0.299	-0.199
				(1.00)		(1.08)	(0.60)
(Standardized Grade – Threshhold for A)^2							-0.402
							(1.76)*
(Standardized Grade – Threshold for A)^3							-0.033
							(0.52)
Constant	-0.018	-0.103	0.141	0.289	0.086	0.210	-0.261
	(0.19)	(0.67)	(0.84)	(0.99)	(0.38)	(0.67)	(0.61)
Observations	384	384	384	384	378	378	378

Absolute value of z statistics in parentheses, **significant at 5%; *significant at 10%.
Columns 5 through 7 include instructor dummies