The Returns to College Persistence for Marginal Students: Regression Discontinuity Evidence from University Dismissal Policies

Ben Ost\textsuperscript{a}, Weixiang Pan\textsuperscript{a} and Doug Webber\textsuperscript{b}

\textsuperscript{a} Department of Economics, University of Illinois at Chicago, 601 South Morgan UH718 M/C144 Chicago, IL 60607, United States
\textsuperscript{b} Corresponding Author. Department of Economics, Temple University, 1301 Cecil B. Moore, Ritter Annex Office 883 Philadelphia, Pennsylvania 19122. Email:douglas.webber@temple.edu phone: (215-204-5025)

Abstract

We estimate the returns to college using administrative data on college enrollment matched to administrative data on earnings. Utilizing the fact that colleges dismiss low-performing students based on exact GPA cutoffs, we use a regression discontinuity design to estimate the earnings impacts of college. Dismissal leads to a short-run increase in earnings and tuition savings, but the future fall in earnings is sufficiently large so that 8 years after the dismissal, persisting students have already recouped their up-front investment and opportunity costs with a rate of return of 4.1%. Though institutional factors allow for the possibility that the running variable is manipulated, we provide several pieces of evidence that this does not drive our results. Furthermore, we estimate bounds under opposite extreme assumptions regarding the form of manipulation and find that even the lower bound estimate suggests substantial earnings benefit to college.

Legal disclaimer: This workforce solution was funded by a grant awarded to the U.S. Department of Labor's Employment and Training Administration. The solution was created by the Center for Human Resource Research on behalf of the Ohio Department of Job and Family Services and does not necessarily reflect the official position of the U.S. Department of Labor. The Department of Labor makes no guarantees, warranties, or assurances of any kind, express or implied, with respect to such information, including any information on linked sites and including, but not limited to, accuracy of the information or its completeness, timeliness, usefulness, adequacy, continued availability, or ownership. This solution is copyrighted by the institution that created it. Internal use, by an organization and/or personal use by an individual for non-commercial purposes, is permissible. All other uses require the prior authorization of the copyright owner.
Introduction

Increasing the rate of college attendance and graduation has become a major policy goal at both the local and national level. President Obama called this an "economic imperative" and launched a variety of initiatives aimed at restoring the United States’ lead in the rate of college attendance.\(^1\) While there is reason to expect that the economic returns to college are high for some students, recent increases in the amount of student debt as well as low rates of student loan repayment have prompted many to question whether college pays off for all students. Critics of Obama's call for expanded higher education attendance argue that certain students are simply not "college material" and may be better off dropping out of school and obtaining more work experience.\(^2\)

In this study, we provide direct evidence on the returns to college for low-performing students. A major obstacle to identifying the returns to college is that comparisons of dropouts and completers are likely contaminated by omitted variable bias. In an ideal experiment, researchers would randomly force certain students to drop out of school while allowing other, similar, students to persist. Our empirical design mimics this experimental ideal by exploiting institutional policies that dismiss students based on their exact grade point average (GPA). We use administrative data from 13 public universities in the state of Ohio to identify students at risk of dismissal based on their past academic performance. We match these students to administrative data on Unemployment Insurance (UI) earnings records to study the impact of university

\(^1\) See President Obama’s “American College Promise” or the higher education proposals from Secretary Clinton/Senator Sanders.
\(^2\) See http://www.slate.com/articles/life/education/2014/03/college_isn_t_for_everyone_let_s_stop_pretending_it_is.html
dismissal on earnings 7-12 years after initial college enrollment. Using a regression discontinuity design, we effectively compare mid-run earnings of students with GPA’s just below the dismissal cutoff to those just above.

The dismissal policies we leverage generate large discontinuities in academic outcomes, which is in line with previous studies examining similar institutional discontinuities (Lindo et al, 2010; Vidal-Fernandez, 2011; Barua and Vidal-Fernandez, 2014). Students just below the cutoff are approximately 16 percentage points less likely to enroll in the following term, complete 0.2 fewer years of school and are 10 percentage points less likely to obtain a BA relative to those above the threshold. These reductions in educational investment correlate with a decrease in weekly earnings such that those just below the GPA cutoff earn approximately 4.8% less than those just above the cutoff.

While enrollment in college has increased substantially over the past several decades, persistence within college has declined markedly over the same period (College Board Education Pays 2013). As such, many policy makers view improving persistence as a key step towards increasing college completion rates. Our study estimates the returns to college among the set of students already enrolled and is thus directly relevant for understanding the value of policies that improve college persistence. Because of the nature of our identification strategy, the estimates we present pertain to low-performing students. While these students are not representative of the typical enrolled student, they are arguably one of the most policy-relevant subgroups, since they are likely to be impacted by policy initiatives aimed at improving persistence.

---

3 We refer to 7-12 years after initial enrollment as the “mid-run” throughout the manuscript.
The primary threat to identification in our context is the possibility that students finely manipulate their GPA in order to narrowly avoid dismissal. This is plausible because dismissal thresholds are publicly known and students could potentially alter their effort in order to manipulate their GPA. Empirically, however, we find relatively little evidence of sorting around the critical threshold. While we cannot definitively rule out manipulation, we view it as an unlikely explanation for our results. Students just below and just above the threshold are nearly identical in terms of their observable characteristics and yet they have dramatically different college persistence rates, college completion rates, and future earnings. For sorting on unobservables to explain our results, this sorting would need to be simultaneously large in magnitude and completely uncorrelated with observable characteristics. We present a number of robustness checks and bounding exercises to bolster this case.

We also produce estimates of the net present value to persisting in college as well as private and social rates of return within the span of our data and projected over the lifecycle. We find that 8 years past the potential dismissal date, students who remained in college had already recouped their schooling investment (in terms of tuition and opportunity costs), with a private internal rate of return of 4.1%. Even if only a portion of the earnings increase persists throughout the lifecycle, this suggests a substantial long-run return on investment for the students in our study.

Though our primary focus is to estimate the reduced form benefits to persisting in college, we also provide evidence on the magnitude of so called “sheepskin” effects of earning a college degree by utilizing variation in the timing of potential dismissal throughout a student’s college career. While the assumptions necessary to identify this
model are fairly strong, they are considerably weaker than those made by past work studying sheepskin effects (e.g. education is treated as exogenous conditional on observable characteristics) (Hungerford and Solon, 1987; Kane and Rouse, 1995; Jaeger and Page, 1996; Liu et al, 2015). We find suggestive evidence that college provides a large sheepskin effect, though our estimates are imprecise, which prevents us from drawing strong conclusions.

**Literature Review**

The literature on the returns to education is far too large to adequately summarize here, see Card (1999) for an excellent review. Although researchers have used many different identification strategies to estimate the returns to schooling in general (e.g. twins, changing compulsory schooling laws, quarter of birth) studies that specifically consider the returns to college utilize a more limited set of strategies. With few exceptions (discussed below), most estimates of the return to college rely on relatively strong assumptions. One strand of this literature relies on the assumption that, conditional on observables, college enrollment and completion are independent of potential outcomes (see Walker and Zhu, 2008 or Webber, 2014 for recent examples). Even when the researcher has access to highly detailed information about individuals’ aptitude and preferences, such as in studies which utilize the National Longitudinal Survey of Youth database, the argument for a causal interpretation of the estimated coefficient associated with education still rests on strong and inherently untestable assumptions. A second strand of the literature uses distance to the nearest college as an instrument for college attendance (see Card, 1995 or Kane and Rouse, 1995). Though arguably less restrictive than assuming that education is exogenous conditional on
observables, the distance-based IV similarly requires fairly strong assumptions since it rules out the possibility that colleges directly alter the surrounding area.

Our study is most closely related to a recent set of studies which use college admission thresholds and a regression discontinuity design\(^4\) to estimate the benefit of college admission. Several recent studies use data from various international contexts, and with the exception of Ockert (2010), find a substantial earnings benefit to gaining admission to a more selective university or course of study (Anelli 2015; Canaan and Mouganie 2015; Hastings, Neilson and Zimmerman 2013; Kirkebøen, Leuven, and Mogstad, 2014). In the US context, Hoekstra (2009) studies the earnings effect of admission to a selective flagship public university (as opposed to a less selective public university). Zimmerman (2014) studies the impact of gaining admission to a low-ranked public university on future earnings. Both of these studies find substantial benefit to gaining admission to the university they study.

We view our study as complementary to this set of studies in that they estimate the returns to college admission whereas we estimate the returns to college persistence among the set of admitted students. Our study is most closely related to Zimmerman (2014) because the students we both study are relatively low performing. That said, the students he studies are on the margin of college admission as opposed to being on the margin of college dismissal after having already completed some postsecondary schooling. In his setting, college admission requires a 3.0 high school GPA while in our

\(^4\) A similar strategy has been used to study academic outcomes, see Van der Klaauw (2002) or Kane (2003) for excellent examples.
setting college dismissal occurs when a student repeatedly performs below the dismissal threshold, set at 2.0 or lower.

**Academic Dismissal Policies**

Nearly all US universities use some form of a probation dismissal policy. Though exact policies vary across universities, in most cases, students on probation receive some academic counseling and are informed that if they fail to raise their GPA in subsequent terms, they will be dismissed from school. Typically, students on probation are allowed between 1 and 3 semesters to rehabilitate their cumulative GPA. In our sample, most schools allow dismissed students to apply for re-entry to the university but require an absence of one calendar year before being considered.

Though all public colleges in Ohio utilize the general probation dismissal policy described above, the specific parameters of these policies vary across schools. Appendix A provides details of the policies used at each public four-year institution in Ohio. The modal policy places students on probation if their cumulative GPA falls below some critical threshold, and dismisses students if they do not raise their GPA above the threshold by the subsequent term. In most cases, students can appeal the dismissal decision and, based on conversations with administrators, these appeals are often times successful. For students who are actually dismissed, the modal waiting period before reapplying is 1 year and readmission is typically not guaranteed.

We use data on the probation dismissal policy at each school to identify the group of students at risk of dismissal. For example, at Akron University, since students have 1 term to rehabilitate their GPA, a student with a cumulative GPA below 2.0 in term t-1 is at risk of dismissal following term t. If his cumulative GPA does not rise above 2.0 in
term t, he faces the possibility of dismissal. For students that we designate as at risk of dismissal, their future enrollment is a potentially discontinuous function of their GPA, whereas enrollment should be a smooth function of GPA for students not currently at risk of dismissal. As such, we focus our regression discontinuity analysis on students who are currently at risk of dismissal. For each student that is at risk of dismissal, we define the running variable based on the policies in place at his/her school. In some cases, the cutoffs differ according to the cumulative number of credits a student has previously earned and our running variable, GPA relative to the institution’s dismissal threshold, is normalized to reflect these different cutoffs for different students.

There are several reasons that students who fail to meet the GPA requirement while at risk of dismissal will not actually be dismissed. First, we code probation and dismissal policies at the university level, but in practice, these policies can be differentially lenient by major or college. Our data do not include any information on actual probation or dismissal status, which means that in some cases we identify students as at risk of dismissal when they were never actually on probation to begin with. Second, all schools in Ohio include provisions through which administrators can grant exceptions due to personal problems such as illness or other adverse shocks. Based on conversations with administrators, this discretion is used frequently and is particularly likely in cases where students are close to the required threshold and for students in their first year of college. Importantly, when administrators grant exceptions to the dismissal policy, they do this by allowing students to enroll despite their low GPAs, not by artificially adding GPA points to help the student rise above the threshold. This distinction is important because it suggests that administrator discretion will contribute to a weaker first-stage
discontinuity, but it will not bias our estimates. Finally, as discussed in Appendix A, our threshold data may be inaccurate for certain schools because we may not have uncovered all policy changes over time. This will further weaken our first-stage, but should not bias our second stage results.

Data

We use administrative data that links college transcripts for every Ohio public university student to UI earnings records in the state of Ohio. The transcript data span the academic years starting in 2000-2010 while the UI earnings data includes weekly earnings from 2003-2012. These data are made available to researchers by the Ohio Educational Research Data Center (OERDC) and include data from the Ohio Workforce Data Quality Initiative (OWDQI).

The wage and employer data come from the Ohio Department of Job and Family Services and include Quarterly Census of Employment and Work (QCEW) enterprise level data as well as worker-level quarterly earnings data. The higher education data includes the universe of 2- and 4-year public college enrollment in Ohio, though in this study, we focus on 4-year institutions. With the exception of federal workers and the self-employed, the UI data covers the universe of workers in Ohio. For some specifications, it is necessary to match quarterly earnings from the UI data to the academic calendar. We match the fall term to the fourth quarter (October-December) and the spring term to the second quarter (April-June). For schools on the quarter system, we match the winter term to the first quarter. We exclude summer terms from the enrollment analysis since most students do not enroll over the summer, but we include summers when calculating annual earnings and when calculating total credits earned.
While these administrative data are in many ways uniquely well suited to answer the research questions of this study, they have several key limitations. First, we cannot observe enrollment at private institutions or at public institutions outside of the state of Ohio. This will contribute to our finding relatively larger educational impacts since students that transfer out of state will be miscategorized as dropouts. However, we are able to track students if they transfer to any of the 38 2- or 4-year colleges in the state of Ohio and these institutions represent more than 75% of students in the state. Second, we cannot distinguish between unemployment, lack of labor force participation, federal employment, self-employment and leaving the state of Ohio. Since we cannot distinguish between these outcomes, we focus our analysis on earnings conditional on having a UI earnings record. We later test for whether differences in the probability of missing earnings data can explain our results and find no evidence that it can.

In addition to restricting the sample to students who are at risk of dismissal, we impose several other sample restrictions. First, we restrict the sample to students who begin at one of the 13 4-year public universities that we study. This is because we cannot calculate exact GPAs or academic standing for transfer students. Second, we focus our attention on students who attempt at least 10 credits in the critical term. Third, we restrict the UI earnings data to payments of at least $500 per quarter and we focus on each worker’s primary employer for each quarter as measured by total quarterly pay from each

---

5 Based on our calculations using the 2012 American Community Survey (2012 was chosen to match our study period), for individuals without earnings in our sample’s age range with at least some college experience there is a 56% probability that they left the state over the past year, an 8% chance that they are self-employed, 4% chance they are employed by the federal government, and a 32% chance that they truly have no income.
employer. Since we cannot observe long-run outcomes for the recent cohorts of students, we focus all our analyses on students who initially enrolled in college at least 6 years prior to the end of our data. This allows us to track all students in our analysis sample for at least 6 years in order to measure outcomes such as total credits completed and 6-year graduation rates. For the mid-run earnings analyses, we can observe earnings as far as 12 years after the term of first enrollment, but this is only the case for a single cohort in our data. Since the UI earnings data go further than the higher education data, we are able to observe earnings 7 years after initial enrollment for all students for whom we observe 6-year academic outcomes.

In order to measure earnings at the latest possible date for as many students as possible, our preferred earnings measure is average weekly earnings in the latest year in which an individual is observed in the data. For the majority of individuals, this is their earnings during 2012, but some individuals are missing earnings for this year since it is increasingly likely to have missing earnings records in later years of the data. For individuals missing earnings in 2012, we use their average weekly earnings from earlier periods if they exist. Since we expect the short-run and mid-run impact of enrollment to go in opposite directions, we focus only on earnings measured at least 7 years after initial college enrollment to construct this measure.

---

6 We drop small quarterly payments because the UI data include any payment from a firm to an individual, even in cases where that payment would not constitute what we normally think of as a job (e.g. legal payment, consulting service, etc.). For related reasons, we also only use an individual’s “dominant job” (highest earning) in earnings calculations in the event they work for multiple employers in the same quarter. These restrictions are standard when using UI data; see Webber (2015) as an example and for further citations.

7 This pattern of increasing missing with age is consistent with that found in other studies using UI earnings records (e.g. Zimmerman (2014) finds similar rates of attrition).
Compared to simply using 2012 earnings as the dependent variable, our preferred earnings measure has the advantage of increasing power and also minimizing potential biases caused by sample attrition. That said, the time at which earnings are measured depends on choices made by the individual, so the measurement timing could be endogenous. In practice, however, there is no discontinuity in the time at which earnings is measured, and all results are unchanged whether or not we control for the age at which earnings are measured.\(^8\)

Table 1 shows descriptive statistics for the overall sample and for the restricted sample of students at risk of being dismissed based on their historical GPAs. In that table, probation is an indicator for whether a student’s GPA falls below the probation threshold, as opposed to being a measure that comes directly from the data. Students at risk of probation are quite different than the general population as they are much more likely to be male and more likely to be black. Not surprisingly, these students have much lower graduation rates and complete far fewer credits on average. The mean age at earnings measured is 28, which is similar to other recent studies using administrative earnings and education data (e.g. Andrews et al., 2014, Zimmerman, 2014).

**Descriptive Analysis**

Given that relatively few studies have utilized data that include both administrative quarterly earnings and administrative enrollment information, we begin by describing the earnings trajectory of college graduates and dropouts in our data. This analysis is in the spirit of Mincer’s (1974) seminal work; however, rather than relying on

\(^8\) The results are also similar if we simply drop those missing earnings in 2012 or use average earnings 7-12 years after initial enrollment.
repeated cross sections measured after schooling is completed, we use panel data including earnings while in school.

Figure 1 shows average log weekly earnings vs quarters since first college enrollment, split according to whether the individual eventually obtains a BA. As discussed in the data section, we show earnings conditional on employment since we cannot identify true non-employment. The difference in the earnings of those who earn and do not earn a BA likely reflects a combination of the causal impact of a BA, selection patterns with respect to who decides to complete a BA and sample selection into employment. Nevertheless, Figure 1 illustrates several interesting patterns.

First, among individuals who eventually obtain a BA, we see a clear seasonal pattern in the first 4 years in which earnings are highest during the third quarter and lowest during the first quarter. Though quarters do not perfectly align with academic terms, the third quarter covers summer vacation for most students. A similar (though weaker) pattern of seasonality is observed for individuals who do not obtain a BA. This pattern lessens after the first year as these students begin to drop out of college. Second, Figure 1 shows that while dropouts out-earn those with a BA in years 2-4, the earnings of dropouts and completers rise at a similar rate over the first 16 quarters, so this gap does not increase over time. Finally, Figure 1 shows that BA completers out-earn dropouts almost immediately after the 16th quarter. Since the vast majority of these students complete the BA in the 16th quarter or later, this suggests that college graduates are able to overtake non-graduates shortly after graduating. Importantly, by the 28th quarter following initial enrollment, there already exists a substantial earnings differential between graduates and non-graduates suggesting that year 7 earnings is likely not too
early in the life course to be looking for possible earnings effects of a college education. Though we can observe earnings as late as 12 years after college for the 2000 cohort, it is useful to know that the earnings gap appears as early as 7 years after initial enrollment since we can observe earnings 7 years after enrollment for all of the cohorts for which we have long-run academic outcomes.

**Empirical model**

Our analysis uses a regression discontinuity design based on enrollment discontinuities caused by academic dismissal policies. We limit our attention to students who are at risk of dismissal based on their past GPAs. For this group, we construct our running variable, $C_{it}$, as

\[ C_{it} = GPA_{it} - cutoff_{it} \]  

(1)

If we perfectly identify at-risk students and universities strictly enforce dismissal policies then all students with $C_{it} < 0$ would be dismissed following term $t$ and none of these students would enroll in term $t+1$. In practice, not all students with GPA’s below the threshold will be dismissed because both the probation and dismissal policies are not rigidly enforced. Hence, our identification strategy can be categorized as a fuzzy (rather than a sharp) regression discontinuity design. Since our data do not include information on whether each student is dismissed, it is not possible to estimate the relationship between our running variable and dismissal. We do observe whether a student re-enrolls in the following term and ultimately, we are more interested in estimating the impact of college enrollment than college dismissal.

While it is possible to estimate the enrollment effect separately for each term, in order to increase power, we stack the data by term and thereby estimate the average effect
across all terms. We explore heterogeneity split according to the timing of dismissal in a subsequent section. We estimate a version of Equation 2 on a series of outcomes.

\[ Y_{t,t+1} = \alpha + \delta C_{i,t} + \beta D_{i,t} + \gamma C_{i,t}D_{i,t} + \varepsilon_{i,t} \]  

(2)

\( Y_{t,t+1} \) denotes enrollment in period \( t+1 \). \( C_{i,t} \) is the running variable defined above and \( D_{i,t} \) represents an indicator for whether \( C_{i,t} \) is below zero or not. We cluster standard errors at the individual level because some individuals could appear in the regression in multiple terms.

**Initial enrollment effect**

Our first step is to explore whether the dismissal policies we study actually impact next term enrollment at the same university. Figure 2 presents clear evidence of a discontinuity at the cutoff, suggesting that our basic empirical approach is viable. The fact that students who fall below the cutoff still appear to enroll at fairly high rates indicates either that the policies are not rigidly enforced or that our measurement of these policies has some error. From the standpoint of identification, neither of these issues will bias estimates since they simply weaken the power of our instrument.

Conversations with administrators suggested that first-year students are often treated more leniently than other students, so we split our first-stage analysis by whether the student has first-year standing at the beginning of term \( t \). Figure 3a shows that for students with first-year standing, there is not a clear discontinuity in terms of next term enrollment. Figure 3b shows the same specification for students with second-year or higher standing and for these students, there is a clear discontinuity in enrollment at the
cutoff.\textsuperscript{9} Being just below the performance cutoff reduces the probability of enrollment in $t+1$ by approximately 15 percentage points.

Given that the dismissal policies do not appear to be enforced for first-year students, we cannot use our empirical approach to study the impact of college persistence for these students.\textsuperscript{10} As such, for all analyses moving forward, we drop the students with first-year standing from our sample. While this sample restriction does not impact internal validity, it does further limit the external validity since we focus on students who make it to at least their second year of college.

\textbf{Specification checks}

The primary threat to identification is the possibility that students finely manipulate their GPA in order to land just above the threshold. In our context, thresholds are publicly known and students could feasibly manipulate their GPA by exerting extra effort (or pleading with professors) in order to obtain higher grades. Theoretically, the type of students who may engage in this sort of manipulation may be particularly motivated and as such, they might have better potential outcomes than their less motivated counterparts. While the institutional incentives to manipulate one’s GPA exist, there are several reasons why this form of manipulation might not occur in practice. First, we restrict the sample to those already on probation so these students have already failed to meet performance standards set forth by the university. This suggests that the type of students we study are perhaps less likely to respond to performance standards

\textsuperscript{9} We have also explored whether the discontinuity generally increases with standing and we did not find this to be the case. The estimated discontinuity is about the same for all students outside of those with first-year standing.

\textsuperscript{10} In results not shown, we verify that there is no discontinuity in earnings for students with first-year standing.
than the typical college student. Second, since appeals of the dismissal decision are frequently granted, students may not perceive the threat of dismissal as something that warrants a response.

While it is useful to consider the likelihood of manipulation from an ex-ante perspective, ultimately, whether manipulation is likely to substantially bias our estimates is an empirical question. We explore the magnitude of potential manipulation both by examining the density of observations surrounding the performance standard and also by testing for whether covariates move smoothly through the threshold. If manipulation occurs, we expect to see too few observations just to the left of the threshold and too many observations just to the right of the threshold. To the extent that this manipulation is correlated with students’ characteristics, the manipulation will also generate discontinuities in predetermined characteristics.

Figure 4a shows the density of the running variable around the threshold. Figure 4a shows clear evidence that students are more likely to obtain a GPA exactly at the cutoff than any other GPA. Though this might seem indicative of manipulation, past work (Zimmerman 2014, Barreca et al 2015) has noted GPAs tend to spike at whole numbers because the number of combinations that result in whole number GPAs are far larger than the number of combinations that result in decimal GPAs. Since the majority of schools use a 2.0 GPA cutoff, if students are more likely to obtain exactly a 2.0 mechanically, this could feasibly explain the large spike in the running variable at zero. To illustrate this point, in Figure 4b we show a simulated histogram where we randomly assign grades to students and plot the resulting GPA distribution. Figure 4b shows a clear spike, just as was observed in Figure 4a. Though we do not view the spike at the
discontinuity necessarily as evidence of manipulation, this large spike does make it
difficult to assess whether students sort around the threshold and as such we follow past
work (e.g. Zimmerman (2014)) and rely on other specification checks to evaluate the
validity of the RD design.

Even if there is no manipulation, as noted by Barreca et al. (2015) this type of
spike has the potential to bias estimates and so we follow their recommendation and
estimate both the typical regression discontinuity and the “donut” RD in which those with
a running variable exactly at zero are excluded from the estimation. The intuition in the
donut RD is that even in the presence of a spike at 2.0, it will still be the case that those
with a 2.01 are a reasonable counterfactual to those with a 1.99. Even if one ignores the
large spike exactly at the threshold, Figure 4a arguably shows some modest signs of
manipulation since there does appear to be a generally higher mass of observations to the
right of the threshold. This could simply be the natural distribution of GPAs (and indeed,
we see a similar bulge in Figure 4b) but it suggests that we need to be particularly
sensitive to manipulation concerns in this context. We consider several pieces of
evidence to help ascertain whether manipulation is likely to be a principal driver of our
results: we test for observable discontinuities, we examine sensitivity to excluding
observations in the immediate vicinity of the threshold and we derive bounds that are
robust to extreme assumptions regarding the form of manipulation. We show these results
later in the manuscript.

To further explore the likelihood that students finely manipulate the running
variable, we examine whether there are discontinuities in observable predetermined
characteristics. In addition to testing for smoothness in each of the individual student
characteristics, we also construct 4 composite variables in which we predict outcomes using all of the covariates and test for smoothness in the predicted outcome. Considering whether there are discontinuities in predicted outcomes is particularly useful because these measures show not only whether there are covariate imbalances, but also whether and in what direction any imbalances may bias estimates.

Figures 5a-5l show our estimates of the discontinuities in each covariate along with the predicted outcomes. Column 1 of Table 2 displays these same estimates. On the whole, there is no evidence of discontinuities in the covariates. Of the 12 specifications shown in column 1, only one is statistically significant at the 10% level and based on visually inspecting Figure 5a, it is clear that this estimated discontinuity is driven by functional fit.11 Importantly, all of the predicted outcomes and key covariates (such as log weekly earnings in period t-1) are smooth through the threshold. There is a clear non-monotonic relationship between certain covariates and the running variable, most notably, credits-earned at time t-1 shows an inverse V shape with the apex near the cutoff. This pattern is expected since students with many credits earned at time t-1 are mechanically less likely to have a cumulative GPA very far below or above the threshold. Since the relationship between credits-earned and the running variable is smooth at the cutoff, this will not create bias, however, it can lead to sharp slope changes right at the cutoff which can contribute to imprecision of the RD estimates. As such, in all of our analysis below, we condition on credits earned at time t-1.

11 Consistent with this view, in results not shown, we find that the estimated discontinuity for credits earned at time t-1 is quite sensitive to bandwidth choice.
Taken together, these results suggest that while we cannot definitively rule out the possibility of manipulation, it is unlikely that it is sufficiently strong so as to be able to explain away the results presented below. In the robustness section, we provide further evidence on this point.

**Results**

We present all of our core results in figure form using a fixed bandwidth of 0.5 grade points. Tables 3 and 4 show these same estimates in table form and examines the robustness of each result to changes in the bandwidth, adding student observable characteristics, or using a “donut” RD.

*Academic Outcomes*

Though we have already established that dismissal reduces enrollment in the short-run, there are several reasons that the impact on an outcome such as BA receipt is not a mechanical function of the disenrollment effect. First, it is possible that the policy only dismisses students who were going to drop out eventually, and thus has no impact on completion. Second, students who are dismissed are permitted to apply for reentry to the university after their dismissal, so being dismissed does not preclude earning a degree at that institution.

To explore the long-run academic impacts empirically, we examine how falling below the threshold impacts total credits earned and the probability of BA receipt at both the initial school and other Ohio schools. Theoretically, we expect that a student’s educational investment likely falls at the dismissing institution, but might rise at other schools.
Figure 6 shows the impact of falling below the cutoff on the total credits earned at one’s starting institution. We find that students below the cutoff earn approximately 5.1 fewer credits, which corresponds to slightly over 1/6 of a full year’s load.\textsuperscript{12} While credits earned at the starting institution falls, it is possible that dismissal increases credits earned at other schools. To investigate this, we change the outcome to be all the credits earned by a student at any public Ohio college outside of their starting institution. We find a relatively small impact (less than one credit at a different institution) in this model (see Figure 7) and in our robustness analysis we find that the estimate is sensitive to bandwidth choice.

Figure 8 shows the impact of falling below the cutoff on the probability of earning a BA from the initial college of enrollment. Falling below the cutoff reduces the probability of graduating by approximately 11 percentage points. Rerunning this analysis for institutions other than the one a student initially enrolled in shows no evidence of a discontinuity (see Figure 9).

While we have no data on enrollment at private schools or public universities outside of Ohio, the fact that we see relatively little movement of students across public institutions within Ohio is suggestive that falling below the cutoff has a real impact on total educational investment. Since Ohio’s higher education system is dominated by public institutions, we suspect that if a large fraction of students were obtaining degrees from other schools as a result of the dismissal, we would observe this effect at public

\textsuperscript{12} Credits are measured based on semester-credit equivalents so the typical school requires 120 credits to graduate. We follow Kane and Rouse (1995) and define a year of education as 30 earned credits so that the typical college degree is 4 years of credits. This is consistent with most labor studies where researchers assume that individuals with a BA have 16 years of education, even if that BA took more than 4 years to complete.
institutions. The 2-year institutions in particular would likely be an attractive institution to transfer to after failing out of a 4-year public institution, but we see no evidence of an increase in AA degrees (see Figure 10). That said, we cannot definitively rule out the possibility that our estimates of the impact on educational investments are too large because students substitute towards private institutions.\textsuperscript{13} To the extent that we are overestimating the impact on educational investment, we will be underestimating the earnings returns to education.

\textit{Impact on Earnings}

Figure 11a shows the reduced form relationship between the running variable and mid-run log weekly earnings. We find that falling below the cutoff decreases weekly earnings by approximately 5.6 percent. The slopes on each side of the discontinuity are not statistically distinguishable from zero. While we cannot differentiate between earnings gains from increased hours and from increased wages, unlike many studies using administrative UI records, we have weekly rather than quarterly earnings and so these estimates cannot be driven by weeks worked.

While our data cover earnings for a majority of the sample, earnings are not observed for students who are self-employed or leave the state. As such, it is possible that some of the earnings effect that we estimate is explainable by differential attrition. Although we find little evidence of differential attrition above and below the critical GPA

\textsuperscript{13} If we collapse the outcomes “BA from starting institution” and “BA from other institutions” into a single BA variable, results are similar to the “BA from starting institution” result. This is not surprising given the zero BA from other institutions effect. We opt to keep these outcomes separate both because it provides a more detailed understanding of completion and because even after collapsing these outcomes, we still do not measure total BA completion given our lack of data on private schools.
cutoff (see Figure 11b), it is possible that even with perfectly balanced attrition rates, our estimates could be biased if the form of attrition differs between the treatment and control groups. For example, suppose college simultaneously increases each person’s propensity to move out of the state and also increases their employment probability by the same amount. If treatment influences attrition non-monotonically, then balanced attrition does not eliminate attrition bias.

An alternative test of whether missing earnings data is likely to bias our estimates is to examine whether we continue to have balanced covariates in the sample for which our later earnings measure is non-missing. Column 2 of Table 2 shows the estimated discontinuities in covariates around the critical threshold when the sample is restricted to the set of students with non-missing mid-run earnings. We find no evidence of discontinuities in any of the covariates. The smoothness of the covariates in the sample with non-missing mid-run earnings suggests that sample selection does not cause those just to the left of the cutoff to be quite different than those just to the right.14

Robustness

For each outcome, we show how the estimates vary according to the bandwidth chosen, whether or not we include covariates, and whether or not we estimate the donut

---

14 While we view the balanced covariates as the strongest evidence regarding sample selection, we have also explored bounding estimates using the set identification strategy presented in Lee (2009). Specifically, we use the estimates of the extent of differential attrition from our preferred specification and make two diametrically extreme assumptions regarding the potential outcomes of those driving the differential attrition. First, we assume that these missing observations would have been at the top of the wage distribution were they observed. Second, we assume that these missing observations would have been at the bottom of the wage distribution were they observed. See Lee (2009) for additional details regarding this procedure. Given that there is little evidence of differential attrition to begin with, it is no surprise that our results are robust to this bounding exercise.
RD suggested by Barrecca et al (2015).\textsuperscript{15} Table 3 shows the robustness of the results for academic outcomes while Table 4 shows the robustness of the results for the labor market outcomes. In Table 3, the estimates appear to be generally invariant to whether or not covariates are included, providing further suggestive evidence that differences in unobservables between those on either side of the cutoff are unlikely to explain our results. The results are similarly robust to implementing the donut RD, suggesting that the mechanical heaping at 2.0 is unlikely to explain our results. Broadly, the estimates are also robust to bandwidth choices, though the smaller bandwidths generally yield somewhat smaller estimates and larger standard errors. Our visual inspection of the 0.75 bandwidths makes clear that a linear fit does not approximate the data well, but the fitted lines for the 0.5 and 0.25 bandwidths both appear to match the figures fairly well.

Table 4 shows that the estimated impact on mid-run earnings is similar across various specifications. The earnings estimates are similar across different bandwidth choices and vary only modestly depending on whether controls are included or exact zeroes are excluded. The second panel of Table 4 shows that after conditioning on covariates or estimating a donut RD, the insignificant “missing earnings” discontinuity documented in Figure 11b becomes statistically significant. Though not as robust across specifications, based on the statistically significant estimates, falling below the cutoff reduces the probability of having missing earnings data by approximately 1.5-2.5 percent, depending on the specification. While this differential attrition has the potential to bias our estimates, the magnitude of the differential attrition is quite small. In order for 2

\textsuperscript{15} Based on the recommendation of Gelman and Imbens (2014), we focus on local linear models. In our context, though, the local linear and polynomial models yield similar estimates.
percent monotonic differential attrition to explain away our earnings estimates, it would be necessary for the potential earnings outcomes of the differential missings to be approximately 250% larger than their observed counterparts. As noted earlier, Table 2 shows that there are no observable differences in the sample that have observable mid-run earnings, making it even less plausible that differential attrition substantially biases our estimated earnings effects.

Though the smaller bandwidth increases the likelihood that the linear fit on either side of the threshold is approximately correct, it also increases the probability that the slope of the lines over-fit noisy data. As such, it is not a priori obvious whether the smaller or larger bandwidth is preferable. We use the 0.5 bandwidth as our preferred specification moving forward, though the results are qualitatively similar for other choices. Furthermore, the 0.5 bandwidth yields more conservative IV estimates since it tends to estimate similar wage effects and larger educational effects. We include covariates to improve precision and we use the donut RD specification because it is found to outperform the standard RD specification in contexts with spikes in the running variable density (Barreca et al 2015).

In addition to establishing the basic robustness of our results, we investigate the likelihood that manipulation drives our estimates from several perspectives. First, we follow Altonji, Elder and Taber (2005) to assess the degree of sorting on unobservables necessary to explain away our estimates. Second, we follow Gerard, Rokkanen and Rothe (2016) to construct bounds that are robust under any form of unobservable sorting. Finally, we follow the recommendation of Barreca et al. (2011) and estimate various donut RDs to assess the likelihood that manipulation drives our estimates. These three
approaches are quite different and all suggest that our estimates are unlikely to be driven by manipulation. See appendix B for a detailed description of each of these approaches and the results.

*IV estimates of the returns to schooling*

The preceding section shows the reduced form estimated impact on both educational outcomes and earnings. In this section, we use the discontinuity as an instrument and estimate the impact of years of schooling on earnings. While computationally, the IV estimate is just the ratio of the earnings estimate to the years of schooling estimate, in order to interpret this as a local average treatment effect, we require the standard IV assumptions.

One potential threat to the exclusion restriction is that dismissal may be psychologically distressing so that in addition to reducing total time in school, it impacts emotional health in a fashion that hurts future earnings. To the extent that it exists at all, we believe this effect is likely small for several reasons. First, we study earnings measured many years after the dismissal so it seems implausible that the direct emotional consequences of dismissal would continue to be important. Second, while dismissal may be emotionally difficult, voluntary dropout due to low performance is likely similarly difficult in many ways. Finally, while dismissed students are given a negative signal about their academic ability, those that just narrowly avoid dismissal are also likely aware of their generally poor performance.

While none are a violation of the exclusion restriction, there are several different channels through which years of schooling may impact earnings. First, increasing years of schooling likely reduces labor market experience and thus the estimated return to
schooling reflects the difference in human capital acquired through school vs work.\textsuperscript{16} We view lost experience as part of the schooling effect as opposed to something we might want to control for. Second, the impact of increasing years of schooling partly operates through increasing BA receipt. Just as with years of employment, an increase in BA receipt does not introduce bias into our estimates of the returns to schooling, but it is important to keep in mind that our estimates include any sheepskin effect averaged over the number of years of college attended.\textsuperscript{17}

Although the increase in experience caused by decreasing years of schooling is not a violation of the exclusion restriction, changes in experience could be a violation of the exclusion restriction if they occur, holding education fixed. For example, some dismissed students may still complete their degree, but take longer to do so and therefore end up with \textit{less} experience than they would have had without the dismissal. Assuming that a BA completed in 6 years has the same labor market return as a BA completed in 4 years, dismissal could act to reduce earnings through lost experience, holding fixed completed education. Relatedly, if experience obtained during time away from school has a smaller return than experience obtained after schooling is completed, effective experience may be lower for dismissed students than for a student with the same number of years in school, but who takes no break.

\textsuperscript{16} This conceptualization of the causal impact of education is somewhat different from the standard Mincer wage regression that simultaneously controls for education and experience.
\textsuperscript{17} This conceptualization is consistent with the literature since most estimates of the returns to schooling include any sheepskin effects. For example, the average return to a year of college is estimated as the total return to college divided by 4, even though the sheepskin effect accrues entirely in the final year.
To investigate the likelihood that dismissal decreases experience for some students, we examine the time from the dismissal event until the date of last enrollment. Importantly, we define this variable as the number of years from the point of potential dismissal until the point of last enrollment – regardless of whether the student is enrolled continuously during this time period. For example, a student who drops out in the term of dismissal and never enrolls again has a value of 0. A student who leaves for 2 years and then takes 3 years to complete the BA has a value of 5. For convenience, in the discussion below, we refer to students whose GPAs fall just above the cutoff and are therefore not dismissed as the control group. Students with GPAs below the cutoff are the treatment group. Treatment is expected to sharply increase the fraction of students who leave in year zero. If dismissal causes some students to take longer to complete their degree, it is possible that treatment will also lead to an increase in the fraction of students who leave in later years. Following Barua and Lang (2016) we test this possibility by considering whether the control group CDF exhibits first-order stochastic dominance (FOSD) over the treatment group CDF.

We show the treatment and control cdfs in Figure 12. Approximately 15% of the control group and 30% of the treatment group have a value of 0. For our purposes, the key question is whether this initial gap reverses in later years. If we had found that a higher proportion of treated individuals were still enrolled in later years compared to the control group, this would suggest that dismissal induces some individuals to leave school at a later date and would be a violation of FOSD. Figure 12 shows that up until 4 years, the treatment group is more likely to have left school than the control group and after that point, they are equally likely to be leaving school.
Unlike many IV contexts, the monotonicity assumption is not innocuous in our case because the dismissal may act as a wake-up call for certain students: students on the path to dropout might be encouraged to complete more years of schooling by being forced to take a break. Unlike violations of the exclusion restriction, violations of monotonicity do not bias our estimates in a homogenous treatment effects world. The IV estimates become misleading, however, if the returns to schooling are very different for compliers compared to defiers. In particular, if the defiers derive no benefit from schooling, their existence reduces the size of the first-stage but leaves the reduced form unchanged, thereby inflating our IV estimates. On the other hand, if on average the defiers derive a similar benefit from schooling as the compliers, the IV estimates will be unbiased. We discuss the monotonicity assumption in greater detail in Appendix C.

Table 5 shows IV estimates of mid-run earnings for 3 different bandwidths. In each specification, our instrument has an F-statistic above conventional rule-of-thumb thresholds. Focusing on the results corresponding to a bandwidth of 0.5, we see that an extra year of schooling is predicted to increase earnings by nearly 25%. While this estimate appears to be much larger than that of past work, this difference is to be expected if sheepskin effects are important. Due to the nature of our identification strategy, the marginal years of education we examine must occur late in an individual’s college career (because we focus on students at risk of dismissal in their second year or later). Thus, to the extent that there is a sheepskin effect of earning a BA separate from continuous human capital accumulation, this effect will be averaged over a smaller number of years in our study compared to most other studies.

*Heterogeneity*
In Table 6, panels A and B, we show the estimated impact of falling below the threshold on long-run academic outcomes for various groups. Panel A shows that there is relatively little heterogeneity across groups in the impact on total credits earned and the standard errors are too large to statistically detect differences across groups. Panel B similarly shows that there is relatively little heterogeneity in the impact on whether individuals earn a BA. The one exception is that the effect on BA receipt is considerably smaller for students with second year standing compared to students with higher standing at the time of the dismissal.

While the heterogeneity in the BA effect could be the result of a differential response to dismissals in earlier and later terms, we suspect that it mostly reflects heterogeneity in the control group. Recall that for students who are unlikely to graduate in any case, the dismissal leads to an earlier dropout but has no impact on whether or not the student graduates. There are many students with second year standing who are unlikely to ever graduate, but conditional on obtaining fourth year standing, most students are expected to graduate. Although it makes sense that dismissal will have a larger impact on the graduation rates of those likely to graduate, it is somewhat surprising that dismissal seems to derail students so close to graduation. One possibility is that these students perceive sheepskin effects as small such that the returns to completion are small conditional on almost completing. Alternatively, it is possible that the students we study are only tenuously connected to their academic institutions and are delicately balancing a variety of factors that allow them to stay in school.\footnote{While it may be difficult to rationalize the lack of persistence in college from a neoclassical perspective, it is worth noting that it is similarly difficult to rationalize why...}
In panel C, we show heterogeneity in the IV estimates of the impact of years of schooling on earnings. Large standard errors for the subgroup analysis means that many of the estimates in panel C are statistically insignificant, but the estimates are positive and large for all groups. While the large standard errors prevent us from statistically distinguishing across groups, in terms of the point estimates, women and disadvantaged minorities have a larger return than men and non-minorities respectively. Though all the estimates are noisy, we observe much larger estimated returns to a year of schooling for those with fourth year standing at the time of dismissal, again consistent with a large sheepskin effect.

*Separating the BA effect from the years of schooling effect*

One unique feature of our identification strategy is that the quasi-experimental variation used to identify the returns to education occurs at different points in time for different individuals; this makes it possible to separately identify the BA effect from the years of schooling effect. Since there is a larger BA effect and smaller credits earned effect for those with later standing, if we see a larger impact on earnings from early dismissals, this suggests that credits earned are more important, whereas if the earnings impact is larger for later dismissals, the sheepskin effect is more important.

Econometrically, we treat years of schooling and BA completion as both endogenous, and our two instruments are whether an individual is below the cutoff and that same variable interacted with standing.

Anecdotally, each term approximately 10% of economics majors fail our upper level economics courses as a result of poor attendance or failure to attempt exams.
In order to separately identify the impact of BA completion and years of schooling, it is necessary to make two assumptions in addition to the standard IV assumptions discussed earlier. First, we assume linearity in the years of schooling effect. Kane and Rouse (1995) also impose this assumption since without it the BA effect could simply reflect a sharp spike in the returns to schooling. Second, we assume that the local average treatment effect of the BA and years of schooling is the same across our instruments. In other words, the compliers with lower standing must on average have the same returns to education as the compliers with higher standing. Without this assumption, we would conflate heterogeneity in the returns to years of schooling with the BA effect. Theoretically compliers with lower standing could have either higher or lower returns to schooling. Compliers with low standing are likely to be academically weaker than compliers with higher standing since the former fail out of college sooner than the latter group. On the other hand, the compliers with higher standing might have relatively weak attachment to schooling since they fail to finish the degree due to the dismissal – even though they are quite close to completion. Although both of the maintained assumptions are fairly strong, these assumptions are considerably weaker than the assumptions used in past work. In particular, in addition to the assumptions that we require, past work assumes that education is exogenous conditional on observables.

Table 7 shows our estimates of the BA and years-of-schooling effect using 3 different bandwidths. Columns 1, 4 and 7 show the IV estimates whereas the other columns show the first-stage estimates. Though the IV estimates are quite similar across the 3 bandwidths, the smallest bandwidth provides less power so the F-statistic on the years of schooling first-stage is only 6.4. As before, we focus on a bandwidth of 0.5 to
balance statistical precision and parameter identification. Columns 5 and 6 show that the negative impact of dismissal on BA receipt is larger in magnitude for those with higher standing, whereas the negative impact on credits completed is smaller in magnitude for those with higher standing. The IV estimates are generally imprecise, but the point estimates are large and positive for the BA effect and positive (although small) for the years of schooling effect.

Taken at face value, our estimates suggest that a student who completes 4 years of schooling and obtains a degree should obtain approximately a 41 percent total return ($0.363+4*0.0139$). It is important to emphasize that these results are likely specific to the sample of low-performing students that we study since we suspect that the human capital benefits of a year of school are smaller for very low-performing students. These students likely skip more classes, are less likely to complete assignments and are more likely to be lost during the classes they do attend.

The large standard errors in Table 7 prevent any strong conclusions, but given that ours is the first study to attempt to separate out the years of schooling effect from the BA effect using credibly identified variation, we view these estimates as a valuable starting point. Interestingly, although our time period, identification strategy, and sample are completely different than Kane and Rouse (1995), our estimates are remarkably similar. They estimate a 23-29% return to the BA degree and an insignificant 2% return to credits/30. The type of data we use in this study is increasingly available in a variety of states, so future research might use a similar empirical methodology combined with a larger sample to potentially obtain more precise estimates.

Net present earnings effect of dismissal
Our empirical design allows us to investigate the entire time path of college effects and identify the causal cross-over point. To implement this analysis, we estimate a series of RD regressions, one for each quarter of earnings following the dismissal event. These regressions estimate the causal effect of dismissal on earnings in quarter t+n. The number of quarters that we observe after dismissal varies by student and depends on three factors. First, students in the later cohorts can be followed for fewer quarters than students in the earlier cohorts since our earnings data end in 2012 for all students. Second, students who are dismissed at higher standing can be observed for fewer quarters following dismissal since the dismissal occurs at a later point. Finally, sample sizes generally fall as we examine earnings further away from the dismissal event as some people leave the state. We can observe the majority of students for up to 24 quarters following the dismissal event but sample sizes drop considerably as we extend this time frame. As such, although we map out the earnings effects for 32 quarters following dismissal, we have the most confidence in the estimates for the first 24 quarters.

In order to facilitate the calculation of net-present value that follows, we estimate these models using actual earnings as the outcome as opposed to logged earnings. Since falling below the GPA threshold only maps to a 15.2 percentage point change in enrollment in t+1, we scale all of our RD estimates by 1/0.152 so that we can interpret our coefficients in terms of the effect of an actual dismissal in period t. For presentational purposes, we plot the earnings path of the control group as well as the change in the earnings path caused by the dismissal. Specifically, for each of the RD

---

19 Note that here, we are measuring quarters relative to the dismissal event as opposed to quarters relative to the term of first enrollment. 24 quarters after dismissal corresponds to ages 27-28 for most students.
regressions, the constant term is an estimate of the baseline earnings for the control group and the constant term plus the discontinuity is an estimate of the earnings for the treatment group.

Figure 13 shows the estimated treatment and control wage paths and demonstrates two main findings. First, earnings rise immediately following the dismissal and earnings for the treatment group are mostly higher than the control group for approximately 3 years following the dismissal. Second, starting approximately 5 years after the initial dismissal, the treatment group’s earnings fall below those of the control group. This earnings gap is fairly constant until around quarter 30. The later quarters are increasingly based on relatively few observations and so we place little weight on these results and do not view the quarter 31 reversed gap as evidence that the earnings consequence of dismissal reverses in that long-run.

Based on the estimates from Figure 13, we can calculate a present discounted value of college dismissal that incorporates both the initial earnings increase as well as the future earnings losses. We assume an annual interest rate of 5%, and calculate three separate NPV statistics: private, social, and social taking into account deadweight loss. This approach follows a similar exercise conducted by Zimmerman (2014).

Along with foregone earnings, the private NPV takes into account the expected tuition and fees that would be paid by students who are not dismissed. To calculate the change in tuition caused by the dismissal, we take our estimate of the effect of dismissal

---

20 The private costs include average tuition rates (net of institutional aid) among Ohio four-year public institutions over the years our sample is enrolled in college. Social costs are measured as the average education-related expenditure at these institutions over our sample frame. Tuition and education-related expenditures were obtained from the Integrated Postsecondary Education Data System (IPEDS).
on total credits earned and multiply this by the average net tuition rate in Ohio over our time period. We find that dismissal saves students $7,650 in expected tuition.

The social NPV includes the average cost of education related expenses per full time equivalent student, which is greater than the private cost because our sample includes only public institutions which are partially subsidized by the state. The expected social costs total $11,190. Finally, the last NPV calculation subtracts off estimated deadweight loss from the prior social NPV. Following Zimmerman (2014), we use the Feldstein (1999) estimate of deadweight loss due to taxation of 30% (applied to the difference between private and social costs).

Figure 14 plots the NPV calculations over the course of our data as well as projections over the lifecycle. By 8 years following possible dismissal, students who avoided dismissal and enrolled the following semester have effectively recouped their investment (both the cost of tuition and the opportunity cost of working less) with a NPV near zero of -$772. The social return is still negative at this point (-$4,227 assuming no deadweight loss due to taxation and -$5263 with deadweight loss). Importantly, we do not view these negative estimates as evidence that college is a poor investment for marginal students since it is not surprising that students do not fully make up for direct and opportunity costs of college before age 30.

In order to obtain lifecycle projections, we assume that the average quarterly return over the last three years of our data persist until age 65. Given that the premium associated with a college degree generally increases over the course of one’s career, our estimates may be conservative. The projected private return through age 65 would be $67,357, with social returns not far behind ($63,902 and $62,865).
Calculating internal rates of return (IRR) help put these returns in a broader context. By 8 years following potential dismissal, the IRR is 4.1%, 0.8%, and -0.2% for private, social, and social/deadweight loss respectively. Looking at projected returns over the entire lifecycle, the IRR calculations increase to 19.2%, 17.1%, and 16.4%.

**Conclusion**

We show that low-performing students who are dismissed from public 4-year colleges in Ohio suffer substantial earnings losses measured 7-12 years after college enrollment. Given the large economic returns to persisting in college for this group, a natural question is why so many similar students drop out. We see four broad categories of potential explanations.

First, it is possible that the dropout decision is not really a decision at all and instead is forced upon students by their university. For example, it might be that given the incoming skill level of students combined with a time constraint, many students cannot satisfy the performance standards and are thus prevented from graduating by probation-dismissal policies.

Second, it is possible that the returns to schooling that we identify are larger than for the typical dropout. Our instrument only alters the behavior of students who would otherwise have stayed in school. In other words, the dismissal only binds for students who view the benefits of school as larger than the costs. As such, it is possible that students who voluntarily drop out of school have a lower return to schooling than the compliers in our study. While Oreopolous (2006) finds evidence that the average treatment effect is similar to the local average treatment effect in the context of high
school students in the United Kingdom, the returns to college have the potential to be much more heterogeneous.

Third, it is possible that despite large economic benefits, dropouts still view the total costs as larger than the total benefits. For example, dropouts could have relatively high discount rates such that the upfront psychic and economic costs of college attendance outweigh the large long-term benefits. Alternatively, if students are risk averse, the certainty of the upfront costs combined with the uncertainty of the long-run benefits could reduce the attractiveness of investing in college. Finally, dropouts may not know the causal economic returns to a college degree given that there remains uncertainty regarding the magnitude of these returns among researchers.

Moreover, even if students view college to have a high rate of return, credit constraints could prevent certain students from persisting (Belley & Lochner, 2007; Goodman, 2010). Though there exists many different sources of loans for college students in the United States, students may be debt averse (Marx and Turner, 2015), may need to finance familial obligations for which they cannot borrow, or may need more loans than the capped amounts in order to sufficiently smooth consumption (Lochner and Monge-Naranjo, 2011).

While it is difficult to distinguish between the forces behind the average dropout decision, among the students we study, there are clear external forces contributing to dropout. While this loss of educational investment likely reduces these students’ total welfare, their dismissal likely has positive externalities so that the total social welfare impact of the dismissal policies may well be positive. In particular, restricting low-performing students from graduating maintains the signaling value of the degree and
provides incentives for all students to succeed. As such, we view our study as
uninformative regarding the total impact of dismissal policies. Instead, our study uses
these dismissal policies to provide evidence on whether low-achieving students benefit
financially from persisting in college. Our estimates suggest that persisting likely
provides a large positive return, even for low-performing students.
References


Figure 1: Average log weekly earnings by BA receipt status

Note: This figure depicts the average log weekly earnings for all Ohio workers who we observe attending a public 4-year university in the state of Ohio during the sample period.
Figure 2: Enrollment discontinuity (full sample)

Note: This figure depicts the probability of enrollment in the subsequent term, by GPA in term t, for students who are predicted to be on academic probation (based on the assignment rules collected by the authors). The running variable is normalized to be 0 at the dismissal threshold.
Figure 3a: Enrollment discontinuity (first-year standing)

Note: This figure depicts the probability of enrollment in the subsequent term for students with first-year standing, by GPA in term $t$, for students who are predicted to be on academic probation (based on the assignment rules collected by the authors). The running variable is normalized to be 0 at the dismissal threshold.

Discontinuity: -0.044* (0.0251)
Figure 3b: Enrollment discontinuity (second-year or higher standing)

Note: This figure depicts the probability of enrollment in the subsequent term for students with second-year or higher standing, by GPA in term t, for students who are predicted to be on academic probation (based on the assignment rules collected by the authors). The running variable is normalized to be 0 at the dismissal threshold.

Discontinuity: -0.152*** (0.0134)
Figure 4a: GPA histogram

Note: This figure plots the distribution of GPA’s for students who are predicted to be on academic probation (based on the assignment rules collected by the authors). The running variable is normalized to be 0 at the dismissal threshold.
Note: This figure plots a simulated distribution of GPA’s in which we randomly assign grades to students and then calculated the resulting GPA distribution. The spike at 2.0 (and at various other points with no practical importance) illustrate that the nature of how GPA’s are calculated leads to heaping, often at integers.
Figures 5a-5l: Tests of covariate balance
Note: The above figures provide tests of covariate balance around our discontinuity threshold.
Figure 6: Discontinuity in total credits earned

Note: This figure plots average total credits earned in college by GPA relative to the dismissal threshold at their institution. The sample includes only students who are predicted to be on academic probation (based on the assignment rules collected by the authors). The running variable is normalized to be 0 at the dismissal threshold.
Figure 7: Discontinuity in credits earned outside of starting institution

Note: This figure plots the total credits earned by an individual outside of their starting institution by GPA relative to the dismissal threshold at their institution. The sample includes only students who are predicted to be on academic probation (based on the assignment rules collected by the authors). The running variable is normalized to be 0 at the dismissal threshold.
Figure 8: Discontinuity in BA receipt from starting institution

Note: This figure plots the average probability of completing a college degree by GPA relative to the dismissal threshold at their institution. The sample includes only students who are predicted to be on academic probation (based on the assignment rules collected by the authors). The running variable is normalized to be 0 at the dismissal threshold.
Figure 9: Discontinuity in BA Degree receipt from a different institution

Note: This figure plots the likelihood of ever receiving a BA degree from an institution other than an individual’s starting institution by GPA relative to the dismissal threshold at their institution. The sample includes only students who are predicted to be on academic probation (based on the assignment rules collected by the authors). The running variable is normalized to be 0 at the dismissal threshold.
Figure 10: Discontinuity in AA Degree receipt

Note: This figure plots the likelihood of ever receiving an AA degree from a 2-year university by GPA relative to the dismissal threshold at their institution. The sample includes only students who are predicted to be on academic probation (based on the assignment rules collected by the authors). The running variable is normalized to be 0 at the dismissal threshold.
Figure 11a: Discontinuity in mid-run earnings

Note: This figure plots average log weekly earnings measured 7-12 years after initial enrollment in college (see the data section for a more detailed description of the dependent variable). The sample includes only students who are predicted to be on academic probation (based on the assignment rules collected by the authors). The running variable is normalized to be 0 at the dismissal threshold.
Figure 11b: Differential attrition around GPA threshold

Note: This figure plots the likelihood of missing earnings 7-12 years after initial college enrollment. The sample includes only students who are predicted to be on academic probation (based on the assignment rules collected by the authors). The running variable is normalized to be 0 at the dismissal threshold.
Figure 12: Cumulative Density Functions for Date of Last Enrollment

Note: This figure plots the cumulative density functions for both the treatment and control groups for the last date of enrollment in any public Ohio institution relative to the potential dismissal date.
Figure 13: Quarter by Quarter RD Estimates

Note: The above figure plots the results of RD estimates performed on each of 32 quarterly earnings variables. “Control” plots only the constant value from each regression, while “treatment” plots the sum of the constant and treatment coefficients.
Figure 14: Net Present Value Calculations Over the Lifecycle

Note: The above figure plots three separate net present value (NPV) calculations within our data frame, and projections outside of our data frame (indicated by a vertical bar at 32 quarters). The details of the three NPV measures are described in the text.
## Table 1: Summary Statistics

<table>
<thead>
<tr>
<th>Variables</th>
<th>Full Student Sample</th>
<th>At-Risk Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ever on probation</td>
<td>0.24</td>
<td>1.00</td>
</tr>
<tr>
<td>Ever at risk of dismissal</td>
<td>0.10</td>
<td>1.00</td>
</tr>
<tr>
<td>Ever dismissed</td>
<td>0.06</td>
<td>0.53</td>
</tr>
<tr>
<td>Female</td>
<td>0.55</td>
<td>0.45</td>
</tr>
<tr>
<td>Black</td>
<td>0.08</td>
<td>0.17</td>
</tr>
<tr>
<td>Age at college entry</td>
<td>19.14</td>
<td>19.02</td>
</tr>
<tr>
<td>BA from starting institution</td>
<td>0.57</td>
<td>0.27</td>
</tr>
<tr>
<td>BA from any Ohio public institution</td>
<td>0.60</td>
<td>0.29</td>
</tr>
<tr>
<td>AA from any Ohio public institution</td>
<td>0.06</td>
<td>0.06</td>
</tr>
<tr>
<td>Total credits earned from starting institution</td>
<td>105.08</td>
<td>76.12</td>
</tr>
<tr>
<td>Total credits earned from any institution</td>
<td>109.32</td>
<td>79.97</td>
</tr>
<tr>
<td>Years to degree</td>
<td>4.17</td>
<td>4.62</td>
</tr>
<tr>
<td>Missing weekly earnings</td>
<td>0.38</td>
<td>0.32</td>
</tr>
<tr>
<td>Weekly earnings</td>
<td>805.85</td>
<td>647.09</td>
</tr>
<tr>
<td>Age when weekly earnings measured</td>
<td>28.80</td>
<td>28.70</td>
</tr>
<tr>
<td>Observations</td>
<td>218,030</td>
<td>21,664</td>
</tr>
</tbody>
</table>

Note: The full student sample is comprised of all students for whom we observe their initial enrollment in a public Ohio institution. The at-risk sample is comprised of only those students who are predicted to be on academic probation based on the rules for each institution collected by the authors.
Table 2: Tests of covariate balance

<table>
<thead>
<tr>
<th>Sample</th>
<th>All at-risk students</th>
<th>Only students with future mid-run earnings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Credits earned at t-1</td>
<td>-1.420* (0.835)</td>
<td>-1.159 (0.978)</td>
</tr>
<tr>
<td>Female</td>
<td>-0.0259 (0.0193)</td>
<td>0.0135 (0.0239)</td>
</tr>
<tr>
<td>Black</td>
<td>-0.00121 (0.0122)</td>
<td>0.0121 (0.0145)</td>
</tr>
<tr>
<td>Age at college entry</td>
<td>-0.0514 (0.0510)</td>
<td>-0.0333 (0.0567)</td>
</tr>
<tr>
<td>Term GPA in term t</td>
<td>0.0247 (0.0213)</td>
<td>0.0183 (0.0252)</td>
</tr>
<tr>
<td>Credits attempted in time t</td>
<td>0.0294 (0.0737)</td>
<td>-0.0135 (0.0854)</td>
</tr>
<tr>
<td>Ln(weekly earnings) term t-1</td>
<td>0.00541 (0.0332)</td>
<td>-0.0141 (0.0363)</td>
</tr>
<tr>
<td>Employed in year t-1</td>
<td>0.0138 (0.0141)</td>
<td>0.0166 (0.0142)</td>
</tr>
<tr>
<td>Predicted mid-run ln(weekly earnings)</td>
<td>0.00653 (0.00583)</td>
<td>0.00448 (0.00676)</td>
</tr>
<tr>
<td>Predicted BA</td>
<td>0.00113 (0.00522)</td>
<td>0.00211 (0.00608)</td>
</tr>
<tr>
<td>Predicted total credits earned</td>
<td>-0.223 (0.455)</td>
<td>-0.230 (0.528)</td>
</tr>
<tr>
<td>Predicted ln(weekly earnings) term t+1</td>
<td>0.000773 (0.00636)</td>
<td>-0.00128 (0.00728)</td>
</tr>
</tbody>
</table>

Observations: 14,071 10,228

Note: The at-risk sample is comprised of only those students who are predicted to be on academic probation based on the rules for each institution collected by the authors. Mid-run earnings are measured 7-12 years after initial enrollment in college, and is described further in the data section. * p<0.10, ** p<0.05, *** p<0.01
Table 3: RD Estimates of Educational Outcomes

<table>
<thead>
<tr>
<th>Bandwidth</th>
<th>.25</th>
<th>.25</th>
<th>.25</th>
<th>.50</th>
<th>.50</th>
<th>.50</th>
<th>.75</th>
<th>.75</th>
<th>.75</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1.295)</td>
<td>(1.248)</td>
<td>(1.306)</td>
<td>(0.997)</td>
<td>(0.962)</td>
<td>(0.994)</td>
<td>(0.870)</td>
<td>(0.841)</td>
<td>(0.864)</td>
</tr>
<tr>
<td></td>
<td>Controls</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>Observations</td>
<td>9163</td>
<td>9163</td>
<td>8711</td>
<td>14618</td>
<td>14618</td>
<td>14166</td>
<td>18045</td>
<td>18045</td>
</tr>
<tr>
<td></td>
<td>R-Squared</td>
<td>0.282</td>
<td>0.328</td>
<td>0.331</td>
<td>0.291</td>
<td>0.328</td>
<td>0.329</td>
<td>0.295</td>
<td>0.327</td>
</tr>
<tr>
<td></td>
<td>BA from starting institution</td>
<td>-0.0989***</td>
<td>-0.107***</td>
<td>-0.103***</td>
<td>-0.108***</td>
<td>-0.117***</td>
<td>-0.117***</td>
<td>-0.134***</td>
<td>-0.142***</td>
</tr>
<tr>
<td></td>
<td>(0.0189)</td>
<td>(0.0186)</td>
<td>(0.0197)</td>
<td>(0.0145)</td>
<td>(0.0141)</td>
<td>(0.0147)</td>
<td>(0.0127)</td>
<td>(0.0122)</td>
<td>(0.0127)</td>
</tr>
<tr>
<td></td>
<td>Controls</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>Observations</td>
<td>9157</td>
<td>9157</td>
<td>8705</td>
<td>14606</td>
<td>14606</td>
<td>14154</td>
<td>18026</td>
<td>18026</td>
</tr>
<tr>
<td></td>
<td>R-Squared</td>
<td>0.0926</td>
<td>0.123</td>
<td>0.124</td>
<td>0.109</td>
<td>0.136</td>
<td>0.138</td>
<td>0.115</td>
<td>0.143</td>
</tr>
<tr>
<td></td>
<td>Credits earned other schools</td>
<td>0.596</td>
<td>0.517</td>
<td>0.493</td>
<td>0.816**</td>
<td>0.764**</td>
<td>0.846**</td>
<td>0.621*</td>
<td>0.710**</td>
</tr>
<tr>
<td></td>
<td>(0.470)</td>
<td>(0.459)</td>
<td>(0.496)</td>
<td>(0.363)</td>
<td>(0.356)</td>
<td>(0.374)</td>
<td>(0.326)</td>
<td>(0.328)</td>
<td>(0.340)</td>
</tr>
<tr>
<td></td>
<td>Controls</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>Observations</td>
<td>9163</td>
<td>9163</td>
<td>8711</td>
<td>14618</td>
<td>14618</td>
<td>14166</td>
<td>18045</td>
<td>18045</td>
</tr>
<tr>
<td></td>
<td>R-Squared</td>
<td>0.00500</td>
<td>0.0159</td>
<td>0.0167</td>
<td>0.00505</td>
<td>0.0153</td>
<td>0.0158</td>
<td>0.00489</td>
<td>0.0161</td>
</tr>
<tr>
<td></td>
<td>BA earned from other schools</td>
<td>-0.000713</td>
<td>-0.000181</td>
<td>-0.00223</td>
<td>-0.000773</td>
<td>-0.000245</td>
<td>-0.00126</td>
<td>-0.000115</td>
<td>0.000707</td>
</tr>
<tr>
<td></td>
<td>(0.00439)</td>
<td>(0.00442)</td>
<td>(0.00488)</td>
<td>(0.00318)</td>
<td>(0.00325)</td>
<td>(0.00337)</td>
<td>(0.00278)</td>
<td>(0.00285)</td>
<td>(0.00296)</td>
</tr>
<tr>
<td></td>
<td>Controls</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>Observations</td>
<td>9157</td>
<td>9157</td>
<td>8705</td>
<td>14606</td>
<td>14606</td>
<td>14154</td>
<td>18026</td>
<td>18026</td>
</tr>
<tr>
<td></td>
<td>R-Squared</td>
<td>0.00246</td>
<td>0.0107</td>
<td>0.0116</td>
<td>0.00242</td>
<td>0.0106</td>
<td>0.0110</td>
<td>0.00219</td>
<td>0.0104</td>
</tr>
<tr>
<td></td>
<td>AA earned from any Ohio public</td>
<td>-0.000741</td>
<td>-0.000322</td>
<td>-0.0107</td>
<td>-0.00982</td>
<td>-0.00641</td>
<td>-0.00107</td>
<td>-0.000697</td>
<td>-0.00353</td>
</tr>
<tr>
<td></td>
<td>(0.00923)</td>
<td>(0.00908)</td>
<td>(0.00979)</td>
<td>(0.00744)</td>
<td>(0.00724)</td>
<td>(0.00749)</td>
<td>(0.00647)</td>
<td>(0.00616)</td>
<td>(0.00637)</td>
</tr>
<tr>
<td></td>
<td>Controls</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>Observations</td>
<td>9163</td>
<td>9163</td>
<td>8711</td>
<td>14618</td>
<td>14618</td>
<td>14166</td>
<td>18045</td>
<td>18045</td>
</tr>
<tr>
<td></td>
<td>R-Squared</td>
<td>0.00864</td>
<td>0.0330</td>
<td>0.0346</td>
<td>0.00744</td>
<td>0.0283</td>
<td>0.0292</td>
<td>0.00656</td>
<td>0.0268</td>
</tr>
</tbody>
</table>

Note: The above table presents estimates of the discontinuity for a variety of educational outcomes around the critical GPA threshold. Controls refer to: sex, race, age, first term GPA, employment status in term t-1, cumulative credit hours earned and attempted in term t-1, institution fixed effects and year fixed effects. Standard errors are clustered at the individual level. * p<0.10, ** p<0.05, *** p<0.01
Table 4: RD Estimates for Earnings

<table>
<thead>
<tr>
<th>Bandwidth</th>
<th>.25</th>
<th>.25</th>
<th>.25</th>
<th>.50</th>
<th>.50</th>
<th>.50</th>
<th>.75</th>
<th>.75</th>
<th>.75</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variable</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mid-run log weekly earnings</td>
<td>-0.0660**</td>
<td>-0.0445*</td>
<td>-0.0497*</td>
<td>-0.0563***</td>
<td>-0.0453**</td>
<td>-0.0485**</td>
<td>-0.0538***</td>
<td>-0.0506***</td>
<td>-0.0530***</td>
</tr>
<tr>
<td>(0.0270)</td>
<td>(0.0250)</td>
<td>(0.0266)</td>
<td>(0.0207)</td>
<td>(0.0193)</td>
<td>(0.0202)</td>
<td>(0.0184)</td>
<td>(0.0172)</td>
<td>(0.0178)</td>
<td></td>
</tr>
<tr>
<td>Controls</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Observations</td>
<td>6448</td>
<td>6448</td>
<td>6132</td>
<td>10228</td>
<td>10228</td>
<td>9912</td>
<td>12533</td>
<td>12533</td>
<td>12217</td>
</tr>
<tr>
<td>R-Squared</td>
<td>0.00966</td>
<td>0.139</td>
<td>0.141</td>
<td>0.00951</td>
<td>0.133</td>
<td>0.134</td>
<td>0.0110</td>
<td>0.134</td>
<td>0.134</td>
</tr>
<tr>
<td>Missing earnings</td>
<td>-0.0117</td>
<td>-0.0238</td>
<td>-0.0187</td>
<td>-0.0165</td>
<td>-0.0265*</td>
<td>-0.0249*</td>
<td>-0.0168</td>
<td>-0.0252**</td>
<td>-0.0244**</td>
</tr>
<tr>
<td>(0.0184)</td>
<td>(0.0176)</td>
<td>(0.0186)</td>
<td>(0.0142)</td>
<td>(0.0136)</td>
<td>(0.0141)</td>
<td>(0.0125)</td>
<td>(0.0119)</td>
<td>(0.0122)</td>
<td></td>
</tr>
<tr>
<td>Controls</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Observations</td>
<td>9163</td>
<td>9163</td>
<td>8711</td>
<td>14618</td>
<td>14618</td>
<td>14166</td>
<td>18045</td>
<td>18045</td>
<td>17593</td>
</tr>
<tr>
<td>R-Squared</td>
<td>0.00192</td>
<td>0.0826</td>
<td>0.0812</td>
<td>0.00291</td>
<td>0.0808</td>
<td>0.0799</td>
<td>0.00388</td>
<td>0.0824</td>
<td>0.0818</td>
</tr>
</tbody>
</table>

Note: The above table presents estimates of the discontinuity in mid-run (7-12 years after initial college enrollment) log weekly earnings and missing earnings around the critical GPA threshold. Controls refer to: sex, race, age, first term GPA, employment status in term t-1, cumulative credit hours earned and attempted in term t-1, age at which earnings are measured, institution fixed effects and year fixed effects. Standard errors are clustered at the individual level. * p<0.10, ** p<0.05, *** p<0.01
### Table 5: IV Estimates of the returns to a year of college education

<table>
<thead>
<tr>
<th>Variable</th>
<th>Credits/30</th>
<th>Bandwidth</th>
<th>F-statistic</th>
<th>Observations</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.311*</td>
<td>0.25</td>
<td>10.56</td>
<td>6252</td>
</tr>
<tr>
<td></td>
<td>0.246**</td>
<td>0.50</td>
<td>24.83</td>
<td>10150</td>
</tr>
<tr>
<td></td>
<td>0.165***</td>
<td>0.75</td>
<td>77.77</td>
<td>12533</td>
</tr>
<tr>
<td>(0.173)</td>
<td>(0.108)</td>
<td>(0.0591)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: The above table presents instrumental variables estimates of the returns to a year of education on future mid-run earnings. The discontinuity in college credits earned generated from institutional dismissal policies serves as the instrument in the above specifications. All specifications include controls for: sex, race, age, first term GPA, employment status in term t-1, cumulative credit hours earned and attempted in term t-1, age at which earnings are measured, institution fixed effects and year fixed effects. Standard errors are clustered at the individual level. * p<0.10, ** p<0.05, *** p<0.01
Table 6 Heterogeneity

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(0.994)</td>
<td>(1.593)</td>
<td>(1.272)</td>
<td>(1.378)</td>
<td>(1.432)</td>
<td>(2.247)</td>
<td>(1.111)</td>
<td>(1.546)</td>
<td>(1.663)</td>
<td>(1.424)</td>
<td></td>
</tr>
<tr>
<td>Sample</td>
<td>All</td>
<td>Women</td>
<td>Men</td>
<td>Selective</td>
<td>Nonselective</td>
<td>Black/Hispanic</td>
<td>Non-Black/Hispanic</td>
<td>2nd year</td>
<td>3rd year</td>
<td>4th+</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>14166</td>
<td>5814</td>
<td>8325</td>
<td>8493</td>
<td>5673</td>
<td>2709</td>
<td>11457</td>
<td>8139</td>
<td>3681</td>
<td>2346</td>
<td></td>
</tr>
<tr>
<td>R2</td>
<td>0.329</td>
<td>0.338</td>
<td>0.329</td>
<td>0.322</td>
<td>0.355</td>
<td>0.355</td>
<td>0.325</td>
<td>0.167</td>
<td>0.157</td>
<td>0.263</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel B</th>
<th>BA Degree</th>
<th>-0.117***</th>
<th>-0.126***</th>
<th>-0.113***</th>
<th>-0.0985***</th>
<th>-0.135***</th>
<th>-0.117***</th>
<th>-0.117***</th>
<th>-0.0762***</th>
<th>-0.114***</th>
<th>-0.210***</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(0.0147)</td>
<td>(0.0234)</td>
<td>(0.0189)</td>
<td>(0.0190)</td>
<td>(0.0226)</td>
<td>(0.0334)</td>
<td>(0.0163)</td>
<td>(0.0188)</td>
<td>(0.0295)</td>
<td>(0.0359)</td>
<td></td>
</tr>
<tr>
<td>Sample</td>
<td>All</td>
<td>Women</td>
<td>Men</td>
<td>Selective</td>
<td>Nonselective</td>
<td>Black/Hispanic</td>
<td>Non-Black/Hispanic</td>
<td>2nd year</td>
<td>3rd year</td>
<td>4th+</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>14154</td>
<td>5809</td>
<td>8318</td>
<td>8489</td>
<td>5665</td>
<td>2707</td>
<td>11447</td>
<td>8139</td>
<td>3680</td>
<td>2335</td>
<td></td>
</tr>
<tr>
<td>R2</td>
<td>0.138</td>
<td>0.148</td>
<td>0.135</td>
<td>0.131</td>
<td>0.163</td>
<td>0.145</td>
<td>0.138</td>
<td>0.100</td>
<td>0.100</td>
<td>0.127</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel C</th>
<th>Earnings</th>
<th>0.247**</th>
<th>0.339**</th>
<th>0.223</th>
<th>0.235**</th>
<th>0.289</th>
<th>0.264</th>
<th>0.207</th>
<th>0.115</th>
<th>0.357*</th>
<th>0.524</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(0.109)</td>
<td>(0.164)</td>
<td>(0.153)</td>
<td>(0.120)</td>
<td>(0.256)</td>
<td>(0.192)</td>
<td>(0.127)</td>
<td>(0.107)</td>
<td>(0.214)</td>
<td>(0.356)</td>
<td></td>
</tr>
<tr>
<td>Sample</td>
<td>All</td>
<td>Women</td>
<td>Men</td>
<td>Selective</td>
<td>Nonselective</td>
<td>Black/Hispanic</td>
<td>Non-Black/Hispanic</td>
<td>2nd year</td>
<td>3rd year</td>
<td>4th+</td>
<td></td>
</tr>
<tr>
<td>F-Stat</td>
<td>24.91</td>
<td>11.22</td>
<td>12.81</td>
<td>19.44</td>
<td>4.9</td>
<td>6.99</td>
<td>17.6</td>
<td>18.35</td>
<td>8.72</td>
<td>5.91</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>9912</td>
<td>4291</td>
<td>5621</td>
<td>5915</td>
<td>3997</td>
<td>1853</td>
<td>8059</td>
<td>5778</td>
<td>2591</td>
<td>1543</td>
<td></td>
</tr>
</tbody>
</table>

Note: The above table presents RD estimates separately by various demographic and educational factors. In addition to the controls noted in equation (2), all specifications include controls for: sex, race, age, first term GPA, employment status in term t-1, cumulative credit hours earned and attempted in term t-1, institution fixed effects and year fixed effects. Standard errors are clustered at the individual level. * p<0.10, ** p<0.05, *** p<0.01
Table 7: Separating credits earned from BA effect

<table>
<thead>
<tr>
<th>Bandwidth</th>
<th>.25</th>
<th>.25</th>
<th>.25</th>
<th>.50</th>
<th>.50</th>
<th>.50</th>
<th>.75</th>
<th>.75</th>
<th>.75</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dependent Variable</td>
<td>Log(earnings)</td>
<td>BA</td>
<td>Credits/30</td>
<td>Log(earnings)</td>
<td>BA</td>
<td>Credits/30</td>
<td>Log(earnings)</td>
<td>BA</td>
<td>Credits/30</td>
</tr>
<tr>
<td>Variable</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BA</td>
<td>0.345</td>
<td></td>
<td></td>
<td>0.363</td>
<td></td>
<td></td>
<td>0.347</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(0.541)</td>
<td></td>
<td></td>
<td></td>
<td>(0.325)</td>
<td></td>
<td></td>
<td>(0.259)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Credits divided by 30</td>
<td>0.0587</td>
<td></td>
<td></td>
<td>0.0139</td>
<td></td>
<td></td>
<td>0.00412</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(0.320)</td>
<td></td>
<td></td>
<td></td>
<td>(0.181)</td>
<td></td>
<td></td>
<td>(0.118)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Below Cutoff</td>
<td>-0.0635</td>
<td>-0.367**</td>
<td>-0.0134</td>
<td>-0.393***</td>
<td>-0.0516</td>
<td>-0.594***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(0.0764)</td>
<td>(0.172)</td>
<td>(0.0569)</td>
<td>(0.128)</td>
<td>(0.0492)</td>
<td>(0.111)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Standing*Below Cutoff</td>
<td>-0.0178</td>
<td>0.0659</td>
<td>-0.0421**</td>
<td>0.0593</td>
<td>-0.0374**</td>
<td>0.0997***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(0.0279)</td>
<td>(0.0528)</td>
<td>(0.0210)</td>
<td>(0.0399)</td>
<td>(0.0183)</td>
<td>(0.0346)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>6129</td>
<td>6129</td>
<td>6132</td>
<td>9908</td>
<td>9908</td>
<td>9912</td>
<td>12209</td>
<td>12209</td>
<td>12217</td>
</tr>
<tr>
<td>R-Squared</td>
<td>0.140</td>
<td>0.120</td>
<td>0.297</td>
<td>0.145</td>
<td>0.132</td>
<td>0.300</td>
<td>0.149</td>
<td>0.140</td>
<td>0.302</td>
</tr>
<tr>
<td>F-Statistic</td>
<td>11.37</td>
<td>6.4</td>
<td>26.41</td>
<td>17.57</td>
<td>50.09</td>
<td>44.56</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Specification</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IV</td>
<td>Stage</td>
<td>Stage</td>
<td>IV</td>
<td>Stage</td>
<td>Stage</td>
<td>IV</td>
<td>Stage</td>
<td>Stage</td>
<td></td>
</tr>
</tbody>
</table>

Note: The above table presents IV estimates of both the return to a year of college and college degree completion. Separate instruments for each educational measure are generated from the fact that students are on academic probation at different points in time, meaning that the interaction between standing and the discontinuity effectively creates multiple experiments. In addition to the controls noted in equation (2), all specifications include controls for: sex, race, age, first term GPA, employment status in term t-1, cumulative credit hours earned and attempted in term t-1, institution fixed effects and year fixed effects. Standard errors are clustered at the individual level. * p<0.10, ** p<0.05, *** p<0.01
Appendix A: Academic Probation Policies

Below we describe the dismissal policies in place in 2011-2012 for each Ohio public 4-year school. We code policies at the institution level which likely introduces some measurement error since standards can differ across departments within an institution. For each school, we spoke to registrar officials to determine the institutions policies and asked about historical changes. While we were not told about any historical policy changes, it is possible that some institutions changed their policy over time in ways that we are not aware of.\textsuperscript{21} To the extent that we miscode some policies, this weakens the strength of our first stage, but it will not bias estimates.

**Akron University**

Students with a cumulative GPA below 2.0 are placed on probation. Students are dismissed after two terms of probation. Dismissed students are not eligible to register for credit work, but they can enroll in noncredit work. Readmission is possible only with specific approval from the Dean. For students that are readmitted, the length of time they must wait before readmission is case specific. Students may appeal both the initial probation placement and the dismissal decision.

**Bowling Green State University**

The academic standards that a student must achieve depend on the number of credits that student has earned. Students with 1\textsuperscript{st}, 2\textsuperscript{nd}, 3\textsuperscript{rd} and 4\textsuperscript{th} year standing face cumulative GPA cutoffs of 1.5, 1.6, 1.8 and 1.9 respectively. Students are placed on probation if their cumulative GPA falls below the cutoff and are dismissed if their cumulative GPA falls below the cutoff after having been on probation. There are also academic warnings based on higher cutoffs, but these warnings carry no actual consequences. Dismissed students cannot enroll in any courses at BGSU for 1 academic year. Students may appeal both the initial probation placement and the dismissal decision.

**University of Cincinnati**

Students are placed on probation only if they have attempted at least 30 credit hours and have a cumulative GPA below 2.0. Once on probation, students must maintain a term

\textsuperscript{21} In addition to discussing the possibility of policy changes with registrar personnel, we searched through historical student handbooks and used the data to empirically search for changes in the size of the enrollment discontinuity over time for each school. The empirical investigation is based on the idea that if a school changed its policy, the enrollment discontinuity should appear only after that policy change. We did not find any indication of changing dismissal policies empirically, which corroborates our findings based on historical student handbooks and conversations with the registrars.
GPA above 2.0 or they are dismissed. Since students are dismissed based on term GPA, we use term GPA as the running variable for students at University of Cincinnati as well as other schools that use the same sort of policy (e.g. Central State University). The academic dismissal is for one academic year and prevents students from enrolling in any classes at the University of Cincinnati. According to the registrar website, “Readmission after the mandatory suspension period is not automatic. Dismissed students must petition for readmission after serving the suspension period. The dismissed student must submit an application for readmission supported by a letter documenting the reasons for the previous academic difficulties”. Students may appeal the dismissal decision but they may not appeal the initial probation placement.

**Cleveland State University**

Students are placed on probation if their cumulative GPA falls below 2.0 and they are not in their first term. Students are also placed on probation if they fail to complete 67% of their courses, but we do not have access to exact course grades so we only classify probationary status based on the GPA cutoff. This misclassification mostly affects students who withdraw from several courses since their GPA may be well above 2.0 but they can still be placed on probation. Students are dismissed if they do not raise their cumulative GPA above a 2.0 in the term following their probation, though there is an appeal process. Dismissed students cannot enroll for 2 consecutive terms after the dismissal. According to the registrar website, “readmission is a lot like applying to college, though in this case information is sought about what the student has done to make him- or herself academically successful upon return.”. The registrar also states that while it is possible to gain admission to another university following dismissal, “Many universities are becoming less willing to accept students who have been suspended from another university, even if they are only applying to be part-time students.”

**Central State University**

The academic standards that a student must achieve depend on the number of credits that student has earned. Students with between 1-20 credits earned must achieve a 1.7. Students with between 20-40 credits earned must achieve a 1.8. Students with between 40-60 credits earned must achieve a 1.9. Students with over 60 credits earned must achieve a 2.0. Students are placed on probation if their cumulative GPA falls below the cutoff and are dismissed if they do not maintain at least 2.0 term GPA while on probation. Dismissed students must leave for at least 1 term following the dismissal. After the term away, students must reapply to the university in order to gain readmission.

**Kent State University**

The academic standards that a student must achieve depend on the number of credits that student has earned. Students with between 16-30 credits earned must achieve a 1.5. Students with between 30-60 credits earned must achieve a 1.7. Students with between 60-90 credits earned must achieve a 1.8. Students with over 90 credits earned must
achieve a 1.9. Dismissed students must leave for at least 12 months following the dismissal. After the time away, students must reapply to the university in order to gain readmission. According to the registrar “Reinstatement after dismissal from Kent State University is neither automatic nor guaranteed. A student may be reinstated only if the student provides convincing evidence of probable academic success if permitted to return to the university. A dismissed student who has previously accumulated a substantial number of credit hours and/or an excessively low GPA should expect that reinstatement is not likely to be approved.” The registrar also specifically stated that the provost has full discretion with dismissals and students can appeal.

**Miami University**

Students are placed on probation only if they have attempted at least 16 credit hours and have a cumulative GPA below 2.0. Students with fewer than 16 credits attempted are given an academic warning which has no specific consequences. Once on probation, students must maintain a term GPA above 2.0 or they are dismissed. Students are only dismissed if they have already attempted 30 credit hours. The academic dismissal is for 2 consecutive semesters and prevents students from enrolling in any classes at the Miami University. While students are guaranteed re-entry following the dismissal, students are not permitted to transfer any credits earned from other institutions during the dismissal period. Students may appeal the dismissal decision.

**Ohio State University**

Relative to the other public schools in Ohio, Ohio State has a much more flexible dismissal policy that does not emphasize sharp GPA thresholds as heavily. Students with a cumulative GPA below 2.0 are generally placed onto academic probation. To avoid dismissal, students must perform adequately in the preceding term where the definition of adequate is determined based on individual student review by academic advisors and an evaluation board. While we do not have data on the exact manner through which every academic advisor judges adequate performance in the previous term, the administrator we spoke with stated that dismissal is unlikely for a student that is very close to the cumulative GPA threshold or for students who earn above a 2.0 in the most recent term. As such, we use term GPA and a threshold of 2.0 following our approach from Cincinnati University. Given the flexibility of Ohio State’s policies, we also explored simply excluding Ohio State from the analysis and the results are robust to this restriction. Dismissed students can apply for readmission after a case-specific waiting period, but readmission is not guaranteed.

**Ohio University**

Students are placed on probation if their cumulative GPA falls below a 2.0. So long as they continue to make progress towards their degree, students can stay on academic probation for a maximum of 4 terms. At the end of the 4th term on probation, students are dismissed and there is no appeal process. After dismissal, students must leave for at least 12 months and then can reapply. Readmission is not guaranteed and in some cases,
students must wait for longer than 12 months to be allowed to reapply. In 2010-2011 there were major changes to the probation-dismissal policies at Ohio University but none of our sample is impacted by these changes.

**Shawnee State University**

The academic standards that a student must achieve depend on the number of credits that student has attempted. Students with between 0-28 credits attempted must achieve a 1.0. Students with between 28-37 credits attempted must achieve a 1.3. Students with between 37-44 credits attempted must achieve a 1.55. Students with between 44-51 credits attempted must achieve a 1.8. Students with between 51-58 credits attempted must achieve a 1.9. Students with over 58 credits attempted must achieve a 2.0. Students are placed on probation if their cumulative GPA falls below the cutoff and are dismissed if they do not maintain at least 2.0 GPA while on probation. Students must leave for at least 12 months following the dismissal. After the time away, students can appeal their dismissal, though lost financial aid cannot be reinstated through appeal.

**University of Toledo**

The academic standards that a student must achieve depend on the number of credits that student has attempted. Students with between 10-20 credits attempted must achieve a 1.0. Students with between 20-30 credits attempted must achieve a 1.5. Students with between 30-40 credits attempted must achieve a 1.7. Students with between 40-50 credits attempted must achieve a 1.8. Students with between 50-60 credits attempted must achieve a 1.9. Students with over 60 credits attempted must achieve a 2.0. Students are placed on probation if their cumulative GPA falls below the cutoff and are dismissed if they do not raise their cumulative GPA to above the threshold while on probation. Students must leave for at least 1 semester following the dismissal. During the suspension period, students can work to remove incompletes from their transcript but course work taken at other educational institutions during the suspension periods is not accepted as transfer credit. As such, students are effectively barred from any for credit educational investment if they want to return to the University of Toledo.

**Wright State University**

Students are placed on probation only if they have attempted at least 12 credit hours and have a cumulative GPA below 2.0. Students are dismissed if their cumulative GPA drops below the 2.0 while on probation. Dismissal is at the discretion of the chief academic officer and these officers will consider the student’s progress towards meeting degree requirements. After dismissal, students are barred from applying for readmission for 1 full year. After the waiting period, students must apply for readmission and this those that are re-admitted may be subject to special requirements. Students that are granted readmission may also petition using the “Fresh Start Rule” to have some low grades earned prior to the dismissal stricken from their cumulative GPA calculation.
Youngstown State University

The academic standards that a student must achieve depend on the number of credits that student has attempted. Students with between 0-32 credits earned must achieve a 1.75. Students with above 30 credits earned must achieve a 2.0. Students are given an academic warning the first time their cumulative GPA falls below the threshold. Students are placed on probation if their cumulative GPA falls below the cutoff a second time and are dismissed if they do not raise their cumulative GPA to above the threshold while on probation. Students must leave for at least 1 semester following the dismissal. Students must petition the dean to be reinstated and reinstatement following the dismissal is not guaranteed. Students that are granted readmission may be subject to certain terms and conditions.
Appendix B

We first consider the degree of selection on unobservables needed to explain away our results. Building on Altonji, Elder and Taber (2005), Oster (Forthcoming) develops a methodology for assessing the degree of selection on unobservables necessary to explain away the result. This methodology is typically used in selection-on-observables designs, but the basic logic transfers to evaluating the likelihood that the RD estimates are driven by unobservable sorting. Oster’s methodology requires that one chooses, based on theory, the maximum R-squared that could be achieved if researchers had data on both observables and unobservables. The most conservative choice is assuming a maximum R-squared of 1. Oster notes that in many cases (such as wage regressions) even a full set of observable and unobservables are unlikely to yield an R-squared of 1 and so she provides a recommended maximum R-squared (see Oster’s paper for details on this process)\(^{22}\).

Applying the Oster (Forthcoming) methodology to our main specifications, we find that unobservable selection would need to be 800 percent higher than the level of observable selection based on the most conservative assumption of a maximum R-squared of 1. Using Oster’s recommended maximum R-squared, selection on unobservables would need to be 233 times stronger than observable selection. The intuition for why unobservable selection would need to be so much larger than observable selection to explain our results is that many of the observed covariates are strong predictors of outcomes, and none of the covariates are discontinuous at the

\(^{22}\) The precise formula used to calculate the degree of unobservable selection necessary to drive estimates to zero can be seen on page 14 of her manuscript.
threshold. Put another way, if sorting on unobservables were large, one would expect to see at least some discontinuities in covariates such as earnings in t-1, GPA in term 1, race or age at college entry since these variables are likely correlated with many important unobservables such as motivation. Any argument that contends that our wage estimates are driven by unobservable differences on either side of the threshold needs to reconcile itself with the fact that there are no discontinuities in these important covariates. The Oster method provides a quantitative assessment of that likelihood.

As a second approach, we adopt a modified form of the set identification strategy developed in Gerard, Rokkanen and Rothe (2016). While the Oster method suggests that unobservable selection would have to be very large to explain away our estimates, the Gerard et al. (2016) approach derives extreme bounds under the assumption that unobservable selection is infinitely more important than observable selection. The intuition for this procedure is to use the CDF of the running variable to quantify the amount of excess density above the threshold and then to apply the Lee (2009) bounding procedure to this excess density.

Denoting $F(C_{i,t})$ as the cdf of the running variable, we first estimate the regression

$$F(C_{i,t}) = \alpha + \delta c_{i,t} + \beta D_{i,t} + \gamma C_{i,t} D_{i,t} + \epsilon_{i,t} \quad (B1)$$

Where all variables are as defined earlier in the text. $\beta$ is an estimate of the amount of excess density just to the right of the threshold. Following the preferred specification from the main analysis, we estimate equation (B1) using the donut RD and a bandwidth
of 0.5. In our context, we find that there is in fact 2.7% excess density just to the right of the threshold. Though fairly small in magnitude, this excess density has the potential to bias estimates if the manipulating students are unobservably different than other students around the threshold. As such, we re-estimate our original RD design, but we alternately exclude either the top or the bottom 2.7% of observations on the right of the threshold. Specifically, we estimate the exact same regression as our preferred specification described earlier, but prior to estimation, we drop the 2.7% of observations to the right of the threshold with the highest (or lowest) outcome. The intuition for this procedure is that if the 2.7% excess observations to the right of the threshold are different from other observations, at worst, the unobservable selection makes it so that they are either the top 2.7% of observations or the bottom 2.7% of observations. These two extreme assumptions therefore provide bounds on the estimate in the presence of this manipulation. Standard errors are obtained by bootstrapping the entire procedure.

As discussed in Lee (2009), it is possible to use predetermined covariates to further narrow the bounds. Rather than dropping the top (bottom) 2.7% of observations overall, we drop the top (bottom) 2.7% of observations within each covariate cell. Covariate cells are constructed by fully interacting all included covariates and therefore require discrete covariates. While it is theoretically possible to include any number of covariates, because covariates are fully interacted to construct cells, we focus on a small

---

23 The estimated CDF discontinuity is smaller when using a 0.25 bandwidth and therefore the bounds are tighter.
subset of covariates for tractability. Specifically, we construct cells by fully interacting school and cohort.24

Rows 3-6 of Table B1 show the results of the basic bounding exercise. For convenience, the first row shows our main estimates recopied from the earlier tables. For all of the academic outcomes, the simple bounding procedure yields a fairly tight range of estimates and the entire range is qualitatively in line with the estimates shown in row 1. For wages, the simple bounded range similarly matches the main estimate qualitatively, though the upper bound is not statistically significant. The next 2 rows show the results of the bounding procedure when we use school-by-cohort cells to help narrow the bounds. This procedure moderately narrows the bounds and so the upper bound wage effect becomes statistically significant.

It is worth reiterating just how extreme these bounds are. The lower bound estimate makes the assumption that 100% of the excess density comes from the very top of the income distribution. This type of assumption is always extreme, but in our case, it is particularly extreme because we find no discontinuities in any of the covariates. As such, the bound estimates are essentially assuming that despite the fact that the excess density is observably similar to other observations, their unobservables are so dissimilar as to make 100% of these observations drawn from the very top/bottom of the income distribution. If we instead assume that these observations come from the 90th/10th percentiles of the distribution, the estimated bounds tighten considerably.

---

24 The second part of the bounding procedure (where the RD is estimated) continues to include all covariates as before.
A final procedure to assessing whether manipulation can explain our estimates is to exclude observations in the immediate vicinity of the threshold (donut RD). This test is discussed at length in Barreca et al. (2011) but the basic idea is that by excluding observations very close to the threshold, we exclude the observations most likely to have manipulated their GPA. What constitutes “very close” is naturally arbitrary so we estimate the model with several differently sized excluded regions. Specifically, we consider excluding observations that are within 0.05, 0.1 or 0.2 grade points of the threshold. As we increase the size of the excluded region, the RD estimates the left and right limits at the threshold based on data that is increasingly distant from this threshold. This additional extrapolation means that functional form errors will be more important. In particular, since we estimate our models linearly, if the true specification is non-linear, our estimates could be biased in an unknown direction. While it is not possible to definitively identify the importance of this functional form bias, it can be evaluated informally by visually inspecting the RD figures as well as observing the robustness of the results as the size of the excluded region changes. If results are robust as the size of the excluded region changes, it either suggests that manipulation (and functional form bias) do not explain our results, or it could be that these two forms of bias are both important and somehow balance each other out.

The last 3 rows of Table B1 show how the results change as we exclude various interior regions. Because sample sizes monotonically fall as we exclude larger regions, standard errors increase substantially, particularly in final row. That said, the point

---

25 In all cases, we still use a 0.5 bandwidth overall so the 0.2 specification corresponds to estimating equation (2) on observations that are either in the range (-0.5, -0.2) or are in the range (0.2, 0.5).
estimates are generally quite stable across these specifications and are slightly larger than the main estimates shown in column 1.
### Table B1: Robustness Tests

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>Enrolled term t+1</th>
<th>Log weekly earnings</th>
<th>Total Credits Earned</th>
<th>BA from starting institution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline RD Estimate</td>
<td>-0.165***</td>
<td>-0.0485**</td>
<td>-5.680***</td>
<td>-0.117***</td>
</tr>
<tr>
<td>(0.0132)</td>
<td>(0.0202)</td>
<td>(0.994)</td>
<td>(0.0147)</td>
<td></td>
</tr>
<tr>
<td>Upper Bound</td>
<td>-0.162***</td>
<td>-0.0254</td>
<td>-4.575***</td>
<td>-0.105***</td>
</tr>
<tr>
<td>(0.0148)</td>
<td>(0.0200)</td>
<td>(1.071)</td>
<td>(0.0147)</td>
<td></td>
</tr>
<tr>
<td>Lower Bound</td>
<td>-0.189***</td>
<td>-0.0955***</td>
<td>-7.274***</td>
<td>-0.131***</td>
</tr>
<tr>
<td>(0.0144)</td>
<td>(0.0195)</td>
<td>(1.056)</td>
<td>(0.0146)</td>
<td></td>
</tr>
<tr>
<td>Upper Bound (cov. adjusted)</td>
<td>-0.162***</td>
<td>-0.0342*</td>
<td>-4.698***</td>
<td>-0.106***</td>
</tr>
<tr>
<td>(0.0148)</td>
<td>(0.0201)</td>
<td>(1.056)</td>
<td>(0.0146)</td>
<td></td>
</tr>
<tr>
<td>Lower Bound (cov. adjusted)</td>
<td>-0.187***</td>
<td>-0.0854***</td>
<td>-6.910***</td>
<td>-0.128***</td>
</tr>
<tr>
<td>(0.0145)</td>
<td>(0.0195)</td>
<td>(1.056)</td>
<td>(0.0146)</td>
<td></td>
</tr>
<tr>
<td>Donut RD .05</td>
<td>-0.178***</td>
<td>-0.0621**</td>
<td>-6.781***</td>
<td>-0.140***</td>
</tr>
<tr>
<td>(0.0167)</td>
<td>(0.0256)</td>
<td>(1.276)</td>
<td>(0.0184)</td>
<td></td>
</tr>
<tr>
<td>Donut RD .1</td>
<td>-0.212***</td>
<td>-0.0638*</td>
<td>-6.991***</td>
<td>-0.150***</td>
</tr>
<tr>
<td>(0.0215)</td>
<td>(0.0342)</td>
<td>(1.697)</td>
<td>(0.0241)</td>
<td></td>
</tr>
<tr>
<td>Donut RD .2</td>
<td>-0.259***</td>
<td>-0.0606</td>
<td>-9.012***</td>
<td>-0.105**</td>
</tr>
<tr>
<td>(0.0402)</td>
<td>(0.0603)</td>
<td>(3.148)</td>
<td>(0.0416)</td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>14166</td>
<td>9912</td>
<td>9912</td>
<td>9908</td>
</tr>
</tbody>
</table>

Note: The first set of bounds perform the “simple” exercise as described in Appendix B. The second set of bounds use school-by-cohort cells to perform the bounding analysis in order to obtain more efficient bounds. All standard errors are bootstrapped to account for estimation error in the first step of the bounding procedure (estimating the CDF discontinuity). The Donut RD rows drop observations close to the cutoff (within .05, .1, or .2) to remove observations which might be more likely to have manipulated their GPA. * p<0.10, ** p<0.05, *** p<0.01
Appendix C

While the monotonicity assumption is often overlooked in many studies, it is not innocuous in our setting (as is discussed in the main text).

Angrist, et al. (1996, p.451), derives the bias from a violation of the monotonicity assumption to be equal to

\[
\frac{P_t(defier)}{P_t(complier)} - \frac{P_t(defier)}{E(Y_t(1) - Y_t(0)|defier) - E(Y_t(1) - Y_t(0)|complier)}
\]

Since the direction of the bias depends on the relative returns to schooling for defiers vs compliers, it is useful to consider from a theoretical perspective which group is likely to derive larger returns from educational investments. In our context, the compliers are individuals who would have completed more years of schooling but did not as a result of the dismissal. Compliers intended to complete additional postsecondary schooling, and thus they expected to obtain substantial benefits from that schooling. That said, being dismissed for one year was sufficient to deter compliers from completing their intended amount of education. As such, the group of compliers likely has larger returns than those who drop out voluntarily but perhaps smaller returns than the typical graduate. The defiers in our context are students who would normally not complete many years of school, but end up persisting because of the dismissal. As such, we suspect that the returns to education for these students cannot be too large (since without the dismissal they would, by definition, not invest heavily in their education) but the returns are large enough so that they eventually do complete more years of schooling. Thus, from an optimal human capital investment point of view, one might expect that the defiers and compliers are fairly similar in terms of their estimated return to schooling. Naturally, the
preceding discussion is a bit ad-hoc and we are aware of no empirical evidence that
would shed light on the relative rates of return to these two types of students.

The bias equation above is likely to be small when 1) the proportion of defiers in
the population is small, and/or 2) when there is a large first stage effect. The first
condition seems intuitively likely, since it is difficult to imagine that there are a large
number of individuals who choose to obtain more education because they were dismissed
from college previously, although this is not directly testable. The second condition, a
large first-stage impact, effectively means the compliers outnumber the defiers to such a
degree that the estimated coefficient is identified primarily, if not entirely, from the
compliers. This condition is given strong support from the results presented in Table 3.

In order to provide further evidence that the monotonicity assumption is not
violated in our setting, we also perform a test suggested by Angrist and Imbens (1995).
We examine the cdf of completed credit hours for treated students and find that it is first-
order stochastically dominated by the corresponding cdf for nontreated students (see
Appendix Figure C1).26 Angrist and Imbens (1995) first presented this as a test for the
monotonicity assumption in the case where the endogenous variable (credit hours in our
case) is multi-valued. Fiorini et al (2013) later argues that first order stochastic
dominance is a necessary, rather than a sufficient condition for monotonicity, but is still
strong suggestive evidence that the monotonicity assumption holds.

26 We restrict this figure to the sample analyzed in the main text (with higher than first
year standing) which is why no students have fewer than 30 credits earned.
Appendix Figure C1: CDF’s of Credits Earned for Treated and Nontreated