DOES BROAD-BASED MERIT AID IMPROVE COLLEGE COMPLETION? EVIDENCE FROM NEW MEXICO'S LOTTERY SCHOLARSHIP

Christopher Erwin
(corresponding author)
Department of Economics
University of New Mexico
Albuquerque, NM 87131
cpe@unm.edu

Melissa Binder
Department of Economics
University of New Mexico
Albuquerque, NM 87131
mbinder@unm.edu

Abstract
We use the natural experiment of a state lottery scholarship to measure the effect of generous financial aid on graduation rates at New Mexico's flagship public university. During the study period, the scholarship program paid full tuition for eight semesters for any state resident earning a 2.5 grade point average in their first semester at any public two-year or four-year college. We find a significant positive completion effect of 10 percentage points (17.9 percent) for academically well-prepared students that is offset by a large negative effect of 11.6 percentage points (38.8 percent) for less-prepared students. We posit that the scholarship program, which effectively erased the difference in tuition at two- and four-year colleges, may have induced weaker students to take their chances on a more prestigious, yet riskier, academic path.

https://doi.org/10.1162/edfp_a_00270

1. INTRODUCTION

The introduction of broad, merit-based college scholarships in the 1990s created a natural experiment for measuring relationships between college costs and academic outcomes. State merit-based scholarships generally fund most, if not all, tuition for qualified resident students. State legislation establishing merit-based scholarships share several common goals: retaining talent in-state, increasing access to higher education by reducing financial burdens, and promoting timely completion. There is considerable variation in initial and continuing eligibility requirements across states. Researchers have cataloged how such programs affect enrollment and course-taking behavior, and, more recently, degree completion. We analyze the effect of the New Mexico Legislative Lottery Scholarship (NMLLS), a uniquely "low-bar" merit-based scholarship, on degree completion.

Since 1993, at least twenty-five states have implemented merit-based scholarships, the first and most studied being Georgia's Helping Outstanding Pupils Educationally, or HOPE, scholarship program.[1] HOPE marked the beginning of what has been a major restructuring of the financial aid landscape in the United States. According to the College Board, from 1993 to 2013, the percentage of total undergraduate state grant aid for which students' financial circumstances were considered decreased from 90 percent to 76 percent. In the 2013–14 academic year, New Mexico was one of thirteen states where this percentage was below 40 percent.[2]

We know more about the relationship between financial aid and enrollment than financial aid and college completion. Different types of financial aid have varying effects on college enrollment. Loans tend to have little to no effect, whereas grants have a positive and significant effect on student enrollment (Linsenmeier, Rosen, and Rouse 2006). Students from low-income families and students of color seem to be most responsive to such aid. Van der Klaauw (2002) demonstrates that students' choice of college is sensitive to financial aid offers. Several studies show a significant and positive relationship between grant aid and student enrollment (Heller 1999; Seftor and Turner 2002; Kane 2003) and a negative relationship between net cost and enrollment (McPherson and Schapiro 1991). The effects of merit-based aid on enrollment have also been well documented. In an experimental setting, Monks (2009) finds large, positive effects of merit aid on enrollment. Studying HOPE, Dynarski (2000) finds that a $1,000 award increased student enrollment by approximately 4 percent. Also studying HOPE, Cornwell, Mustard, and Sridhar (2006) find that the program increased student enrollment by 6 percent. In New Mexico, Binder, Ganderton, and Hutchins (2002) and Binder and Ganderton (2004) find that although the NMLLS boosted enrollment at four-year colleges in New Mexico, the effect appears to be driven by additional enrollment of students who otherwise would have attended college out-of-state.

The NMLLS was specifically designed to increase access to higher education and encourage students to finish high school. The bill's sponsor, Senator Michael Sanchez, discussed the impetus for the program, noting:

1. See Sjoquist and Winters (2012) for a complete list of the twenty-five states.
2. See College Board (2015), figures 28A, 28B.

> when I went to high school, I saw a lot of my friends, a lot of other individuals, who had to drop out of school to either go on to the service to help provide for their families or work on family farms or get some kind of job to help their families out in this area.
>
> . . .we just thought that, what is an incentive to try to keep people in school? . . . Talking to different people . . . it was always a matter of well, "why should we finish high school because even if we graduate from high school, we're not going to be able to afford to go to college." (Ness 2008, p. 36)

Making higher education widely accessible is certainly a noble objective, but effective merit-based aid programs should also increase degree completion. Although there is likely a positive productivity signal sent by those whose highest level of education is "some college, no degree," research suggests that the returns to such attainment are far exceeded by earning a bachelor's degree (Arrow 1973; Jaeger and Page 1996). Degree completion is associated with better health, increased earnings, and overall happiness (Card 1999; Cuñado and de Gracia 2012; Frey 2018, pp. 13–20).

In this paper we examine how the NMLLS affects college completion at the University of New Mexico (UNM) by exploring changes in completion rates before and after the implementation of the scholarship for eligible resident students and a matched sample of nonresident (and therefore ineligible) students. Estimates reveal no significant overall effect of the program on completion rates. However, we do find large and statistically significant completion effects after disaggregating by high school performance. Academically well-prepared eligible freshmen are 10 percentage points (17.9 percent) more likely to graduate within six years, compared with ineligible peers with similar high school grade point averages (GPAs). Less academically prepared freshmen are approximately 11.6 percentage points (38.8 percent) less likely than their ineligible peers to graduate within six years. These opposite responses dampen the overall effect of the NMLLS. Further decomposition by family income suggests that low-income students likely drive this pattern.

Findings are informative to states with existing broad-based merit scholarships and those contemplating launching programs of their own. Because the NMLLS covers all tuition over our sample period[3] for many high school graduates, effectively removing price differentials between universities, our research also informs recent proposals to make college "free" for students with family incomes under $125,000.[4] Our results support the idea that removing price as a signal in higher education markets may skew students' college-going decisions, resulting in increased "overmatching" (see Arcidiacono, Aucejo, and Hotz 2016, for example).

The paper proceeds as follows: Section 2 discusses existing literature regarding merit aid and college completion, and introduces the NMLLS; section 3 presents a

3. As of the 2017–18 academic year, the NMLLS only covers approximately 60 percent of tuition at UNM.
4. Hillary Clinton and Senator Bernie Sanders both proposed similar versions of this policy during the 2016 presidential campaign. Sanders advocated free tuition at all public universities and colleges, whereas Clinton advocated for a $125,000 household income cap (see Kreighbaum 2017). New York recently launched the Excelsior Scholarship, making tuition at all SUNY and CUNY two- and four-year colleges free for residents from families with annual incomes up to $125,000 (New York State 2017).

theoretical model of college persistence; section 4 describes the data; section 5 summarizes the empirical approach; section 6 discusses main findings and robustness checks; section 7 discusses other explanations for patterns we find in the results; and section 8 concludes.

2. FINANCIAL AID AND STUDENT OUTCOMES

The natural experiment of lottery-financed merit-based aid programs provides a promising avenue for determining the relationship between aid and college completion. Analyzing statewide educational attainment data, Sjoquist and Winters (2012, 2015) found no difference in college attainment for those exposed to lottery scholarship programs. Using a similar methodology, Jia (2019) found that program features matter: Lower initial scholarship eligibility requirements increased two-year degree attainment, and funding generosity increased the completion of a bachelor's degree.

Scott-Clayton (2011) found completion effects of 9.4 percentage points (59 percent) for students just above an ACT cutoff for West Virginia's lottery-funded PROMISE scholarship program, compared with students just below. Using similar strategies, Bruce and Carruthers (2014) and Welch (2014) found no program effect for Tennessee's lottery scholarship. The discrepancy between these studies may arise from differences in student characteristics. Because of differences in program requirements, all students in Scott-Clayton's sample have high school GPAs of 3.0 or higher whereas students in Bruce and Carruthers's and Welch's samples have high school GPAs below 3.0.[5] It may be that only stronger students are able to respond to merit requirements. A high rate of scholarship loss supports this supposition. For example, only 50 percent of students who initially earn the PROMISE scholarship retain it for four years of college. It also bears noting that Scott-Clayton's large 9.4 percentage point (59 percent) completion effect at four years declines to 4.5 percentage points (12 percent) at five years. It is therefore possible that the scholarship program improves time to degree without changing eventual college completion. This would explain why Sjoquist and Winters (2012, 2015) find no population graduation effect.

Castleman and Long (2016) examine the effect of the need-based Florida Student Access Grant (FSAG), which awards $1,300 annual grants to students whose family's Expected Family Contribution falls below an annually determined cutoff, with no additional academic restrictions for grant receipt in the first year. Because the Expected Family Contribution is generated from information provided by students to the Free Application for Federal Student Aid (FAFSA) according to an opaque algorithm, and because the cutoff is determined each year and is not publicized,[6] it is unlikely that students manipulated their FAFSA responses to become eligible. Students just above and just below this cutoff were therefore likely to have differed only by grant receipt,

5. Because Scott-Clayton did not limit the sample to those who took the ACT only once, her marginal program students were able to manipulate their test score, so those above differed in unobserved characteristics, like ambition. But even though Bruce and Carruthers limit their sample to students who took the ACT only once, they face a similar situation—students just below the cutoff sample who did not retest might contain a higher proportion of low-ambition students, relative to those just above the cutoff who had less incentive to retest. Thus selection alone is unlikely to explain the discrepancy between the studies.

6. We were unable to find any reference to the cutoff online, except as reported in Castleman and Long, and that figure is more than ten years old.

providing an opportunity to test the effect of need-based financial aid on college outcomes. Castleman and Long find that students just below the cutoff for the FSAG in the 2000–01 school year are 4.6 percentage points (22 percent) more likely to earn a bachelor's degree in six years than those just above the cutoff. In this case, the effect on graduation persists over time: It is 3.2 percentage points (20 percent) at five years and 5.2 percentage points (21 percent) at seven years, the longest period reported.

At least one study implicates financial aid in worse college outcomes. In their study of Massachusetts' Adams Scholarship, a program providing tenth graders in the top quartile of a state standardized test with tuition waivers to attend public in-state colleges, Cohodes and Goodman (2014) find award eligibility decreases the likelihood of obtaining a degree within six years by 2.5 percentage points (4 percent). The mechanism for this perverse outcome appears to be the diversion of students from higher-quality private to lower-quality public institutions. The authors conclude that students are willing to sacrifice significant college quality in response to scholarship receipt.

Mixed evidence for lottery program effects may be a result of their broad base in terms of income (so that many recipients are not financially constrained) and relatively narrow base in terms of merit (so that many recipients are likely to succeed in college anyway). We are therefore particularly interested in the effect of NMLLS on lower income and higher-ability students, the group that saw the greatest benefit from the FSAG program.

NMLLS Program Details

The NMLLS, established by the New Mexico Legislature in 1996, first became available to students in fall 1997. New Mexico residents qualify for the NMLLS if they earn a high school diploma or general educational development (GED) equivalency in New Mexico and enroll at a public postsecondary institution in the first regular fall or spring semester after high school graduation. Most state lottery scholarship programs reward high school achievement and begin with the first semester of college enrollment. In New Mexico, however, students become eligible for full tuition at any of the sixteen qualified public two- or four-year colleges after they complete a full-time course load (at least 12 credits) with a 2.5 GPA or higher in their first college semester. To encourage students to try for the scholarship, New Mexico colleges offer students "Bridge to Success" scholarships, which completely or mostly offset tuition in their first semester. In the period examined, students could receive the award for up to eight semesters, provided they enroll full-time, continuously, and maintain a cumulative 2.5 GPA. Only 58 percent of first semester students over 1997–99 met NMLLS requirements, and only 30 percent remained eligible at the end of their second year.

Before the NMLLS, New Mexico nearly exclusively awarded financial aid based on need. According to a 1994 National Association of State Student Grant & Aid Programs report, New Mexico devoted an average of $222 per full-time equivalent (FTE) undergraduate student in financial aid in the 1993–94 academic year. Of the $222 total per FTE, only $3 (1.4 percent) was merit-based. By contrast, in 2000, New Mexico allocated $687 per undergraduate FTE, with $368 (54 percent) being merit-based. It appears that the NMLLS not only supplemented rather than supplanted student aid, but drastically changed the student aid landscape throughout the state.

Table 1. Full-Time Resident Tuition at all New Mexico Legislative Lottery Scholarship—Eligible Institutions

Institution	Program Length (Years)	Tuition and Fees ($)
New Mexico Institute of Mining and Technology	4	7,000
University of New Mexico	4	6,950
New Mexico State University	4	6,729
Western New Mexico University	4	6,644
Eastern New Mexico University	4	5,630
New Mexico Highlands University	4	5,550
New Mexico Military Institute	2	5,179
Northern New Mexico College	4	5,112
San Juan College	2	1,773
Central New Mexico Community College	2	1,340
Clovis Community College	2	1,324
Santa Fe Community College	2	1,196
New Mexico Junior College	2	1,158
Luna Community College	2	968
Southwestern Indian Polytechnic University	2	730

Notes: Figures present tuition and fees for one academic year taking 15 credit hours per semester. For two-year institutions, it is assumed the student is within the community college district, where applicable.

Source: Institutional financial aid Web sites.

Compared with states with similar programs, NMLLS eligibility requirements are relatively "low-bar." For example, Georgia's HOPE scholarship requires students to graduate high school with a 3.0 cumulative GPA and maintain a 3.0 GPA in college.[7] Eligibility for Tennessee's HOPE scholarship requires minimum ACT/SAT scores in addition to the 3.0 high school GPA requirement. Renewal requires a 2.75 minimum overall GPA after attempting 24 and 48 credit hours, and requires a 3.0 minimum overall GPA at 72- and 96-credit hour reviews.[8] Florida's Bright Futures Scholarship has three levels of merit-based awards, each with varying high school GPA, standardized test scores, and community service requirements.[9]

If financial constraints are binding for students, then the NMLLS should have the desired effect of increasing the proportion of students meeting the 2.5 cumulative GPA and 12 credit hours continuous enrollment requirements, thus increasing completion rates. But if other constraints, such as academic preparation, are also binding, the scholarship could have the opposite effect, reducing completion rates for marginal students induced to enroll at the state's flagship university who otherwise would have enrolled at a less prestigious university, a two-year program at a community college, or perhaps not have enrolled in college at all. With price signals in the market for higher education removed, some students may choose to embark on a more prestigious, yet riskier, academic path—one that maximizes the "worth" of the scholarship (i.e., that which covers the largest cost). Consider full-time tuition at all sixteen participating public institutions in New Mexico, as depicted in table 1. A student better matched at Santa Fe Community

7. Georgia Student Finance Commission (2019).
8. TN Higher Education Commission & Student Assistance Corporation (2019).
9. Office of Student Financial Assistance (2019).

college may decide to attend UNM instead simply because the scholarship covers more costs, the degree carries more prestige, and thus the NMLLS is "worth" more at the state's flagship university.

3. MERIT AID AND PERSISTENCE

We model students' college persistence behavior (and ultimately their decision to graduate) using a multistage investment model adapted from Bettinger (2004). Students decide to enroll in college if they perceive in the initial period that the discounted stream of future benefits exceeds the expected discounted value of college costs. Benefits are based on the earnings differential between those with college and high school degrees,

$$E_0\left[\sum_{t=TTD+1}^{T^*} \delta^{t-1}\left(w_{coll} - w_{hs}\right)\right], \tag{1}$$

where E_0 is a student's expectation before beginning school ($t = 0$), $\delta \equiv \frac{1}{1+r}$ is the student's discount factor, r is the discount rate, and students expect to work for T^* years following graduation. TTD is the expected number of years it takes to earn a degree, with students beginning work in the following period. College graduates earn w_{coll} and high school graduates earn $w_{hs.}$ The cost of college is:

$$E_0\left[\sum_{t=1}^{TTD} \delta^{t-1}\left\{w_{hs}T_{s_{it}} + \gamma F_t - A_{it}\left(e_{it}\left(a_{i,t-1}, T_{s_{it}}\right)\right)\right\}\right], \tag{2}$$

where $T_{s,it}$ is the fraction of time student i dedicates to studying or attending class of the total time available for working or studying in time t; this fraction, multiplied by the high school graduate's earnings, represents the opportunity cost of attending college. F_t is tuition and fees in time t, where the parameter γ is the proportion of tuition and fees for which the student is responsible, as college attendance is commonly covered by one's parents. A_{it} is non-loan financial aid available to student i in time t. At least some of the available aid is merit-based, and so A_{it} is increasing in student effort per quality credit (credit times the four-point GPA) earned, e_{it}, which in turn is increasing in academic skill acquired in the previous period, $a_{i,\,t-1}$, and the time dedicated to studying, $T_{s,it}$.

Before enrolling, the student has some idea of how to divide time between studying and work, as well as how much effort is needed to maintain the offered financial aid package. Upon enrolling, students expect the benefits of earning a degree to outweigh the costs. Once enrolled, however, they may discover that they overestimated their academic preparation, underestimated the effort required to earn college credits, or both. As a result, more time devoted to school may be needed, which raises opportunity cost, or may result in lower grades than expected, thereby resulting in higher direct costs because merit aid is rescinded. In either scenario, costs have risen, and are now more likely to exceed the benefits of continuing in college.

The model predicts countervailing effects of broadly available merit aid on degree receipt. Because the scholarship reduces the cost of attendance, more students will attend and complete college. Simultaneously, students who are induced to attend college due to lower cost may overestimate their ability and underestimate the effort required to earn the NMLLS, so may be more likely to drop out. Academic preparation is central to understanding how students respond to such financial aid. As the NMLLS effectively removes price signals across public in-state institutions, students may seek to maximize the value of the scholarship by pursuing a degree from an institution that they feel carries the most value—typically a more expensive or reputable institution. This may result in some students overmatching (e.g., attending a school for which they are academically underprepared), leading to higher attrition rates for these students.

To provide an incentive to graduate in a timely manner, the NMLLS was available to students for only eight semesters following the bridging semester. However, the scholarship also only required a 12 credit-hour load for a student to be considered full-time. Thus, the incentive to graduate in nine semesters was countered by the incentive to maintain a 2.5 GPA, which would be harder to do with a higher credit load. Students thus face a tradeoff between losing the scholarship if they fail to meet the renewal requirements, and facing higher direct and opportunity costs at the end of their programs if they take longer than nine semesters to graduate. If the opportunity cost, including the risk of losing the scholarship, of a 15-credit per semester course load that would produce a degree in nine semesters exceeds the cost of the 20 or so credits not covered by the scholarship, and the added opportunity cost of delayed full-time work, then the program may not effectively encourage timely completion. Students whose families are funding their college education may not respond to the semester cap.

4. DATASET

To estimate completion effects, we use administrative data for all first-time, full-time entering freshmen at UNM before and after the implementation of the NMLLS. UNM enrolls over 20,000 students each year in the City of Albuquerque, the largest metropolitan area of the state with over 500,000 residents. UNM is nearly an open-enrollment institution. Our data include sociodemographic information (age, race, ethnicity, gender, family income, declined to state race-ethnicity), high school academic performance (high school GPA, standardized test scores, indication of remedial coursework at UNM), and college academic performance by semester (credits earned, college GPA, date of graduation). Data are complete with the exception of family income and high school GPA. We only have family income for FAFSA-filing students, constituting 51 percent of students. For those who did not submit a FAFSA, we assume their family income is sufficiently high (i.e., ≥$40,000) as to not qualify for the Federal Pell Grant Program. This assumption is supported by the 1996–97 Federal Pell Grant End-of-Year Report showing that less than 3 percent of Pell Grant recipients had family income in excess of $40,000 (USDOE 1997). This assumption is not perfect. King (2004) estimates that in 2000 over 10 percent of all Pell Grant–eligible students did not fill out a FAFSA. If the analysis in King holds for our dataset, then we would incur systematic measurement bias in the family income variable—some lower income students would be incorrectly placed in the higher income category. Because we find evidence that low-income students drive the completion effects we detect, measurement error would likely only serve to dampen point estimates for low-income regressions. We are missing high school GPA for home-schooled students, a small portion of

Table 2. Student Characteristics Before and After Initiation of the New Mexico Legislative Lottery Scholarship Program

	Residents			Nonresidents		
Variable	Before	After	Difference	Before	After	Difference
Graduate in 4 years	0.124	0.103	-0.021^{***}	0.166	0.154	-0.012
Graduate in 4.5 years	0.209	0.189	-0.020^{**}	0.232	0.211	-0.021
Graduate in 5 years	0.360	0.332	-0.028^{**}	0.311	0.297	-0.014
Graduate in 6 years	0.460	0.420	-0.040^{***}	0.359	0.319	-0.040
High school GPA	3.348 (0.502)	3.273 (0.471)	-0.075^{***}	3.284 (0.521)	3.300 (0.503)	0.016
ACT	22.677 (3.860)	22.176 (3.887)	-0.501^{***}	22.668 (3.995)	22.861 (4.096)	0.193
Remedial	0.244	0.290	0.046^{***}	0.140	0.227	0.087^{***}
Income < $40K	0.230	0.205	-0.025^{***}	0.164	0.162	-0.002
Female	0.564	0.565	0.001	0.528	0.545	0.017
Hispanic	0.390	0.375	-0.015	0.180	0.166	-0.014
Native American	0.038	0.045	0.007	0.047	0.051	0.004
Asian	0.049	0.037	-0.012^{**}	0.047	0.026	-0.021^{*}
Black	0.022	0.022	<0.001	0.077	0.080	0.003
No. of Observations	2,495	6,307		379	649	

Notes: Standard deviations are in parentheses.

Source: Freshmen Tracking System, Office of Institutional Analytics, University of New Mexico.

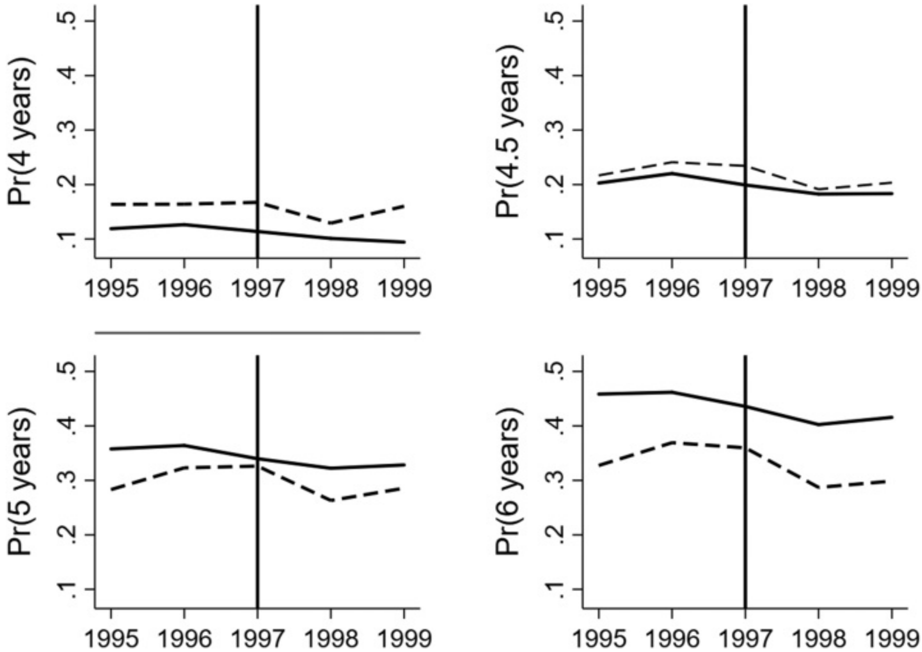
***Significant at the 1% level; **significant at the 5% level; *significant at the 10% level.

matriculating students at UNM. For these students, we assigned them the mean high school GPA of 3.28.

We concentrate on the years 1995 to 1999, bounding the policy change by two years before and three years after implementation. These years encompass the largest economic expansion in the United States since World War II. During this period, labor market conditions in New Mexico were gradually tightening but remained relatively stable, so we need not worry much that broad economic conditions are driving the results. In 1995, admissions requirements at UNM were raised, thus 1995 represents the earliest cohort we can reasonably include in the analysis. In 2000, the First-Year Learning Communities program was launched, aiming to increase freshman retention by providing additional training on adjusting to life in college. Although launched in 2000, the program began in earnest in 2001. It is possible this program may have differentially impacted resident and nonresident students. For this reason, 1999 is the latest cohort appropriate to include in the analysis. As we show in section 6, results are similar when we expand the sample period to include students in the 2000 cohort. To our knowledge, there were no other concurrent policy changes at the high school or postsecondary level in New Mexico over the 1995–99 period that would have differentially impacted enrollment and/or completion for residents and nonresidents.

In our preferred specification, we compare recent high school graduates from New Mexico (who are NMLLS eligible) with those from out of state (who are not eligible, but who experience the same campus environment), while excluding foreign students.

Table 2 compares summary statistics for resident and nonresident students before and after the implementation of the NMLLS. It appears that the composition of



Notes: The plots show the likelihood of degree completion for incoming cohorts over the period 1995 to 1999. Solid lines represent resident students and dashed lines represent nonresident students. The vertical bars at 1997 mark the implementation of the New Mexico Legislative Lottery Scholarship.

Figure 1. Pre-Post Trends in the Probability of Graduating, By Residency

these groups changed across pre- and post-treatment periods. In years before the implementation of the NMLLS, resident students had higher high school GPAs and ACT composite scores compared with years following the implementation of the NMLLS. Moreover, students matriculating after implementation were more likely to take remedial coursework at UNM. These changes are statistically significant, suggesting that the NMLLS may have induced students with weaker academic preparation to enroll at UNM. Table 2 also shows that residents were less likely to come from lower income families following implementation of the NMLLS, another indication of a compositional effect. The academic achievement of nonresident students did not change significantly following implementation of the NMLLS, according to high school GPA and composite ACT scores.

Although several statistically significant differences exist between resident and nonresident students in terms of high school GPA, composite ACT scores, remedial coursework, family income, race, and ethnicity, this does not necessarily threaten the validity of our difference-in-differences model of completion if the common trends assumption holds. The identifying assumption of the difference-in-differences model is that pre-treatment trends in the outcome variable be similar in trajectory across the treatment and control groups. As a visual check of this identifying assumption, figure 1 presents pre-treatment trends in graduation rates for residents and nonresidents between 1995 and 1999. We are particularly interested in six-year graduation rate trends, a standard measure of completion.[10] Visual inspection supports the validity of a difference-in-differences identification strategy examining six-year graduation rates. A more rigorous method of testing the common trends assumption is presented in Autor (2003). Following this strategy, we specify a flexible difference-in-differences model by interacting the resident dummy variable with cohort dummy variables, producing a model allowing for treatment at different time periods. This model can be expressed as

$$Prob\,(Graduate_{ist}) = \gamma_s + \lambda_t + \sum_{j=-m}^{q} \beta_j D_{st}\left(t = k + j\right) + X_{ist}\delta + \varepsilon_{ist}, \tag{3}$$

where i denotes the student, s denotes residency status, and t denotes cohort year. The variable D_{st} is the binary treatment indicator and k is the year the treatment started ($k =$ 1997 in our case). X_{ist} contains controls for race, ethnicity, gender, family income, remedial coursework in college, high school GPA, and standardized test scores. In equation 3, m and q are the number of leads and lags of the treatment effect included. We include two leads and two lags in our tests, defining 1999 as the reference cohort.

Testing the common trends assumption using equation 3 requires examining whether

$$\beta_j = 0\ \forall j < 0. \tag{4}$$

In other words, the common trends assumption holds when the coefficients on all leads of the treatment are zero. This specification also has the advantage of informing whether estimated treatment effects occur in multiple post-treatment time periods, fade away with time, or remain constant, for example. Tests are conducted using ordinary least squares and are presented in Appendix A (which is available in a separate online appendix that can be accessed on *Education Finance and Policy*'s Web site at www.mitpressjournals.org/doi/suppl/10.1162/edfp_a_00270). Results provide evidence that the common trends assumption holds, as estimated coefficients on all leads are not statistically different from zero.

Our data include 8,802 resident students, 6,307 of whom enrolled during the post-NMLLS period and were eligible for the Bridge to Success Scholarship. Of these, 2,753 met cumulative GPA and credit attainment requirements to begin the NMLLS in their second semester. Table 3 documents the number of students who maintain the scholarship in the second through ninth semester. It is apparent that scholarship loss was quite common. Of the 2,753 students who qualified for the NMLLS, approximately 70 percent were still eligible for the NMLLS going into their third year.

5. EMPIRICAL MODEL

We conduct difference-in-differences matching estimation on the propensity score to mitigate any observable differences between resident and nonresident students. Our approach uses kernel matching, a one-to-many matching technique that assigns larger

10. For degree-earning students entering UNM between 1994 and 1999, average time to degree was 4.79 years, with a standard deviation of 0.66 years.

Table 3. New Mexico Legislative Lottery Scholarship (NMLLS) Student Attrition, 1997–99

Semester	Residents Eligible	Percent Remaining
2	2,753	100
3	2,324	84.4
4	2,141	77.8
5	1,926	70.0
6	1,792	65.1
7	1,680	61.0
8	1,617	58.7
9	1,558	56.6

Notes: We consider the sample of resident students who met cumulative GPA and credit requirements in their first semester to qualify for the NMLLS.

Source: Office of Institutional Analytics, University of New Mexico.

weights to control units closer in propensity score. The general form of the matching estimator is given by

$$\Delta^{DDME} = \frac{1}{n_{1t}} \sum_{i \in I_{1t} \cap S_p} \left\{ Y_{1ti} - \sum_{j \in I_{0t} \cap S_p} W\left(i, j\right) Y_{0tj} \right\} - \frac{1}{n_{1t'}} \sum_{i \in I_{1t'} \cap S_p} \left\{ Y_{1t'i} - \sum_{j \in I_{0t'} \cap S_p} W\left(i, j\right) Y_{0t'j} \right\}, \quad (5)$$

where n_{1t}, $n_{1t'}$ are the number of treated cases before and after the inception of the NMLLS, S_p is the common support region, and I_{0t}, $I_{0t'}$, I_{1t}, $I_{1t'}$ are the resident and nonresident groups before and after the NMLLS. Graduation rates for resident and nonresident students are given by Y_{1t}, Y_{0t}, $Y_{1t'}$, $Y_{0t'}$. The function $w(i, j)$ denotes the weight given to jth case, where $\sum_j w(i, j) = 1$ and $0 < w(i, j) < 1$. The weighting function $w(i, j)$ is given by

$$w\left(i, j\right) = \frac{K\left[\hat{l}\left(x_j\right) - \hat{l}\left(x_i\right)\right]}{\sum_{j \in I_{0t} \cap S_p} K\left[\hat{l}\left(x_j\right) - \hat{l}\left(x_i\right)\right]}, \quad (6)$$

where K is the Epanechnikov kernel function and $\hat{l}(\cdot) \equiv \ln(\frac{\hat{p}(\cdot)}{1-\hat{p}(\cdot)})$ is the fitted linearized propensity score from a logistic regression model estimated by maximum likelihood. We use linearized propensity scores as they are more likely to have a distribution that is approximately normal. Treatment effects, Δ^{DDME}, are calculated using kernel-weighted least squares according to equation 6. Robust standard errors are reported. The propensity score model includes all covariates in levels, as well as several quadratic terms.[11]

11. We conduct a sequential search for quadratic terms to include in the propensity score model. We start by estimating logistic models that include all terms in levels and one of all possible quadratic terms. We then calculate the likelihood ratio statistic for the null hypothesis that the most recently added quadratic term has a coefficient of zero. We select for inclusion the quadratic term with the highest test statistic over 2.71, corresponding

Table 4. Estimated Parameters for Propensity Score Model of New Mexico Legislative Lottery Scholarship (NMLLS) Data, 1995–99

Variable	Estimate	Standard Error
HSGPA	0.612	0.947
ACT	0.251**	0.104
Remedial	−0.972	0.785
Income < $20K	0.089	0.169
Income < $40K	0.210*	0.119
Female	1.346***	0.416
Hispanic	0.793***	0.124
Native American	−0.221	0.180
Asian	−2.300*	1.203
Black	−5.732***	0.974
Declined to state race-ethnicity	−0.084	0.292
ACT*Black	0.210***	0.048
Female*White	−0.339**	0.156
Act^2	−0.011***	0.002
ACT*Female	−0.049***	0.018
HSGPA^2	−0.424***	0.139
Female*Black	0.571*	0.317
Act*HSGPA	0.089***	0.026
HSGPA*Remedial	0.551**	0.253
Remedial*Asian	1.118**	0.532
HSGPA*Asian	0.667*	0.349
Constant	−2.346	2.110
Observations		9,830

Notes: Standard errors are in parentheses. Propensity scores are estimated using a logistic model. The variable *Declined to state race-ethnicity* is equal to one if the student declined to state their race-ethnicity, and zero otherwise. HSGPA = high school grade point average.

***Significant at the 1% level; **significant at the 5% level; *significant at the 10% level.

Results of the propensity score model are presented in table 4. It is important to note that although the propensity score model may seem awkward in that it predicts the immutable condition of being a New Mexico resident, it is not essential that the propensity score model have a meaningful interpretation. Instead, the validity of the propensity score model rests on how well it balances covariates across treatment and control groups (Imbens and Rubin 2015; Imbens 2015).

Having a small group of nonresident students relative to resident students has implications for our estimates. To increase the precision of our estimated treatment effects, and to avoid imposing functional form where possible, we choose to conduct kernel density matching.[12] This method has the advantage of lower variance because more information is used. On the other hand, it may result in an increase in bias due

to a z-statistic of 1.645. We then add this covariate to the "baseline" model and repeat this process until all the remaining likelihood ratio statistics are below the threshold of 2.71.

12. There are 8,802 resident students and only 1,028 nonresident students in the sample. One-to-many matching allows us to proceed without a significant loss in information. For example, if we were to conduct a simple nearest-neighbor matching procedure without replacement, estimates would (at most) be based on 1,028 matches, or 2,056 observations, which constitute approximately 21 percent of the sample.

to the potential for considering "bad" matches. Although the further the observations are in terms of propensity score, the less weight is given to the potential bad match, this makes adequate overlap a necessary condition for the validity of this method.

In our analysis, we limit matching to those individuals whose propensity scores lie in the common support region, which is over 99.7 percent of the original sample. We do not trim observations from the analysis. As a sensitivity analysis, we estimate effects using various fixed bandwidths, *h*, for the kernel function. Importantly, the choice of bandwidth also involves a bias-variance trade-off. Smaller bandwidths consider a smaller portion of the pool of control observations, and thus use less information, which tends to reduce bias (from being less likely to consider poor matches) while increasing sampling variance. In order to assess the effectiveness of the matching procedure, several tests are conducted following Imbens and Rubin (2015), although they are modified for difference-in-differences matching with repeated cross sections. An explanation of these tests and their results are presented in Appendix B (available in the online appendix).

A power analysis in online Appendix C shows that most models we estimate have sufficient power to detect a 5 percent change in completion rates at the 5 percent significance level. Models limiting the sample to students from low-income families are substantially underpowered, however. The reader should thus exercise caution in interpreting results when the sample is limited in this way. Models limited to less academically prepared students also fail to meet the accepted standard of 80 percent power. Underpowered regressions are less likely to detect meaningful program effects, even if they do exist. Evidence of low power is seen in results for students from low-income families—there exist several coefficients large in magnitude that do not achieve statistical significance. Although meaningful completion effects may exist in these cases, all we can conclude is that we cannot reject the null hypothesis that the coefficient is not statistically different from zero.

In addition to estimating the overall effect of the NMLLS, we are also interested in whether program effects differ depending on academic preparation. We explore this possibility by estimating separate models on students above and below the mean high school GPA.[13] We disaggregate further by family income to examine program effects for students whose financial constraints are more likely binding. Robustness checks using the 1995–2000 cohorts and various smoothing parameters are discussed in section 6. Additionally, we estimate models of cumulative credits earned and time to degree to examine whether apparent completion effects are driven by changes in student course-taking behavior. Lastly, we explore regression discontinuity design in estimating completion effects of the NMLLS.

6. RESULTS

Means and normalized differences after kernel matching are presented in table 5. Comparing means before and after the NMLLS, it appears that the matching algorithm performed well in balancing covariates. Normalized differences for pre- and post-NMLLS periods are near zero, with the largest normalized difference (−0.122) far below one-quarter of a standard deviation unit in absolute value. We produce these statistics by

13. Results are similar when we split the sample around the median high school GPA.

Table 5. Means and Normalized Differences after Kernel Matching, Full Sample, 1995–99

	Pre-NMLLS			Post-NMLLS		
Variable	Resident	Nonresident	ND	Resident	Nonresident	ND
HSGPA	3.34	3.28	0.114	3.27	3.33	−0.122
Composite ACT	22.70	22.36	0.085	22.18	22.32	−0.033
Remedial	0.24	0.22	0.051	0.29	0.30	−0.030
Income < $40,000	0.22	0.21	0.024	0.20	0.22	−0.055
Female	0.56	0.54	0.048	0.56	0.59	−0.045
Hispanic	0.39	0.41	−0.031	0.38	0.36	0.029
Native American	0.04	0.05	−0.035	0.04	0.05	−0.022
Asian	0.04	0.04	0.012	0.04	0.03	0.056
Black	0.02	0.03	−0.031	0.02	0.02	0.017

Notes: Means are from Epanechnikov kernel matching performed with a bandwidth of $h = 0.2$. Normalized differences (ND) are calculated by taking the difference average covariate values by residency status and dividing by a measure of standard deviation. NMLLS = New Mexico Legislative Lottery Scholarship.

Table 6. New Mexico Legislative Lottery Scholarship Graduation Effects by Years since First Enrollment, Kernel Matching, 1995–99

		Graduation Rates by Years Since First Enrollment			
Group	Observations	4	4.5	5	6
Full sample	9,804	−0.035 (0.026)	−0.028 (0.030)	−0.030 (0.036)	−0.014 (0.039)
$\bar{Y}$		0.109	0.195	0.340	0.431
GPA ≤ 3.28	4,780	−0.016 (0.027)	−0.033 (0.038)	−0.100* (0.053)	−0.116** (0.055)
$\bar{Y}$		0.042	0.098	0.217	0.299
GPA > 3.28	5,019	−0.014 (0.041)	0.025 (0.047)	0.065 (0.052)	0.100* (0.053)
$\bar{Y}$		0.174	0.289	0.459	0.559
GPA > 3.78	1,886	0.035 (0.081)	0.102 (0.083)	0.083 (0.088)	0.096 (0.089)
$\bar{Y}$		0.264	0.393	0.573	0.669

Notes: Robust standard errors are reported in parentheses. Estimates are from difference-in-differences kernel matching performed with a bandwidth of $h = 0.2$ using the Epanechnikov kernel function. We report estimates for students with below average or average high school grade point averages (GPAs) (≤3.28), above average high school GPAs (>3.28), and high school GPAs greater than one standard deviation above the mean (>3.78). $\bar{Y}$ denotes resident baseline graduation rate by high school performance and years since first enrollment.

**Significant at the 5% level; *significant at the 10% level.

academic preparation as well, finding a similar pattern, although differences were slightly higher when considering students more than one standard deviation above the mean high school GPA. Overall, normalized differences suggest excellent balance in covariates following kernel matching.

Table 6 presents results of the difference-in-differences kernel matching estimation. Note that we find little evidence of an overall program effect. Point estimates are near zero and do not approach statistical significance. These estimates, however, mask large program responses that appear when we divide the sample by academic preparation. Considering students with below-average high school GPA, we find a negative

Table 7. New Mexico Legislative Lottery Scholarship Graduation Effects, Kernel Matching, Family Income < \$40,000, 1995–99

		Graduation Rates by Years Since First Enrollment			
Group	Observations	4	4.5	5	6
Full sample	2,003	−0.032 (0.038)	−0.034 (0.047)	−0.060 (0.063)	−0.049 (0.077)
$\bar{Y}$		0.084	0.145	0.267	0.368
GPA ≤ 3.28	978	−0.007 (0.014)	−0.044 (0.039)	−0.202** (0.079)	−0.203** (0.088)
$\bar{Y}$		0.034	0.074	0.156	0.237
GPA > 3.28	1,018	0.026 (0.087)	0.066 (0.098)	0.207 (0.118)	0.234* (0.136)
$\bar{Y}$		0.132	0.213	0.374	0.494
GPA > 3.78	361	−0.078 (0.192)	0.009 (0.215)	0.130 (0.270)	−0.138 (0.279)
$\bar{Y}$		0.197	0.290	0.472	0.606

Notes: Robust standard errors are reported in parentheses. Estimates are from difference-in-differences kernel matching performed with a bandwidth of $h = 0.2$ using the Epanechnikov kernel function. We report estimates for students with below average or average high school grade point averages (GPAs) (≤3.28), above average high school GPAs (>3.28), and high school GPAs greater than one standard deviation above the mean (>3.78). $\bar{Y}$ denotes resident baseline graduation rate by high school performance and years since first enrollment.

**Significant at the 5% level; *significant at the 10% level.

completion effect for six-year graduation of 11.6 percentage points (38.8 percent). Students with an above average high school GPA are 10 percentage points (17.9 percent) more likely to graduate within six years compared with similar nonresident students. Effects are significant at the 5 and 10 percent levels, respectively. Thus, although we are certain of a negative completion effect for less academically prepared recipients, we remain cautious in concluding a significant positive completion effect for more academically prepared recipients. The NMLLS did not have a meaningful impact on the likelihood of graduating within six years for the most academically prepared students.

Table 7 presents results of the matching estimation performed on low-income students, defined as those coming from households with less than \$40,000 in annual income. We again find little evidence of completion effects in the aggregate, but see meaningful effects when disaggregating by student ability. For low-income, low-achieving students, we estimate a large decrease in completion within six years. For low-income, higher-achieving students, we find a large increase in completion within six years, significant at the 10 percent level. As shown in table 8, we find no significant completion effects for students from higher income households. It appears that our results at higher levels of aggregation may be driven by students from families where financial constraints are binding.

Alternative Bandwidths, Cohorts, and Control Groups

We test results for sensitivity to the choice of the smoothing parameter, or bandwidth, in our kernel matching algorithm. Specifically, we estimate models using bandwidths $h = \{0.1, 0.2, 0.3\}$ with the sample stratified by academic preparedness and family income, as in our main results. These results are presented in online Appendix D. Table D.1

Table 8. New Mexico Legislative Lottery Scholarship Graduation Effects, Kernel Matching, Family Income ≥ $40,000, 1995–99

		Graduation Rates by Years Since First Enrollment			
Group	Observations	4	4.5	5	6
Full sample	7,792	−0.039 (0.030)	−0.034 (0.035)	−0.031 (0.040)	−0.016 (0.042)
$\bar{Y}$		0.116	0.208	0.359	0.448
GPA ≤ 3.28	3,791	−0.017 (0.036)	−0.045 (0.042)	−0.052 (0.052)	−0.069 (0.056)
$\bar{Y}$		0.044	0.104	0.233	0.315
GPA > 3.28	3,997	−0.022 (0.045)	0.012 (0.051)	0.017 (0.056)	0.045 (0.057)
$\bar{Y}$		0.185	0.310	0.482	0.576
GPA > 3.78	1,525	0.049 (0.083)	0.119 (0.086)	0.090 (0.091)	0.137 (0.098)
$\bar{Y}$		0.281	0.418	0.598	0.685

Notes: Robust standard errors are reported in parentheses. Estimates are from difference-in-differences kernel matching performed with a bandwidth of $h = 0.2$ using the Epanechnikov kernel function. We report estimates for students with below average or average high school grade point averages (GPAs) (≤3.28), above average high school GPAs (>3.28), and high school GPAs greater than one standard deviation above the mean (>3.78). $\bar{Y}$ denotes resident baseline graduation rate by high school performance and years since first enrollment.

presents results for our estimates by graduation semester and academic preparation. Overall, we find a similar pattern relative to our preferred specification. That is, we find no evidence of completion effects in the aggregate but find a positive relationship between academic preparation and degree completion. Tables D.2 and D.3 also broadly agree with the results of our preferred specification: Significant program effects are confined to those from families with lower incomes—the same divergent effects are detected, but effect sizes are significantly larger in absolute value.

As was mentioned earlier, we do not include cohorts beyond 1999 in our preferred specification due to the launch of the First-Year Learning Community program in 2000. However, anecdotal evidence suggests this program did not begin in earnest until 2001. For this reason, Appendix E (available in the online appendix), compares results from 1995–1999 cohorts and 1995–2000 cohorts. Tables E.1 through E.3 reveal a similar pattern of program effects compared to our preferred specification.[14]

Although not reported, we estimate simple pre–post models of completion using qualified UNM resident students before the implementation of the NMLLS as the control group. Estimates are produced via logistic regression, where the coefficient of interest is on a dummy variable equal to one in years when the NMLLS was in

14. We also used New Mexico residents who delayed enrollment (and were therefore not eligible for NMLLS) as a control group. These nontraditional students are likely to differ in unobservable ways from students who entered college right away, especially given the large tuition penalty for delaying enrollment once the NMLLS was in place. A student who missed out on the scholarship by delaying enrollment might have less maturity or some difficulty to overcome before starting college, characteristics that would also make completion less likely. Indeed, in models that use nontraditional students as the control group, we find unrealistically large program effects of 27.4 percentage points (76.3 percent) overall for students from lower income families. Program effects for high-achieving, low-income students are estimated to be 46 percentage points (93.9 percent). These effects likely tell us more about the negative chances of students who were unable to enroll in college right away, than the positive impact on graduation.

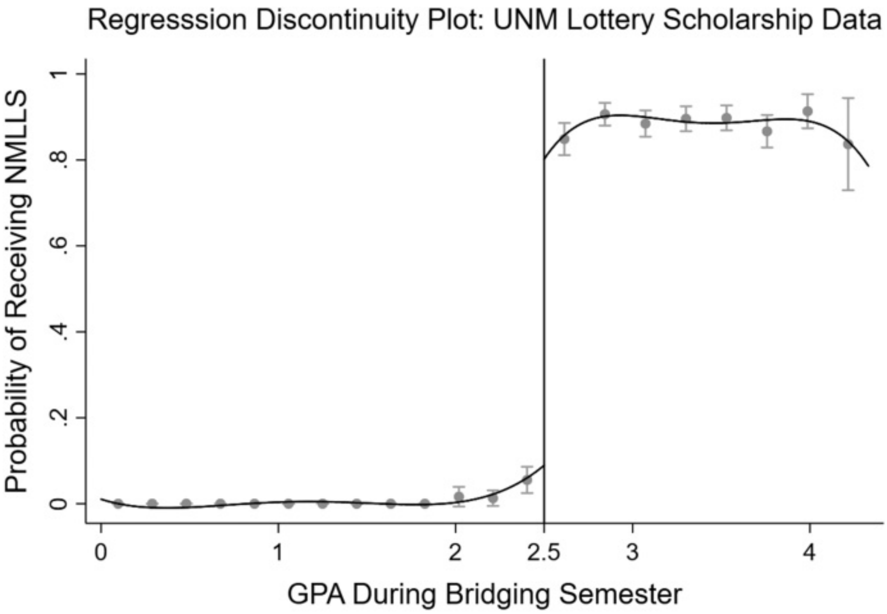
place, and zero otherwise. We assume model errors are independent across cohorts, yet correlated within cohorts, thus standard errors are clustered at the cohort level. Because these models do not account for any trends over time, they are limited in this respect. However, these models do not rely on nonresident students as the control group and thus provide insight into whether estimated program effects are an artifact of the data. Results of simple pre–post models reveal a significant 1.6 percentage point (3.6 percent) decline in completion rates overall, with a 3.3 percentage point (10.4 percent) decline for low-achieving high school students. This offers evidence that the negative completion effects we estimate for some NMLLS recipients in preferred specifications are not due to model misspecification.

Regression Discontinuity Design Approach

We explore exploiting eligibility requirements of the NMLLS to estimate whether the program had any meaningful effect on degree completion at UNM. Recall that students are eligible for the NMLLS if they are a New Mexico resident, have lived in the state for at least one year, graduated from high school or earned their GED in New Mexico, immediately enroll in a qualified public institution by the next fall semester, and meet credit hour and college GPA requirements during the bridging semester. The NMLLS requires that students complete at least 12 credit hours during the bridging semester with a minimum 2.5 GPA. Accordingly, we limit the sample to all students who meet the NMLLS eligibility requirements, apart from the bridging semester GPA requirement, and compare students just above the 2.5 threshold to students just below. We find the regression discontinuity approach appealing because it is simple, objective, and requires little information. It is also relatively straightforward to verify with visual checks, easy to interpret estimates, and easy to perform falsification tests. In sum, in many ways it is a cleaner approach than difference-in-differences matching estimation.

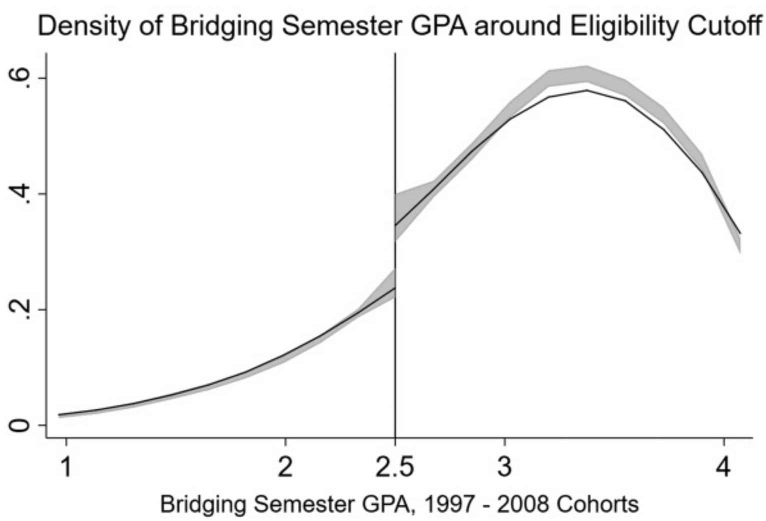
Because participation in the NMLLS is not strictly a deterministic function of college GPA (i.e., other funding sources such as academic or athletic scholarships are prioritized above NMLLS funds, for example), we appeal to a fuzzy regression discontinuity approach using a sample of resident students from 1997–2008. This approach only requires that there is a significant jump in the probability of treatment assignment above the cutoff of the running variable, bridging semester GPA in our case. In figure 2, we visually inspect the jump in the probability of NMLLS funding by bridging semester GPA. The jump between the quadratic fitted lines below and above the threshold is below one (approximately 80 percent), so fuzzy regression discontinuity seems appropriate in this context.

However, we fail to pass a critical identification test for regression discontinuity studies—that is, cutoff manipulation. Figure 3 plots the density of the running variable, here the bridging semester GPA. Ninety-five percent confidence intervals are shown in gray. As is evident, there is a statistically significant discontinuity in the density of the running variable around the NMLLS eligibility cutoff. It appears that some students manipulate this eligibility cutoff by perhaps taking easier courses or dropping courses when a poor grade is expected. Because regression discontinuity may be thought of as random assignment in the neighborhood of the cutoff, this provides evidence of students nonrandomly selecting into treatment and control groups. Table 9



Notes: Points depict the within-bin sample average of New Mexico Legislative Lottery Scholarship receipt probability by bridging semester GPA. A quadratic fit has been added below and above the cutoff at 2.5. The number of bins is calculated using the mimicking-variance evenly spaced method using spacing estimators. The uniform kernel is used to construct the global polynomial estimators. The plot provides visual evidence of the appropriateness of a fuzzy regression discontinuity approach.

Figure 2. Jump in Treatment Probability Around the Bridging Semester Grade Point Average (GPA) Cutoff



Notes: Here we present the density of the running variable (bridging semester GPA) around the NMLLS eligibility cutoff with 95 percent confidence intervals shown in gray. Cohorts between 1997 and 2008 are included. This plot was constructed using a local cubic approximation. The uniform kernel is used to construct the global polynomial estimators. The plot reveals a statistically significant discontinuity in the running variable density around the eligibility cutoff.

Figure 3. Bridging Semester Density Around the New Mexico Legislative Lottery Scholarship (NMLLS) Grade Point Average (GPA) cutoff, 1997–2008

Table 9. Testing for Manipulation of the Bridging Semester Grade Point Average (GPA) Cutoff for New Mexico Legislative Lottery Scholarship (NMLLS) Eligibility, 1997–2008 Cohorts

	Bandwidths		Effective Observations		Conventional Test		Robust Test	
$h_- \neq h_+$	Left	Right	Left	Right	T	p-value	T	p-value
$T_2(h_1)$	0.511	0.524	1,748	4,591	5.263	<0.001	3.381	<0.001
$T_3(h_2)$	0.489	0.445	1,591	3,806	4.109	<0.001	4.245	<0.001
$T_4(h_3)$	0.653	0.616	1,977	5,433	3.096	0.002	4.895	<0.001
$h_- = h_+$								
$T_2(h_1)$	0.524	0.524	1,750	4,591	5.270	<0.001	3.822	<0.001
$T_3(h_2)$	0.445	0.445	1,569	3,806	4.090	<0.001	4.564	<0.001
$T_4(h_3)$	0.616	0.616	1,903	5,433	3.095	0.002	5.193	<0.001

Notes: Here we report results from manipulation tests following Cattaneo, Jansson, and Ma (2017) examining 1997–2008 cohorts at University of New Mexico. The subscript on T denotes the order of the local polynomial used to construct the bias-corrected density point estimators. The subscript on h denotes the order of the local polynomial used to construct the density point estimates. A uniform kernel was used to construct local polynomial estimators. We perform tests with identical and different data-driven bandwidths. Conventional and robust test statistics examine the null hypothesis of continuity in the bridging semester GPA around the NMLLS eligibility cutoff.

presents results of formal manipulation tests using local polynomial density estimators following Cattaneo, Jansson, and Ma (2017). The null hypothesis of these tests is continuity in the density of the running variable around the bridging semester GPA cutoff. We strongly reject the null hypothesis of density continuity around the GPA cutoff under varying assumptions.

7. OTHER POSSIBLE EXPLANATIONS FOR THE RESULTS WE FIND

Program Anticipation

If there were anticipatory effects of the NMLLS, this would violate identifying assumptions of the difference-in-differences estimator and would lead to biased results. The passage of the lottery scholarship occurred in March 1996 and the policy was instituted approximately seventeen months later, giving New Mexico students and families two semesters to anticipate the policy change and modify their behavior. This could have resulted in some students taking easier high school course loads to ensure high school graduation and ultimately NMLLS eligibility. Such students would be less prepared for higher education than their peers but would still be NMLLS-eligible. This narrative is consistent with the proposal that the NMLLS incentivized marginal students to enroll at UNM who may have otherwise not enrolled in college or attended a two-year college instead.

Considering whether out-of-state families acted on the anticipated policy, the time between passage and implementation likely did not afford a long enough window to move to New Mexico and establish program eligibility (at least for the inaugural year) due to (1) the requirement of living in New Mexico for at least one year and (2) the high costs associated with moving to another state, especially with a student currently attending a high school outside of New Mexico. In either case, we do not detect any indication of anticipatory effects evidenced by results from flexible difference-in-differences models in online Appendix A.

Confounding Factors

A massive increase in enrollment at UNM accompanied the NMLLS (2,495 to 6,307 resident students). One possible confounding factor is an increase in wealth. Lovenheim and Reynolds (2013) show that greater housing wealth both increases the likelihood of enrollment at public flagship universities relative to non-flagship schools and directly increases the likelihood of college completion. The authors find a $10,000 increase in real housing wealth increases the relative likelihood of attending a flagship university by 2 percent, and the overall completion likelihood by 1.8 percent. Simple accounting reveals that this is not likely a significant driver of our results. Over the study period real housing prices in New Mexico increased by a scant 0.5 percent.[15] If we assume a (very) conservative median home price of $215,000 in 1994, this only translates into approximate 0.2 percent increases in both relative flagship enrollment and overall college completion likelihood.[16] Moreover, real personal income only increased by 5.5 percent over the same period, an annualized growth rate just above 1 percent per year, so it is unlikely that any broad increase in overall wealth drove increases in resident enrollment after the launch of the NMLLS.[17] We also consider labor market conditions as a potential confounding factor. As we mention above, because our sample period spans the longest continuous period of economic growth in the United States since WWII, broad economic conditions are unlikely to be driving the enrollment effect of the NMLLS. Further, because UNM is a de facto open enrollment institution, changes in selectivity are not likely to confound the analysis (Binder and Ganderton 2004).

Our difference-in-differences matching strategy controls for selection bias stemming from both initial observable differences between residents and nonresidents, as well as observable changes in the composition of these groups over time. Consider four groups in the analysis as defined by residency status and whether the NMLLS was in place. Selection bias from these two sources is addressed by propensity score weighting so that groups closely resemble residents in the post-NMLLS period in terms of observable characteristics. In other words, both residents and nonresidents in the pre-NMLLS period, as well as nonresidents in the post-NMLLS period, observed to be considerably different from residents in the post-NMLLS period, are given a smaller weight. Residents in the post-NMLLS period are assigned a weight of one.[18] Propensity score weighting does not account for any changes in unobservable characteristics over time, which is a limitation of this, and any study that relies on repeated cross-sections.

Congestion

Another possible explanation for the patterns we see are capacity constraints and congestion at UNM. If the large increase in NMLLS-qualified students forced the university

15. See https://fred.stlouisfed.org/series/ATNHPIUS10740Q. July 1995 to July 1999 is examined. Prices are adjusted for inflation using Bureau of Labor Statistics' Consumer Price Index less shelter measure (Series CUUR0000SA0L2, available from https://data.bls.gov/cgi-bin/srgate).
16. The $215,000 figure is the median listed home price in Albuquerque according to Zillow.com as of 30 June 2017 (obtained from https://www.zillow.com/albuquerque-nm/home-values/).
17. Calculations using Bureau of Economic Analysis' annual personal income estimates for 1994 and 1999 (see https://www.bea.gov/itable/) and the Bureau of Labor Statistics' annual CPI-U estimates for 1994 and 1999 (see https://data.bls.gov/cgi-bin/srgate).
18. See Stuart et al. (2014) for a thorough treatment of combining difference-in-differences models with propensity score weighting.

to admit fewer nonresidents and be more selective in their criteria, then we would expect higher "quality" nonresident cohorts post-NMLLS. This would result in model results being biased downwards. However, according to university officials, in 1996, the year prior to the implementation of the NMLLS, the university was at approximately 50 percent capacity. Accordingly, they did not experience any "bottlenecks" in terms of class size, advising capacity, waitlists for classes, and so on, after the lottery scholarship launched.[19] If there was congestion at UNM post-NMLLS, this would likely increase students' time to degree, which we find no evidence to support.

Another point that merits mention is the funding mechanism under which New Mexico public institutions of higher education operate. New Mexico universities are funded using an enrollment formula. That is, the more students are enrolled, the more state dollars are allocated to the institution. This provides an incentive for institutions to compete with one another on the basis of enrollment. There is no de jure limit on the number of additional students UNM may enroll in each semester, so it is likely that the university simply absorbed this additional enrollment without crowding out other groups, such as nonresidents and low-income students.

Student Course-Taking Behavior

Because the incentive to graduate in nine semesters is countered by the incentive to maintain a 2.5 GPA under the NMLLS, we are concerned that students may have responded to the NMLLS by altering their course-taking behavior. Specifically, one might expect the NMLLS to incentivize students to take fewer credits in order to increase their likelihood of continued eligibility. If this is the case, then estimates may reflect a lengthening of time to degree, not necessarily lower completion rates on behalf of less academically prepared students. Because of this concern, we construct difference-in-differences matching estimates of cumulative credits earned since enrollment. These estimates are presented in online Appendix F.

Estimates in Appendix F present the effects of the NMLLS on cumulative credits earned using the same matching procedure as models of college completion. Table F.1 provides no evidence of a change in credits attempted overall. Significant positive course-taking effects are detected for academically well prepared students. Notably, while effects display the expected negative sign for less academically prepared students at UNM, they are not statistically different from zero. We find evidence that the NMLLS incentivized better-prepared students to take more credits, where effects range from approximately 1.6 percentage points (6.5 percent) after the first year to 16.4 percentage points (16.1 percent) at the six-year mark. We find these effects to be largely driven by students from low-income families. These results refute the notion that the NMLLS resulted in marginally prepared students completing degrees at a slower pace. We also directly test this hypothesis by estimating difference-in-differences matching estimates using semesters to graduation as the outcome. We find no evidence of any change in time to degree associated with the NMLLS program.

19. Interview with Dr. Terry Babbitt, Vice President of the Enrollment Management Division, conducted 19 April 2017.

8. CONCLUSIONS

We examine the effect of an exceptionally generous and low-bar merit-based scholarship on college completion. We estimate variants of the difference-in-differences model using qualified resident students as the treatment group and a matched sample of ineligible nonresident students as the control group. The common trends assumption is supported both visually and empirically. The sample is stratified by academic preparation and family income to see which, if any, subgroups are driving completion effects. We conduct kernel matching and examine its success through rigorous testing. A flexible difference-in-differences model is estimated to verify that program effects are limited to treatment years. Sensitivity to cohorts included, as well as the smoothing parameter used in the matching alogirthm, are reported. We also estimate models of credit accumulation and time to degree completion, in addition to exploring the validity of a regression discontinuity approach in estimating completion effects of the NMLLS.

Our analysis reveals a divergent effect of the NMLLS: More academically prepared high school students are more likely to graduate in six years than their nonresident counterparts, whereas the opposite is true for less academically prepared recipients of the NMLLS. These countervailing results mask completion effects of the NMLLS in the aggregate. We find positive completion effects for those with above average high school GPA similar in magnitude to those in the literature, and negative effects for lower-achieving scholarship recipients, consistent with discouragement from raising expectations for marginal students who otherwise would not have attended college, or at least a four-year college.

Results suggest low-income, high-achieving high school students benefit from the NMLLS, and lower-achieving students do not. The latter may be explained by overmatching at UNM, where marginally prepared students who would have otherwise chosen to pursue an easier course of study at another institution, or not attend college at all, attend the state's flagship university because the scholarship renders it more affordable. Discouragement may also play a role. Students who lose the scholarship may expect higher costs and a lower likelihood of completion, and so may be more likely to drop out, than nonresidents with similar academic performance.

The main conclusion we draw from our analysis is that setting the bar too low in terms of merit aid may be detrimental to the success of the least academically prepared students. The promise of generous financial aid tied to seemingly modest academic criteria may actually worsen college persistence for students with weaker academic preparation. When price signals in the market for higher education are removed, as is the case with the NMLLS, many students may choose to attend institutions for which they are a poor match (i.e., are less academically prepared than their peers).

Since its inception in 1997, the NMLLS has seen significant changes. Starting in the 2014–15 academic year, the scholarship was capped at seven semesters (plus the initial bridging semester) and initial and renewal credit requirements were increased from 12 to 15 credits earned per semester. A statewide budget crisis in 2017 resulted in the legislature making major cuts to the NMLLS—whereas the scholarship paid 100 percent of tuition over our study period, the program only covers approximately 60 percent of tuition as of the 2017–18 academic year. The 2017 Regular Session saw the passage of SB 420, which allows students to take a "gap" year after high school and still remain eligible for the NMLLS. It is not clear how recent program changes will affect student course-taking and persistence at UNM. A decline in scholarship generosity will reduce access to higher education in New Mexico, but may be necessary given the constant financial pressure the Lottery Scholarship Fund faces. Raising the bar in terms of initial eligibility and renewal requirements sends a signal to high school students that they are expected to work harder than before, which may result in more efficient spending (i.e., less funding of marginally prepared students who ultimately drop out) and shorter time to degree. Allowing for a “gap” year is sure to increase program costs for an already financially troubled program.

Considering the poor financial health of the NMLLS, it may be time to narrow the program and prioritize funding for certain students. Our results suggest that completion effects may be driven by low-income families. Adding a family-income cap or some other type of need-based component would reduce overall program costs and target spending toward those students who seem most responsive to the NMLLS. A need-based component would also make the NMLLS more politically tenable, as it is often slated as a major regressive tax in New Mexico.

In general, further research is needed to investigate how to increase degree completion, not merely enrollment, while avoiding harming less academically prepared students. One potential remedy may be to pair lottery scholarship funds with stronger academic supports, such as additional mandatory advising, mid-semester check-ups, or an additional one-credit mandatory course on topics such as scholarship details and strategies for academically suriving the freshman year. Another potentially promising reform would be to tie program eligibility to high school performance rather than college performance. Having initial and renewal requirements tied to college performance provides incentives for undesired student behaviors, such as padding GPAs with easier coursework or taking fewer credits so as to increase the likelihood of continued scholarship eligibility.

Our results inform recent proposals to make college free for a large proportion of students, whether at the state or national level. Our findings suggest that such proposals, which effectively remove price differentials between public colleges, may distort students’ college choice decisions. In order to maximize the value of such scholarships, students may increasingly overmatch by choosing more prestigious public colleges for which they are underprepared. Future research would benefit from taking a closer look at subsamples of cohorts to examine whether a fuzzy regression discontinuity approach would be appropriate for intitial years of the scholarship—before other institutional changes took place.

ACKNOWLEDGMENTS

We thank the University of New Mexico’s Office of Institutional Analytics for making data available for this project. We are also indebted to Sandy Baum, Rodney Andrews, Alok Bohara; seminar participants at the University of New Mexico, APPAM, AEFP, and SEA; as well as anonymous referees for helpful comments. We are solely responsible for any errors in this analysis.

REFERENCES

Arcidiacono, Peter, Esteban M. Aucejo, and Joseph V. Hotz. 2016. University differences in the graduation of minorities in STEM fields: Evidence from California. *American Economic Review* 106(3): 525–562.

Arrow, Kenneth J. 1973. Higher education as a filter. *Journal of Public Economics* 2(3): 193–216.

Autor, David H. 2003. Outsourcing at will: The contribution of unjust dismissal doctrine to the growth of employment outsourcing. *Journal of Labor Economics* 21(1): 1–42.

Bettinger, Eric. 2004. How financial aid affects persistence. In *College choices: The economics of where to go, when to go, and how to pay for it*, edited by Caroline M. Hoxby, pp. 207–238. Chicago: University of Chicago Press.

Binder, Melissa, and Philip T. Ganderton. 2004. The New Mexico lottery scholarship: Does it help minority and low-income students? In *State merit scholarships and racial inequality*, edited by Donald E. Heller and Patricia Marin, pp. 101–122. Cambridge, MA: The Civil Rights Project, Harvard University.

Binder, Melissa, Philip T. Ganderton, and Kristin Hutchens. 2002. Incentive effects of New Mexico's merit-based state scholarship program: Who responds and how? In *Who should we help? The negative social consequences of merit scholarships*, edited by Donald. E. Heller and Patricia Marin, pp. 41–56. Cambridge, MA: The Civil Rights Project, Harvard University.

Bruce, Donald J., and Celeste K. Carruthers. 2014. Jackpot? The impact of lottery scholarships on enrollment in tennessee. *Journal of Urban Economics* 81:30–44.

Card, David. 1999. The causal effect of education on earnings. In *Handbook of labor economics*, Volume 3, Part A, edited by Orley Ashenfelter and David Card, pp. 1801–1863. Amsterdam: Elsevier Science.

Castleman, Benjamin L., and Bridget Terry Long. 2016. Looking beyond enrollment: The causal effect of need-based grants on college access, persistence, and graduation. *Journal of Labor Economics* 34(4): 1023–1073.

Cattaneo, Matias D., Michael Jansson, and Xinwei Ma. 2017. Simple local polynomial density estimators. Unpublished paper, University of California, Berkeley.

Cohodes, Sarah R., and Joshua S. Goodman. 2014. Merit aid, college quality, and college completion: Massachusetts' Adams scholarship as an in-kind subsidy. *American Economic Journal: Applied Economics* 6(4): 251–285.

College Board. 2015. *Trends in student aid* 2015. Available http://trends.collegeboard.org/sites/default/files/trends-student-aid-web-final-508-2.pdf. Accessed 29 January 2016.

Cornwell, Christopher, David B. Mustard, and Deepa J. Sridhar. 2006. The enrollment effects of merit-based financial aid: Evidence from Georgia's HOPE program. *Journal of Labor Economics* 24(4): 761–786.

Cuñado, Juncal, and Fernando Perez de Gracia. 2012. Does education affect happiness? Evidence for Spain. *Social Indicators Research* 108(1): 185–196.

Dynarski, Susan. 2000. Hope for whom? Financial aid for the middle class and its impact on college attendance. NBER Working Paper No. 7756.

Frey, Bruno S. 2018. *Economics of happiness*. SpringerBriefs in Economics. Cham, Switzerland: Springer International Publishing AG.

Georgia Student Finance Commission. 2019. *Eligibility for the HOPE Scholarship*. Available https://www.gafutures.org/hope-state-aid-programs/hope-zell-miller-scholarships/hope-scholarship/eligibility/. Accessed 30 July 2019.

Heller, Donald. E. 1999. The effects of tuition and state financial aid on public college enrollment. *Review of Higher Education* 23(1): 65–89.

Imbens, Guido W. 2015. Matching methods in practice: Three examples. *Journal of Human Resources* 50(2): 373–419.

Imbens, Guido W., and Donald. B. Rubin. 2015. *Causal inference for statistics, social, and biomedical sciences: An introduction*. New York: Cambridge University Press.

Jaeger, David A., and Marianne E. Page. 1996. Degrees matter: New evidence on sheepskin effects in the returns to education. *Review of Economics and Statistics* 78(4): 733–740.

Jia, Ning. 2019. Heterogeneous effects of merit scholarship programs: Do program features matter? *Applied Economics* 51(27): 2963–2979.

Kane, Thomas J. 2003. A quasi-experimental estimate of the impact of financial aid on college-going. NBER Working Paper No. 9703.

King, Jacqueline E. 2004. Missed opportunities: Students who do not apply for financial aid. Washington, DC: American Council on Education Issue Brief.

Kreighbaum, Andrew. 2017. Sanders keeps focus on free college. *Inside Higher Ed*, 4 April.

Linsenmeier, David M., Harvey S. Rosen, and Cecilia Elena Rouse. 2006. Financial aid packages and college enrollment decisions: An econometric case study. *Review of Economics and Statistics* 88(1): 126–145.

Lovenheim, Michael F., and C. Lockwood Reynolds. 2013. The effect of housing wealth on college choice: Evidence from the housing boom. *Journal of Human Resources* 48(1): 1–35.

McPherson, Michael S., and Morton Owen Schapiro. 1991. Does student aid affect college enrollment? New evidence on a persistent controversy. *American Economic Review* 81(1): 309–318.

Monks, James. 2009. The impact of merit-based financial aid on college enrollment: A field experiment. *Economics of Education Review* 28(1): 99–106.

Ness, Eric C. 2008. *Merit aid and the politics of education*. New York: Routledge Press.

New York State. 2017. *Tuition-free degree program: The Excelsior Scholarship*. Available https://www.ny.gov/programs/tuition-free-degree-program-excelsior-scholarship. Accessed 1 November 2017.

Office of Student Financial Assistance. 2019. *Florida student scholarship & grant programs*. Available https://www.floridastudentfinancialaidsg.org/SAPBFMAIN/SAPBFMAIN. Accessed 30 July 2019.

Scott-Clayton, Judith. 2011. On money and motivation: A quasi-experimental analysis of financial incentives for college achievement. *Journal of Human Resources* 46(3): 614–646.

Seftor, Neil S., and Sarah E. Turner. 2002. Back to school: Federal student aid policy and adult college enrollment. *Journal of Human Resources* 37(2): 336–352.

Sjoquist, David L., and John V. Winters. 2012. Building the stock of college-educated labor revisited. *Journal of Human Resources* 47(1): 270–285.

Sjoquist, David L., and John V. Winters. 2015. State merit-based financial aid programs and college attainment. *Journal of Regional Science* 55(3): 364–390.

Stuart, Elizabeth A., Haiden A. Huskamp, Kenneth Duckworth, Jeffrey Simmons, Zirui Song, Michael E. Chernew, and Colleen L. Barry. 2014. Using propensity scores in difference-in-differences models to estimate the effects of a policy change. *Health Services and Outcomes Research Methodology* 14(4): 166–182.

TN Higher Education Commission & Student Assistance Corporation. 2019. *Tennessee HOPE Scholarship*. Available https://www.tn.gov/collegepays/money-for-college/tn-education-lottery-programs/tennessee-hope-scholarship.html. Accessed 30 July 2019.

U.S. Department of Education (USDOE). 1997. *1996–1997 Federal Pell Grant Program end-of-year report*. Available https://www2.ed.gov/finaid/prof/resources/data/pell-historical/pell-eoy-1996-97.pdf. Accessed 30 July 2019.

Van der Klaauw, Wilbert. 2002. Estimating the effect of financial aid offers on college enrollment: A regression–discontinuity approach. *International Economic Review* 43(4): 1249–1287.

Welch, Jilleah G. 2014. HOPE for community college students: The impact of merit aid on persistence, graduation, and earnings. *Economics of Education Review* 43:1–20.