

College Merit-Aid and Academic Renewal Thresholds: Evidence from Interactions of Discontinuous Rules

Benjamin Bleatings*

This Version: Thursday 4th March, 2021, Most Recent Version: [Link](#)

PRELIMINARY DRAFT:

PLEASE DO NOT CITE, SHARE, OR CIRCULATE

Abstract

Merit-based financial aid programs are amongst the largest sources of financial aid for college, yet estimated effects may be biased by students sorting into treatment due to known academic cutoffs. This paper estimates how likely students are to meet minimum academic renewal requirements, in a state where sorting into treatment has been shown, by using an unmanipulated birthdate cutoff created by compulsory kindergarten entry laws. Fuzzy regression discontinuity shows being born slightly after the cutoff, 14 years before the college program begins, leads to a 49.6 percentage point (pp) increase in scholarship receipt. Difference in discontinuity estimates, accounting for potential relative age and season of birth confounding, find students are 24 pp more likely to exceed GPA and completed credits renewal requirements. The results suggest GPA or credits thresholds are equally effective, advancing the literature on behavioral aspects of large merit-aid programs and informing their design.¹

Keywords: Merit-Aid Scholarship, Difference-in-Discontinuity

JEL Classification: H42, H75, I22

*Benjamin Bleatings, Ph.D. Candidate, Department of Economics, West Virginia University, Morgantown WV 26506. E-mail: bbleming@mix.wvu.edu

¹I thank my adviser, Adam Nowak, and my dissertation committee, Dan Grossman, Scott Cunningham, Bryan McCannon, and Brad Humphreys. I thank Bill Even, John Winters, Daniel Hungerman, and Gregorio Caetano for many helpful comments and suggestions. I thank seminar participants at Eastern Economic Association 2020, Southern Economic Association 2020, West Virginia University Brown Bag 2020, Southern Utah University, Kansas University, Wayne State University, and the Computational Justice Lab at Claremont Graduate University. I thank Liz Reynolds for graciously providing detailed administrative data.

1 Introduction

College graduates make \$1 million more in lifetime earnings on average than non-college graduates (Carnevale, Cheah, & Hanson, 2015); however, college costs remain a primary obstacle to college (Cowan, 2016; Solis, 2017). To combat financial constraints, college financial aid has grown substantially.² West Virginia spends \$40 million yearly (West Virginia Higher Education Policy Commission, 2014) on merit-based aid and Georgia had, from 1993-2011, 1,383,718 students receive merit aid.³ Figure A.1 shows 27 states with merit-aid programs.⁴ Abbott, Gallipoli, Meghir, and Violante (2019) suggests that expanding ability-tested grants is superior to labor supply tax cuts or expanding student loans, but merit-aid is criticized for being ineffective (Ortagus, Kelchen, Rosigner, & Voorhees, 2018) and a transfer to the wealthy (C. Cornwell & Mustard, 2007; Dynarski, 2004).

To inform debate about merit-aid and compare it to other interventions, it is crucial to get accurate treatment effects for the many outcomes merit-aid affects (C. Cornwell & Mustard, 2007; Cowan & White, 2015; Goodman, 2008; Scott-Clayton, 2011; Scott-Clayton & Zafar, 2019; Sjöquist & Winters, 2014, 2015b, 2015c). Since causal effects are used for comparing the welfare effects of government policies, accurate estimates have significant implications for understanding how merit-aid compares to other government policies (Hendren & Sprung-Keyser, 2020). A seemingly attractive research design to evaluate the causal effects of merit-aid is regression discontinuity (RD), because there are defined academic cutoffs which discontinuously change the probability of treatment. Using entry test (ACT, SAT) scores and/or high school GPA cutoffs is similar to test cutoffs used in the original RD study (Cook, 2008; Thistlethwaite & Campbell, 1960).

²College grant aid per full-time enrolled student has doubled from \$4,360 in 1998 to \$9,520 in 2018 (College Board, 2019).

³Georgia data accessed at: https://web.archive.org/web/20111006161755/https://www.gsfcc.org/GSFCNEW/SandG.facts.CFM?guid=&returnurl=http://www.gacollege411.org/Financial_Aid_Planning/HOPE_Program/Georgia_s_HOPE_Scholarship_Program_Overview.aspx. This citation required the use of waybackmachine.com.

⁴This is the classification given by Dynarski (2004), Heller and Marin (2004), The Brookings Institution, and State Administrative Websites. 9 states offer a generous large program. To be considered large, it means a large amount of students are covered and/or a large portion of expenses are covered.

However, an important difference between current merit-aid research and Thistlethwaite and Campbell (1960) is that academic cutoffs for many merit-aid scholarships are known. Known academic cutoffs present a challenge to the RD design, because individuals can manipulate scores into receiving treatment and there is a large economic incentive to do so.⁵ Indeed, students strategically manipulate academic scores into being just high enough to receive treatment, by retaking entry exams such as the ACT, in Tennessee, Florida, and West Virginia (Bruce & Carruthers, 2014; Scott-Clayton & Zafar, 2019; Zhang, Hu, Sun, & Pu, 2016). This selection into treatment is not consistent with identifying continuity assumptions typically used for RD (Dong, 2018; Hahn, Todd, & Van der Klaauw, 2001).

This paper estimates the effects of West Virginia's merit-aid program, that at the time covered the entirety of students' tuition and fees, on reaching academic thresholds to renew the scholarship. Academic renewal thresholds have been an important design element of merit-aid programs (C. M. Cornwell, Lee, & Mustard, 2005; Jia, 2019) and are comparable to performance goal-setting schemes (Beattie, Laliberté, Michaud-Leclerc, & Oreopoulos, 2019; Clark, Gill, Prowse, & Rush, 2020). West Virginia's merit-aid program, the Promise Scholarship, has known academic cutoffs; so, the RD design in this paper exploits variation in college scholarship receipt due to being born on opposite sides of the kindergarten entry cutoff date. This date is set by statewide compulsory schooling laws in place 14 years before the college scholarship begins in 2002. By being born on opposite sides of the cutoff, students enter in different kindergarten cohorts which affects their college cohort and eligibility to receive the scholarship.

This research design is less vulnerable to running variable manipulation, because birthdate is realized 16 years before the scholarship announcement and 19 years before it begins. Because the variables that lead to differences in scholarship receipt are realized 20 years before the scholarship begins, it is implausible that future students or their parents manipulated birthdates to increase their likelihood of becoming eligible for the scholarship. Since the cutoff determines both scholarship eligibility and age relative to the rest of ones' cohort, the prior year's date cutoff is used to examine

⁵Two academic criteria are entry test scores and high school GPA. Entry tests can be retaken and high school GPA can be inflated through easier courses being taken or idiosyncratic high school decisions.

and remove potential confounding due to relative age. Fusing compulsory entry laws as variation in education (Black, Devereux, & Salvanes, 2011; Clark & Royer, 2013) and a birthdate-based discontinuity in financial aid for college (Denning, 2019), this design uses early education laws as a natural experiment in college grant receipt.⁶ This design addresses two pervasive challenges, selection into treatment and sample, in the voluminous, active quasi-experimental literature on merit-aid program evaluation.

First, internal validity is investigated. The kindergarten cohort birthdate cutoff creates variation in college cohort and college scholarship receipt. Academically eligible students born slightly after the kindergarten cohort birthdate cutoff are 53.7 percentage points (pp) more likely to enter in the first eligible college cohort and 49.6 pp more likely to receive the scholarship than academically eligible students born slightly before the cutoff.⁷ Additionally, the treatment and control groups are reasonably similar, because density tests suggest birthdate is unmanipulated (McCrary, 2008) relative to the the prior year's birthdate cutoff (Grembi, Nannicini, & Troiano, 2016).

This alternative research design reveals new insights into the effects of merit-aid scholarships on academic outcomes. There is no difference in the likelihood of reaching pace-based (credit hours) or distinction-based (GPA) performance goals. Point estimates indicate students who receive the scholarship are 24 to 30 pp more likely to hit both the credits and GPA renewal thresholds in their first year of college.

There are heterogeneous effects on exceeding academic renewal thresholds by gender and academic credentials. Results suggest that those above the median ACT score are more likely to reach the credits threshold than those below it. Also, students below the median high school GPA are more likely to reach the GPA thresholds. Finally, there is no difference between men and women for exceeding renewal credits in each of the first 3 years; however, men are more likely to complete the GPA requirement in each of their first 3 years, but women are not.

⁶A direction for future research in financial aid noted by Dynarski and Scott-Clayton (2013) is analyzing how college financial aid interacts with other programs.

⁷Students who are sophomores in college when the program begins (i.e. begin college in 2001) are not retroactively eligible.

One way these estimates may diverge from a causal treatment effect of the scholarship is due to relative age confounding at the cutoff (Deming & Dynarski, 2008); however, relative age effects are found to be empirically negligible in this context.⁸ Relative age changes discontinuously at the pre-scholarship cutoff and scholarship eligibility does not. Specifically, students are 14 pp less likely to be male, score 0.84 points lower on the ACT Science subsection, and score 83.31 points higher on their combined SAT scores at the pre-scholarship cutoff, which is hardly indicative of systematic relative age effects on pre-treatment covariates.

Second, exceeding the birthdate cutoff only has positive effects on 1 of 4 group's academic performance in college, those who are academically eligible and born just beyond the post-scholarship cutoff (i.e. those who become eligible for the scholarship by exceeding the cutoff).⁹ Finally, subtracting the intent-to-treat effects of relative age from being just beyond the pre-scholarship cutoff from the local average RD treatment effect, using a difference in discontinuities estimate (Grembi et al., 2016), has nearly no effect on the point estimates. These estimates are also not vulnerable to concerns about how mothers' characteristics differ across seasons of birth, because there is no reason to believe there is a difference in mothers' characteristics on either side of the pre-scholarship cutoff compared to the post-scholarship cutoff (Buckles & Hungerman, 2013).

Students must finish both requirements to renew the scholarship and research in Georgia has found merit-aid reduces academic performance (C. M. Cornwell et al., 2005), so the large discrepancy in prior work (Scott-Clayton, 2011) warrants additional attention. The most closely related prior work finds the merit-aid scholarship increases the likelihood of hitting pace-based (credit hours) performance was 24 pp and distinction-based (GPA) was 8 pp (Scott-Clayton, 2011). The results in this paper suggest that this difference is a limitation of methods used, rather than a substantive behavior difference.¹⁰ Policies based on this apparent difference in behavioral responses to different measures of academic performance are likely misguided. Furthermore, it supports the

⁸For a review of relative ages effect on mostly education outcomes, see Peña (2017).

⁹There is no effect for those who are academically ineligible at either cutoff and there is no effect for academically eligible students born just beyond the pre-scholarship cutoff.

¹⁰Only the widest bandwidths with uniform kernels are able to replicate the differences across GPA and credits thresholds.

conclusions of C. M. Cornwell et al. (2005), which suggests that the reason Georgia's merit-aid reduced academic performance was it did not limit the number of semesters that aid could be used.

These results are important given the growth of interest in behavioral interventions within education (Gneezy, Meier, & Rey-Biel, 2011; Lavecchia, Liu, & Oreopoulos, 2016; Levitt, List, Neckermann, & Sadoff, 2016). By setting a threshold for renewal, the program provides a reference point (Kahneman & Tversky, 1979) and this point conceivably represents a goal. Those who fail out of college are less likely to set goals (Beattie et al., 2019).

However, the performance goal set by the minimum renewal threshold might be expected to be ineffective without the incentive of financial aid. Without financial incentives, performance goals have been found to be less effective than task-based goals for college students (Clark et al., 2020). Similarly, without accompanying support, financial aid for performance has also been found to be less effective (J. Angrist, Lang, & Oreopoulos, 2009). These results suggest that a large financial incentive for completing a minimum level of academic performance can effectively increase academic performance. Indeed, if students complete the 30 credits renewal requirement each year, they will have obtained the minimum number of credits, 120, required to graduate in 4 years. While there are justified targeting concerns with merit-aid, it is effective at bolstering academic performance in West Virginia.

Beyond policy relevance of merit-aid, this research fits the broad direction of applied microeconomics. The alternative source of variation is consistent with estimating treatment effects using natural experiments (J. D. Angrist & Pischke, 2010).¹¹ In contrast to methodological approaches to addressing heterogeneity in treatment effects (Hsu & Shen, 2019), monotonicity (Bertanha & Imbens, 2019), or extrapolating away from cutoffs in RD designs (J. D. Angrist & Rokkanen, 2015; Dong & Lewbel, 2015), this paper estimates a “more average” effect using a different natural experiment. By using a birthdate running variable, the estimated effects are not local to an academic cutoff; estimates are local near a date, making generalizations away from the cutoff more likely to be appropriate.

¹¹The new source of variation provides a new way to form a counterfactual to solve the fundamental problem of causal inference (Holland, 1986).

Exact birthdate is continuous (while ACT is more discrete) which is consistent with identifying assumptions typically used (Hahn et al., 2001). The continuous running variable is less vulnerable to error due to specification bias caused by discrete variables not being arbitrarily close to the cutoff (Lee & Card, 2008) or choice of standard errors (Kolesár & Rothe, 2018). The estimates may be more heterogeneous, because potential students have less exact control of birthdate compared to entry test (ACT/SAT) scores. Higher running variable variance (as is arguably the case with a birthdate running variable compared to an entry test score) implies a greater mixture of students affected by the birthdate cutoff (Bloom, 2012; Lee, 2008) meaning more groups represented by estimates.

Finally, the empirical strategy in this paper, using Kindergarten birthdate cutoffs to investigate a program beginning with a later grade, has started to become more regularly applied (Canaan, 2020; Takaku & Yokoyama, 2021). This strategy is appealing, because it approximates a natural experiment. It compares students who are born a few days apart and that unknowingly at the time sorted them into eligible and not eligible groups. Also, it can be widely applied, because compulsory laws are widespread and many programs begin with a specific cohort. Developing this strategy can result in a new, rigorous method for evaluating causal effects of programs in which individuals are treated beginning with a specific cohort and compulsory schooling entry birthdate cutoff laws existed when said individuals entered school.¹²

This paper extends prior work with this empirical strategy in three ways. First, it uses exact date of birth as the running variable while previous work has been limited to birth month-year. Second, it explicitly models the process for how kindergarten birthdate cutoffs can be used to estimate effects of programs that begin in a grade beyond kindergarten, noting specifically that this strategy depends on grade progression. This basic model can be used to inform assessments of how appropriate this strategy is for other contexts and/or programs and who the compliers are

¹²Also note that in general applied economics has recently seen a massive increase in exploiting quasi-random variation from historical sources, such as exploration routes (Duranton & Turner, 2011), colonization (Acemoglu, Johnson, & Robinson, 2001), bombing (Dell & Querubin, 2018), and policy changes (Lleras-Muney & Shertzer, 2015).

([J. D. Angrist, Imbens, & Rubin, 1996](#)). Finally, this is the first paper to apply this empirical strategy to the debate surrounding the effects of merit-based aid for college.

The following section discusses related prior research. Then, the programs' rules and resulting variation are formally modeled. Next the variation is validated as being viable and internally valid. Finally, methods and results are presented, followed by a brief conclusion.

2 Related Literature

This section begins by reviewing students' response to incentives created by the academic renewal thresholds that are common for merit-aid programs. Then, the research designs that are commonly used to estimate effects of merit-aid are outlined. The section concludes with comments on alternative strategies to address the econometric issues caused by students manipulating academic scores into treatment.

2.1 Behavior Responses to Academic Renewal Thresholds

There are several important aspects of merit-aid program design, an important one is academic renewal thresholds. There are discrepancies across different states' merit-aid programs in how merit-aid affects pace-based academic performance. In Georgia, students were allowed to use their aid for 180 credits and that reduced the likelihood of completing a "full-load" by 9.3% and reduced the credits completed by 1 ([C. M. Cornwell et al., 2005](#)). When there is a credits time limit and academic renewal thresholds required for renewal, students would drop courses to ensure the academic thresholds were exceeded.

Contrast Georgia's credit limit with West Virginia, where students can only use aid for 8 contiguous semesters. When the limit on funds is determined by semesters, not credit hours, students are much more likely to reach academic renewal thresholds. The puzzle that remains for academic renewal thresholds in West Virginia is why are students 8 pp more likely to complete GPA require-

ments and 24 pp more likely to complete credits requirements ([Scott-Clayton, 2011](#))?¹³ After all, students need to hit both renewal thresholds to renew the scholarship. Furthermore, does changing the time limit for receiving merit-aid from credits to semesters switch the sign from negative to modestly positive (8 pp) or to a large positive (24 pp) or is it actually different for different academic thresholds?

Another behavioral response to academic renewal thresholds is major choice. Students might choose easier majors to reduce the risk of meeting academic renewal thresholds. Students choose easier majors in Georgia ([Sjoquist & Winters, 2015a](#)) and nationwide ([Sjoquist & Winters, 2015b](#)), to make meeting the thresholds easier. However, students have not been shown to be less likely to choose STEM majors in West Virginia ([Scott-Clayton, 2011](#)).

2.2 Research Designs

There are two primary sources of variation that have been used to estimate the impacts of merit-aid: cohort-based and academic eligibility based. The academic eligibility design exploits small differences in test scores and/or GPA along with academic eligibility cutoffs. However, in many states these academic eligibility cutoffs are known, leading to students manipulating academic scores into being eligible which is inconsistent with the oft-used continuity assumptions for identification.

The cohort-based variation exploits the fact that students who enter college before merit-aid programs begin are not eligible retroactively. This leads to two different designs, which depend on data availability. The first design estimates differences-in-differences models, where the control group is states that do not begin the program and estimated effects are intent-to-treat (ITT) if individual-level data is used.

The ITT effects can be scaled to a local average treatment effect if it is observed who actually receives the scholarship as in [Scott-Clayton \(2011\)](#) and [Scott-Clayton and Zafar \(2019\)](#). In this setup, entering in an eligible cohort is used as an instrumental variable for scholarship receipt.

¹³In West Virginia, students are found to increase their total credits in their first year by 1.8 and there are no significant impacts on GPA ([Scott-Clayton, 2011](#)).

Scholarship receipt is then used to explain individual-level outcomes. One limitation of this approach is that the “enter in an eligible cohort” instrumental variable is not necessarily defined at the individual level, it is defined at the cohort level. Seen from this perspective, a cohort-level instrumental variable is being used to explain individual-level outcomes which may be less than ideal.¹⁴ Furthermore, the discrete instrument makes it unclear who the compliers are.

2.3 Previous Strategies for Addressing Academic Score Manipulation

In order to estimate causal effects of merit-aid by addressing academic score manipulation, there are two strategies that have been seen in the literature so far. The first strategy is to use the first entry test score instead of the best entry test score ([Bruce & Carruthers, 2014](#)). There is no manipulation in the first score, because students have not retaken the test yet. The limitation of this approach is that it requires additional data on whether the score recorded is the first or best, which may not be possible with many datasets.¹⁵

The second strategy for addressing academic entry manipulation is to use unknown academic cutoffs. If the academic scores are not known, students cannot sort into treatment by manipulating their scores just beyond the threshold. One state with ex ante unknown cutoffs for merit-aid is Massachusetts ([Cohodes & Goodman, 2014](#); [Goodman, 2008](#)). The limitation of using ex ante unknown cutoffs is that it limits the states that can be studied. Limited states with unknown academic cutoffs forces researchers to argue about whether results in Massachusetts are externally valid for states like West Virginia. This may be a difficult position, because West Virginia’s population likely differs substantially in relevant ways from Massachusetts. Even still, regression discontinuity estimates from academic qualification thresholds may not generalize to students far away from cutoffs without additional strong assumptions ([J. D. Angrist & Rokkanen, 2015](#)).

¹⁴It is shown in [Scott-Clayton \(2011\)](#) that collapsing to cohort-level means, the effect remains statistically significant.

¹⁵In fact, it is not explicit in studies that use administrative state datasets whether the test scores are first score or best score or something else like combined across multiple attempts. It could be that some schools report the best score and others report the first score. Combining across attempts-a process known as superscoring- is how this university assigns institutional scholarships.

3 Data

The data is student-level of the 2000-2002 cohorts at a large, public university in West Virginia.¹⁶

The sample consists of only in-state students who are first-time freshman. Exact birthdate, gender, race, high school GPA, entry test scores, and academic outcomes up to the third year of college are observed. Table A.1 presents summary statistics of the sample.

Scholarship Eligibility The data has information on 3 requirements required to be eligible for the scholarship:

1. Earn at least a 3.0 GPA in high school,
2. ACT Composite score ≥ 21 OR SAT Combined Score $\geq 1,000$,
3. One must be a college freshman in the 2002 cohort or after.

As the program is merit-based, 1 and 2 are requirements related to academic proficiency. To simplify, requirements 1 and 2 are combined into a single variable, academic eligibility, defined as

$$Acad. Eligible_i = 1[(HS\ GPA \geq 3) \cap (ACT\ Comp \geq 21 \cup SAT\ Comb \geq 1,000)],$$

in which *HS GPA* stands for high school GPA, *ACT Comp* stands for ACT composite score, and *SAT Comb* stands for SAT combined score. The academic requirements change for subsequent cohorts, but the change is not relevant for this analysis, since there is only 1 eligible cohort used.¹⁷

Exact Birthdate and Cutoffs Exact birthdate is observed. According to standard practice, it is centered so that day = 0 is the first birthdate that is expected to be discontinuously more likely to receive the scholarship. The cutoff at day = 0 is called the post-scholarship cutoff, because being

¹⁶ After multiple rounds of data requests over multiple years, the West Virginia Higher Education Policy Commission (WVHEPC) ultimately denied the data request for statewide data.

¹⁷ There are also subscore requirements, minimum scores on each section- math, science, reading, verbal, etc.- that are not incorporated into the analysis.

born on or after this date makes one more likely to receive the scholarship, as detailed in Section 4. Furthermore, students born at day = -365 are more likely to be the oldest in their cohort, but not more likely to receive the scholarship. Because it is a cutoff that does not determine scholarship receipt, it is referred to as the pre-scholarship cutoff throughout this paper.

3.1 Outcomes: Academic Performance Based Renewal

The dependent variables, y , are binary indicators for whether a student finished the minimum required credits or GPA in their first year. Students must finish at least 30 credits in each year to renew their scholarship.¹⁸ In their first year, they must obtain at least a 2.75 GPA.¹⁹ By focusing on the first year, it ensures that results are not driven by selection out of the sample as students drop out of college. While continuous outcomes are interesting, binary outcomes for exceeding renewal requirements have been used before (Scott-Clayton, 2011), performance based funding is criticized (Ortagus et al., 2018), performance based goals have been shown to be less effective than task based goals (Clark et al., 2020), and the binary indicators have implications for program design (C. M. Cornwell et al., 2005).

3.1.1 Outcomes of Interest By Cohort

Figure A.3 visualizes how a differences in differences analysis would estimate intent to treat effects of the program by plotting outcomes by cohort and academic eligibility.²⁰ There seems to be a positive intent to treat effect for both likelihood of completing credits (Figure A.3A) and GPA (Figure A.3B). However, a worrisome problem with estimating treatment effects this way is that, due to manipulation of test scores, those who would have had a 20, now have a 21 and find themselves in the academically eligible group. Furthermore, estimating the treatment effects this way

¹⁸They have the Fall, Spring, and Summer semester to do so.

¹⁹For every year afterwards they must obtain at least a 3.0 GPA.

²⁰It would be possible to use out of state students as a comparison group like Scott-Clayton (2011), but I do not have data on out of state students.

is unable to leverage the data fully, specifically it does not use the information of who received the scholarship which is in the data.

Possible Biases From Academic Score Manipulation The effect of this academic score sorting is ambiguous on the direction it biases the treatment effects. On one hand, students may study more and thus increase their scores which may bias estimates upward. However, students did not have this incentive before and so students would not have retaken tests. Thus they would have still had a 20 in cohorts before the program begins. If they sort into treatment, but are not more academically able, then the treatment effects of the scholarship are biased downwards, because the academically eligible are less able on average. The next section describes an alternative source of variation to circumvent relying on mismeasured and manipulated academic eligibility as variation.

4 Kindergarten Cohort and College Aid Receipt

This section discusses the discontinuous rules and related factors that lead to discontinuous college scholarship receipt at the kindergarten birthdate cutoff, 14 years before the scholarship begins. The main insight builds on scholarship receipt depending on college entry cohort. Specifically, under a mild grade progression assumption, college entry cohorts depend predictably on kindergarten entry cohorts.

If college entry cohort depends predictably on kindergarten cohort, and the scholarship depends on college entry cohort, then this is an alternative source of variation to estimate effects of merit-aid programs on individuals. Conveniently, kindergarten entry cohort depends discontinuously on birthdate in relation to kindergarten birthdate cutoff due to state compulsory entry laws. While at least two other papers ([Canaan, 2020](#); [Takaku & Yokoyama, 2021](#)) have used this logic to investigate treatments that begin with a specific cohort, this is the first paper that makes the underlying theory responsible for the discontinuous variation mathematically explicit.

Equations 1, 2, and 3 outline the system of equations which leads to the variation in scholarship receipt exploited in this paper,

$$\text{Promise Scholarship} = f(\text{College Cohort}), \quad (1)$$

$$\text{College Cohort} = g(\text{Kindergarten Cohort}), \quad (2)$$

$$\text{Kindergarten Cohort} = h(\text{Birthdate Relative to Cutoff}). \quad (3)$$

Equations 1 and 3 are known, discontinuous rules about scholarship eligibility and kindergarten cohort entry respectively. Equation 2 is related to grade progression. By substitution,

$$\text{Promise Scholarship} = f(g(h(\text{Birthdate Relative to Kindergarten Cutoff}))). \quad (4)$$

Each of these equations is discussed below.

First, Equation 1 represents the rule that to receive the scholarship, students must enter college after the scholarship program begins. Equation 1B expands Equation 1 into its discontinuous form,

$$P(\text{Scholarship} = 1) = \begin{cases} q' + k' & \text{if Entry Cohort} \geq 2002 \\ q' & \text{if Entry Cohort} < 2002. \end{cases} \quad (1B)$$

The constant k' can be considered how important being in an eligible entry cohort is for scholarship receipt. Prior research suggests $k' > 0$, meaning that entry cohort is an important eligibility factor determining scholarship receipt (Scott-Clayton, 2011; Scott-Clayton & Zafar, 2019).

Next, Equation 3 represents the compulsory kindergarten entry birthdate cutoff rule. This is a rule set by laws that were in place in 1988 in West Virginia. Students who have not turned 5 by this date, September 1, are required to enter kindergarten in the next cohort, effectively requiring

a birthday, BD , of a certain date. Equation 3B expands Equation 3 into its discontinuous form for the 1989 kindergarten cohort,

$$P(\text{Kindergarten Cohort} = 1989) = \begin{cases} p' + c' & \text{if } BD \geq \overline{BD}_{y=1988} \\ p' & \text{if } BD < \overline{BD}_{y=1988}. \end{cases} \quad (3B)$$

The equation suggests that with a large enough c' , students are discontinuously more likely to enter kindergarten in 1989 if born after the cutoff date, $\overline{BD}_{y=1988}$, for entering kindergarten in 1988. The only way c' could equal 0, invalidating the design, is that the cutoff laws are not enforced. In practice, the cutoff is imperfectly enforced and costly to circumvent, as discussed in Section B.²¹

Finally, Equation 2 represents a grade progression function. In order to link Equation 1 to Equation 3 in a systematic and predictable way for estimating effects of the college scholarship, Equation 2 must take a specific functional form for a large enough share of the population. The assumed functional form of g is

$$\text{College Cohort} = \text{Kindergarten Cohort} + 13, \quad (2B)$$

that is students finish one grade per year (kindergarten through 12th grade) and then enter college. This is a reasonable assumption for the purposes of estimating the effects of merit-based aid, because it is unlikely that students who receive merit-based aid are being held back.

²¹Since kindergarten entry cohort is not observed, it is impossible to estimate c' (the degree of imperfect enforcement) using the available data.

²²There are 13 grades between kindergarten and college. For a program beginning with a different year, 13 would be a different number. For example, if the program began in 6th grade like Canaan (2020), then 13 would be 6.

Kindergarten entry cohort and grade progression are not observed, but the assumption in Equation 2B can be implicitly supported by the data. Under the one grade per year progression assumption in Equation 2B, Equation 3B implies,

$$P(\text{College Cohort} = 2002) = \begin{cases} p + c & \text{if } BD \geq \overline{BD}_{y=1988} \\ p & \text{if } BD < \overline{BD}_{y=1988}. \end{cases} \quad (5)$$

If $c \neq 0$, then it implies that Equation 2B is realistic for a large enough share of the population. The reason c' , which previously represented only how strictly the compulsory law was enforced, is now c , is because c also depends on how large of a share of the population follows the one grade per year assumption. A final practical matter which might push c towards 0, to the detriment of the design, is immigration to West Virginia from other states, with different cutoffs, in grade school.^{23,24,25} Since beginning in the eligible year matters (Scott-Clayton, 2011; Scott-Clayton & Zafar, 2019), Equation 5 implies

$$P(\text{Scholarship} = 1) = \begin{cases} q + k & \text{if } BD \geq \overline{BD}_{y=1988} \\ q & \text{if } BD < \overline{BD}_{y=1988}. \end{cases} \quad (6)$$

If $c \neq 0$ and beginning in the first year of the program is an important enough requirement, then it is expected that $k \neq 0$. Without additional restrictions, $c \neq k$, because academic eligibility is also an important criterion for eligibility.

Finally, it is worth noting that $p = 0$ and $c = 1$ and/or $q = 0$ and $k = 1$ are not necessary (or sufficient) conditions for this design to be valid. Some students may be held back at some point

²³This could be remedied if state of birth was observed; however, migration will cause estimates to be conservative. It is likely this is not much of problem, given the average annual growth rate of the population of West Virginia from 1950-2000 was -0.12%.

²⁴The scholarship has a rule that a student must gain resident status at least 1 year in advance, so students cannot migrate into West Virginia to receive the scholarship.

²⁵(Lleras-Muney, 2002, pg. 407) notes a similar point in footnote 36, going further to point out that inter-state mobility was probably not related to compulsory entry laws, which is likely also the case for this paper. Further, Lleras-Muney (2002) notes that mobility is low in the US in the period of study (Card & Krueger, 1992). Mobility into West Virginia is likely also low. Finally, it is highly unlikely that someone moved into West Virginia early in grade school to receive the scholarship.

in grade school, receive a waiver to enter in a later cohort, or migrate from another state in grade school which would make $p > 0$ and/or $c < 1$, which will almost certainly affect q and k also. Since it is likely that $(p + c) - p < 1$, a fuzzy regression discontinuity design is used instead of a sharp design. All the terms in Equations 5 and 6 are observed, so p , c , q , and k are non-parametrically estimated, around $\overline{BD}_{y=1988}$ (and $\overline{BD}_{y=1987}$), in Section 5.1.

4.1 Scholastic and Birthdate Cutoff Comparisons

For a 10% random sample, Figure 2 plots how scholarship receipt depends on birthdate and academic credentials. There are few orange o's and many blue x's to the left of the birthdate cutoff and vice-versa to the right of the cutoff in Figures 2A and 2B. This suggests that c and $k > 0$.

Figure 2A compares the ACT cutoff to the birthdate cutoff. There are less individuals just below the ACT cutoff for those born after the birth cutoff, which is consistent with findings in Scott-Clayton (2011) and Scott-Clayton and Zafar (2019). There are a limited number of observations below the cutoff and many above the cutoff which is not ideal for RD, because it suggests students manipulate scores into treatment.²⁶ The limited number of observations near the ACT cutoff is also not good for RD, because it needs many observations near the cutoff.

Figure 2B compares the HS GPA cutoff to the birthdate cutoff. There are very few observations near the GPA cutoff before or after the birthdate cutoff, meaning it is unlikely the GPA cutoffs are appropriate for use in an RD design. There are many students who have near perfect/perfect high school GPAs on either side of the date cutoff. This shows the advantage of using a non-academic cutoff, it is possible to compare students of similar academic ability before and after the cutoff.²⁷

²⁶A formal investigation of density on opposite sides of the cutoff is performed in Section 5.2.

²⁷One more interesting takeaway message from Figure 2B is there is only 3 students who have a HS GPA below 2.5 born after the cutoff date. There are 6 students below 2.5 in (-730, -365) and 9 students below 2.5 in (-365,0). In the full sample, these numbers are 93/1801 students and in (-365,0) there are 93/1796, in (0, 365) there are 59/1868. This is likely related to high school grade inflation and/or crowd out of the less academically able students.

4.2 Threats to Identification

There are two main threats to identification. The first is relative age effects, which is addressed by using the prior year's cutoff which does not affect scholarship receipt. Relative age is estimated using the pre-scholarship cutoff and found to be empirically negligible. The second threat is there being another policy that begins with the cohort that begins Kindergarten in 1989. A search for such policies finds no policies that affect those discontinuously on either side of the birth cutoff.

These are discussed in detail in Section C.

5 Internal Validity of Research Design

This section investigates internal validity of using the proposed kindergarten cutoff as variation in scholarship receipt. First, the first stage is tested, specifically examining whether small differences in birthdate lead to different cohort year of entering college (c' in Equation 5) and differential likelihood of scholarship receipt (k' in Equation 6.) Second, implications of the identification assumption are examined in two ways. Density tests investigate smoothness in of number of observations at the date cutoff. Second, in the vein of a randomized controlled trial (RCT), pre-treatment covariate balance at the birthdate cutoffs are examined.

5.1 First Stage

The logic underlying the research design is that birthdate relative to kindergarten entry birthdate cutoffs predicts college entry cohort. Since scholarship eligibility is cohort-based, the birthdate cutoff also leads to variation in scholarship receipt. This section, describing Figure 3, tests whether birthday relative to the kindergarten birthdate cutoff causes differences in college entry cohort and scholarship receipt.

5.1.1 Effect of Kindergarten Birthdate Cutoff on College Entry Cohort

The proposed research design requires a discontinuous increase in likelihood of beginning college 13 years later at the kindergarten entry birthdate cutoffs for beginning kindergarten. Figure 3A plots the likelihood of entering college in 2001, which is 1 cohort prior to the scholarship beginning. The cutoff is at birthdate = -365, because this pre-scholarship cutoff is on the same exact date in the prior year.

Students born just to the left of the cutoff in Figure 3A, who are therefore expected to enter college in 2000 if they finish one grade per year, have an approximately 20% chance of entering college in 2001. As the students birthdates are further to the left of the cutoff, it becomes less likely that they enter in 2001, eventually reaching 0 after a few months. The leading explanation is that students just at the cutoff are more likely to successfully evade the cutoff in some way, so they can enter in a later kindergarten cohort.

Students born just to the right of the cutoff in Figure 3A, who are therefore expected to enter college in 2001 if they finish one grade per year, have an approximately 70% chance of entering college in 2001. This probability grows as students are born further to right, until reaching a 100% chance after a few months. This probability falls again after a few months, because the data also contains students who enter college in 2002. Using a flexible local-linear regression with a triangular kernel and MSE-optimal bandwidths to estimate the difference in the conditional expectation function on either side of the cutoff, there is a difference of 52.6 percentage points at the cutoff.

Figure 3C inspects whether the kindergarten cutoff at birthdate = 0 determines whether a student enters college in 2002, the first year of the scholarship. This is the expected college cohort given expected kindergarten cohort and regular grade progression. Just like at the prior cutoff, about 20% of students to the left of the cutoff enter later than expected and about 75% enter in their expected cohort.²⁸ Using the same local-linear method for calculating the difference, the

²⁸Because there is no data on the 2003 cohort, the probability of entering in 2002 stays at 1 and does not exhibit the same reduction in probability far from the cutoff like in Figure 3A.

kindergarten entry cutoff leads to an approximately 53.3 pp discontinuity in college entry cohort for entering in the first eligible college cohort. That is, $c = .533$.

No Evidence Students Delay College Entry One concern is that, because the scholarship is announced in 1999 and the program begins in 2002, students have an incentive to delay the beginning of college; however, this is unlikely for two reasons. First, the optimal decision on delay depends on comparing the value of the scholarship to the cost of delaying entering the labor market after college by 1 year as discussed in [C. M. Cornwell et al. \(2005\)](#). Second, the year before the scholarship begins has a 52.6 pp difference at the cutoff, while the year after the scholarship has a 53.3 pp cutoff which are not statistically different. Also, academically eligible students, who the strongest incentive to delay and are shown in Figure [3E](#), are the same as the aggregate sample, with a 53.7 pp difference in entering in the expected cohort at the cutoff.

5.1.2 Effect of Kindergarten Birthdate Cutoff on Scholarship Receipt

Figure [3B](#) shows the effect of exceeding the kindergarten birthdate cutoff, which by grade progression would have students enter in 2001 which is 1 year before the scholarship begins, on scholarship receipt. Students born immediately to the left and right of this cutoff do not receive this scholarship. This is expected, because being born after the cutoff leads one to be expected to enter college in the cohort before the scholarship begins.

Figure [3D](#) shows the effect of being born to the right of the kindergarten birthdate cutoff, which by grade progression would have students enter in 2002 which is the first year the scholarship begins, on scholarship receipt. Students to immediately to the left of the cutoff receive the scholarship with about a 10% probability, while students to the right of the birthdate cutoff receive the scholarship with about a 50% probability.²⁹ Similar to the probability of entering in a given cohort, the probability falls to the left of the cutoff as get farther from it and increases to the right of the cutoff as one gets farther from it. Using the same local-linear methods, the difference in

²⁹There is substantial variance in the probabiltiy of receiving the scholarship to the right of the cutoff, with 40% in the lowest bin receiving and over 80% in the highest bin having receiving it.

scholarship receipt at the cutoff is 35.5 pp, that is $k = .355$, which is adequate in terms of a first stage effect on likelihood of receiving treatment. Also $c > k$, because Figure 3D does not account for a the academic eligibility criterion for merit-based aid.

Figure 3F repeats Figure 3D for students who are academically eligible. Students immediately to the left of the cutoff receive the scholarship with 20% probability, while those to the right receive it with 70% probability.^{30,31} The difference in scholarship receipt, k , at the cutoff is 49.6 pp, which is much closer to the 53.7 pp difference in Figure 3E.³² After accounting for academic eligibility, $c \rightarrow k$.³³ In summary, the kindergarten cutoff is a strong enough predictor of college cohort ($c > 0$) and scholarship receipt ($k > 0$) to be used in a RD design.

5.2 Density Around ACT Composite Score and Birthdate Cutoffs

The identifying assumption for the fuzzy RD research design used in this paper is local smoothness of potential outcomes at the cutoff (Hahn et al., 2001).³⁴ This assumption cannot be directly tested; however, one way to support the viability of this assumption is to test the density of the number of observations, around the birthdate cutoff (McCrory, 2008). The null hypothesis of the density test is a smooth distribution around the cutoff, so failing to reject is implicit support of the identifying assumption (McCrory, 2008).

The intuition is that if the distribution of observations is not smooth around the cutoff, then students who manipulated the running variable may be different from those who didn't (observably or unobservably). If students on opposite sides of the cutoff differ in their ability to manipulate the running variable, then it may inappropriate to assume that one side is an appropriate counterfactual

³⁰Also, the variance on the right side of the cutoff decreases by including information on academic eligibility. Whereas there was a range of about .4, there is now only a range of about .25.

³¹The probability of receiving the scholarship does get close to 1 after enough dates after the cutoff. It may not get all the way to 1, due to other types of aid replacing the Promise scholarship.

³²Figures A.4 and A.5 show that those to the right of the cutoff receive at least 1 additional year of the scholarship at the cutoff, because some students lose their scholarship (Carruthers & Özek, 2016).

³³The reason $c \neq k$ is due to other unobserved tertiary factors, such as ACT subscore and HS GPA subject requirements.

³⁴Other identifying assumptions include local randomization (Cattaneo, Frandsen, & Titiunik, 2015), which allows the use of randomization inference methods for regression discontinuity.

for the other. A failure to reject is evidence students are not manipulating the running variable into being eligible for treatment.³⁵ Ex-ante, this is a believable proposition, because students and parents did not know that being born slightly later would lead to being eligible for the scholarship. Something that could cause this test to fail is an increase in enrollment, which has been found in prior research (C. Cornwell, Mustard, & Sridhar, 2006; Scott-Clayton, 2011). However, to fail the test would require that enrollment increase right at the cutoff, relative to the pre-scholarship cutoff (Grembi et al., 2016).

ACT Composite Score Figure 4A shows there is a relatively smooth distribution of observations around the ACT cutoff value of 21 in the pre-scholarship cohorts. There are nearly the same number of students (about 280) who score 20 as 21 and less students score 23 and 24 than 19 or 20. After a score of 21, there is a steady decline in the number of people who receive each subsequent score. Figure 4B shows there is no longer a smooth distribution of ACT scores around the cutoff value of 21 for the first cohort of students who can receive a Promise scholarship if they get at least a 21.³⁶ In the first cohort of the scholarship beginning more students receive a 21-24 than a 19 or 20. This is unsurprising, because students know the cutoff and there is a large economic incentive to retake tests until they exceed it. The findings in 4A and 4B replicate the ACT score manipulation in Scott-Clayton (2011) in a subsample of the state-wide data. The result suggests that using ACT score as a cutoff in an RD design is not consistent with the identification assumption that potential outcomes are smooth locally (Hahn et al., 2001).

5.2.1 Birthdate

Figures 4C and 4D show the birthdate density around the pre-scholarship and post-scholarship kindergarten cutoffs. There are 26 bins that contain 14 days (2 weeks) on either side of the cutoffs. In contrast to Figures 4A and 4B, there are many more bars in total, because the underlying running

³⁵On the other hand, rejecting the null hypothesis is not a sufficient condition for running variable manipulation.

³⁶A formal density test at the ACT cutoff is not presented here, because the discrete running variable, ACT composite score, leads to econometric issues with the test (Frandsen, 2017) .

variable is continuous. Continuous running variables are favorable, compared to discrete variables, for RD designs.

Figure 4C inspects the density around the birthdate cutoff, before the cutoff matters for scholarship receipt. There is a larger number of observations immediately to the right of the cutoff than immediately to the left. For 3 bins of 2 weeks to the right of the cutoff, there are at least 80 students and the 3 bins to the left have 75 or less students. Importantly this is not due to the scholarship, because the scholarship does not come into force until the next cohort. As the birthdates get farther from the cutoff, the increase in density levels out.

Figure 4D inspects the density around the birthdate cutoff, when the cutoff does matter for scholarship receipt. There is a similarly sized discontinuity in the distribution at the cutoff to Figure 4C when the cutoff does not matter for scholarship eligibility. For 3 bins to the right of the cutoff, there are about 75 students and for 3 bins to the left of the cutoff there are about 65 students.

As distance to the right of the cutoff increases, the difference in the distribution does not level out as much as Figure 4C. Since a goal of the law is to reduce state brain-drain (Sjoquist & Winters, 2014) and prior work finds an increase in total students (Scott-Clayton, 2011), this is not entirely unexpected. However, it is also not inconsistent with the identifying assumptions, which would require a discontinuous density right at the cutoff relative to the prior year.³⁷ One possible interpretation is that the scholarship had no effect on in-state attendance for those who are the oldest in their cohort.

Figures 4E and 4F present formal density tests of the distributions of number of observations around the cutoffs before it matters for scholarship receipt and when it does. The exact method for the implementation follows Cattaneo, Jansson, and Ma (2019). The histograms in Figures 4C and 4D required a choice for the bin width, but these density tests do not require pre-binning. Bin widths are selected based on asymptotic mean square error (AMSE).

Figure 4E shows the test of the density at the cutoff, before the cutoff determines scholarship receipt. The point estimate of the density to the left of the cutoff is .00144 and the point estimate on

³⁷There are 1,791 students to left and 1,792 students to the right of the cutoff in Figure 4C. There are 1792 to the left and 1861 to the right in Figure 4D.

the right is .00185, a difference of 0.00041. The null hypothesis of smoothness is not rejected, with a p-value of 0.1618 which suggests there is no effect of relative age on enrolling at this university prior to the beginning of the scholarship.

Figure 4F repeats the same procedure for the cutoff that determines scholarship receipt. The point estimate left of the cutoff is 0.00133 and the point estimate to the right is 0.0014, a difference of 0.00007. The null hypothesis of smoothness is not rejected, with a p-value of 0.8131, which suggests there is no effect of the scholarship on enrolling, at the birth cutoff, at this university.³⁸ This test suggests there is no manipulation of birthday or selection into the sample at the cutoff.

5.3 Pre-Treatment Covariate Balance

One issue identified in the density test is that density no longer levels off after a few weeks past the cutoff, so this section investigates whether this results in students who are different along observable pre-treatment covariates close to either side of the cutoffs. If the average of pre-treatment covariates at the pre-treatment covariates is the same at the cutoff, then it also supports the identifying continuity assumptions. The standard method, that is used below, for detecting differences in pre-treatment covariates is to substitute the pre-treatment covariates for the outcomes of interest, then fit local linear regressions on both sides of the cutoff.

5.3.1 Scholarship and Pre-Scholarship Cutoffs

Table 1 presents results from pre-treatment covariate balance tests.³⁹ Columns 1 and 4 present the results from the post-scholarship cutoff. Columns 2 and 5 present results from the pre-scholarship cutoff does not matter for scholarship receipt.

Scholarship Year In column 1, there are no statistically significant effects of being born just to the right of the cutoff on any outcome- male, ACT composite score, ACT English, or ACT Reading

³⁸When attempting to impose strong assumptions on the higher order derivatives and cumulative distribution, which increase the power of the test, the optimal bandwidths exceed the limit of the data for the post-scholarship cutoff (Cattaneo, Jansson, & Ma, 2019).

³⁹Figure A.6 shows plots that visualize these tests.

at the post-scholarship cutoff. Column 4 shows that there are no statistically significant effects of being born after the cutoff on SAT combined score, ACT Math, or ACT Science sections. Column 4, Panel A does show that students born after the cutoff in the year of the scholarship have a high school GPA that is 0.13 points higher than those to the left of the cutoff. This is a small effect, as it is only 1/3 of the difference between an A- and an A or an A and an A+ and only 1/2 to 1/4 of a standard deviation in high school GPA.

Pre-Scholarship In column 2, there are no statistically significant effects of being just beyond the cutoff, in the year before the cutoff, on ACT composite score (Panel B), ACT english score (Panel C), or ACT reading (Panel D). Panel A, column 2 does show that being just beyond the cutoff, in the year before the scholarship, shows a 14 pp decline in probability of being male. In column 5, there are no statistically significant effects of being beyond the cutoff on high school GPA (Panel A) or ACT math (Panel C). Those beyond the pre-scholarship cutoff have an 83.31 point increase in SAT Combined score, which is statistically significant at 95% confidence. However, their ACT science scores are 0.84 points lower, which is statistically significant at 90% confidence.

No Existence of Relative Age Effects on Pre-Treatment Covariates There is no consistent evidence that students on either side of the cutoff are different. Students treated by relative age at the pre-scholarship cutoff score higher on the SAT, but score lower on the ACT science scores. If the story about relative age leading to either more or less qualified students, it would be expected for the cutoff to affect all academic outcomes in the same direction.⁴⁰

5.3.2 Did Scholarship Affect the Difference at the Cutoffs?

Columns 3 and 6 contain the most informative results for assessing pre-treatment covariate balance. These columns show the difference between the two cutoffs. Because there is a possibility that these covariates are discontinuous due to relative age effects, before the scholarship, strong

⁴⁰Another possibility is that students who take the SAT and ACT are materially different in some unspecified way, although this seems unlikely.

evidence of pre-treatment covariate imbalance requires that the scholarship cutoff widens this discontinuity beyond where it was previously. Sometimes, the estimate reported in columns 3 and 6 is not the exact difference of the preceding columns, because the bandwidth is constrained to a common number that also ensures there is no overlap of birthdates (and students) used in the estimation of the two discontinuities.

Of all estimates in column 3 and 6, the only one that is statistically significant is Panel A, column 3. The estimate in Panel A, column 3 estimates that the scholarship increased the probability of being male by 17 pp at the cutoff. While this is less than ideal, it is important to note that this is driven by a statistically significant and large coefficient on the pre-cutoff. ACT math and ACT science are nearly 10% significant, suggesting the post-cutoff increases pre-treatment academic ability relative to the pre-cutoff. These effects are also driven primarily by the pre-cutoff.

Finally, there are no statistically significant differences between the pre and post cutoffs on either high school GPA (Panel A, column 6) or ACT composite score (Panel B, column 3). High school GPA is only 0.06 points higher at the post cutoff relative to the pre cutoff. The ACT composite score is only 0.47 points higher at the post scholarship cutoff compared to the pre scholarship cutoff.

It is not supported by the data that there are massive relative age related differences (columns 1, 2, 4, and 5) at the cutoffs. It is also not supported that there are large differences in college attendance decisions (i.e. highly qualified students choosing to attend in-state instead) at the cutoffs that are induced by the scholarship beginning (columns 3 and 6) . The pre-treatment covariate balance tests are thus consistent with the identifying assumption.

6 Methods

Fuzzy regression discontinuity is used to estimate the effect of scholarship receipt on academic performance. It exploits psuedo-random assignment of the scholarship, which is a result of state-wide kindergarten compulsory entry age assigning students to different cohorts based on small

differences in birth-date. The local average treatment effect of the scholarship is estimated via a two-stage approach.

The first stage follows,

$$S_i = \lambda + \psi After_i + f(Birthdate_i) + \epsilon_i, \quad \forall |Birthdate_i| < h, \quad (7)$$

in which subscript i stands for individual and c stands for cohort. The effect of birthdate is a flexible local linear regression that, in both regressions, is fit separately on either side of the cutoff. S stands for scholarship receipt. $After$ is a dummy variable that equals 1 if a student enters in cohort 2002, the first cohort that students were eligible. The coefficient ψ has been shown to be large, around .4 or .5 in Section 5.1.

To construct the Wald estimate of the local average treatment effect the reduced form is estimated by the following equation,

$$y_i = \alpha + \beta After_i + f(Birthdate_i) + e_i, \quad \forall |Birthdate_i| < h. \quad (8)$$

The estimated intent to treat (ITT) effect of the scholarship is β , the observed difference between the regressions on either side of the cutoff, at the cutoff. The Wald local average treatment effect of the scholarship is $\tau = \frac{\beta}{\psi}$.

Bandwidth and Kernel While the local linear regression approach above does not require specifying functional form, it does require selecting a bandwidth (h) and kernel for weighting. Using smaller bandwidths reduces bias and increases variance, so the mean-square error (MSE) optimal bandwidths are used, which optimally trades off bias and variance (Imbens & Kalyanaraman, 2012). A triangular kernel, which puts more weight on closer observations, is preferred, because it gives point estimates with optimal properties when used in tandem with MSE-optimal bandwidths (Cattaneo, Idrobo, & Titiunik, 2019, pg. 44).⁴¹

⁴¹Results are similar when varying the kernel and bandwidth and also to using bias-correction procedures (Calonico, Cattaneo, & Titiunik, 2014).

6.1 Identification

Next, assumptions for τ to represent the causal effect of receiving the scholarship on academic performance are considered. First, the standard instrumental variables assumptions are described. Next, regression discontinuity assumptions are discussed.

First Stage, Independence, and Exclusion Equations 7 and 8 outline a fuzzy regression discontinuity approach, and since treatment does not go from 0 to 1 at the cutoff, the usual instrumental variable assumptions are required.⁴² For the design to be valid, there must be a first stage, which is discontinuous change in the likelihood of scholarship receipt at the date cutoff. Section 5.1 establishes that this assumption is met. The exclusion restriction requires small differences in birthday affect outcomes only through scholarship receipt and independence requires that small differences in birthday are exogenous to scholarship receipt. As discussed in Section C, there are some concerns that these are violated; however, Sections 5.2 and 5.3 find no empirical evidence that this is the case.

Monotonicity Two other usual assumptions of instrumental variables are monotonicity and the Stable Treatment Unit Value Assumption (SUTVA.) Monotonicity simply means that being to the right of the date cutoff does not reduce the likelihood of entering in the later cohort and subsequently receiving the scholarship for any individual student which is reasonable. Furthermore, this builds on Scott-Clayton (2011), because the compliers that the estimated effects are valid for are students who finish one grade per year in this approach.

Stable Unit Treatment Value Assumption Next, SUTVA assumes that the potential outcomes of students to the left of the cutoff are not affected by the treatment of students to the right of the cutoff. This assumption is quite reasonable for year 1 outcomes, but when students on either side of the cutoff are in school at the same time, one cohort being treated may cause students in the other cohort to also enroll in more credits to be on pace with peers. This is part of the reason to

⁴²Dong (2018) requires less assumptions.

focus on the first year outcomes. Another reason to focus on 1st year outcomes is SUTVA assumes equal dosages of treatment, which is violated for long-term outcomes, because students lose their scholarship if they fall below the scholarship, as noted by Carruthers and Özek (2016).⁴³

Local Smoothness The identifying assumption is local smoothness of potential outcomes (Dong, 2018; Hahn et al., 2001).⁴⁴ Under local smoothness of potential outcomes, it is valid to use units on either side of the cutoffs as counterfactuals for one another and calculate the treatment effect as the difference in outcomes at the cutoff.⁴⁵ Researchers support local smoothness using the McCrary Density test (McCrary, 2008) and ensuring that pre-treatment covariates are balanced for treatment and control groups. The tests presented in Sections 5.2 and 5.3 are consistent with local smoothness of potential outcomes.

7 Main Results

First, reduced form plots and estimates show the relationship between the date cutoff and academic outcomes. Then, the local linear regression discontinuity estimates are presented. Next, difference in discontinuity results confirm that the estimated effects are driven by scholarship receipt.

7.1 Reduced Form Effects

First, the reduced form effects are visualized and the intent to treat effects are visualized. The intent-to-treat effects are estimated first, because they are the numerator of the local average treatment effect. If the intent-to-treat effects are not zero at the post-scholarship cutoff for academically eligible students, are zero at the post-scholarship cutoff for academically ineligible, and are zero at

⁴³Figures A.4 and A.5 show that many lose their scholarship, but there is still a discontinuity when considering the treatment being 4 years of scholarship.

⁴⁴Figure A.8 shows the identification assumption. To be a causal effect, potential outcomes must be smooth at the cutoff.

⁴⁵These potential outcomes are not observed, so this assumption is not directly testable. This is the essence of the “fundamental problem of causal inference,” (Holland, 1986, pg. 947).

the pre-scholarship cutoff for both types of students, then it is suggestive that the estimated effects are driven by scholarship receipt.

7.1.1 Reduced Form Plots

Figure 5 plots the reduced form relationships for academically eligible students at the pre and post-scholarship cutoffs. In Figures 5A and 5C, there is not obviously distinguishable effect of exceeding the pre-scholarship cutoff on outcomes. This suggests there is no detectable effect of being the oldest in one's cohort on the outcomes of interest.

At the post-scholarship cutoff shown in Figures 5B and 5D, there are more distinguishable effect of exceeding the cutoff on outcomes. In Figure 5B there is an increase in from around 50% before the cutoff to near 70% after the cutoff. In Figure 5D the mean before is around 60% and the mean after is around 70%. In Figures 5B and 5D the outcomes increase as the distance to the cutoff increases, which is consistent with how the scholarship receipt grows after the scholarship.⁴⁶

7.1.2 Intent to Treat Estimates

Table 2 presents ITT estimates, from Equation 8, across academically eligible and academically ineligible students for the pre and post scholarship cuotffs. Column 1, the estimated ITT effect of exceeding the post-scholarship cutoff for academically eligible students, is consistently estimated to be statistically significant at 90% confidence and in line with the approximate sizes seen in Figures 5A and 5C. Column 2 estimates the same effect, except at the pre-scholarship cutoff where relative age effects may be present but scholarship effects are not. Just as seen in Figures 5A and 5C, there is no statistically significant effects of exceeding the pre-scholarship cutoff on exceeding the renewal credits or GPA thresholds. Column 3 shows the difference in the estimated ITT effects across the cutoffs, which is between 0.196 and 0.114, but often not statistically significant due to the increased imprecision that comes from estimating two separate discontinuities.⁴⁷

⁴⁶This exercise is repeated for academically ineligible students in Figure A.9.

⁴⁷Column 3 restricts the bandwidths, h , to 160 to ensure that no observations are used in the estimation of both discontinuities.

Column 4 repeats the estimation of ITT effects for academically ineligible students at the post-scholarship cutoff and there are no statistically significant estimates, which is consistent with the effects being driven by scholarship receipt and not relative age or time trends. Column 5 repeats the estimation of ITT effects for academically ineligible students at the pre-scholarship cutoff and finds no statistically significant effects. Column 6 finds the difference at the pre and post scholarship cutoffs for academically ineligible students is between -0.048 and 0.104 and never statistically significant.⁴⁸ Figure 5 and Table 2 are strong support that estimates are not driven by relative age or season of birth confounding, they are driven by scholarship receipt primarily.

7.2 Local Average Treatment Effects from Fuzzy Regression Discontinuity

Table 3 presents estimates of scholarship receipt on first year academic outcomes. There are varying choices of kernel, sample, covariates, coefficient adjustments, and bandwidth selection criterion to account for several justifiable ways of estimating the effects. Table 3, Panel A focuses on the likelihood of completing 30 credits in ones' first year. In column 1, no bias-correction or robustness corrections to the estimates are done and the scholarship increases the likelihood of completing the scholarship by 36 percentage points, which is significant above 95 percent confidence. In column 2, the bias-correction and robustness adjustments are done (Calonico, Cattaneo, Farrell, & Titiunik, 2017), and the results are essentially unchanged.

In columns 1 and 2 (in Panel A and B) academically ineligible students are included, making the control group less ideal. In column 3, academically ineligible students are dropped and the effect size decreases to 27 percentage points, probably due to a more academically gifted control group being utilized. Additionally, the statistical significance drops to 90%. In column 4, academic controls for HS GPA and entry test (ACT/SAT) scores are included to compare individuals of similar scores.⁴⁹ The effect size drops to 0.24, which is the exact point estimate Scott-Clayton (2011)

⁴⁸Also, the positive estimate of the difference are driven by a large negative at the pre-scholarship cutoff for academically ineligible students.

⁴⁹Non-academically eligible students are dropped before academic controls are included, because covariate inclusion in RD assumes that the controls are additively separable from the running variable (Calonico, Cattaneo, Farrell, & Titiunik, 2019). Dropping non-academically eligible students increases the plausibility of this assumption.

reports, but the coefficient is no longer statistically significant at conventional levels ($p < 0.128$). Column 5 uses CER-optimal bandwidths, which can be more desirable for inference, but less desirable for point estimates. The estimated effect does not change, but the standard error grows, reducing the p-value.

Columns 6 and 7 use a uniform kernel, because it minimizes asymptotic variance under certain conditions (Cattaneo, Idrobo, & Titiunik, 2019, pg. 44). The difference in scholarship receipt at the cutoff increases noticeably with the distance from the cutoff for a few months before flattening, so triangular kernels place more weight on potentially unaffected observations. Column 6 uses the MSE-optimal bandwidth criterion with a uniform kernel, the effect size remains similar, but the standard error drops and the effect is statistically significant above 90% ($p < 0.085$). In column 7, a uniform kernel is used with the coverage error rate optimal bandwidths, the effect size grows to a 31 percentage point increase. The standard error also grows, causing this estimate to remain statistically significant at 94.6 percent confidence. The point estimates are similar to Scott-Clayton (2011), sometimes slightly larger depending on bandwidth, sample, controls, and correction.

Table 3 Panel B replicates Panel A, using the likelihood of meeting the renewal GPA of 2.75 as the dependent variable. Column 1 shows the scholarship increases the likelihood of hitting the GPA threshold for renewal by 31 percentage points. The effect is statistically significant above 95%. Columns 2-4 add parameter corrections, subsample to academically eligible students, and include academic controls; all changes leave the effect size and statistical significance almost unchanged. Column 5 uses the CER-optimal bandwidth and finds a 33% larger effect. This point estimate should be interpreted cautiously as the point estimate does not have optimal point estimate properties (Cattaneo, Idrobo, & Titiunik, 2019, pg. 44).

Columns 6 and 7 of Table 3, Panel B use the uniform kernel. In column 6, the coefficient is 33% smaller, but the standard error is reduced. Still, the coefficient is only statistically significant at 9% confidence. In column 7, a uniform kernel is used with the CER-optimal bandwidths and the estimate becomes significant at 95 percent confidence. and the effect grows back to near the

previously estimated effects. These estimates are similar to the increased likelihood of completing the renewal credits that are presented in Panel A of Table 3, in contrast to Scott-Clayton (2011).⁵⁰

7.3 Difference in Discontinuity Estimates

Second, difference in discontinuity is used, which takes advantage of the cutoff being the exact same date in the prior year. Under certain assumptions relative age can be differenced out of the treatment effect estimate. To address the confounding effect of differential relative age at the entry date cutoff is difference-in-discontinuity is used (Canaan, 2020; Grembi et al., 2016; Smith, 2016).⁵¹ The effect of relative age at the cutoff is present in the years before the scholarship becomes available. To remove the confounding effect of relative age, one simply subtracts off the effect at the birth cutoff in prior periods

$$T^{DD} \equiv (Y^- - Y^+) - (\tilde{Y}^- - \tilde{Y}^+),$$

where \tilde{Y} are the outcomes at the cutoff in period before the cutoff determines scholarship receipt. To implement this estimator, the data is formatted so that 6 months on either side of each cutoff are related to that particular cutoff. Effectively, the difference in discontinuity estimate is the local average treatment effect (τ , shown in Table 3) minus the ITT effect ($\beta_{\text{pre-cutoff}}$), at the pre-scholarship cutoff, from Equation 8 (shown in Table 2, column 2).

7.3.1 Identification Assumptions

A new assumption is the effect of relative age, in the case of no treatment, is constant over time. This is similar to the parallel trends assumption for differences-in-differences, with the important difference that this assumption is local. Under typical IV assumptions (due to the fuzzy cutoff), local smoothness, and local “parallel trends”, T^{DD} is a causal effect of receiving the scholarship.

⁵⁰Table A.3 estimates parametric functions that assume functional forms for the birthdate running variable and find similar results when dropping birthdates farther away from the cutoff.

⁵¹This estimator differs from other RD estimators that exploits longitudinal data, because the variation it uses is cross-sectional, not within-unit. Other estimators include fixed effect RD estimators (Pettersson-Lidbom, 2012), first-difference RD estimator (Lemieux & Milligan, 2008), and dynamic RD designs (Cellini, Ferreira, & Rothstein, 2010).

7.3.2 Assumption to Interpret T^{DD} as Local Average Treatment Effect

One more assumption is required for T^{DD} to represent the local causal effect of receiving the scholarship. The final assumption is that the effect of receiving the scholarship does not depend on relative age. Stated differently, there must be no interaction between relative age effects and the effects of the scholarship. There must be additive separability between relative age and receiving the scholarship.

7.3.3 Results

Figure 6 presents difference in discontinuity estimates of the effect of the scholarship. Figure 6A plots the difference in discontinuity estimates on credits for different bandwidths and kernels. For all combinations of bandwidth and kernel the point estimates of receiving the scholarship, due to exceeding the birthdate cutoff and netting out potential relative age effects, is 0.24. Most of the estimates are statistically significant at 90% confidence. This is nearly the exact same estimate for credits as [Scott-Clayton \(2011\)](#).

Figure 6B plots difference in discontinuity estimates on GPA for different bandwidths and kernels. For most combinations of kernels and bandwidths, the estimated effects are much closer to 0.24 than 0.08, as in [Scott-Clayton \(2011\)](#). Most estimates are also statistically significant at 90% confidence. The only kernel and bandwidth that comes close to estimating the 0.08 pp effect that is seen in [Scott-Clayton \(2011\)](#) is the uniform kernel using students that are 160 days away from the cutoff on both sides. This is the evidence that suggests the large difference previously estimated is due to functional form limitations of parametric 2SLS, where the instrument is solely cohort of entry.

7.4 Heterogeneous Results

Finally, there may be heterogeneous effects by subgroups. One subgroup analysis is by academic credentials, because using the ACT cutoff for an RD does not permit estimating effects for those

near the top of the academic ability distribution. The second subgroup analysis is by gender, because there are gender gaps in education.

By Academic Credentials Figure 7 investigates effects for academically eligible students who are above and below the median ACT (24) and HS GPA (3.71) (for academically eligible students). The estimates are noisy, due to small sample sizes; however, some differences in point estimates do arise by above/below median academic scores. First, in Figure 7A, there is a larger point estimate for exceeding credits for those who are above median ACT than those below (or equal to) median ACT. Second, in Figure 7D, there is a larger point estimate for exceeding GPA for those who are below the median high school GPA. These are important results for beginning to untangle the heterogeneity of program effects across those with different academic credentials.

By Gender Finally, Figure 8 shows difference in discontinuity effects by gender. The outcomes are binary for whether or not the students finished the GPA or credits renewal requirement in each of their first 3 years. Figure 8A shows there is no difference between males and females of finishing the credits requirement for 3 years. Figure 8B shows males are more likely to complete the GPA requirement for 3 years, but women are not.⁵² These estimates are important in the context of the known gender gaps in education.

8 Conclusion

Merit-aid programs have grown significantly recently; however, they may be poorly targeted and compete with other programs for scarce state funding. It is difficult to estimate the causal effects of merit-aid, because students have an incentive to sort themselves into treatment, whether it be through retaking academic scores or by attending in-state when they would not have without merit-aid. Prior solutions to estimating causal effects are not applicable to West Virginia, because the academic cutoffs are known and no data has been able to distinguish between first and best scores.

⁵²Figures A.10 and A.11 shows estimates by year by gender.

This paper proposes, formalizes, and validates an alternative empirical strategy that provides a way to estimate causal effects of merit-aid which is compatible with a wider range of datasets. This alternative variation compares students born on slightly different days, 19 years before the scholarship begins, who have vastly different probabilities of receiving the scholarship due to Kindergarten cohort cutoff dates and regular grade progression. Students are not more likely to sort into treatment at this cutoff, unlike academic cutoffs, and are similar along observable pre-treatment characteristics. Furthermore, there are no relative age effects that are detectable at this cutoff.

Receiving a merit-aid scholarship leads to an equally high likelihood, 24 percentage points, of completing both the credits and GPA renewal requirements unlike prior work that is constrained by functional form assumptions ([Scott-Clayton, 2011](#)). The scholarship bolsters the credits completed by those above the median ACT score more than those below it and bolsters the GPA performance of those with lower high school GPAs. Finally, the scholarship leads to men completing the GPA requirements for 3 years, but not women.

There are two primary limitations. The first is that the dataset does not contain all students in West Virginia, so it is possible that differences are driven by sample composition. The second is that there relatively few observations observed at the cutoff, so statistical power is limited, especially in sub-group analyses. Both limitations are addressable with additional data from other universities in West Virginia and in other merit-aid states that have known academic cutoffs.

This research has significant policy implications. Academic renewal thresholds were found to be ineffective in Georgia ([C. M. Cornwell et al., 2005](#)) and effective in West Virginia ([Scott-Clayton, 2011](#)). This research finds that both credits and GPA renewal thresholds are equally effective, when there is a limit on the number of semesters that the aid can be used. Other financial incentives for academic performance can use defined thresholds as a way to increase credits completed and GPA.

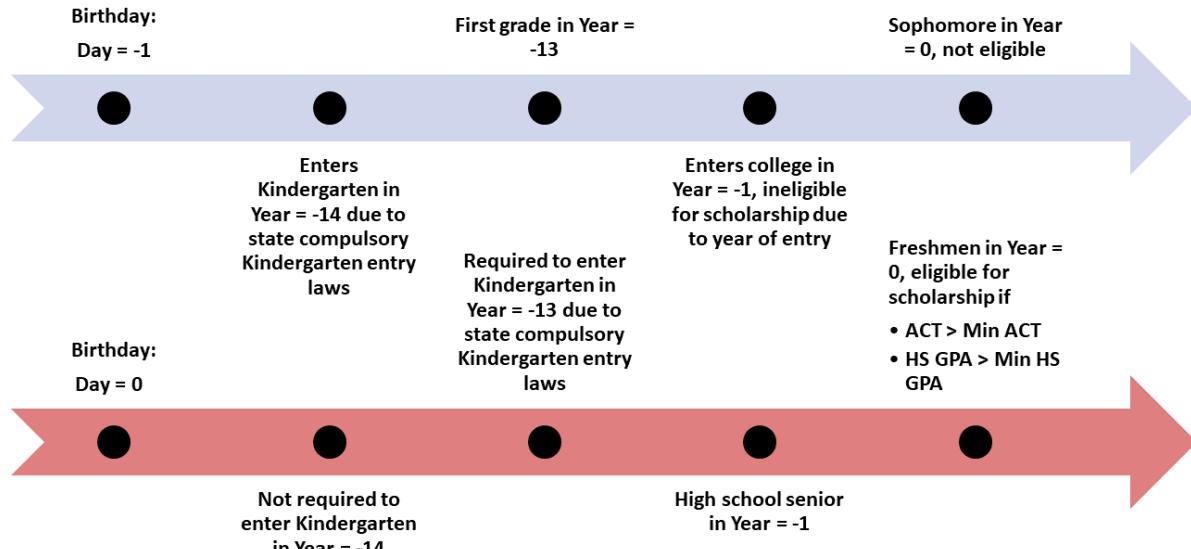
There are a number of concerns for future research. First, as mentioned above, is this similarity across types of academic performance the same for other college students in the same state. Second, since the cutoff adds both a goal and financial incentive, is there a way to separately identify

the behavioral impact of the goal without financial aid? Third, knowing the difference between c and c' could help identify how much attrition from the expected grade path occurs. This would add to the knowledge about who the compliers are. Fourth, is this same research design applicable to other programs and/or other states? If so, this research design could help advance the knowledge of causal impacts of education programs.

Figure 1: Illustration of Scholarship Differentially Assigned by Small Changes in Birth Date, Compulsory Kindergarten Entry Date Cutoff Laws, and Grade Progression

(A): Kindergarten Cutoff that Matters for Scholarship Receipt

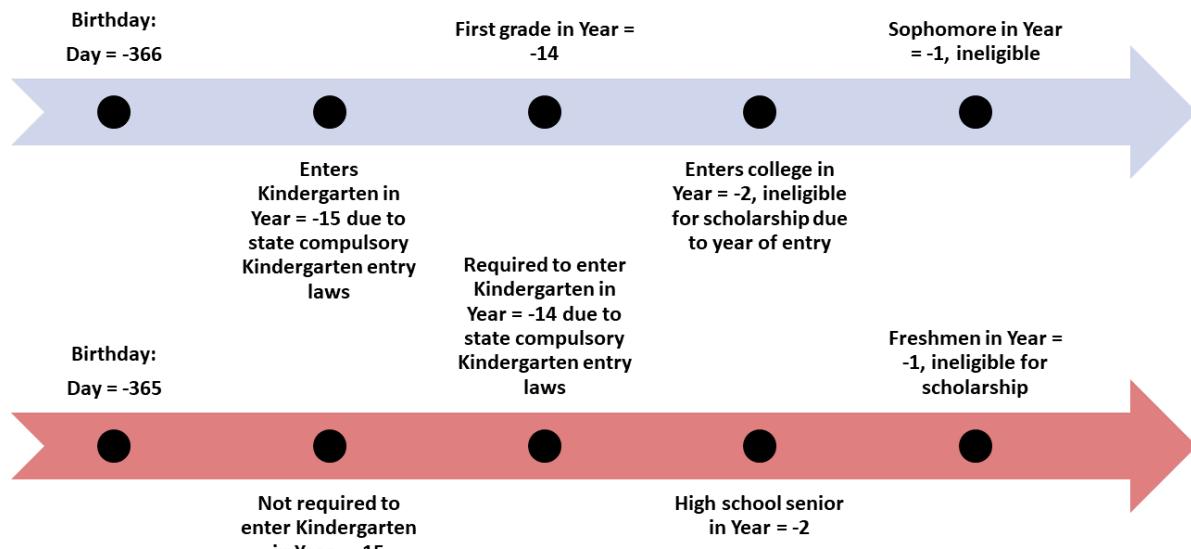
Control Group 1



Treated by Scholarship and Relative Age

(B): Kindergarten Cutoff that Doesn't Matter for Scholarship Receipt

Control Group 2

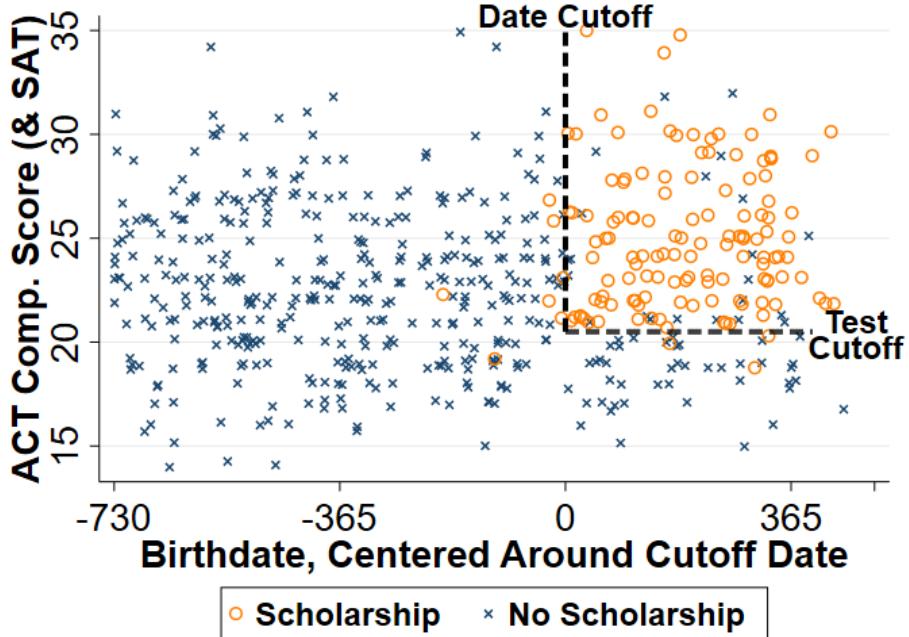


Treated by Relative Age

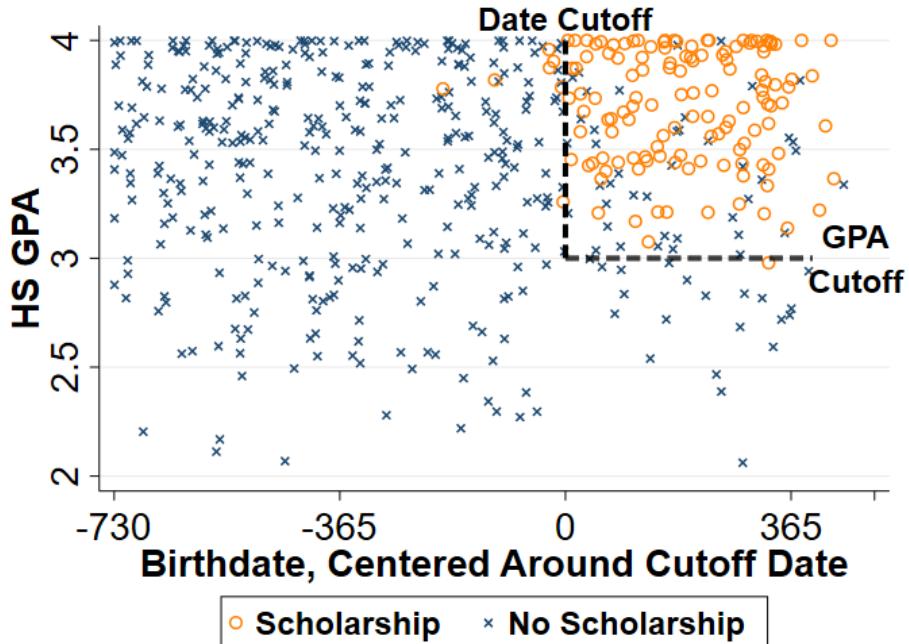
Note: Year 0 stands for initial year of this college scholarship. Day = -1 is a student born right before the birthdate cutoff to enter kindergarten in the first cohort that is eligible for the college scholarship if they finish 1 grade per year. Day = 0 is a student born right after the cutoff which allows them to enter in the kindergarten cohort that will enter college in the first college eligible cohort, finishing 1 grade per year.

Figure 2: Eligibility Measures and Scholarship Receipt

(A): Birthdate, Entry Test Score, and Scholarship



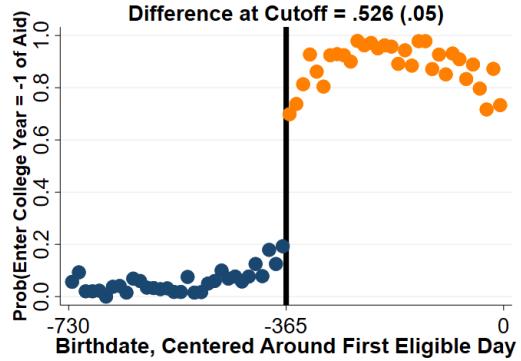
(B): Birthdate, HS GPA, and Scholarship



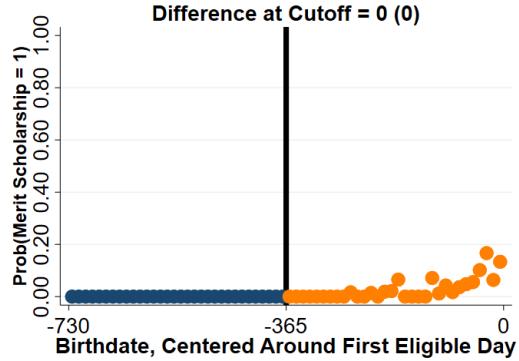
Note: 10% random sample. N = 559. Birthdates centered at 0, the first eligible date for the scholarship. The cutoff ACT score is 20.5. SAT score converted to ACT using the conversion chart at [Edwards \(2020\)](#). The cutoff HS GPA is 3.0.

Figure 3: Kindergarten Entry Date Cutoffs on Cohort Entry Year and Scholarship Receipt

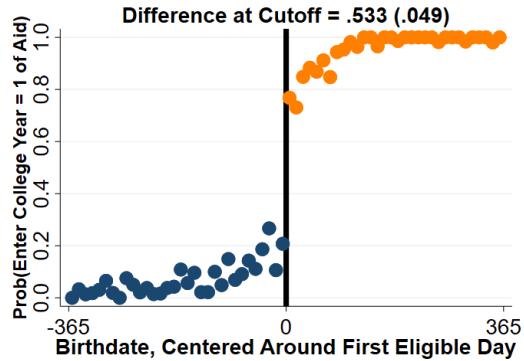
(A): Pre-Scholarship Cutoff on Enter in 2001



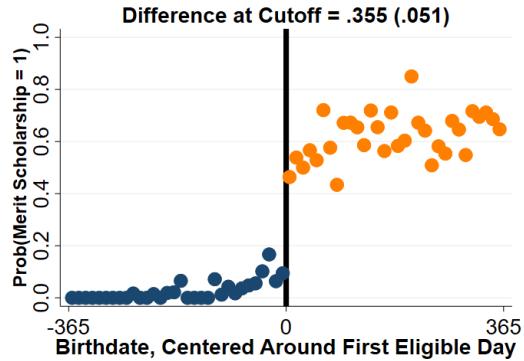
(B): Pre-Scholarship Cutoff on Scholarship Receipt



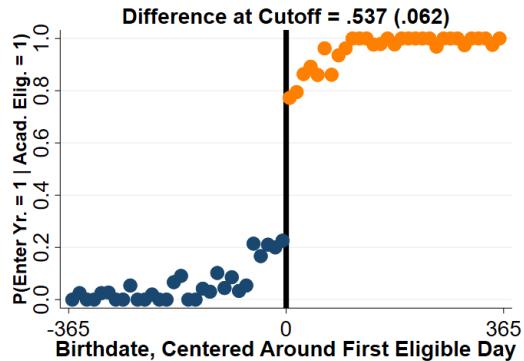
(C): Post-Scholarship Cutoff on Enter in 2002



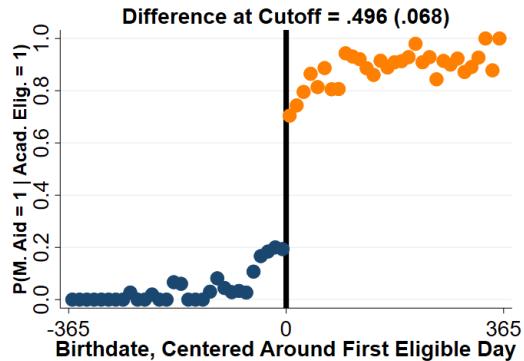
(D): Post-Scholarship Cutoff on Scholarship Receipt



(E): Post-Scholarship Cutoff on Enter in 2002, Academically Eligible Students



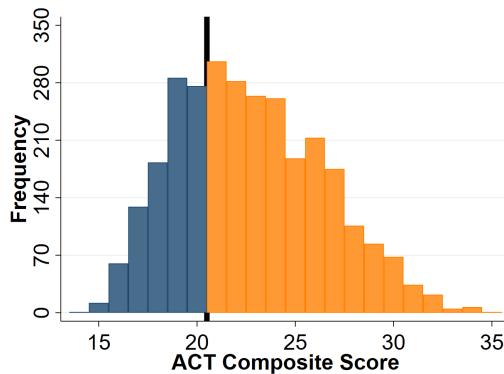
(F): Post-Scholarship Cutoff on Scholarship Received, Academically Eligible Students



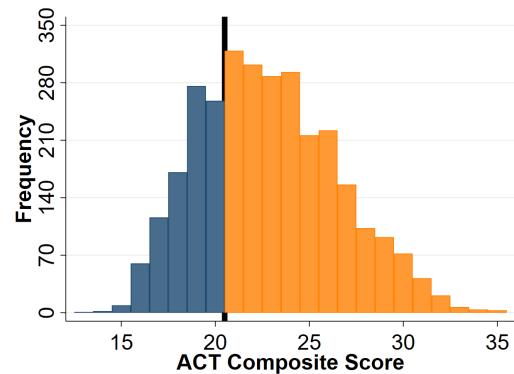
Note: Plots show the conditional mean within birthdate bins. Bin-length = 11.375 days. The black vertical line indicates cutoff birthdates. The birthdate is centered at the post-scholarship cutoff birthdate, meaning 0 is the first birthdate of increased likelihood of scholarship treatment. In Panels A, C, and E, the y-variable is 1/0 for being a first time freshman in the first year of the scholarship. In Panels B, D, and F, the y-variable is mean of receiving at least 1 year of financial merit-based aid. In Panels A and B, n = 3,606. In Panels C and D, n = 3,669. In Panels E and F, n = 2,341. The difference at the cutoff is estimated using local linear regression with a triangular kernel and a two-sided mean-square error (MSE) optimal bandwidth.

Figure 4: Comparing Academic and Birthdate Running Variables' Continuity Around Cutoffs

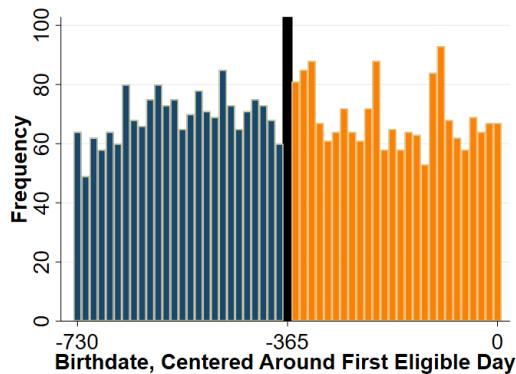
(A): ACT Scores, Two Cohorts Pre-Scholarship



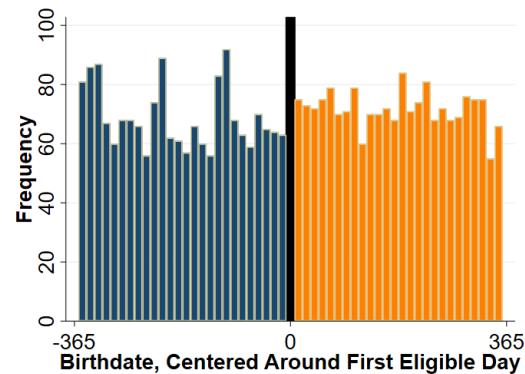
(B): ACT Scores, First Cohort of Scholarship



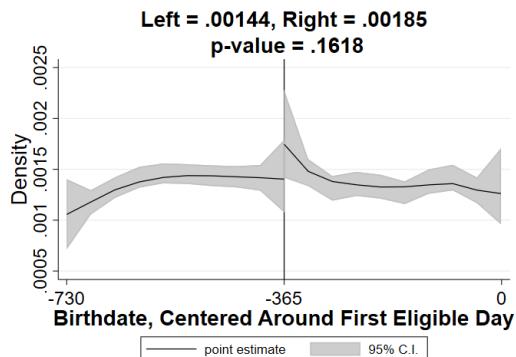
(C): Birthdate Density Around Cutoff, Before Scholarship



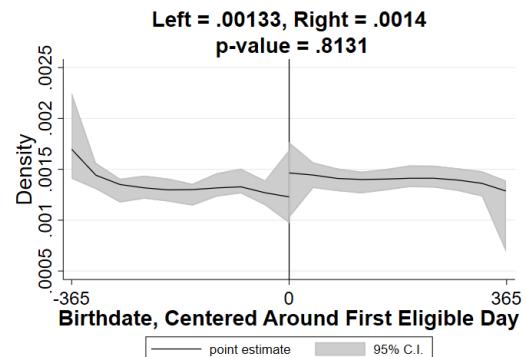
(D): Birthdate Density Around Cutoff, For Scholarship



(E): McCrary Density Test, Pre-Scholarship



(F): McCrary Density Test, First Scholarship Cohort



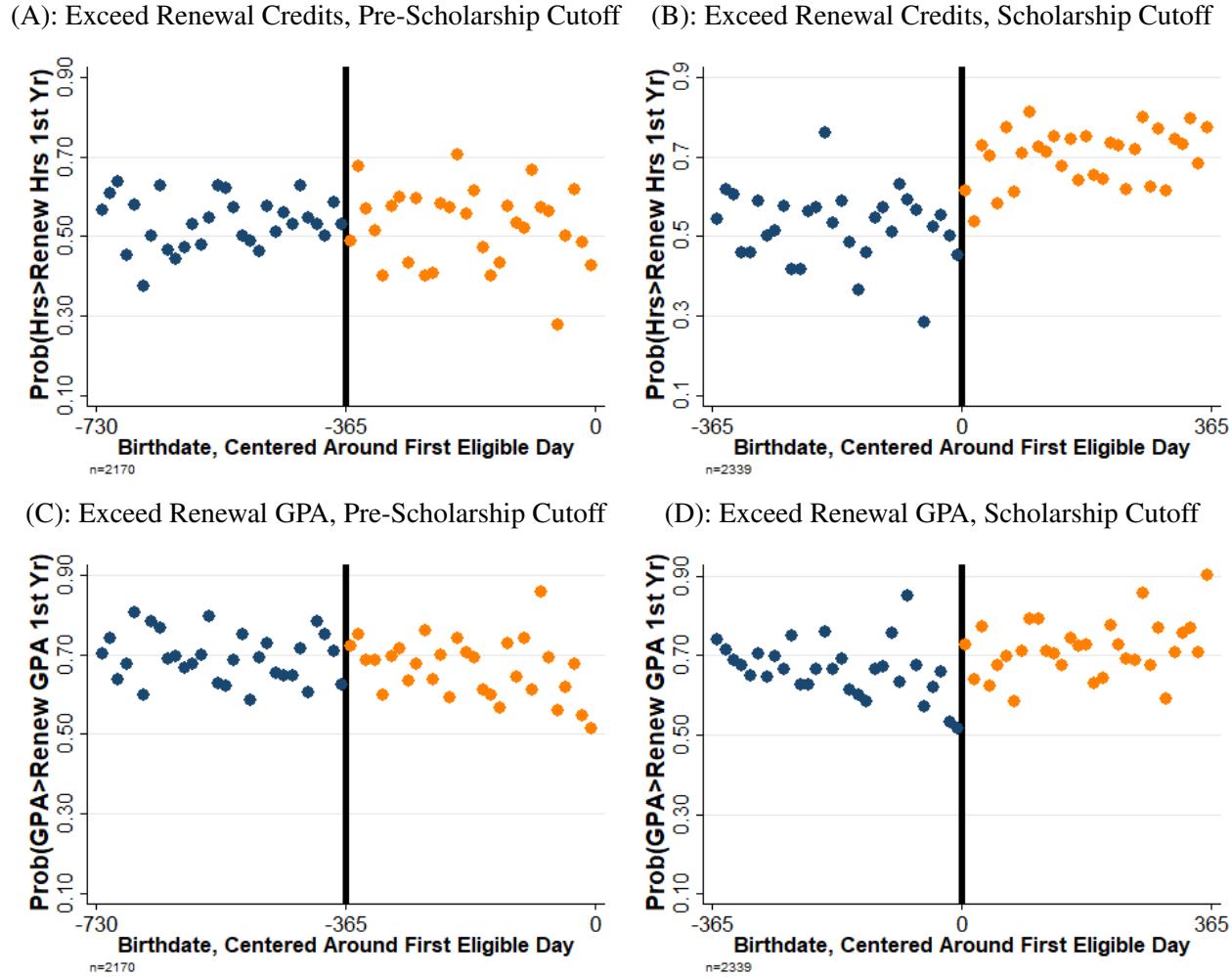
Note: Panels A and B show the frequency of each ACT Composite score realization. Panel A uses the 2000 and 2001 cohorts. Panel B uses the 2000 cohort. Birthdate 0 is the cutoff for the post-scholarship cutoff and birthdate -365 is the pre-scholarship cutoff. Panel C ($n= 3,600$) shows the distribution of those born within 1 year of the pre-scholarship cutoff date. Panel D ($n= 3,666$) shows the distribution of birthdate for those born within a year of the post-scholarship cutoff date. Each bar in Panels C and D represent 2 weeks of birthdates. Panels E and F are density tests at the pre and post-scholarship birthdate cutoffs. The estimates are calculated according to [Cattaneo, Jansson, and Ma \(2019\)](#). Left is the point estimate to the right of the cutoff and right is point estimate to the right, both are bias-corrected. The p-value is from a test for whether density to the left equals the density to the right and is robust. Failure to reject indicates that the density is smooth across the cutoff, consistent with identifying assumptions ([McCrary, 2008](#)).

Table 1: Pre-Treatment Covariate Balance

Panel A						
	Dep. Var.: Male			Dep. Var.: HS GPA		
	(1) Post	(2) Pre	(3) Diff	(4) Post	(5) Pre	(6) Diff
After Cutoff	0.03 (0.06)	-0.14** (0.06)	0.17** (0.08)	0.13** (0.05)	0.07 (0.05)	0.06 (0.07)
Bandwidth	149	154	150	189	166	160
Observations	1489	1569	3038	1861	1669	3231
Panel B						
	Dep. Var.: ACT Composite			Dep. Var.: SAT Combined		
	(1) Post	(2) Pre	(3) Diff	(4) Post	(5) Pre	(6) Diff
After Cutoff	-0.05 (0.44)	-0.44 (0.44)	0.47 (0.62)	40.59 (33.68)	83.31** (37.06)	-42.76 (53.54)
Bandwidth	167	156	160	190	180	160
Observations	1362	1299	2661	337	319	591
Panel C						
	Dep. Var.: ACT English			Dep. Var.: ACT Math		
	(1) Post	(2) Pre	(3) Diff	(4) Post	(5) Pre	(6) Diff
After Cutoff	0.01 (0.55)	0.05 (0.53)	-0.02 (0.73)	0.11 (0.48)	-0.73 (0.52)	0.90 (0.77)
Bandwidth	146	146	160	198	173	160
Observations	1199	1199	2661	1616	1427	2661
Panel D						
	Dep. Var.: ACT Reading			Dep. Var.: ACT Science		
	(1) Post	(2) Pre	(3) Diff	(4) Post	(5) Pre	(6) Diff
After Cutoff	-0.31 (0.60)	-0.35 (0.58)	0.09 (0.84)	0.00 (0.41)	-0.84* (0.49)	0.91 (0.63)
Bandwidth	179	153	160	207	137	160
Observations	1446	1282	2661	1683	1132	2661

Note: * $p<0.1$, ** $p<0.05$, *** $p<0.01$. Coefficients are from local linear regressions with triangular kernels and measure the discontinuity at the cutoff. The coefficients are not bias-corrected and the SE's are not robust. Mean-square error optimal bandwidths are used. The post-scholarship cutoff ($d=0$) cutoff is the cutoff that determines scholarship eligibility. The pre-scholarship ($d= -365$) cutoff is the cutoff on the same day in the prior year and does not determine scholarship eligibility. Difference may be slightly different than the difference of the two coefficients, due to different bandwidth lengths. Bandwidth is two-sided, meaning that many observations on both sides of the cutoff.

Figure 5: Effect of Being Born After Birthdate Cutoffs on Likelihood of Exceeding Renewal Thresholds in Freshman Year, Academically Eligible Students Only



Note: Plots show the conditional mean within birthdate bins. Bin-length = 11.375 days. The black vertical line represents cutoff birthdates. The birthdate is centered at the first birthdate where increased scholarship receipt is expected, the post-scholarship cutoff is day = 0. Panel A shows the effect of being born later than the pre-scholarship cutoff on the likelihood of exceeding the freshman credits renewal threshold (≥ 30) in one's freshman year. Panel B shows the effect of being born later than the post-scholarship cutoff on the likelihood of exceeding the credits renewal threshold in one's freshman year. Panel C shows the effect of being born later than the pre-scholarship cutoff on exceeding the freshman year GPA renewal threshold (2.75). Panel D shows the effect of being born later than the post-scholarship cutoff on exceeding the freshman year GPA renewal threshold.

Table 2: Intent to Treat Effect on Meeting Academic Renewal Threshold By Academic Eligibility

Panel A: Credits, Conventional						
	Academically Eligible			Academically Ineligible		
	(1) Post	(2) Pre	(3) Diff	(4) Post	(5) Pre	(6) Diff
After Cutoff	0.135*	0.011	0.129	0.055	-0.070	0.104
	(0.070)	(0.070)	(0.097)	(0.076)	(0.076)	(0.102)
Bandwidth	161	153	160	168	128	160
Observations	1021	926	1995	599	512	1236
Panel B: Credits, Bias-Corrected and Robust SE's						
	Academically Eligible			Academically Ineligible		
	(1) Post	(2) Pre	(3) Diff	(4) Post	(5) Pre	(6) Diff
After Cutoff	0.122	0.019	0.114	0.051	-0.081	0.087
	(0.084)	(0.084)	(0.148)	(0.091)	(0.089)	(0.143)
Bandwidth	161	153	160	168	128	160
Observations	1021	926	1995	599	512	1236
Panel C: GPA, Conventional						
	Academically Eligible			Academically Ineligible		
	(1) Post	(2) Pre	(3) Diff	(4) Post	(5) Pre	(6) Diff
After Cutoff	0.100*	0.006	0.127	0.049	0.036	0.001
	(0.057)	(0.063)	(0.092)	(0.085)	(0.084)	(0.120)
Bandwidth	231	151	160	180	141	160
Observations	1490	909	1995	629	564	1236
Panel D: GPA, Bias-Corrected and Robust SE's						
	Academically Eligible			Academically Ineligible		
	(1) Post	(2) Pre	(3) Diff	(4) Post	(5) Pre	(6) Diff
After Cutoff	0.116*	0.006	0.196	0.038	0.064	-0.048
	(0.067)	(0.075)	(0.133)	(0.102)	(0.096)	(0.170)
Bandwidth	231	151	160	180	141	160
Observations	1490	909	1995	629	564	1236

Note: * p<0.1, ** p<0.05, *** p<0.01. Coefficients are from local linear regressions with triangular kernels and measure the discontinuity at the cutoff. Mean-square error optimal bandwidths are used. The post-scholarship cutoff ($d=0$) cutoff is the cutoff that determines scholarship eligibility. The pre-scholarship ($d= -365$) cutoff is the cutoff on the same day in the prior year and does not determine scholarship eligibility. Difference may be slightly different than the difference of the two coefficients, due to different bandwidth lengths. Bandwidth is two-sided, meaning that many observations on both sides of the cutoff.

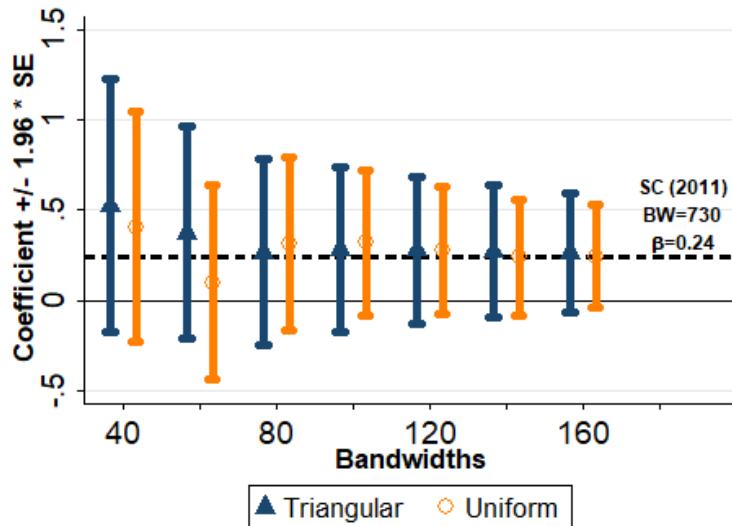
Table 3: Local Linear Estimates of Scholarship on Meeting First Year Credits and GPA Renewal Thresholds

Panel A: Dep. Var.: Renew Credits							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Scholarship	0.36** (0.15)	0.35** (0.18)	0.27* (0.16)	0.24 (0.16)	0.24 (0.19)	0.23* (0.13)	0.31* (0.16)
Corrections	-	B-C, R	B-C, R	B-C, R	B-C, R	B-C, R	B-C, R
Sample	All	All	Ac. Elg.	Ac. Elg.	Ac. Elg.	Ac. Elg.	Ac. Elg.
Bandwidth	141	141	135	134	91	133	90
Eff. Obs. Left	680	680	395	395	248	393	248
Eff. Obs. Right	740	740	471	471	319	462	319
Obs-Left	1970	1970	1081	1081	1081	1081	1081
Obs-Right	1979	1979	1356	1356	1356	1356	1356
Acad. Controls			X	X	X	X	X
BW Criterion	MSE	MSE	MSE	MSE	CER	MSE	CER
Kernel	Tri	Tri	Tri	Tri	Tri	Uni	Uni
Panel B: Dep. Var.: Renew GPA							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Scholarship	0.31** (0.15)	0.34** (0.18)	0.34** (0.16)	0.30** (0.15)	0.40** (0.19)	0.20 (0.13)	0.32** (0.16)
Corrections	-	B-C, R	B-C, R	B-C, R	B-C, R	B-C, R	B-C, R
Sample	All	All	Ac. Elg.	Ac. Elg.	Ac. Elg.	Ac. Elg.	Ac. Elg.
Bandwidth	151	151	140	142	96	138	93
Eff. Obs. Left	720	720	404	408	268	398	261
Eff. Obs. Right	777	777	491	500	336	482	327
Obs-Left	1970	1970	1081	1081	1081	1081	1081
Obs-Right	1979	1979	1356	1356	1356	1356	1356
Acad. Controls			X	X	X	X	X
BW Criterion	MSE	MSE	MSE	MSE	CER	MSE	CER
Kernel	Tri	Tri	Tri	Tri	Tri	Uni	Uni

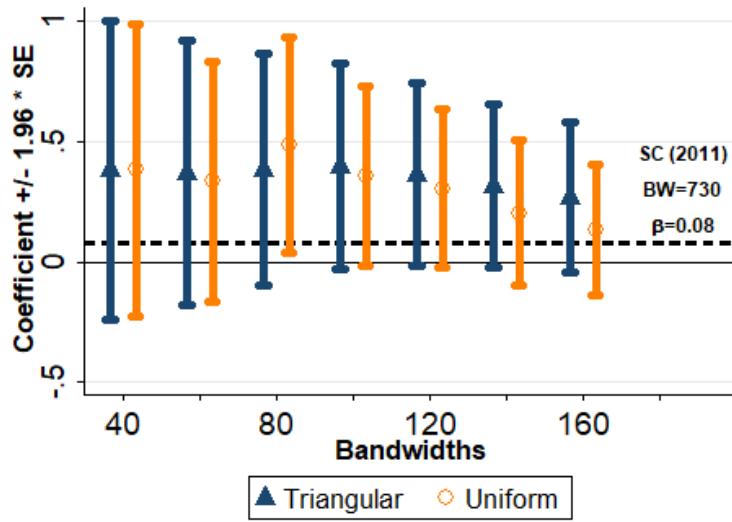
Note: * p<0.1, ** p<0.05, *** p<0.01. Estimates displayed are the second stage Wald estimator from a fuzzy regression discontinuity design using birthdate in relation to Kindergarten entry date as the cutoff/instrument. The estimate is the effect of receiving the Promise scholarship. In Panel A, the dependent variable is a dummy variable for complete 30 credits (renewal amount) in first year of college. In Panel B, the dependent variable is a dummy variable for receive at least a 2.75 GPA (renewal GPA) in first year of college. In all each column and panel, the first stage cutoff has a strong effect on scholarship receipt, consistent with 3D. In column 1, conventional estimates are reported. In columns 2 and beyond the bias-correction, robust estimates are reported (Calonico et al., 2017). In columns 1 and 2, the full sample is used. Columns 3 and beyond use only academically eligible students. From column 4 onwards, high school GPA and entry test (ACT/SAT) score are included. In columns 1-4 and 6, the bandwidth selection procedure optimizes mean square error (MSE). In columns 5 and 7, the bandwidth selector optimizes coverage error rate. In columns 1-5 a triangular kernel is used. In columns 6 and 7 a uniform kernel is used.

Figure 6: Coefficient Estimates from Difference in Discontinuity on Freshman Outcomes

(A): Probability of Completing Renewal Credits



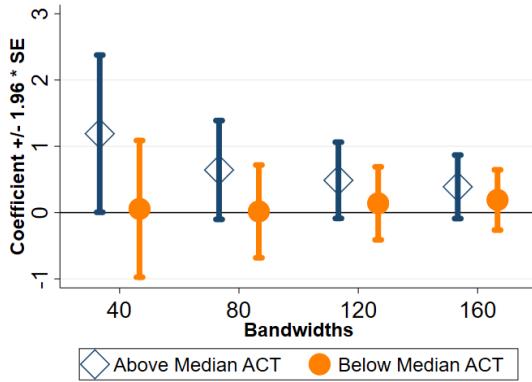
(B): Probability of Completing Renewal GPA



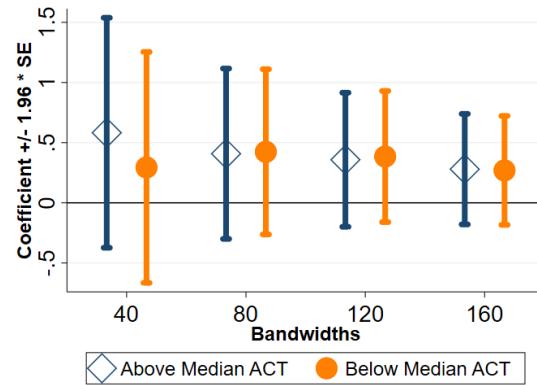
Note: Coefficients represent “conventional”, not bias-corrected or robust for MSE-optimal, treatment estimates from difference-in-discontinuity estimates (Grembi et al., 2016). Diff-in-discontinuity is the RD estimate from the kindergarten entry date cutoff before it determined the scholarship receipt, but did determine relative age subtracted from the RD estimate in where it did determine scholarship receipt. This approach differences out the relative age effect under fairly mild assumptions. The different bandwidths indicate the number of days used from the respective cutoffs. Triangular and uniforms kernels used. Panel B legend valid for both. All regressions subsample to academically eligible students meaning $[(ACT \geq 21 \text{ OR } SAT \geq 1000) \text{ AND } HS\ GPA \geq 3.0]$ and include controls for HS GPA and ACT composite score (w/ SAT converted to ACT). Panel A dependent variable is exceed renewal credits in freshman year. Panel B dependent variable is renewal GPA in freshman year. The horizontal dashed lines represent the Scott-Clayton (2011) point estimates which correspond approximately to 730 day bandwidths, since Scott-Clayton (2011) uses 2 cohorts before and after.

Figure 7: Heterogeneity by Academic Credentials

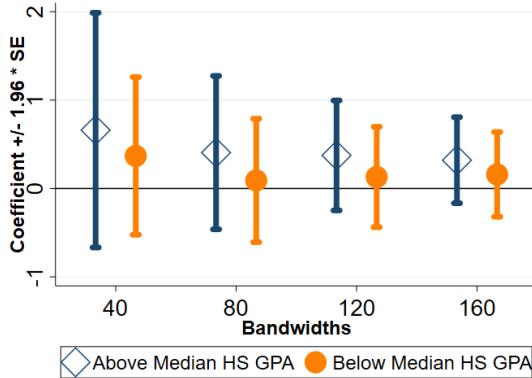
(A): Credits by ACT



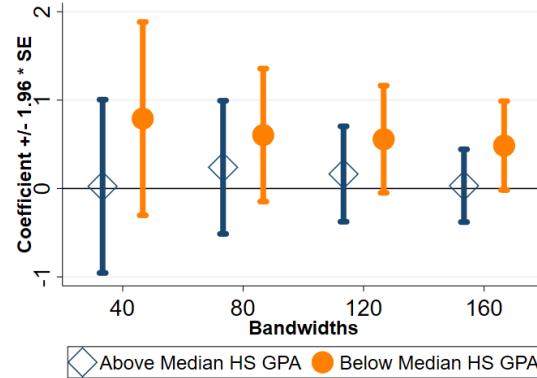
(B): GPA by ACT



(C): Credits by HS GPA

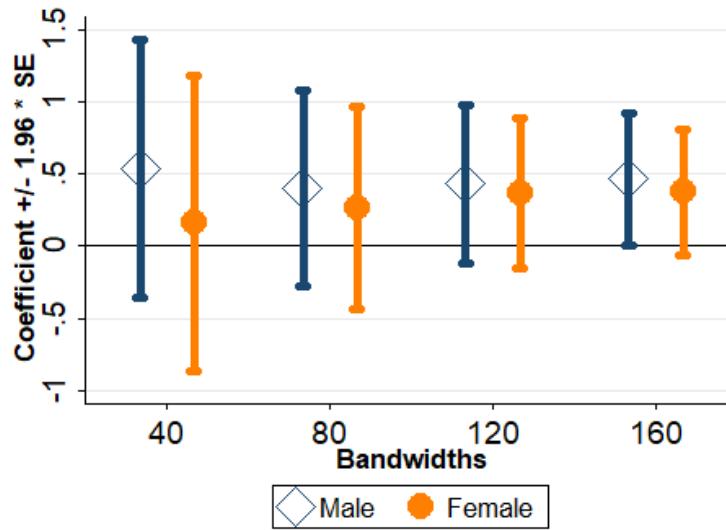


(D): GPA by HS GPA

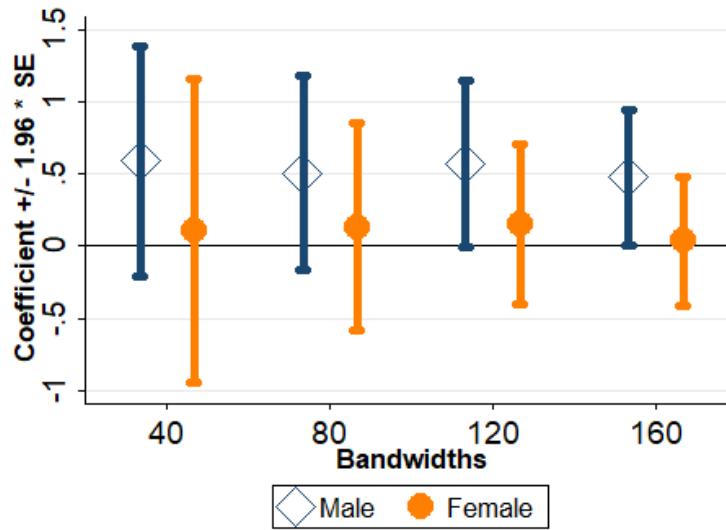


Note: Difference in discontinuity estimates at different bandwidths plotted with 95% confidence intervals. The estimates are conventional. The local linear regression uses a triangular kernel. In Panels A and C, the dependent variable is exceeding the freshman credits renewal threshold. In Panels B and D, the dependent variable is exceeding the freshman GPA renewal threshold. The median HS GPA in this sample, for those who are academically eligible, is 3.71. The median ACT score in this sample is 24. The below group contains the median in all figures.

Figure 8: Coefficient Estimates from Difference in Discontinuity on 3 Year Outcomes, by Gender
 (A): Probability of Completing Renewal Credits



(B): Probability of Completing Renewal GPA



Note: Coefficients represent “conventional”, not bias-corrected or robust for MSE-optimal, treatment estimates from difference-in-discontinuity estimates (Grembi et al., 2016). Diff-in-discontinuity is the RD estimate from the kindergarten entry date cutoff before it determined the scholarship receipt, but did determine relative age subtracted from the RD estimate in where it did determine scholarship receipt. This approach differences out the relative age effect under fairly mild assumptions. The different bandwidths indicate the number of days used from the respective cutoffs. Triangular and uniforms kernels used. Panel B legend valid for both. All regressions subsample to academically eligible students meaning $[(ACT \geq 20.5 \text{ OR } SAT \geq 1000) \text{ AND } HS\ GPA \geq 3.0]$ and include controls for HS GPA and ACT composite score (w/ SAT converted to ACT). Panel A dependent variable is exceed renewal credits in freshman year. Panel B dependent variable is renewal GPA in freshman year. The horizontal dashed lines represent the Scott-Clayton (2011) point estimates which correspond approximately to 730 day bandwidths, since Scott-Clayton (2011) uses 2 cohorts before and after.

References

- Abbott, B., Gallipoli, G., Meghir, C., & Violante, G. L. (2019). Education policy and intergenerational transfers in equilibrium. *Journal of Political Economy*, 127(6), 2569–2624.
- Acemoglu, D., Johnson, S., & Robinson, J. A. (2001). The colonial origins of comparative development: An empirical investigation. *American Economic Review*, 91(5), 1369–1401.
- Angrist, J., Lang, D., & Oreopoulos, P. (2009). Incentives and services for college achievement: Evidence from a randomized trial. *American Economic Journal: Applied Economics*, 1(1), 136–63.
- Angrist, J. D., Imbens, G. W., & Rubin, D. B. (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association*, 91(434), 444–455.
- Angrist, J. D., & Pischke, J.-S. (2010). The credibility revolution in empirical economics: How better research design is taking the con out of econometrics. *Journal of Economic Perspectives*, 24(2), 3–30.
- Angrist, J. D., & Rokkanen, M. (2015). Wanna get away? Regression discontinuity estimation of exam school effects away from the cutoff. *Journal of the American Statistical Association*, 110(512), 1331–1344.
- Beattie, G., Laliberté, J.-W. P., Michaud-Leclerc, C., & Oreopoulos, P. (2019). What sets college thrivers and divers apart? A contrast in study habits, attitudes, and mental health. *Economics Letters*, 178, 50–53.
- Bertanha, M., & Imbens, G. W. (2019). External validity in fuzzy regression discontinuity designs. *Journal of Business & Economic Statistics*, 1–39.
- Black, S. E., Devereux, P. J., & Salvanes, K. G. (2011). Too young to leave the nest? The effects of school starting age. *The Review of Economics and Statistics*, 93(2), 455–467.
- Bloom, H. S. (2012). Modern regression discontinuity analysis. *Journal of Research on Educational Effectiveness*, 5(1), 43–82.
- Bruce, D. J., & Carruthers, C. K. (2014). Jackpot? The impact of lottery scholarships on enrollment in Tennessee. *Journal of Urban Economics*, 81, 30–44.

- Buckles, K. S., & Hungerman, D. M. (2013). Season of birth and later outcomes: Old questions, new answers. *Review of Economics and Statistics*, 95(3), 711–724.
- Calonico, S., Cattaneo, M. D., Farrell, M. H., & Titiunik, R. (2017). rdrobust: Software for regression-discontinuity designs. *The Stata Journal*, 17(2), 372–404.
- Calonico, S., Cattaneo, M. D., Farrell, M. H., & Titiunik, R. (2019). Regression discontinuity designs using covariates. *Review of Economics and Statistics*, 101(3), 442–451.
- Calonico, S., Cattaneo, M. D., & Titiunik, R. (2014). Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica*, 82(6), 2295–2326.
- Canaan, S. (2020). The long-run effects of reducing early school tracking. *Journal of Public Economics*, 187, 104206.
- Card, D., & Krueger, A. B. (1992). Does school quality matter? Returns to education and the characteristics of public schools in the United States. *Journal of Political Economy*, 100(1), 1–40.
- Carnevale, A. P., Cheah, B., & Hanson, A. R. (2015). *The economic value of college majors* (Tech. Rep.). Center on Education and the Workforce.
- Carruthers, C. K., & Özak, U. (2016). Losing HOPE: Financial aid and the line between college and work. *Economics of Education Review*, 53, 1–15.
- Cattaneo, M. D., Frandsen, B. R., & Titiunik, R. (2015). Randomization inference in the regression discontinuity design: An application to party advantages in the us senate. *Journal of Causal Inference*, 3(1), 1–24.
- Cattaneo, M. D., Idrobo, N., & Titiunik, R. (2019). *A practical introduction to regression discontinuity designs: Foundations*. Cambridge University Press.
- Cattaneo, M. D., Jansson, M., & Ma, X. (2019). Simple local polynomial density estimators. *Journal of the American Statistical Association*, 1–7.
- Cellini, S. R., Ferreira, F., & Rothstein, J. (2010). The value of school facility investments: Evidence from a dynamic regression discontinuity design. *The Quarterly Journal of Economics*, 125(1), 215–261.

- Clark, D., Gill, D., Prowse, V., & Rush, M. (2020). Using goals to motivate college students: Theory and evidence from field experiments. *Review of Economics and Statistics*, 102(4), 648–663.
- Clark, D., & Royer, H. (2013). The effect of education on adult mortality and health: Evidence from Britain. *American Economic Review*, 103(6), 2087–2120.
- Cohodes, S. R., & Goodman, J. S. (2014). Merit aid, college quality, and college completion: Massachusetts' adams scholarship as an in-kind subsidy. *American Economic Journal: Applied Economics*, 6(4), 251–85.
- Cohodes, S. R., Grossman, D. S., Kleiner, S. A., & Lovenheim, M. F. (2016). The effect of child health insurance access on schooling: Evidence from public insurance expansions. *Journal of Human Resources*, 51(3), 727–759.
- College Board. (2019). *Average aid per student over time*. <https://research.collegeboard.org/trends/student-aid/figures-tables/average-aid-student-over-time>.
- Cook, T. D. (2008). “Waiting for life to arrive”: A history of the regression-discontinuity design in psychology, statistics and economics. *Journal of Econometrics*, 142(2), 636–654.
- Cornwell, C., & Mustard, D. B. (2007). Merit-based college scholarships and car sales. *Education Finance and Policy*, 2(2), 133–151.
- Cornwell, C., Mustard, D. B., & Sridhar, D. J. (2006). The enrollment effects of merit-based financial aid: Evidence from georgia's hope program. *Journal of Labor Economics*, 24(4), 761–786.
- Cornwell, C. M., Lee, K. H., & Mustard, D. B. (2005). Student responses to merit scholarship retention rules. *Journal of Human Resources*, 40(4), 895–917.
- Cowan, B. W. (2016). Testing for educational credit constraints using heterogeneity in individual time preferences. *Journal of Labor Economics*, 34(2), 363–402.
- Cowan, B. W., & White, D. R. (2015). The effects of merit-based financial aid on drinking in college. *Journal of Health Economics*, 44, 137–149.

- Dell, M., & Querubin, P. (2018). Nation building through foreign intervention: Evidence from discontinuities in military strategies. *The Quarterly Journal of Economics*, 133(2), 701–764.
- Deming, D., & Dynarski, S. (2008). The lengthening of childhood. *Journal of Economic Perspectives*, 22(3), 71–92.
- Denning, J. T. (2019). Born under a lucky star? Financial aid, college completion, labor supply, and credit constraints. *Journal of Human Resources*, 54(3), 760–784.
- Dong, Y. (2018). Alternative assumptions to identify LATE in fuzzy regression discontinuity designs. *Oxford Bulletin of Economics and Statistics*, 80(5), 1020–1027.
- Dong, Y., & Lewbel, A. (2015). Identifying the effect of changing the policy threshold in regression discontinuity models. *Review of Economics and Statistics*, 97(5), 1081–1092.
- Duranton, G., & Turner, M. A. (2011). The fundamental law of road congestion: Evidence from us cities. *American Economic Review*, 101(6), 2616–52.
- Dynarski, S. (2004). The new merit aid. In *College choices: The economics of where to go, when to go, and how to pay for it* (pp. 63–100). University of Chicago Press.
- Dynarski, S., & Scott-Clayton, J. (2013). *Financial aid policy: Lessons from research* (Tech. Rep.). National Bureau of Economic Research.
- Edwards, H. (2020). *Official act to sat (new 1600 and old 2400) conversion charts*. <https://blog.prepscholar.com/act-to-sat-conversion>.
- Frandsen, B. R. (2017). Party bias in union representation elections: Testing for manipulation in the regression discontinuity design when the running variable is discrete. In *Regression discontinuity designs*. Emerald Publishing Limited.
- Gneezy, U., Meier, S., & Rey-Biel, P. (2011). When and why incentives (don't) work to modify behavior. *Journal of Economic Perspectives*, 25(4), 191–210.
- Goodman, J. (2008). Who merits financial aid? *Journal of Public Economics*, 92(10-11), 2121–2131.
- Grembi, V., Nannicini, T., & Troiano, U. (2016). Do fiscal rules matter? *American Economic Journal: Applied Economics*, 1–30.

- Hahn, J., Todd, P., & Van der Klaauw, W. (2001). Identification and estimation of treatment effects with a regression-discontinuity design. *Econometrica*, 69(1), 201–209.
- Heller, D. E., & Marin, P. (2004). State merit scholarship programs and racial inequality. *Civil Rights Project at Harvard University (The)*.
- Hendren, N., & Sprung-Keyser, B. (2020). A unified welfare analysis of government policies. *The Quarterly Journal of Economics*, 135(3), 1209–1318.
- Holland, P. W. (1986). Statistics and causal inference. *Journal of the American Statistical Association*, 81(396), 945–960.
- Hsu, Y.-C., & Shen, S. (2019). Testing treatment effect heterogeneity in regression discontinuity designs. *Journal of Econometrics*, 208(2), 468–486.
- Imbens, G., & Kalyanaraman, K. (2012). Optimal bandwidth choice for the regression discontinuity estimator. *The Review of Economic Studies*, 79(3), 933–959.
- Jia, N. (2019). Heterogeneous effects of merit scholarships: Do program features matter? *Applied Economics*, 51(27), 2963–2979.
- Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2), 263–291.
- Kolesár, M., & Rothe, C. (2018). Inference in regression discontinuity designs with a discrete running variable. *American Economic Review*, 108(8), 2277–2304.
- Lavecchia, A. M., Liu, H., & Oreopoulos, P. (2016). Behavioral economics of education: Progress and possibilities. In *Handbook of the economics of education* (Vol. 5, pp. 1–74). Elsevier.
- Lee, D. S. (2008). Randomized experiments from non-random selection in US House elections. *Journal of Econometrics*, 142(2), 675–697.
- Lee, D. S., & Card, D. (2008). Regression discontinuity inference with specification error. *Journal of Econometrics*, 142(2), 655–674.
- Lemieux, T., & Milligan, K. (2008). Incentive effects of social assistance: A regression discontinuity approach. *Journal of Econometrics*, 142(2), 807–828.
- Levitt, S. D., List, J. A., Neckermann, S., & Sadoff, S. (2016). The behavioralist goes to school:

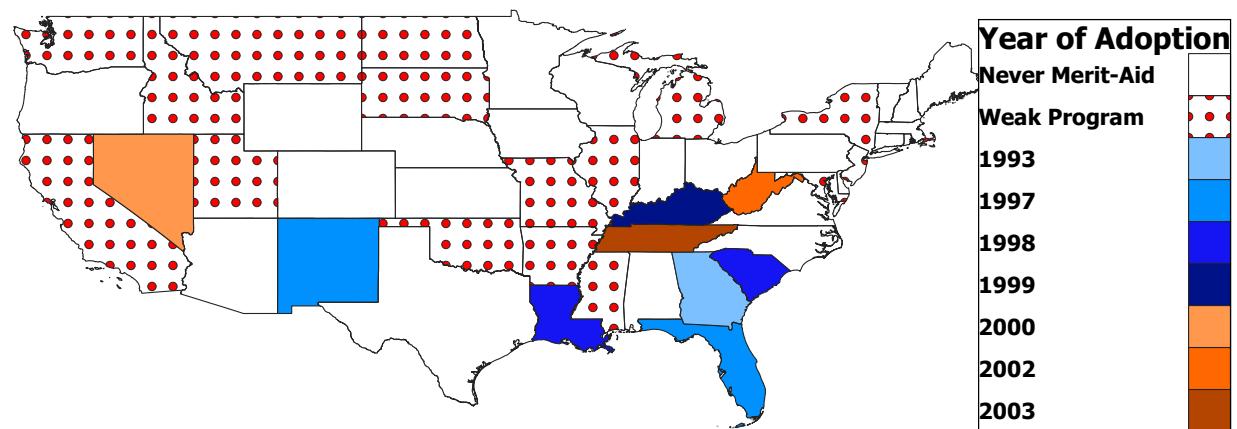
- Leveraging behavioral economics to improve educational performance. *American Economic Journal: Economic Policy*, 8(4), 183–219.
- Lleras-Muney, A. (2002). Were compulsory attendance and child labor laws effective? An analysis from 1915 to 1939. *The Journal of Law and Economics*, 45(2), 401–435.
- Lleras-Muney, A., & Shertzer, A. (2015). Did the americanization movement succeed? an evaluation of the effect of english-only and compulsory schooling laws on immigrants. *American Economic Journal: Economic Policy*, 7(3), 258–90.
- Matta, R., Ribas, R. P., Sampaio, B., & Sampaio, G. R. (2016). The effect of age at school entry on college admission and earnings: A regression-discontinuity approach. *IZA Journal of Labor Economics*, 5(1), 9.
- McCrory, J. (2008). Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of Econometrics*, 142(2), 698–714.
- Miller, S., & Wherry, L. R. (2019). The long-term effects of early life medicaid coverage. *Journal of Human Resources*, 54(3), 785–824.
- Ortagus, J., Kelchen, R., Rosigner, K., & Voorhees, N. (2018). Performance-based funding in American higher education: A systematic synthesis of the intended and unintended consequences. *Educational Evaluation and Policy Analysis*, 53(2), 299–329.
- Peña, P. A. (2017). Creating winners and losers: Date of birth, relative age in school, and outcomes in childhood and adulthood. *Economics of Education Review*, 56, 152–176.
- Pettersson-Lidbom, P. (2012). Does the size of the legislature affect the size of government? evidence from two natural experiments. *Journal of Public Economics*, 96(3-4), 269–278.
- Scott-Clayton, J. (2011). On money and motivation a quasi-experimental analysis of financial incentives for college achievement. *Journal of Human Resources*, 46(3), 614–646.
- Scott-Clayton, J., & Zafar, B. (2019). Financial aid, debt management, and socioeconomic outcomes: Post-college effects of merit-based aid. *Journal of Public Economics*, 170, 68–82.
- Sjoquist, D. L., & Winters, J. V. (2014). Merit aid and post-college retention in the state. *Journal of Urban Economics*, 80, 39–50.

- Sjoquist, D. L., & Winters, J. V. (2015a). The effect of georgia's hope scholarship on college major: a focus on stem. *IZA Journal of Labor Economics*, 4(1), 1–29.
- Sjoquist, D. L., & Winters, J. V. (2015b). State merit aid programs and college major: A focus on STEM. *Journal of Labor Economics*, 33(4), 973–1006.
- Sjoquist, D. L., & Winters, J. V. (2015c). State merit-based financial aid programs and college attainment. *Journal of Regional Science*, 55(3), 364–390.
- Smith, A. C. (2016). Spring forward at your own risk: Daylight saving time and fatal vehicle crashes. *American Economic Journal: Applied Economics*, 8(2), 65–91.
- Solis, A. (2017). Credit access and college enrollment. *Journal of Political Economy*, 125(2), 562–622.
- Takaku, R., & Yokoyama, I. (2021). What the covid-19 school closure left in its wake: evidence from a regression discontinuity analysis in Japan. *Journal of Public Economics*, 195, 104364.
- Thistletonwaite, D. L., & Campbell, D. T. (1960). Regression-discontinuity analysis: An alternative to the ex post facto experiment. *Journal of Educational Psychology*, 51(6), 309.
- West Virginia Higher Education Policy Commission. (2014). *Final report of promise scholarship ad-hoc advisory committee*. <http://www.wvhepc.edu/wp-content/uploads/2014/01/PROMISEfinalreport.pdf>.
- West Virginia Legislature. (2021). *West Virginia Code*. <http://www.wvlegislature.gov/wvcodeentire.htm>.
- Whaley, M. (1985). The status of kindergarten: A survey of the states.
- Wherry, L. R., Miller, S., Kaestner, R., & Meyer, B. D. (2018). Childhood medicaid coverage and later-life health care utilization. *Review of Economics and Statistics*, 100(2), 287–302.
- Zhang, L., Hu, S., Sun, L., & Pu, S. (2016). The effect of Florida's Bright Futures Program on college choice: A regression discontinuity approach. *The Journal of Higher Education*, 87(1), 115–146.

Online Appendix

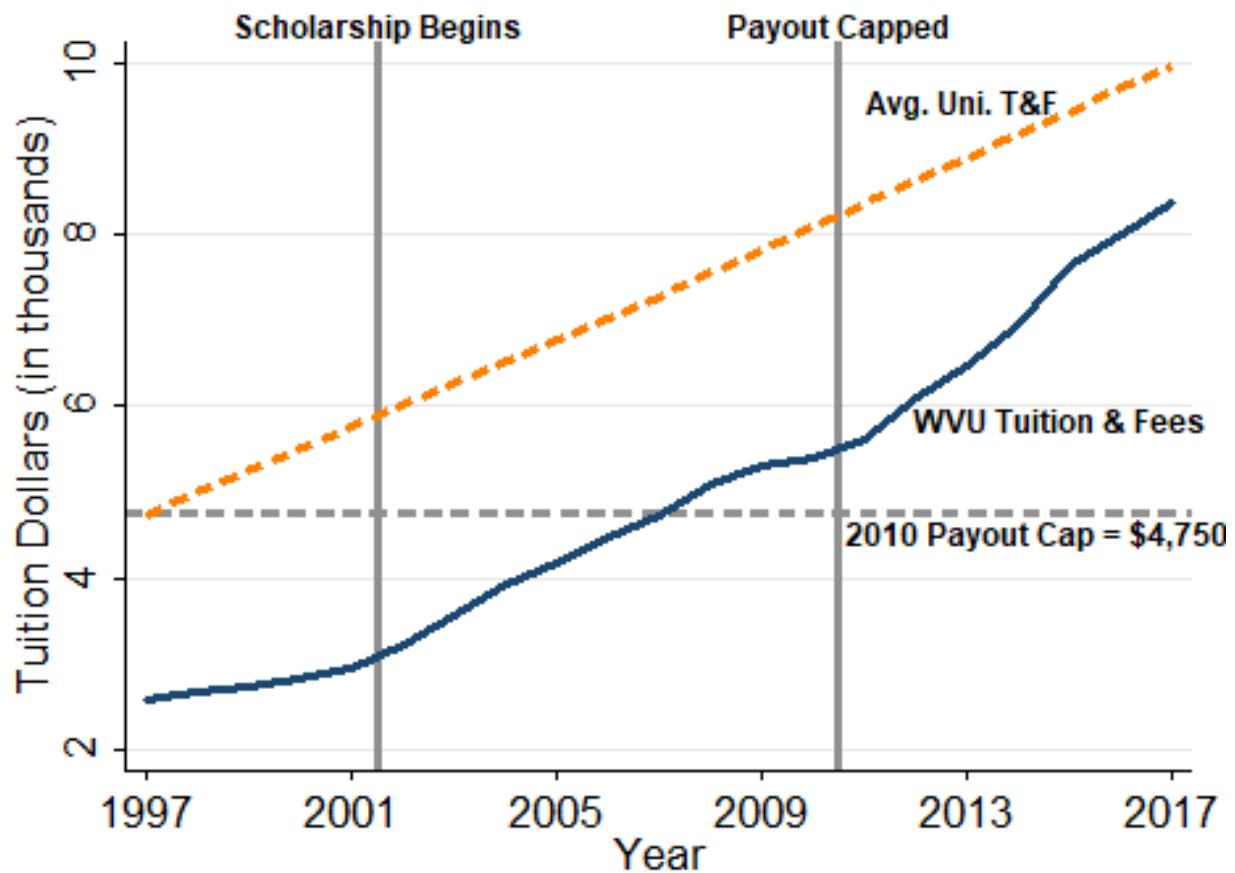
A Additional Tables and Figures

Figure A.1: Merit-Aid Adoption by State and Time



Note: Data comes from [Dynarski \(2004\)](#), [Heller and Marin \(2004\)](#), and various state agencies. Weak program means the program does not cover many students or is not generous. Weak programs date of beginning not shown.

Figure A.2: Trends in Cost of College and West Virginia Promise Scholarship Program



Note: West Virginia University tuition and fees come from West Virginia University website. Average university tuition and fees comes from College Board. There are 3 observations, in 1997, 2007, and 2017. It is linearly interpolated for other years. Data on WVU T&F from <https://institutionalresearch.wvu.edu/institutional-research/statistical-information/tuition-and-fees>.

Table A.1: Descriptive Statistics By Academic and Temporal Eligibility

Panel A: Acad Elig., Born Day ≥ 0

	Mean	Median	Std Dev	Min	Max	Count
At Least 1 Year Scholarship	0.89	1	0.31	0	1	1356
Freshman Credits	29.16	31	7.46	0	47	1356
Freshman Renewal Credits	0.70	1	0.46	0	1	1356
Freshman GPA	3.05	3	0.70	0	4	1335
Freshman Renew GPAs	0.72	1	0.45	0	1	1356
High School GPA	3.67	4	0.28	3	4	1356
Act Composite Scores	24.71	24	2.97	21	35	1114
Sat Combined Scores	1195.83	1190	125.60	1000	1560	242

Panel B: Acad Elig., Born Day < 0

	Mean	Median	Std Dev	Min	Max	Count
At Least 1 Year Scholarship	0.04	0	0.20	0	1	1081
Freshman Credits	27.63	30	8.31	0	53	1081
Freshman Renewal Credits	0.52	1	0.50	0	1	1081
Freshman GPA	2.97	3	0.74	0	4	1055
Freshman Renew GPAs	0.66	1	0.47	0	1	1081
High School GPA	3.67	4	0.29	3	4	1081
Act Composite Scores	24.72	24	2.83	21	34	874
Sat Combined Scores	1176.81	1170	119.17	1000	1550	207

Panel C: Not Acad Elig., Born Day ≥ 0

	Mean	Median	Std Dev	Min	Max	Count
At Least 1 Year Scholarship	0.05	0	0.21	0	1	617
Freshman Credits	21.86	25	9.88	0	49	617
Freshman Renewal Credits	0.24	0	0.43	0	1	617
Freshman GPA	2.21	2	0.79	0	4	585
Freshman Renew GPAs	0.30	0	0.46	0	1	617
High School GPA	3.04	3	0.41	2	4	617
Act Composite Scores	19.56	19	2.23	13	30	519
Sat Combined Scores	1003.06	980	130.40	780	1420	98

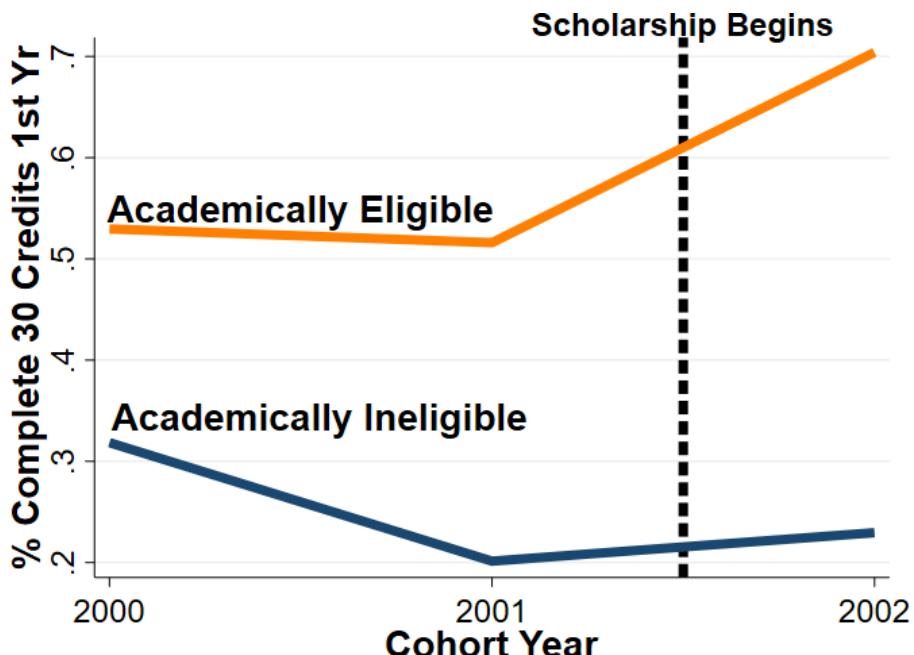
Panel D: Not Acad Elig., Born Day < 0

	Mean	Median	Std Dev	Min	Max	Count
At Least 1 Year Scholarship	0.00	0	0.05	0	1	784
Freshman Credits	21.52	25	10.01	0	41	784
Freshman Renewal Credits	0.20	0	0.40	0	1	784
Freshman GPA	2.26	2	0.79	0	4	730
Freshman Renew GPAs	0.33	0	0.47	0	1	784
High School GPA	3.01	3	0.46	1	4	784
Act Composite Scores	19.40	19	2.48	13	32	663
Sat Combined Scores	974.96	970	111.48	750	1260	121

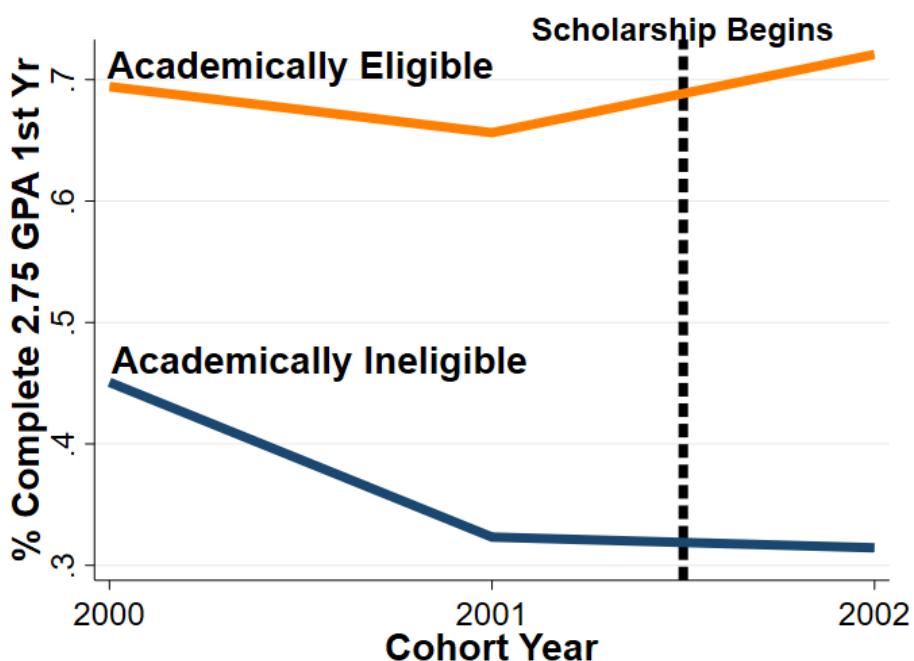
Note: Student-level observations. In-state students at WVU. First-time freshman only. Only cohorts 2001 and 2002. Freshman renewal GPA = 2.75. Freshman renewal credits = 30. Academic eligibility obtained by (HS GPA > 3) AND (ACT Composite > 20 OR SAT Composite > 1000). Temporal eligibility if born on or after cutoff date. Panel A includes academically eligible and temporally eligible students. Panel B includes academically eligible, temporally ineligible. Panel C includes non-academically eligible, temporally eligible. Panel D includes academically ineligible and temporally ineligible students.

Figure A.3: 1st Year Renewal Likelihood by Cohort

(A): First Year Credit Renewal

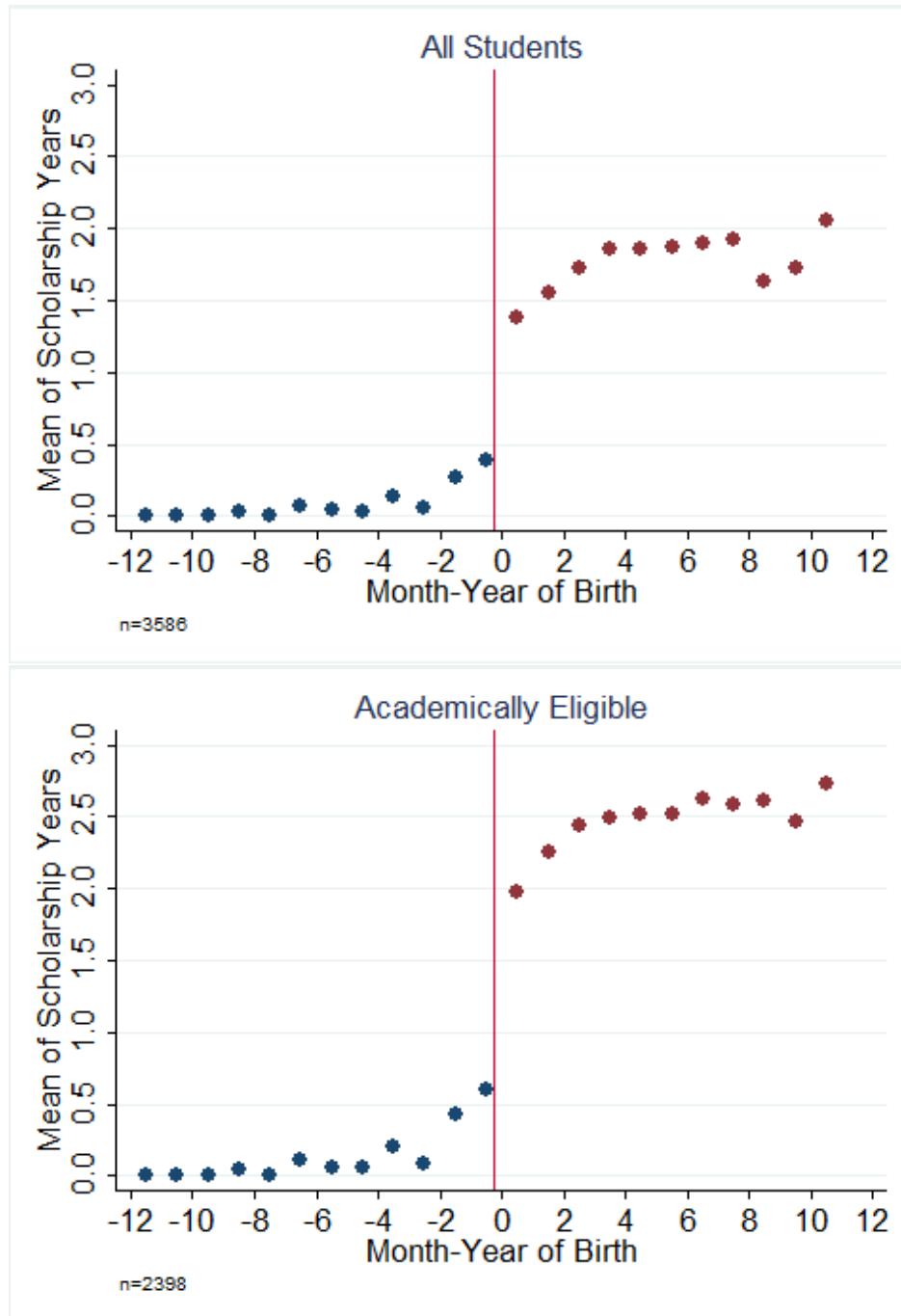


(B): First Year GPA Renewal



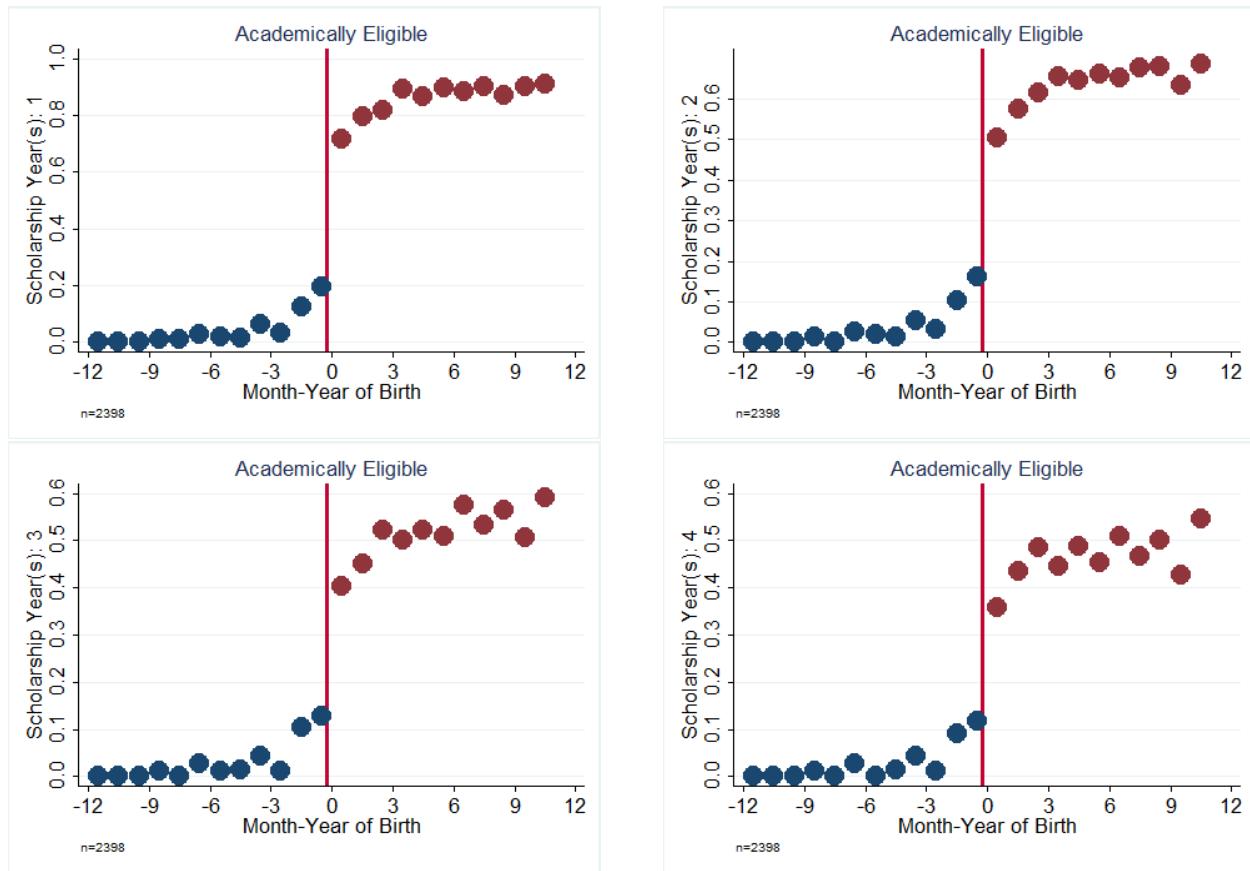
Note: A shows the percent of students who exceed the number of completed credits (30) required for scholarship renewal, once the scholarship begins, by cohort and academic eligibility. B shows the percent of students who exceed the GPA requirement required for scholarship renewal, once the scholarship begins, by cohort and academic eligibility.

Figure A.4: Effect of Birth Month on Years of Scholarship



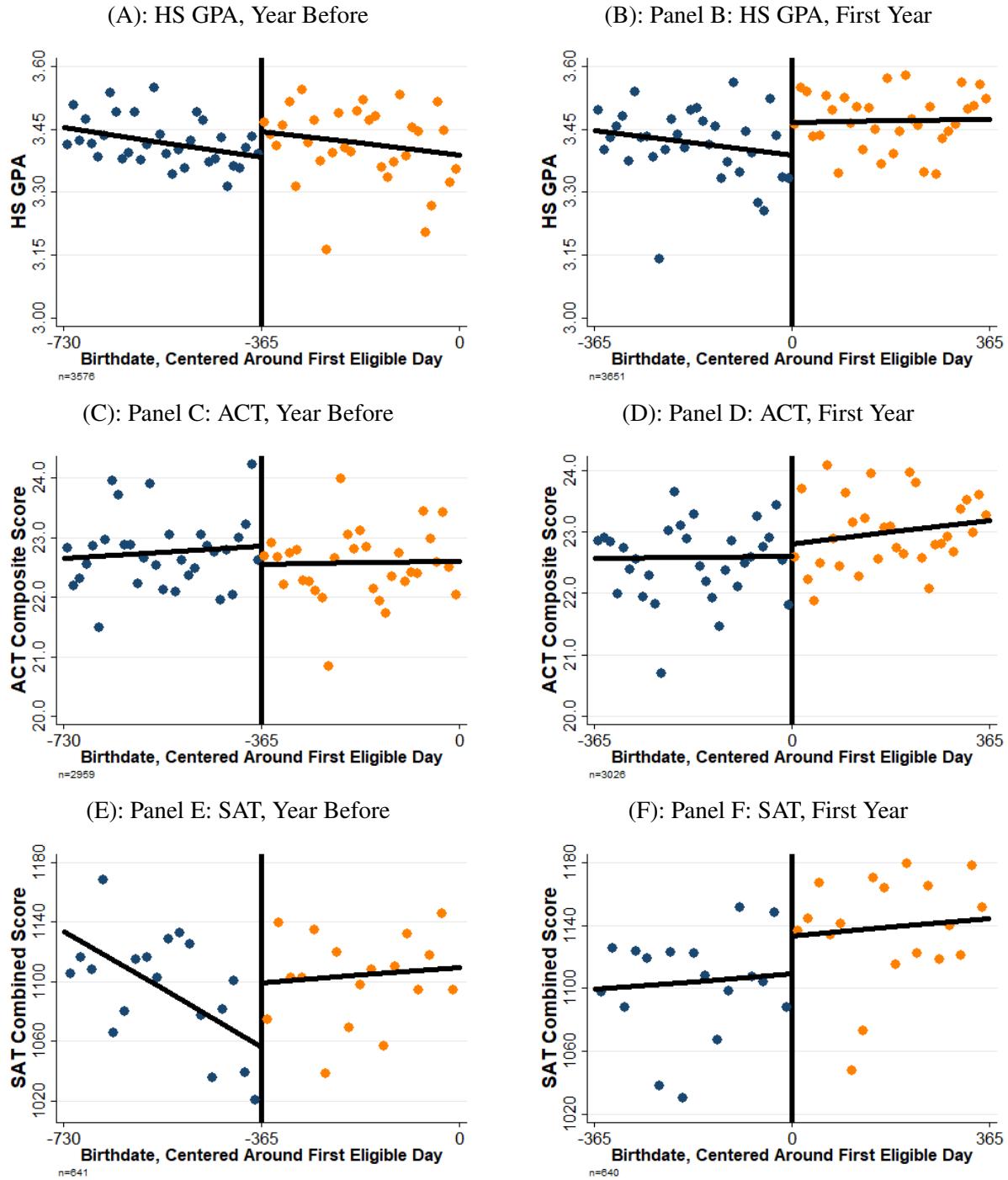
Note: Academic eligibility determined by high school GPA and ACT score.

Figure A.5: Relationship between Birth Date and Years of Scholarship Received, Academically Eligible



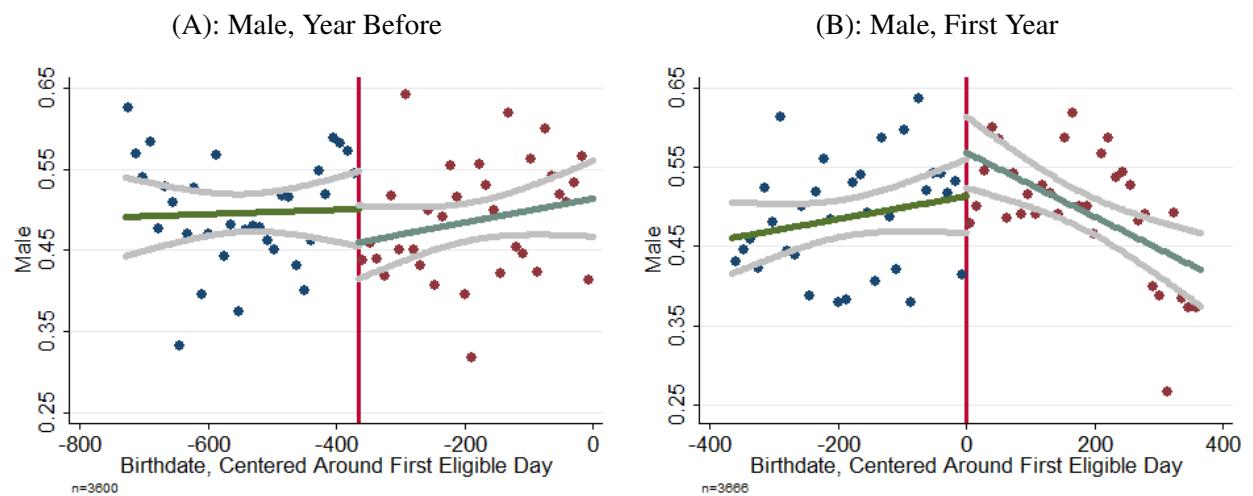
Note: Each successive graph uses a more restrictive criteria for the likelihood of scholarship. The y-axis measures likelihood of receiving scholarship for 1-4 years, 4 being the maximum number of years allowed by the policy.

Figure A.6: Pre-Treatment Covariate Continuity at Date Cutoff



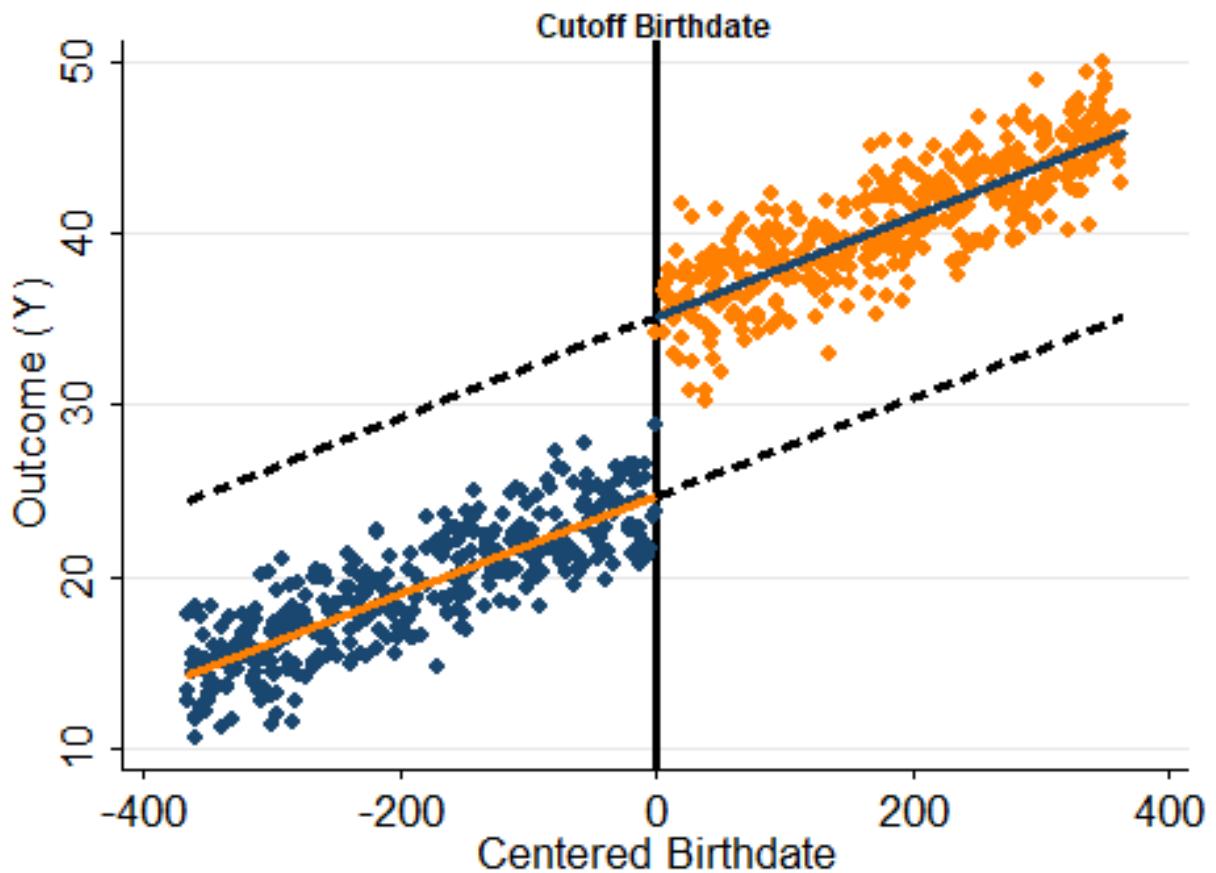
Note: Red line represents cutoff birthdate. Panels A, C, and E use the year prior to the scholarship being available cutoff, 365 days before 0. Panels B, D, and F use the year where the cutoff determines eligibility. Points are means conditional on point being within a bandwidth (bw) of h of a given birthdate. Plot A and B bw = 11.375. Plots C and D bw = 11.74. Plots E and F bw = 20.11. Bandwidths chosen as default histogram option. Specifically: number of bins chosen as: $\min(\sqrt{N}, \frac{10\ln(N)}{\ln(10)})$, where N is number of observations.

Figure A.7: Tests of Gender Continuity at Date Cutoff



Note: Red line represents cutoff birthdate. Panels A, C, and E use the year prior to the scholarship being available cutoff, 365 days before 0. Panels B, D, and F use the year where the cutoff determines eligibility. Points are means conditional on point being within a bandwidth (bw) of h of a given birthdate. Plot A and B bw = 11.375. Plots C and D bw = 11.74. Plots E and F bw = 20.11. Bandwidths chosen as default histogram option. Specifically: number of bins chosen as: $\min(\sqrt{N}, \frac{10\ln(N)}{\ln(10)})$, where N is number of observations.

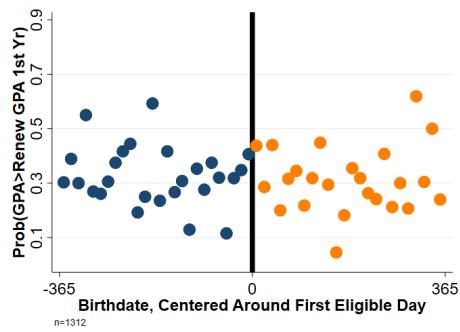
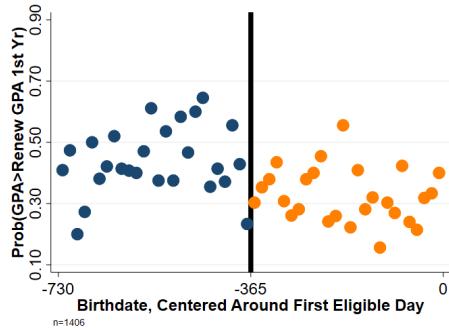
Figure A.8: Regression Discontinuity Identification Assumption: Local Smoothness of Potential Outcomes



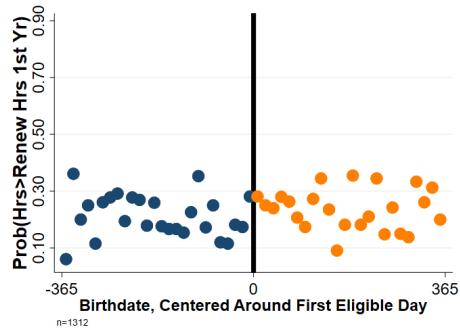
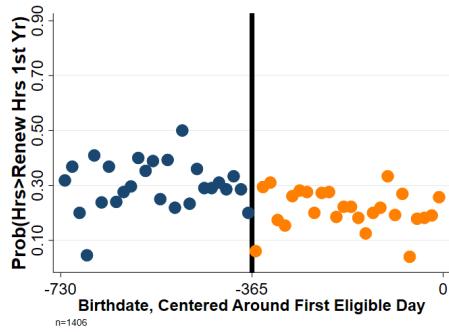
Note: Data is author-simulated. Picture depicts the assumption regression discontinuity, local smoothness, makes to estimate causal effects. Under local smoothness of potential outcomes, units just above and below make a valid counterfactual for one another.

Figure A.9: Reduced Form Effect of Birthdate on Likelihood of Exceeding Renewal Thresholds in Freshman Year, Academically Ineligible Students

(A): Exceed Renewal GPA, Pre-Scholarship Cutoff (B): Exceed Renewal GPA, Scholarship Cutoff



(C): Exceed Renewal Credits, Pre-Scholarship Cutoff (D): Exceed Renewal Credits, Scholarship Cutoff



Note: Red line represents cutoff birthdate. Panels A & C use the year prior to the scholarship being available cutoff, 365 days before 0. Panels B & D use the year where the cutoff determines scholarship eligibility. Points are means conditional on point being within a bandwidth (bw) of h of a given birthdate. Plot A bw = 14, 15.125. Plot B bw = 14. Plot C bw = 11.75. Plot D bw = 12.14. Plots A-D Bandwidths chosen as default histogram option. Specifically: number of bins chosen as: $\min(\sqrt{N}, \frac{10\ln(N)}{\ln(10)})$, where N is number of observations.

Table A.2: Local Linear Estimates of Scholarship on First Year Credits and GPA

Panel A: Dep. Var.: Continuous Credits		(1)	(2)	(3)
Scholarship		4.87*	5.61*	0.54
		(2.91)	(3.39)	(2.49)
Corrections	-	B-C, R	B-C, R	
Sample	All	All	Ac. Elg.	
Bandwidth	134	134	133	
Eff. Obs. Left	658	658	393	
Eff. Obs. Right	706	706	462	
Obs-Left	1970	1970	1081	
Obs-Right	1979	1979	1356	
Panel B: Dep. Var.: Continuous GPA		(1)	(2)	(3)
Scholarship		0.34	0.39*	0.51**
		(0.21)	(0.23)	(0.22)
Corrections	-	B-C, R	B-C, R	
Sample	All	All	Ac. Elg.	
Bandwidth	193	193	146	
Eff. Obs. Left	868	868	407	
Eff. Obs. Right	978	978	500	
Obs-Left	1865	1865	1055	
Obs-Right	1925	1925	1335	

Note: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Estimates displayed are the second stage effect of scholarship from fuzzy regression discontinuity design using birthdate in relation to Kindergarten entry date as the cutoff/instrument. The estimate is the effect of receiving the Promise scholarship. In all panels, the first stage cutoff has a strong effect on scholarship receipt. In all results, bandwidths are calculated using , in which a common bandwidth is assumed on each side. In columns 1 and 2, the full sample is used. In column 3, only students who exceed the academic cutoffs are included in the estimation sample. In columns 2 and 3, the bias-correction and standard error adjustment in [Calonico et al. \(2017\)](#) are used, because of the additional bias from using MSE-optimal bandwidths. Panel A the dependent variable is a dummy for finishes 30 credits (the renewal requirement) freshman year. Panel B the dependent variable is the number of continuous credits completed freshman year. Panel C the dependent variable is a dummy for exceeds a 2.75 GPA (the renewal requirement) freshman year. In Panel D the dependent variable is continuous GPA.

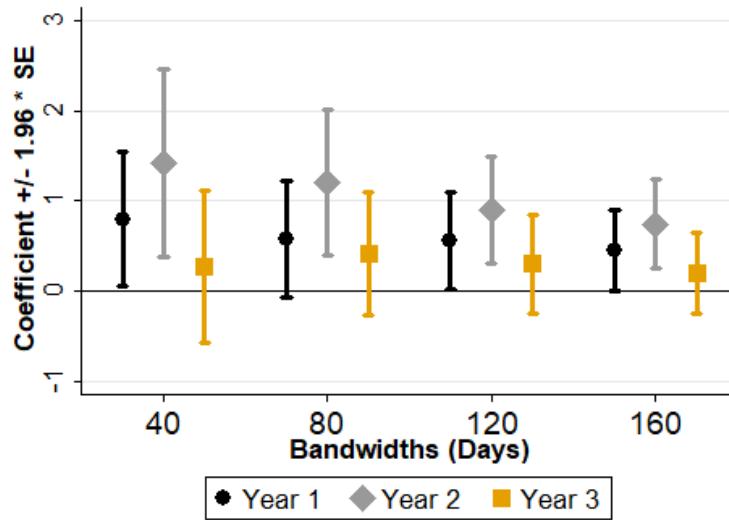
Table A.3: Scholarship on Freshman Academic Renewal Threshold Completion, Parametric

	(1) +/- 365 Days	(2) +/- 180 Days
Panel A: Credits, Linear		
	(1)	(2)
At Least 1 Year Scholarship	0.17*** (0.03)	0.27*** (0.05)
Observations	2437	1140
Kleibergen-Papp F	4273.3	1646.5
Panel B: Credits, Quadratic		
	(1)	(2)
At Least 1 Year Scholarship	0.21*** (0.04)	0.26*** (0.05)
Observations	2437	1140
Kleibergen-Papp F	2605.5	1924.4
Panel C: GPA, Linear		
	(1)	(2)
At Least 1 Year Scholarship	0.04 (0.03)	0.12** (0.05)
Observations	2437	1140
Kleibergen-Papp F	4273.3	1646.5
Panel D: GPA, Quadratic		
	(1)	(2)
At Least 1 Year Scholarship	0.06* (0.04)	0.12** (0.05)
Observations	2437	1140
Kleibergen-Papp F	2605.5	1924.4

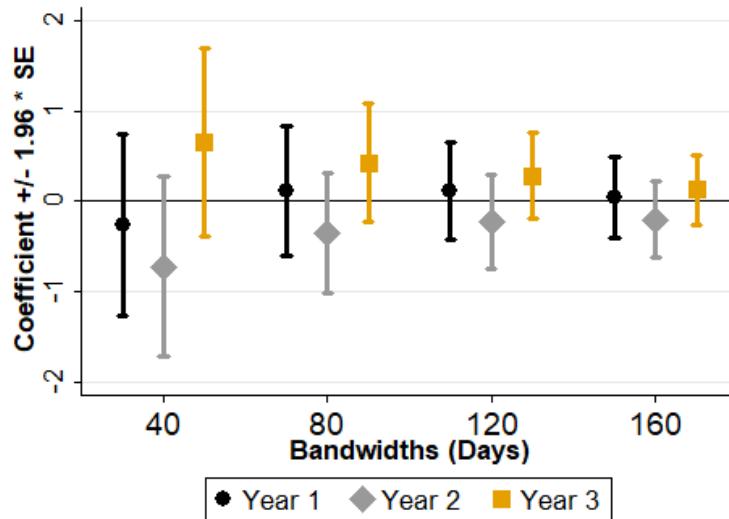
Note: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are robust to heteroskedasticity. Coefficients are from 2SLS regressions, representing the difference at the birthdate cutoff when a functional form for birthdate is assumed globally. In Panel A, exceeding the completed credits (30) in freshman year required for renewal is the dependent variable and linear regressions are fit on either side of the cutoff. In Panel B, the same dependent variable is used as A, but the functional form of birthdate is assumed to be quadratic. In Panel C, the dependent variable is exceeding the freshman GPA (2.75) required to renew the scholarship and linear regressions are fit on either side of the birthdate cutoff. Panel D uses the same dependent variable as Panel C and assumes the functional form of birthdate is quadratic. Finally, column 1 uses 365 days on either side of the cutoffs. Column 2 uses 180 days on either side of the cutoffs.

Figure A.10: GPA Estimates by Year by Gender

(A): Probability of Completing Renewal GPA, Male



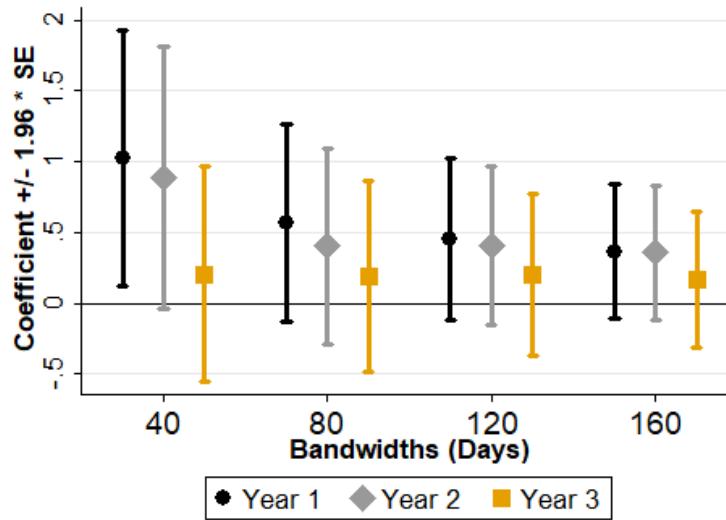
(B): Probability of Completing Renewal GPA, Female



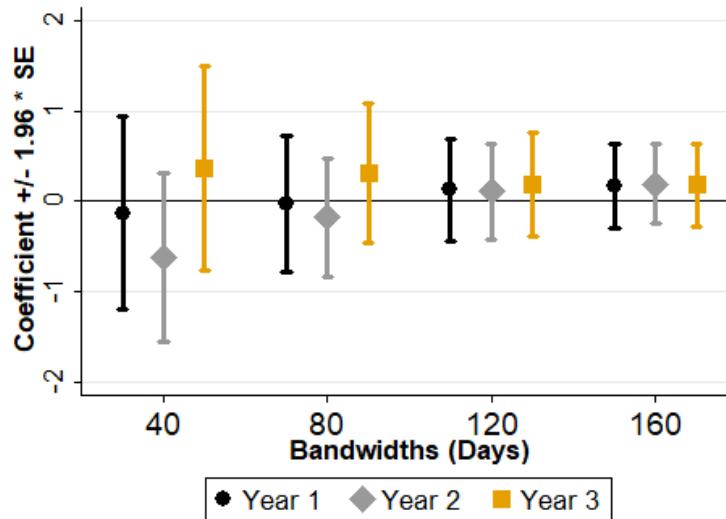
Note: Dependent variable is: finish renewal GPA in that year only. Panel A is male. Panel B is female. $N \leq 524$ in Panel A. $N \leq 541$ in Panel B. Coefficients represent ‘conventional’, not bias-corrected or robust for MSE-optimal, treatment estimates from difference-in-discontinuity estimates (Grembi et al., 2016). All regressions subsample to academically eligible students meaning $[(ACT \geq 20.5 \text{ OR } SAT \geq 1000) \text{ AND HS GPA} \geq 3.0]$ and include controls for HS GPA and ACT composite score (w/ SAT converted to ACT).

Figure A.11: Credits Estimates by Year by Gender

(A): Probability of Completing Renewal Credits, Male



(B): Probability of Completing Renewal Credits, Female



Note: Dependent variable is: finish renewal credits in that year only. Panel A is male. Panel B is female. $N \leq 524$ in Panel A. $N \leq 541$ in Panel B. Coefficients represent “conventional”, not bias-corrected or robust for MSE-optimal, treatment estimates from difference-in-discontinuity estimates (Grembi et al., 2016). All regressions subsample to academically eligible students meaning $[(ACT \geq 20.5 \text{ OR } SAT \geq 1000) \text{ AND } HS \text{ GPA} \geq 3.0]$ and include controls for HS GPA and ACT composite score (w/ SAT converted to ACT).

B Compulsory Kindergarten Entry Laws

The compulsory entry laws relevant to this paper are birth date cutoffs for when children must begin kindergarten. These laws vary across states and times. Since Promise begins in 2002 and most students enter college after completing grades K-12, the relevant entry laws were in effect during the 1980's.

There was change in the entry date set by West Virginia's laws over time. In 1982, the rule changed to students who were 5 by September 1 of the current school year were supposed to enter kindergarten. Additionally, in 1988 the law was changed again which means that students in year after are sorted at a different date cutoff, November 1. Thus, the 1988 law cutoff is the last time that September 1 is the cutoff.

The kindergarten date cutoff determining year of college cohort entry depends on how strictly the cutoff is enforced (i.e. whether students born later are allowed to enter early).⁵³ ([Whaley, 1985](#), pg. 20) states, “local districts may permit earlier entry if a child shows readiness based on a test of basic skills.” Due to the potential for students to enter early it is highly improbable that the cutoff changes the probability of entering in the first eligible cohort from 0 to 1, thus dictating a fuzzy RD research design. However, ([Whaley, 1985](#), pg. 20) also states, “Complaints occur about the fact one district may allow early entry, while a neighboring one may not,” suggesting that the laws are enforced to a certain extent. West Virginia compulsory age of kindergarten entry does not apply to students that enter kindergarten in another state, then move into West Virginia.

⁵³If the cutoff is not enforced strictly enough, then the cutoff will not have explanatory power for cohort of entry or scholarship receipt.

C Threats to Identification

There are two main threats to identification. The first is relative age effects, which are addressed by estimating how exceeding the prior year's cutoff, which does not affect scholarship receipt, affects pre-treatment covariates, density, and outcomes. The second is there being another policy that begins with the cohort that begins Kindergarten in 1989.

C.1 Relative Age

An important concern with using a birthdate cutoff in kindergarten entry as variation in scholarship receipt is that the effect of the scholarship is confounded by how a student's position in the age distribution of their cohort also affects outcomes. This effect, of age relative to one's cohort, is called relative age effects and these have been shown to affect many outcomes ([Peña, 2017](#)). If one is the oldest in their cohort when they enter kindergarten, then they are the most cognitively developed and this can have a non-trivial difference on outcomes.

Beyond relative age effects, parents have also become aware of this phenomenon and may try to delay the beginning of their child starting school if they are able ([Deming & Dynarski, 2008](#)). Often knowledge or ability to delay is correlated with income or education which could further confound the effect of the cutoff. Prior work establishes that relative age does matter for probability of admission to college ([Matta, Ribas, Sampaio, & Sampaio, 2016](#)) in Brazil.

Relative Age in This Context In this study, relative age could affect the estimated effect of the scholarship in at least two ways. One is a confounding effect where students that are more cognitively advanced when they entered kindergarten have accumulated higher academic ability. Higher academic ability would bias the effect of the scholarship upwards. However, another way relative age could affect outcomes is through selection effects. Students right at the cutoff may be more cognitively advanced and therefore able to attend more prestigious colleges out of state. This would bias the estimates downwards, so the effect of relative age on academic outcomes in this data is ambiguous. While it is possible that students and parents are manipulating their

age relative to their cohort, it is implausible that they are purposefully manipulating into college scholarship receipt. Manipulation into scholarship is implausible, because the college scholarship is not announced until 9 years after this kindergarten cohort begins and begins 19 years after the students are born.

Estimating Relative Age Confounding To ensure estimated effects are not due to relative age confounding, this study uses the pre-scholarship cutoff. Students aged 5 on August 31, 1987 are required to enter kindergarten in 1987, but those who turn 5 on September 1, 1987 enter kindergarten in 1988. Thus students born on September 1, 1982 are treated by being the oldest in their class when they begin in 1988. Importantly, these students are not treated by the scholarship. In Section 5, students' pre-treatment covariates and the density around the pre-scholarship cutoff (September 1, 1987) will be compared to the post-scholarship cutoff (September 1, 1988).

C.2 Other Cohort-Based Policies

A threat to identification is that there is some other confounding policy that begins for children born just beyond the cutoff point (i.e. on or after September 1, 1983). This could be a federal government policy or some other specific policy that affects children in the 1989 kindergarten cohort in West Virginia. To investigate these possibilities, two sources are searched for such programs: 1) the unified welfare analysis across programs in [Hendren and Sprung-Keyser \(2020\)](#) and the record of the West Virginia state legislature.

Policies from a Unified Welfare Analysis ([Hendren & Sprung-Keyser, 2020](#)) There are 133 government policies that are compared in [Hendren and Sprung-Keyser \(2020\)](#). These policies are searched for policies that change discontinuously for students on either side of the birth cutoff, September 1, 1983, for entering Kindergarten in 1989 in West Virginia. While many policies affect students in the 1980's, most do not discontinuously affect those who are born on opposite sides of the cutoff.

Throughout the 1980's there was significant expansions of federal health insurance programs, particularly Medicaid. One policy change was the extension of Medicaid to low-income children who were born after September 30, 1983. This expansion increased education and health care utilization ([Wherry, Miller, Kaestner, & Meyer, 2018](#)); however, this eligibility only changes discontinuously for students who are born 30 days from the cutoff ([Cohodes, Grossman, Kleiner, & Lovenheim, 2016](#)). Second, there was an extension of Medicaid to pregnant mothers which raised high school completion for babies who were in utero at the time ([Miller & Wherry, 2019](#)). Nevertheless, while this program changed in the 1980's, it did not affect those on either side of the 9/1/1983 birthdate cutoff differently. This is true for all other health insurance expansions at this time.

West Virginia Board of Education Policies Next, the West Virginia State Legislature Code, Chapter 18- Education, is searched for potentially confounding policies that apply only to those to the right of the birthdate cutoff ([West Virginia Legislature, 2021](#)). Find – programs that begin with the 1989 cohort. Also, search 18A- School Personnel, 18B- Higher Education, 18C- Student Loans, Scholarships , 5B- Economic Development Act of 1985, 8/14-”...Special School Zone...”, 21- Labor.⁵⁴ Find – programs that begin with the 1989 cohort. Importantly, these programs are not found to have a large effect on pre-treatment covariates in Section [5.2](#) or density in Section [5.3](#).

⁵⁴18C, Article 7 is Promise.