Spring 5-12-2018

Transitions from higher education to the labor market: merit aid, time to degree signals, and major choice

Christopher P. Erwin
University of New Mexico

Christopher Patrick Erwin

Candidate

Economics

Department

This dissertation is approved, and it is acceptable in quality and form for publication:

Approved by the Dissertation Committee:

Melissa Binder, Chairperson

Alok Bohara

Xiaoxue Li

Nancy Lopez

TRANSITIONS FROM HIGHER EDUCATION TO THE LABOR MARKET: MERIT AID, TIME TO DEGREE SIGNALS, AND MAJOR CHOICE

BY

CHRISTOPHER PATRICK ERWIN

B.S., Economics, Arizona State University, 2006
M.A., Economics, California State University, Fullerton, 2011
M.A., Economics, University of New Mexico, 2013

DISSERTATION

Submitted in Partial Fulfillment of the
Requirements for the Degree of

Doctor of Philosophy
Economics

The University of New Mexico
Albuquerque, New Mexico

May 2018

DEDICATION

This work is dedicated to my canyoneering crew members, too numerous to name, for without our thrilling, sometimes hair-raising, adventures in the backcountry I would have surely burned out long ago. You have been the best of friends.

ACKNOWLEDGEMENTS

My sincerest thanks and heartfelt appreciation goes out to my committee chair Dr. Melissa Binder, who has served as my patient mentor, instructor, advisor, co-author, confidant, and friend during my time at the University of New Mexico. The researcher and scholar I have become is largely due to her confidence in my abilities and the countless hours of time she invested in developing my skills as an economist.

I will forever be thankful to my professor and former department chair Dr. Janie Chermak for her generous support and guidance in navigating a couple difficult situations during the initial years of my program. Her kindness and accommodation is much appreciated.

I also thank my committee members Dr. Alok Bohara, Dr. Nancy Lopez, and Dr. Xiaoxue Li for their valuable feedback and recommendations. Their guidance has undoubtedly strengthened this work.

I wish to thank the team at the Center for Education Policy Research at the University of New Mexico for giving me the opportunity to work on interesting and challenging projects that are exceedingly important to the health of families and children in the State of New Mexico. The team has been generous in supporting me and ensuring I always had a place on campus to call home.

Finally, I would like to thank my wonderful family for supporting my goals and aspirations during my doctoral program. It has been a long, challenging road and I could not have done it without their encouragement.

TRANSITIONS FROM HIGHER EDUCATION TO THE LABOR MARKET: MERIT AID, TIME TO DEGREE SIGNALS, AND MAJOR CHOICE

by

Christopher Patrick Erwin

B.S., Economics, Arizona State University, 2006
M.A., Economics, California State University, Fullerton, 2011
M.A., Economics, University of New Mexico, 2013
Ph.D., Economics, University of New Mexico, 2018

ABSTRACT

Human capital production is central to economic wellbeing from a national perspective: it improves productivity, spurs technological innovation, and promotes sustainable economic growth over time. Equally important, investments in human capital are central to economic wellbeing at the individual level. College graduates tend to earn more money, are more employable, are better able to manage economic downturns, and have even been shown to have better health. Moreover, for many students, the college experience is an important lesson in living independently, developing social and professional networks, and generally taking on more responsibility in one's life. It has been said that pursuing a degree in higher education is the largest investment one can make for the future. This work employs a variety of empirical strategies to better understand how students make choices regarding college: where to attend, what to study, and whether to work during college or not, for example.

Chapter 1 provides background on two contentious issues in American higher education: the changing structure of financial aid and the longstanding trend of lengthening time to baccalaureate degree in the United States. Traditional financial aid has included subsidized and non-subsidized loans, need-based financial aid (e.g., Federal Pell Grant Program), and targeted merit-based financial aid, such as academic or athletic scholarships. However, the early 1990s saw the advent of a new type of financial aid: broad-based, state merit-based aid scholarships. I provide background information on how such scholarships are generally structured, paying special attention to New Mexico's program, the New Mexico Legislative Lottery Scholarship (NMLLS). Chapter 1 then discusses longstanding trends in the time it takes undergraduate students to earn bachelor's degrees in the United States. Focus is targeted at the potential costs of this phenomenon to state budgets, institutions, and the students themselves. Chapter 1 concludes by discussing responses to this trend by lawmakers and higher education officials, and whether such responses are warranted.

Using a rich administrative data set from New Mexico's flagship university, Chapter 2 examines whether the NMLLS resulted in any meaningful change in 4-, 5-, and 6-year completion rates comparing qualified resident students to nonqualified nonresident students before and after the program became effective. Propensity score matching is performed to mitigate any observable differences between resident and nonresidents. We find no overall completion effects in the aggregate, but do find economically meaningful divergent completion effects by academic preparation. It appears that results may be driven by students from families whose financial constraints are binding.

Chapter three investigates whether college graduates suffer a wage penalty in the labor market for taking longer than traditional standards to complete an undergraduate degree. We view this as a test of whether employers view time to degree as a productivity signal in the labor market. According to human capital theory, there should be no wage penalty if students are accumulating the same amount of credits at a slower rate. However, if employers view lengthened time to degree as a negative productivity signal, then one would expect a wage penalty. Previous literature has found large and statistically significant wage penalties associated with lengthened time to degree, but we are not convinced they have adequately addressed the endogeneity of time to degree in the student wage equation. When we address endogeneity by controlling for institutional quality, student ability, and instrumenting for the student's own time to degree with the average time to degree at their institution, we find no evidence that employers view lengthened time to degree as a negative productivity signal.

Chapter 4 uses the same data set and methods as Chapter 2, but examines students' choice of majors. I argue that one potential unintended consequence of broad-based merit scholarships is to discourage students from attempting more difficult majors, such as STEM, in order to maximize the likelihood of scholarship retaining. This may be particularly true for marginally academically prepared students. I find no evidence that either the likelihood of first majoring in a STEM field or earning a degree in STEM is affected in the aggregate by the enactment of the NMLLS. Statistically significant effects emerge when disaggregating by academic preparation, however—academically less prepared students are less likely to pursue a STEM major as a result of the NMLLS, while more academically prepared students are more likely to first major in a STEM field. This is in accordance with the previous literature as well as the theoretical model offered in the paper.

Chapter 5 concludes, with a summary of main results from Chapters 2 through 4, with special attention being paid to the policy implication of these findings. Chapter 5 also offers suggestions for future work.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1: Introduction: Merit aid, student decisions, and employer responses

This work focuses on two major issues which have garnered significant attention in higher education over the previous two decades. The first is the changing structure of financial aid in the United States. Traditionally, students have either financed higher education out of their own pockets (or their parents' pockets) or through loans, need-based financial aid, or academic or athletic scholarships. Since the early 1990s, however, a new type of funding mechanism has been popularized: state merit aid scholarships. Such scholarships tend to be very generous in that they generally cover all or nearly all of a qualifying students direct college costs. State merit aid scholarships are broad in the sense that they are generally available to all in-state resident students meeting some sort of academic criteria. Many of these scholarship are at least partially funded by proceeds from state lottery ticket sales. Such scholarships are meant to increase access to higher education for students from families that would otherwise not be able to afford college, or may not be motivated to finish high school since they otherwise assume they would not be able to afford college afterwards.

Currently, at least 27 states have some form of merit aid scholarship, each with varying initial and renewal requirements. As an example of a broad-based, generous state merit aid scholarship, consider the first and most studied scholarship of its kind, Georgia's Helping Outstanding Pupils Educationally scholarship, or HOPE. The program was launched in 1993 and provides full tuition and fees to qualifying students. Initial qualifications include being a state resident of Georgia and graduating from high school with a minimum 3.0 cumulative grade point average (GPA). Renewal of the scholarship requires maintaining a cumulative 3.0 GPA in college, and students may continue to receive the award for up to seven years, until they attempt 127 credit hours, or complete a bachelor's degree—whichever comes first. As of 2010, approximately 43 percent of all undergraduates in the University of Georgia System were receiving HOPE (Sjoquist and Winters, 2015a). Despite the popularization of state merit aid scholarships such as HOPE, many questions remain regarding possible unintended consequences of generous financial aid tied to modest academic requirements.

The second major phenomenon that has garnered significant attention from university officials, lawmakers, and even state and federal governments is lengthening time to baccalaureate degree in the United States. Since the 1970s, the time it takes a student to complete a bachelor's degree has steadily risen. For example, 58 percent college graduates from the 1972 high school class graduated in four years or less; this figure dropped to 44 percent for the 1992 high school class (Bound and Turner, 2010). More recent data suggest this trend has continued. This has caused alarm for several reasons, including the straining of resources at universities, inefficient spending of state and federal appropriations, and suboptimal labor market outcomes for students taking longer than the traditional four years to complete their undergraduate studies. Recently, states have been taking matters into their own hands by proposing punishments for those that do not remain on-track to graduate within four years. In 2012 alone, five state legislatures passed laws aimed at reducing time to degree. In the wake of all the attention paid to baccalaureate time to degree by policymakers, it is worth asking whether all of the concern is justified.

Chapter 2 focuses on New Mexico's state merit aid scholarship, the New Mexico Legislative Lottery Scholarship (NMLLS). The NMLLS is unique in that it has the lowest academic requirements of any state merit aid scholarship. To qualify, one only needs to be a state resident, graduate from a New Mexico high school, and immediately enroll in any one of 16 qualified in-state institutions. In-state institutions then automatically award a "Bridge to Success Scholarship" that pays all tuition and fees in the student's first semester. If the student earns a minimum 2.5 GPA and completes at least 12 credit hours in this first semester, they are then eligible for the NMLLS. The NMLLS pays full tuition and fees for up to four additional years beyond the "bridging" semester as long as students maintain a cumulative 2.5 GPA and complete at least 12 credit hours per semester. Beyond its low academic requirements, the NMLLS is also unique in that it is the only state merit aid scholarship where initial eligibility requirements are based on college—not high school—performance.

Using a rich administrative data set that covers the population of students over a ten-year period, we examine how this "low-bar" state merit aid scholarship promotes graduation. This is an empirical question from the outset: one would expect that relaxing financial constraints for students, perhaps even affording them the opportunity to not have to work during college and spend more time studying, would result in higher graduation rates. However, low-bar nature of initial and renewal eligibility requirements may incentivize some students to attend a college they otherwise would not have without the scholarship. If the scholarship results in an influx of marginally academically prepared students, then perhaps graduation rates would suffer. I attempt to answer this question by using a sophisticated difference-in-differences matching estimator using qualified resident students as the treatment group and nonresident (and therefore nonqualified) students as the control group. Propensity score matching is performed to mitigate any observable differences between residents and nonresidents, and the success of the matching algorithm is examined using recent statistical diagnostic tests developed by Imbens and Rubin (2015). I also examine the feasibility of using regression discontinuity design to examine the relationship between merit aid and college completion, which itself is an implicit test of whether students are able to strategically "game" NMLLS eligibility requirements. Although results suggest no overall completion effects resulting from the NMLLS, meaningful divergent effects appear when disaggregating the sample by academic preparation—more academically prepared students exhibited positive completion effects while less academically prepared students exhibited negative completion effects.

Chapter 3 revisits previous literature to examine the latest trends in increasing time to baccalaureate degree in the United States. Using a restricted nationally-representative longitudinal study from the Department of Education, we ask whether students that overshoot the traditional four-year graduation mark are penalized in the labor market. Another way of looking at this research question is: does baccalaureate time to degree serve as a productivity signal in the United States? From one perspective, human capital theory predicts that time to degree should not matter if students are acquiring the same amount of human capital over a longer period of time (assuming human capital does not depreciate). From another perspective, employers may view lengthier time to degree as a negative productivity signal. If so, and assuming wages reflect productivity, there would be a wage penalty associated with longer time to degree. Testing for this is not straightforward, and the previous literature is not able to overcome endogeneity issues. Because variables such as student ability[1] and school quality are correlated with both time to degree and future earnings, this is an identification problem that needs to be addressed. I address this endogeneity problem using instrumental variables techniques and provide an update to previous literature on this subject. We find that we are able to replicate wage penalties found in previous studies which appeal to ordinary least squares, but instrumental variables reveals no such wage penalties for lengthened time to degree. These results suggest that previous studies of the relationship between time to degree and wages suffer from significant bias.

Chapter 4 revisits the NMLLS with another question about unintended consequences. Merit aid tied to academic requirements, in general, may result in the perverse outcome of students pursuing easier courses of study in order to maximize their chances of scholarship retention. Given the nation's preoccupation with promoting degrees in science, technology, engineering, and mathematics (STEM), I ask whether the NMLLS dissuades scholarship recipients from pursuing majors in STEM and whether the scholarship results in lower STEM degree production. Previous literature on the subject finds either null or negative effects of state merit aid on STEM degree production (Cornwell *et al.*, 2006; Sjoquist and Winters, 2015a). Since the matching algorithm from Chapter 2 does not consider any outcomes—only covariates—and is considered successful, we appeal to the same difference-in-differences matching estimator and diagnostic testing procedures. The main contribution of this chapter is that we use a rich administrative data that allows for a much more detailed analysis relative to previous literature. In particular, we test for major choice effects in the aggregate, but also disaggregate by academic preparation, something not attempted in previous studies. Results reveal no effect of the NMLLS on either first majoring in a STEM field or eventually earning a STEM degree, however divergent effects are again found by academic preparation—less academically prepared students are less likely to first declare a major in STEM, while more academically prepared students are more likely to do so.

In Chapter 5, main conclusions and policy implications are revisited and extended discussion is provided. Particular attention will be paid to the scientific contributions from each chapter, and how these contributions fit into the broader field of the economics of education. This chapter also discusses policy implications and methodological limitations, and concludes by offering direction for future research in the above areas.

The main argument in this dissertation is that unintended consequences may occur when we design policies in higher education to broadly subsidize students or to restrict their behavior. Giving generous amounts of merit aid to students that are perhaps marginally qualified for postsecondary studies may result in the unintended consequence of promoting overmatching in higher education, which sets up some students for failure. Such programs also have the consequence of changing students' choice of major, and therefore ultimately their career paths. Although the intentions are well-placed, the outcomes are sometimes suboptimal. This is true of proposals to punish students that wish to take a nontraditional path to a baccalaureate degree. Students are rational actors in the economy, and spending time and money legislating their college choices is an inefficient allocation of resources. Sometimes the best remedy is no remedy at all.

Notes

[1] We use the term "ability" not to describe the innate characteristics of the student, but to instead describe a constructed measure, such as academic preparation, for example. As such, it is not intended to provide negative connotation for any students included in the sample.

Chapter 2: Does broad-based merit aid improve college completion? Evidence from New Mexico's lottery scholarship

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 GPA in their first semester at any public 2-year or 4-year college. We find a significant positive completion effect of 9.4 percentage points (16.8 percent) for academically well-prepared students that is offset by a nearly equal and opposite negative effect for less prepared students. We posit that the scholarship program, which effectively erased the difference in tuition at 2- and 4-year colleges, may have induced weaker students to take their chances on a more prestigious, yet riskier, academic path.

2.1 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 the 1993, at least 25 states have implemented merit-based scholarships, the first and most studied being Georgia's Helping Outstanding Pupils Educationally, or HOPE, scholarship program.[2] HOPE marked the beginning of what has been a major restructuring of the financial aid landscape in America. 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 2013-2014 academic year, New Mexico was one of 13 states where this percentage was below 40 percent.[3]

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, while grants have a positive and significant effect on student enrollment (Linsenmeier *et al.* 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 are sensitive to financial aid offers. Several studies show a significant and positive relationship between grant aid and student enrollment (Seftor and Turner, 2002; Kane, 2003; Heller, 2009) 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 four percent. Also studying HOPE, Cornwell *et al.*

(2006) find the program increased student enrollment by 6 percent. In New Mexico, Binder and Ganderton (2002, 2004) find that while the NMLLS boosted enrollment at four-year colleges in New Mexico, the effect appears to be driven by additional enrollment of students that 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:

> …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, pp. 36)

While making higher education widely accessible is certainly a noble objective, 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; Schafer *et al.*, 2013).

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 9.4 percentage points (16.8 percent) more likely to graduate within six years, compared to ineligible peers with similar high school GPAs. Less academically-prepared freshmen are approximately 8.7 percentage points (27.5 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[4] 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.[5] 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 *et al*., 2016, for example).

The paper proceeds as follows: Section 2.2 discusses existing literature regarding merit-aid and college completion, and introduces the NMLLS; Section 2.3 presents a theoretical model of college persistence; Section 2.4 describes the data; Section 2.5 summarizes the empirical approach; Section 2.6 discusses main findings and robustness checks; Section 2.7 discusses other explanations for patterns we find in the results; and Section 2.8 concludes.

2.2 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, 2015b) found no difference in college attainment for those exposed to lottery scholarship programs. Using a similar methodology, Jia (2017) 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 cut-off for West Virginia's lottery-funded PROMISE scholarship program, compared to 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 while students in Bruce and Carruthers' and Welch's sample have high school GPAs below 3.0.[6] 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) point 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, 2015b) find no population graduation effect.

Castleman and Long (2013) examine the effect of the need-based Florida Student Access Grant (FSAG), which awards $1,300 annual grants to students whose family's EFC falls below an annually determined cutoff, with no additional academic restrictions for grant receipt in the first year. Because the EFC is generated from information provided by students to the FAFSA according to an opaque algorithm, and because the cutoff is determined each year and is not publicized,[7] 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, 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 2000-2001 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 10th graders in the top quartile of a state standardized test with tuition waivers to attend public in-state colleges, Cohodes and Goodman (2014) find that 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.

2.2.1 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 following 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 16 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 maintained a cumulative 2.5 GPA. Only 58 percent of first semester students over 1994-1999 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 1993-1994 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 the NMLLS not only supplemented rather than supplanted student aid, but drastically changed the student aid landscape throughout the state.

Compared to 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.[8] 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.[9] 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.[10]

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 16 participating public institutions in New Mexico as depicted in Table 2.1. A student better matched at Santa Fe Community 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.

2.3 Merit aid and college 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,

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

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:

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

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 the student is responsible for, 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 outweigh the costs. Once enrolled, however, they may discover 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 which 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.

2.4 Data set

We use administrative data for all first-time, full-time entering freshmen at UNM before and after the implementation of the NMLLS to estimate completion effects. 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 socio-demographic 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 that did not submit a FAFSA, we assume that 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 1995-1996 Federal Pell Grant End-of-Year Report showing that less than two percent of Pell recipients had family income in excess of $40,000.[11] This assumption is not perfect. King (2004) estimates that in 2000 over ten percent of all Pell-eligible students did not fill out a FAFSA.[12] If the analysis in King (2004) holds for our data set, 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 matriculating students at UNM. For these students, we assigned them the mean high school GPA of 3.28.

We concentrate on the years 1994 to 1999, bounding the policy change by three years before and after implementation. These years encompass the largest economic expansion in the U.S. 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. To our knowledge, there were no concurrent policy changes at the high school or postsecondary level in New Mexico over the 1994-1999 period which would have differentially impacted enrollment and/or completion for residents and nonresidents. As discussed below, having an equal number of years before and after the lottery scholarship is advantageous given our identification strategy. As we show in Section 6, results are similar when we expand the sample period to include additional student cohorts.

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 2.2 compares summary statistics for resident and nonresident students before and after the implementation of the NMLLS. It appears the composition of 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 to 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 2.2 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 improved following implementation of the NMLLS, according to HSGPA 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 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 2.1 presents pre-treatment trends in graduation rates for residents and nonresidents between 1994 and 1999. We are particularly interested in six-year graduation rate trends, a standard measure of completion.[13] 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

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

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 which 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. Models report robust standard errors. In equation (3), m and q are the number of leads and lags of the treatment effect included. We include two leads and three lags in our test, defining 1994 as the reference cohort.

Testing the common trends assumption using (3) requires examining whether

$$\beta_j = 0 \; \forall j < 0. \qquad (4)$$

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 for the four graduation outcomes using ordinary least squares and results are presented in Tables 2.3 through 2.6. Results provide evidence that the common trends assumption holds for all of our specifications, as estimated coefficients on all leads are not statistically different from zero.

Our data include 10,022 resident students, 6,307 of which enrolled during the post-NMLLS period and were eligible for the Bridge to Success Scholarship. Of these, 2,664 met cumulative GPA and credit attainment requirements to begin the NMLLS in their second semester. Table 2.7 documents the number of students that maintain the scholarship in the second through ninth semester. It is apparent scholarship loss was quite common. Of the 2,664 students that qualified for the NMLLS, approximately 30 percent were still eligible for the NMLLS going into their third year.

2.5 Empirical model

We conduct difference-in-differences matching estimation on the propensity score to mitigate any observable differences between resident and nonresident students. Chabé-Ferret (2015) conducts Monte Carlo simulations using experimental job training program data from LaLonde (1986), finding that difference-in-differences matching is superior to simple difference-in-differences estimation (i.e., no matching) in replicating experimental results when the model is symmetric (i.e., there are an equal number of pre- and post-treatment periods) and matching is performed on time-invariant characteristics, two conditions our model satisfies. Our approach uses kernel matching, a one-to-many matching technique that assigns larger weights to control units closer in propensity score. The general form of the matching estimator is given by

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

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

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

where K is the Epanechnikov kernel function and $\hat{l}(\cdot) \equiv \ln\left(\frac{\hat{p}(\cdot)}{1-\hat{p}(\cdot)}\right)$ 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.[14] Results of the propensity score model are presented in Table 2.8. It is important to note that while 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. In order to increase the precision of our estimated treatment effects, and to avoid imposing functional form where possible, we choose to conduct kernel density matching.[15] This method has the advantage of lower variance since more information is used. On the other hand, it may result in an increase in bias due 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 99.5 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 2.A.

A power analysis in Table 2.9 shows that most models we estimate have sufficient power to detect a five percent change in completion rates at the five 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 actually 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.[16] We disaggregate further by family income in order to examine program effects for students whose financial constraints are more likely binding. Robustness checks using varying cohorts and smoothing parameters are discussed in Section 2.6. 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.

We note that while difference-in-differences models hinge on the comparability of pre-treatment trends in outcomes across residents and nonresidents, combining difference-in-differences methods with propensity score matching controls for compositional changes in groups over time (Stuart *et al.* 2014). It is also worth noting that regressions control for high school achievement and standardized test scores, the main indication of compositional change. Also, because UNM is a *de facto* open enrollment institution, changes in selectivity are not likely to confound the analysis (Binder and Ganderton, 2004). We agree that compositional change occurred, but this does not threaten the validity of the treatment effects we find.

2.6 Results

Means and normalized differences after kernel matching are presented in Table 2.10. 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 (-.122) far below one-quarter of a standard deviation unit in absolute value. We produce these statistics by 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 2.11 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 completion effect for six-year graduation of 8.7 percentage points (27.5 percent). Students with above average high school GPA are 9.4 percentage points (16.8 percent) more likely to graduate within six years compared to similar nonresident students. Effects are significant at the five and ten percent-levels, respectively. Thus, while 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 2.12 presents results of the matching estimation performed on low-income students, defined as 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 ten percent-level. As shown in Table 2.13, 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.

2.6.1 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 Appendix 2.B. Table 2.B1 presents results for our estimates by graduation semester and academic preparation. In Table 2.B1, larger bandwidths result in larger effect sizes and smaller standard errors. 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 2.B2 and 2.B3 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. Appendix 2.C presents results using different sets of freshmen cohorts. Although a bit noisier than robustness checks using alternative bandwidths, we see a similar pattern of completion rates emerge as compared to our preferred specification.[17]

We also 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 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 some sort of 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.

2.6.2 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 a student is 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 which meet the NMLLS eligibility requirements with the exception of 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 (FRD) approach using a sample of resident students from 1997 - 2008. FRD 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. We visually inspect the jump in the probability of NMLLS funding by bridging semester GPA in Figure 2.2. The jump between the quadratic fitted lines below and above the threshold is below one (approximately 80 percent), so FRD seems appropriate in our context.

However, we fail to pass a critical identification test for regression discontinuity studies: cutoff manipulation. Figure 2.3 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 some students manipulate this eligibility cutoff by perhaps taking easier courses or dropping courses when a poor grade is expected. Since 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 2.14 presents results of formal manipulation tests using local polynomial density estimators following McCrary (2008) and Cattaneo *et al.* (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.

2.7 Other possible explanations for the patterns we find

2.7.1 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 17 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 Tables 2.3 to 2.6.

2.7.2 Confounding factors

A massive increase in enrollment at UNM accompanied the NMLLS (3,715 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 that a $10,000 increase in real housing wealth increases the relative likelihood of attending a flagship university by two percent and the overall completion likelihood by 1.8 percent. Simple accounting reveals 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.[18] 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.[19] Moreover, real personal income only increased by 5.5 percent over the same period, an annualized growth rate less than one 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.[20] 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.

2.7.3 Congestion

Another possible explanation for the patterns we see are capacity constraints and congestion at UNM. This would explain what we see in Table 2.2—where the academic preparation and standardized test scores of nonresidents increase by a statistically significant amount after program implementation. If the large increase in NMLLS-qualified students forced the university 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, *et cetera*, after the lottery scholarship launched.[21] 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 which 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 enrolled, the more state dollars are allocated to the institution. This provides an incentive for institutions to compete with one another on the basis on enrollment. There is no *de jure* limit on the number of additional students UNM may enroll in a given semester, so it is likely the university simply absorbed this additional enrollment without crowding out other groups, such as nonresidents and low-income students.

2.7.4 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. Due to this concern, we construct difference-in-differences matching estimates of cumulative credits earned since enrollment. These estimates are presented in Appendix 2.D.

Estimates in Appendix 2.D present the effects of the NMLLS on cumulative credits earned using the same matching procedure as models of college completion. Table 2.D1 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 2 percentage points (7 percent) after the first year to 15 percentage points (15 percent) at the six-year mark. We find that these effects are 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.

2.8. 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 compared to their nonresident counterparts, while 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 that otherwise would not have attended college, or at least a four-year college.

Results suggest that low-income, high achieving high school students benefit from the NMLLS, while lower-achieving students do not. The latter may be explained by overmatching at UNM, where marginally prepared students that 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 that 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.

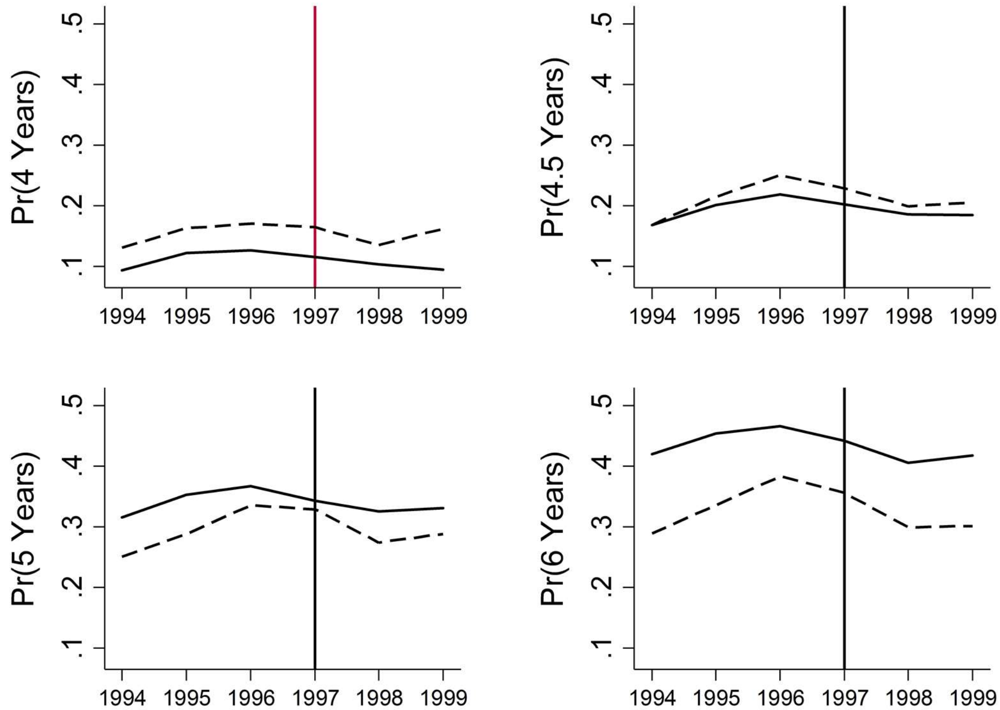
The main conclusion that we draw from our analysis is that setting the bar too low in terms of merit aid may be detrimental to the success of the least academically prepared students. The promise of generous financial aid tied to seemingly modest academic criteria may actually worsen college persistence 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 2014-2015 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 2017-2018 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 that 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 towards those that 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.

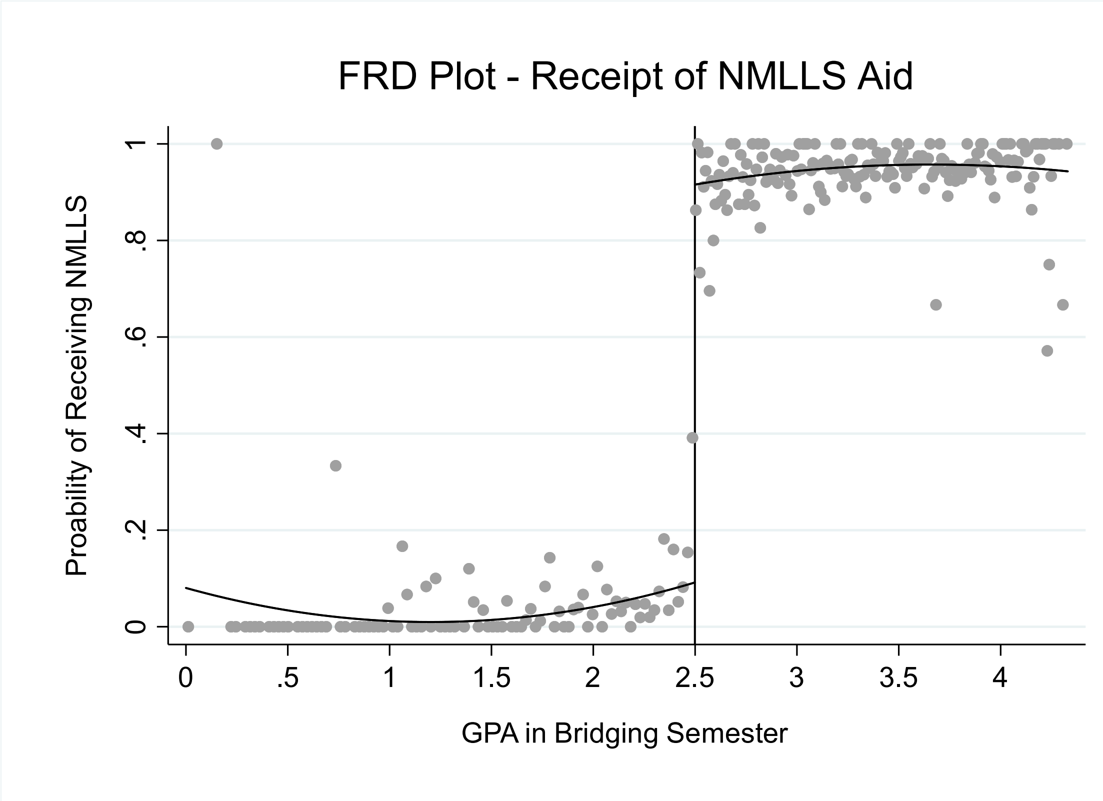
In general, further research is needed to investigate how to increase degree completion, not merely enrollment, while avoiding harming less academically prepared students. One potential remedy may be to pair lottery scholarship funds with stronger academic supports such as additional mandatory advising, mid-semester check-ups, or an additional one-credit mandatory course on topics such as scholarship details and strategies for academically suriving the freshmen year. Another potentially promising reform would be to tie program eligibility to high school performance rather than college performance. Having intitial 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.



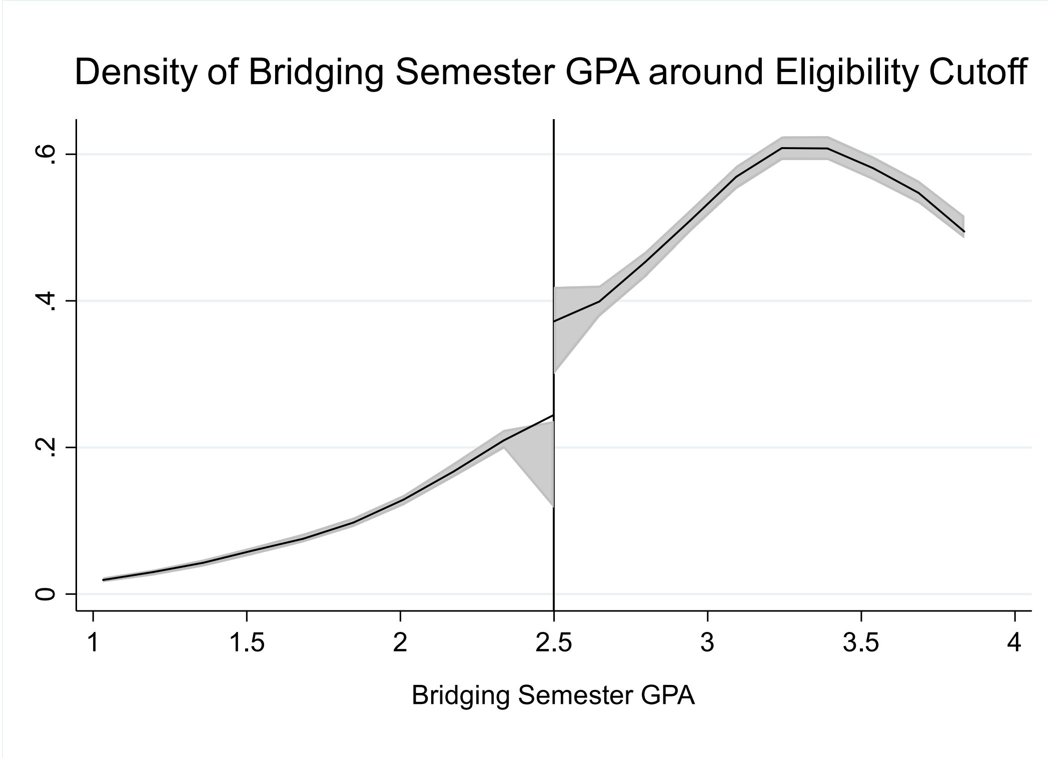
Note: The plots above show the likelihood of degree completion for incoming cohorts over the period 1994 to 1999. Solid lines represent resident students while dashed lines represent nonresident students. The vertical bars at 1997 mark the implementation of the New Mexico Legislative Lottery Scholarship.

Figure 2.1 Pre-post trends in the probability of graduating, by residency



Note: Points depict the within-bin sample average of NMLLS 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 (FRD) approach.

Figure 2.2 Jump in treatment probability around the bridging semester GPA cutoff



Note: Figure 2.3 presents the density of the running variable (bridging semester GPA) around the NMLLS eligibility cutoff with 95 percent confidence intervals in shown in gray. 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 2.3 Bridging semester density around the bridging semester GPA cutoff

Table 2.1 Full-time resident tuition at all NMLLS-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
Mesalands Community College	2	1,990
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 Institute	2	730

Source: Institution financial aid department websites. Accessed 28 March 2017. Figures present tuition and fees for one academic year taking fifteen credit hours per semester. For two-year schools it is assumed the student is within the community college district, where applicable.

Table 2.2 Student characteristics before and after initiation of the NMLLS program

	Residents			Nonresidents		
Variable	Before	After	Diff.	Before	After	Diff.
Grad. in 4 Yrs	.114	.103	-.011*	.153	.154	.001
Grad. in 4.5 Yrs	.195	.189	-.006	.210	.211	.002
Grad. in 5 Yrs	.345	.332	-.013	.290	.297	.008
Grad.in 6 Yrs	.447	.420	-.027***	.334	.319	-.015
HSGPA	3.312 (.502)	3.273 (.471)	-.038***	3.233 (.532)	3.300 (.503)	.067**
ACT	22.530 (3.834)	22.176 (3.887)	-.354***	22.317 (4.109)	22.861 (4.096)	.544**
Remedial	.264	.290	.026***	.164	.227	.063***
Income < $40K	.230	.205	-.025***	.155	.162	.007
Female	.571	.565	-.006	.526	.545	.019
Hispanic	.386	.375	-.011	.147	.166	.020
Native	.043	.045	.002	.041	.051	.010
Asian	.047	.037	-.010**	.034	.026	-.008
Black	.021	.022	.002	.082	.080	-.002
Observations	3,715	6,307		587	649	

Source: Freshmen Tracking System, Office of Institutional Analytics, UNM. ***, **, and * represent statistical significance at the 1, 5, and 10 percent-levels, respectively. Standard deviations are in parentheses.

Table 2.3 Common trends assumption test, 1994-1999

	Graduation Rates by Years since First Enrollment			
Leads and Lags	4 Years	4 ½ Years	5 Years	6 Years
NMLLS $_{t-2}$	-.004	-.014	-.006	-.018
	(.035)	(.040)	(.045)	(.047)
NMLLS $_{t-1}$	.005	-.021	-.025	-.041
	(.036)	(.042)	(.046)	(.049)
NMLLS $_{t0}$	.012	-.001	-.024	-.019
	(.035)	(.040)	(.045)	(.047)
NMLLS $_{t+1}$	.027	.010	.012	.00007
	(.034)	(.039)	(.044)	(.047)
NMLLS $_{t+2}$	-.004	.006	.004	.010
	(.034)	(.038)	(.043)	(.045)
R^2	.0872	.1020	.1131	.1173
Observations				11,258

Robust standard errors are reported in parentheses. Ordinary least squares estimates for all students entering UNM between 1994 – 1999 given. Reported coefficients are on interactions between cohort years and a resident dummy variable. Models include resident and cohort dummies as well as controls for race, ethnicity, standardized test scores, high school GPA, gender, and family income. The period t_0 is 1997, the year the NMLLS was implemented. *, **, and *** denote statistical significance at the 10, 5, and 1 percent-level, respectively.

Table 2.4 Common trends assumption test, 1994-1999, HSGPA ≤ 3.28

	Graduation Rates by Years since First Enrollment			
Leads and Lags	4 Years	4 ½ Years	5 Years	6 Years
NMLLS $_{t-2}$	.003 (.033)	.009 (.038)	.005 (.052)	.005 (.057)
NMLLS $_{t-1}$	-.023 (.038)	-.031 (.048)	-.019 (.057)	-.034 (.064)
NMLLS $_{t0}$	-.037 (.037)	-.064 (.047)	-.092 (.058)	-.102 (.062)
NMLLS $_{t+1}$	.005 (.031)	-.020 (.042)	-.073 (.055)	-.111* (.060)
NMLLS $_{t+2}$	.015 (.029)	-.002 (.039)	-.039 (.053)	-.032 (.056)
R^2	.0177	.0306	.0435	.0520
Observations				5,502

Robust standard errors are reported in parentheses. Ordinary least squares estimates for all students entering UNM between 1994 – 1999 with less than or equal to average high school GPAs given. Reported coefficients are on interactions between cohort years and a resident dummy variable. Models include resident and cohort dummies as well as controls for race, ethnicity, standardized test scores, high school GPA, gender, and family income. The period t_0 is 1997, the year the NMLLS was implemented. *, **, and *** denote statistical significance at the 10, 5, and 1 percent-level, respectively.

Table 2.5 Common trends assumption test, 1994-1999, HSGPA > 3.28

	Graduation Rates by Years since First Enrollment			
Leads and Lags	4 Years	4 ½ Years	5 Years	6 Years
NMLLS $_{t-2}$	-.005 (.065)	-.034 (.071)	-.013 (.076)	-.038 (.077)
NMLLS $_{t-1}$	.022 (.063)	-.023 (.070)	-.038 (.074)	-.052 (.075)
NMLLS $_{t0}$	.047 (.060)	.038 (.065)	.020 (.071)	.036 (.072)
NMLLS $_{t+1}$	.046 (.060)	.028 (.066)	.080 (.071)	.089 (.072)
NMLLS $_{t+2}$	-.007 (.060)	.018 (.064)	.044 (.069)	.043 (.070)
R^2	.0685	.0663	.0651	.0620
Observations				5,756

Robust standard errors are reported in parentheses. Ordinary least squares estimates for all students entering UNM between 1994 – 1999 with above average high school GPAs given. Reported coefficients are on interactions between cohort years and a resident dummy variable. Models include resident and cohort dummies as well as controls for race, ethnicity, standardized test scores, high school GPA, gender, and family income. The period t_0 is 1997, the year the NMLLS was implemented. *, **, and *** denote statistical significance at the 10, 5, and 1 percent-level, respectively.

Table 2.6 Common trends assumption test, 1994-1999, HSGPA > 3.78

	Graduation Rates by Years since First Enrollment			
Leads and Lags	4 Years	4 ½ Years	5 Years	6 Years
NMLLS $_{t-2}$	.067 (.119)	-.012 (.122)	.032 (.128)	-.011 (.130)
NMLLS $_{t-1}$	-.035 (.124)	-.143 (.127)	-.066 (.133)	-.083 (.133)
NMLLS $_{t0}$	.148 (.121)	.161 (.123)	.198 (.131)	.220* (.134)
NMLLS $_{t+1}$	.003 (.123)	-.021 (.125)	.068 (.129)	.072 (.131)
NMLLS $_{t+2}$	.051 (.118)	.031 (.121)	.089 (.126)	.053 (.129)
R^2	.0608	.0623	.0594	.0451
Observations				2,112

Robust standard errors are reported in parentheses. Ordinary least squares estimates for all students entering UNM between 1994 – 1999 with high school GPAs greater than one standard deviation above the mean are given. Reported coefficients are on interactions between cohort years and a resident dummy variable. Models include resident and cohort dummies as well as controls for race, ethnicity, standardized test scores, high school GPA, gender, and family income. The period t_0 is 1997, the year the NMLLS was implemented. *, **, and *** denote statistical significance at the 10, 5, and 1 percent-level, respectively.

Table 2.7 NMLLS student attrition, 1994-1999

Semester	Residents Eligible	Percent Remaining
2	2,664	100.0%
3	2,249	84.4%
4	2,017	75.7%
5	1,863	69.9%
6	1,734	65.1%
7	1,629	61.1%
8	1,568	58.9%
9	1,510	56.7%

Source: Office of Institutitonal Analystics, University of New Mexico. We consider the sample of resident students that met cumulative GPA and credit requirements in their first semester to qualify for the NMLLS.

Table 2.8 Estimated parameters for propensity score model of NMLLS data, 1994-1999

Variable	Estimate	Std. Error
HSGPA	1.729**	.724
ACT	.498***	.090
Remedial	.891***	.118
Income < 20K	.268*	.158
Income < 40K	.160	.108
Female	1.670***	.367
Hispanic	1.865***	.550
Native American	1.884**	.923
Asian	.032	.207
Black	-5.729***	1.155
Declined to state race-ethnicity	-.108	.282
ACT^2	-.013***	.002
ACT*Black	.141***	.045
Female*White	-.571***	.146
$HSGPA^2$	-.461***	.116
ACT*Female	-.053***	.016
ACT*HSGPA	.059***	.020
Remedial*Asian	1.147**	.505
GPA*Black	.546	.339
ACT*Native	-.082**	.041
Female*Native	-.608*	.317
HSGPA*Hispanic	-.312*	.165
Constant	-7.711***	1.600
Observations		11,258

Standard errors are in parentheses. ***, **, and * correspond to statistical significance at the 1, 5, and 10 percent levels, respectively. Propensity scores are estimated using a logistic model. Forty-nine observations were dropped following estimation of the propensity score to ensure overlap, leaving 11,209 observations. The variable *Declined to state race-ethnicity* is equal to one if the student declined to state their race-ethnicity, and zero otherwise.

Table 2.9 Power calculation for regressions on subgroups, 1994-1999

	Residents			Nonresidents			Power
Group	Mean	SD	Obs.	Mean	SD	Obs.	
All Family Incomes							
Full Sa	0.43	0.495	9978	0.327	0.469	1231	0.995
HSGP	0.299	0.458	4901	0.199	0.4	572	0.654
HSGP	0.557	0.497	5077	0.438	0.497	657	0.985
HSGP	0.661	0.474	1879	0.509	0.501	226	0.872
Family Income < $40,000							
Full Sa	0.366	0.482	2104	0.276	0.448	192	0.437
HSGP	0.236	0.425	1037	0.181	0.387	94	0.153
HSGP	0.492	0.5	1067	0.379	0.488	95	0.381
HSGP	0.602	0.49	374	0.5	0.509	30	0.231
Family Income ≥ $40,000							
Full Sa	0.448	0.497	7874	0.336	0.474	1034	0.985
HSGP	0.316	0.465	3864	0.2	0.4	471	0.584
HSGP	0.575	0.495	4010	0.449	0.498	561	0.968
HSGP	0.676	0.468	1505	0.51	0.501	196	0.814

Note: We calculate the power to detect a five percent change in completion rates at the five percent significance level. This is done by testing the difference-in-differences estimator in a linear probability model of six-year completion rates.

Table 2.10: Means and normalized differences after kernel matching, full sample, 1994-1999

	Pre-NMLLS			Post-NMLLS		
Variable	Res.	Nonres.	ND	Res.	Nonres.	ND
HS GPA	3.31	3.27	0.088	3.27	3.33	-0.122
Composite ACT	22.56	22.37	0.047	22.19	22.58	-0.099
Remedial	0.26	0.24	0.032	0.29	0.28	0.012
Income < $40,000	0.22	0.21	0.04	0.2	0.21	-0.032
Female	0.57	0.58	-0.009	0.56	0.59	-0.063
Hispanic	0.39	0.39	-0.019	0.37	0.36	0.025
Native	0.04	0.04	0.001	0.05	0.05	-0.03
Asian	0.04	0.03	0.058	0.04	0.03	0.019
Black	0.02	0.02	-0.023	0.02	0.02	0.018

Means are from Epanechnikov kernel matching performed with a bandwidth of $h = .2$. Normalized differences (ND) are calculated by taking the difference average covariate values by residency status and dividing by a measure of standard deviation.

Table 2.11: NMLLS graduation effects by years since first enrollment, kernel matching, 1994-1999

		Graduation Rates by Years since First Enrollment			
Group	Obs.	4 Years	4 ½ Years	5 Years	6 Years
Full Sample	11,209	-.035 (.023)	-.030 (.027)	-.024 (.033)	-.019 (.035)
$\bar{Y}$		.114	.195	.345	.447
GPA ≤ 3.28	5,473	-.015 (.022)	-.035 (.029)	-.069* (.040)	-.087** (.043)
$\bar{Y}$		.044	.094	.222	.316
GPA > 3.28	5,734	-.022 (.036)	.016 (.043)	.069 (.048)	.094* (.050)
$\bar{Y}$		.176	.284	.453	.561
GPA > 3.78	2,105	.031 (.070)	.082 (.072)	.093 (.080)	.107 (.081)
$\bar{Y}$		.246	.359	.542	.642

Robust standard errors are reported in parentheses. *, **, and *** denote statistical significance at the 10, 5, and 1 percent-level, respectively. Estimates are from difference-in-differences kernel matching performed with a bandwidth of $h = .2$ using the Epanechnikov kernel function. We report estimates for students with below average or average high school 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 the baseline graduation rate by high school performance and years since first enrollment.

Table 2.12: NMLLS graduation effects, kernel matching, family income < $40,000, 1994-1999

		Graduation Rates by Years since First Enrollment			
Group	Obs.	4 Years	4 ½ Years	5 Years	6 Years
Full Sample	2,296	-.011 (.038)	-.012 (.043)	-.037 (.060)	-.020 (.070)
$\bar{Y}$		.085	.139	.276	.377
GPA ≤ 3.28	1,131	.003 (.011)	-.018 (.032)	-.197** (.081)	-.202** (.090)
$\bar{Y}$		.030	.064	.180	.257
GPA > 3.28	1,162	.022 (.076)	.037 (.082)	.161 (.100)	.200* (.115)
$\bar{Y}$		.136	.207	.363	.486
GPA > 3.78	404	-.091 (.148)	-.032 (.168)	.151 (.221)	.054 (.228)
$\bar{Y}$		.164	.270	.478	.629

Robust standard errors are reported in parentheses. *, **, and *** denote statistical significance at the 10, 5, and 1 percent-level, respectively. Estimates are from difference-in-differences kernel matching performed with a bandwidth of h = .2 using the Epanechnikov kernel function. We report estimates for students with below average or average high school 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 the baseline graduation rate by high school performance and years since first enrollment.

Table 2.13: NMLLS graduation effects, kernel matching, family income ≥ \$40,000, 1994-1999

		Graduation Rates by Years since First Enrollment			
Group	Obs.	4 Years	4 ½ Years	5 Years	6 Years
Full Sample	8,908	-.035 (.027)	-.028 (.031)	-.015 (.037)	-.010 (.039)
$\bar{Y}$		.122	.212	.365	.467
GPA ≤ 3.28	4,335	-.018 (.028)	-.040 (.036)	-.035 (.045)	-.055 (.048)
$\bar{Y}$		.048	.104	.234	.335
GPA > 3.28	4,571	-.027 (.041)	.011 (.047)	.033 (.052)	.058 (.052)
$\bar{Y}$		.187	.307	.480	.583
GPA > 3.78	1,701	.055 (.075)	.109 (.077)	.096 (.084)	.130 (.085)
$\bar{Y}$		.269	.384	.560	.646

Robust standard errors are reported in parentheses. *, **, and *** denote statistical significance at the 10, 5, and 1 percent-level, respectively. Estimates are from difference-in-differences kernel matching performed with a bandwidth of h = .2 using the Epanechnikov kernel function. We report estimates for students with below average or average high school 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 the baseline graduation rate by high school performance and years since first enrollment.

Table 2.14. Testing for manipulation of the bridging semester GPA cutoff for NMLLS eligibility

	Bandwidths		Effective Obs.		Conv. Test		Robust Test	
	left	right	left	right	*T*	*p-value*	*T*	*p-value*
$h_- \neq h_+$								
$T_2(h_1)$	0.511	0.524	1748	4591	5.263	< .001	3.831	< .001
$T_3(h_2)$	0.489	0.445	1591	3806	4.109	< .001	4.245	< .001
$T_4(h_3)$	0.653	0.616	1977	5433	3.096	0.002	4.895	< .001
$h_- = h_+$								
$T_2(h_1)$	0.524	0.524	1750	4591	5.270	< .001	3.822	< .001
$T_3(h_2)$	0.445	0.445	1569	3806	4.090	< .001	4.564	< .001
$T_4(h_3)$	0.616	0.616	1903	5433	3.095	0.002	5.193	< .001

Note: Here we present results from manipulation test following McCrary (2008) and Cattaneo *et al.* (2017). 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 staistics test the null hypothesis of continuity of the bridging semester GPA around the NMLLS eligibility cutoff.

Notes

[2] See Sjoquist and Winters (2012) for a complete list.

[3] The College Board, Trends in Student Aid 2015, Figure 28A and Figure 28B. Retrieved January 29, 2016 from http://trends.collegeboard.org/sites/default/files/trends-student-aid-web-final-508-2.pdf.

[4] As of the 2017-2018 academic year, the NMLLS only covers approximately 60 percent of tuition at UNM.

[5] Hillary Clinton and Sen. Bernie Sanders both proposed similar versions of this policy during the 2016 presidential campaign. Sen. Sanders advocated free tuition at all public universities and colleges, while Clinton advocated for a $125,000 household income cap. Retrieved 22 Aug 2017 from https://www.insidehighered.com/news/2017/04/04/sanders-democratic-colleagues-introduce-new-free-college-bill. 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. Retrieved 1 November 2017 from https://www.ny.gov/programs/tuition-free-degree-program-excelsior-scholarship.

[6] 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 scores, so that 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 re-test 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.

[7] We were unable to find any reference to the cutoff online, except as reported in Castleman and Long, and that figure is more than 10 years old.

[8]Georgia Student Finance Commission, GACollege411, Georgia Hope Scholarship Program Overview. Retrieved May 29, 2013 from https://secure.gacollege411.org/Financial_Aid_Planning/HOPE_Program/Georgia_s_HOPE_Scholarship_Program_Overview.aspx.

[9]Tennessee Student Assistance Corporation, Tennessee Hope Scholarship. Retrieved May 29, 2013 from http://www.tn.gov/collegepays/mon_college/hope_scholar.htm.

[10]Florida Department of Education, Office of Student Financial Aid, Florida Student Scholarship and Grant Programs, Chart of Eligibility and Award Criteria. Retrieved May 29, 2013 from http://www.floridastudentfinancialaid.org/ssfad/PDF/BFEligibilityAwardChart.pdf.

[11]1995-1996 Federal Pell Grant Program End-of-Year Report, U.S. Department of Education, online at https://www2.ed.gov/finaid/prof/resources/data/pell-historical/pell-eoy-1995-96.pdf (accessed 26 March 2017).

[12] King, Jacqueline E. "Missed Opportunities: Students who do not Apply for Financial Aid," American Council on Education Issue Brief, 2004. Online at http://www.soe.vt.edu/highered/files/Perspectives_PolicyNews/10-04/2004FAFSA.pdf (accessed 1 April 2017).

[13] For degree earning students entering UNM between 1994 and 1999, average

time to degree was 4.79 years with a standard deviation of .66 years.

[14] 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 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.

[15] There are 9,979 resident students and only 1,233 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, estimates would (at most) be based on 1,233 matches, or 2,466 observations, which constitutes approximately 22 percent of the sample.

[16] Results are similar when we split the sample around the median high school GPA.

[17] We also used New Mexican residents who delayed enrollment (and were therefore not eligible for NMLLS) as a control group. These non-traditional 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 non-traditional

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.

[18] US. Federal Housing Finance Agency, All-Transactions House Price Index for Albuquerque, NM (MSA), Federal Reserve Bank of St. Louis. Retrieved September 22, 2016 from https://fred.stlouisfed.org/series/ATNHPIUS10740Q. July 1994 to July 1999 is examined. Prices are adjusted for inflation using BLS's CPI less shelter measure (Series CUUR0000SA0L2), retrieved 18 Aug 2017 from https://data.bls.gov/cgi-bin/srgate.

[19] The $215,000 figure is the median listed home price in Albuquerque according to Zillow.com as of June 30, 2017. Obtained 18 Aug 2017 at https://www.zillow.com/albuquerque-nm/home-values/.

[20] Calculations using Bureau of Economic Analysis' annual personal income estimates for 1994 and 1999 (SA1 Personal Income Summary: Personal Income, Population, Per Capita Personal Income) and the Bureau of Labor Statistics' annual CPI-U estimates for 1994 and 1999 (Series CUUR0000SA0). Retrieved 18 Aug 2017 from https://www.bea.gov/itable/ and https://data.bls.gov/cgi-bin/srgate, respectively.

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

Chapter 3: Wage Effects of Baccalaureate Time to Degree in the United States

Only 42 percent of students earning baccalaureate degrees in the United States graduate within four years, compared to 53 percent three decades ago. Despite this shift, and plenty of concern about potential harm to students on behalf of state legislators and university officials, we know very little about whether delayed graduation carries a labor market penalty. Researchers examining time to degree using cross-sectional data report a negative relationship between time to degree and earnings, which presumably reflects ability. Increases in time to degree, however, cannot reasonably be linked to lower ability over time. Using two nationally-representative longitudinal studies, we proxy for student ability and instrument for time to degree, and find no evidence of a labor market penalty for delayed graduation. Moreover, the potential loss of earnings from later post-graduation entry into the labor market may easily be countered by higher earnings during school for those who take longer to finish. Together, these findings suggest that taking longer to complete college is not necessarily a problem that needs fixing.

3.1 Introduction

Most college graduates in the United States spend more than four years earning a baccalaureate degree, a fact that has drawn alarm from some researchers, policymakers, and media outlets. Among the proposed remedies are higher penalties for withdrawing from courses, course credit pricing penalizing students taking fewer than 15 credits, and endorsement of “lockstep” programs that restrict student choice in courses, making it more difficult to change majors.[22] In 2016, the Obama administration proposed two significant changes to the federal Pell Grant program. The first provision would have provided approximately 700,000 students "making real progress toward on time graduation" with an additional $1,915 on average to help pay for college and complete their degrees faster." The second provision, dubbed the "on-track Pell bonus," would have raise the maximum Pell award by $300 for approximately 2.3 million students that take 15 credits per semester in an academic year, a policy meant to encourage the receipt of a bachelor's degree in four years.[23] Critics contend that such policies overload students that necessarily work during college, or that enter higher education marginally prepared for college-level courses. Kinsey and Goldrick-Rab (2015) provide evidence that tying additional need-based aid to academic performance may only serve to slow down students, which may take fewer classes the next semester if their current semester GPA suffers. Support for such incentives remains broad, however. Backers of these measures cite the high cost to students of delaying entry into the labor market, particularly how lengthened time to degree may encourage students to take on additional debt. We are interested in whether "delayed" students incur wage penalties beyond opportunity costs associated with solely taking longer to obtain a degree. This is a contemporary issue: in 2012 alone, Idaho, Indiana, Mississippi, Missouri, and New Mexico passed legislation aimed at reducing time to degree at their public universities.

Time to degree has increased dramatically in recent decades. In the 1970s, 53 percent of college graduates earned their degrees in four years. Twenty years later, only 39 percent had done so. For non-top-50 public universities the decline was even steeper—from 50 percent in the 1970s to 29 percent in the 1990s. Researchers posit the trend cannot be explained by changes in student preparedness or composition, and instead find evidence that decreased resources per student at less selective public universities and students working more during college are likely causes (Bound *et al.,* 2012). Students who work more while enrolled in college have lower opportunity costs and it may be that the reduced opportunity costs compensate for delayed entry into the labor market. Students have always had the option of taking more credits and not changing majors, so it bears exploring whether going through college more slowly might be a good strategy, rather than a mistake.

We address this question in two steps. First, we examine a simple human capital model to explore under what circumstances combining part-time work and a 5- or 6-year path to degree attainment might be optimal. Second, we ask whether longer time to degree is penalized in the labor market. For the latter question, human capital theory holds that additional years of education increases the productivity of workers, thereby affording higher wages in the labor market. Under human capital, if students complete the same amount of credits over a longer period of time, then time to degree should have no direct effect on wages. Yet, if time to degree serves as a productivity signal to employers, then those finishing sooner versus later may be valued as being more productive in the labor market.

Several researchers report a negative association between earnings and time to degree, which they attribute to student ability. This association alone does not rule out the human capital hypothesis, since the real test is whether workers with the same ability, but different time to degree, are compensated differently. We perform this test by controlling for ability and instrumenting for time to degree with the institutional average.

We believe our instrument is plausible because institutional policies and norms surely affect a student's college trajectory, but should have no bearing on labor market rewards apart from the institution's quality, which we also control for.

Our findings suggest that concern over delayed graduation may be misplaced. Under plausible assumptions about hours worked in college, the return to a college degree, and discount rates, students may come out ahead when they work while earning a degree in five or six years. Addressing endogeneity using Two Stage Least Squares (2SLS), we find time to degree has no association with near-term labor market earnings.

3.2 Simple model of human capital

We appeal to a simplified discrete multi-stage human capital investment problem similar to Turner (2004). The framework is modified to examine the circumstances under which students rationally prefer a mixture of part-time work (i.e., 30 hours per week) and part-time school (i.e., six years to graduation), to working 10 hours per week, attending college full-time, and graduating in the "normal time" of four academic years.

Define Y_{HS} and Y_C as earnings before and after college completion, respectively. For ease of exposition, we assume students in this scenario may work 30 hours per week, earning $\frac{3}{4}Y_H$ each year, and attend school part-time for six years, paying F annually in direct costs. Students may also choose to work 10 hours per week, attend school full-time, and graduate in four years. We assume that the costs of part-time and full-time enrollment are equal. Students prefer to earn a baccalaureate degree in six years while working part time over the traditional four-year path to degree attainment if

(1) $$\frac{3}{4}\sum_{t=1}^{6}\frac{Y_{HS}}{(1+r)^t}+\sum_{t=7}^{T}\frac{Y_C}{(1+r)^t}-\sum_{t=1}^{6}\frac{F}{(1+r)^t}>\frac{1}{4}\sum_{t=1}^{4}\frac{Y_{HS}}{(1+r)^t}+\sum_{t=5}^{T}\frac{Y_C}{(1+r)^t}-\sum_{t=1}^{4}\frac{F}{(1+r)^t}$$

holds, a condition that may be reduced to

(2) $$\frac{4[Y_C+F]}{Y_{HS}} < \frac{2(1+r)^6+(1+r)^2-3}{r(r+2)}.$$[24]

Whether equation (2) holds depends on parameters in the maximization problem, including student risk preferences, direct college costs, and the returns to a baccalaureate degree. All else equal, students are more likely to pursue a nontraditional (and longer) path to college completion when 1) discount rates are high, 2) the returns to a college degree are lower, and 3) the direct costs of schooling are relatively low. For example, using figures from the Bureau of Labor Statistics, let Y_{HS} = \$34,600 and Y_C = \$57,800.[25] To demonstrate how time to degree may vary, we choose one non-top 50 public university (University of New Mexico) and one top 50 public university (University of Washington) where F is equal to \$7,146 and \$10,974, respectively. Assuming a fairly standard discount rate of $r = .05$, the simplified model results in the UNM student preferring to take six years while working part-time over the traditional four-year path working ten hours per week. Because direct college costs are larger, the UW student prefers the traditional four-year path under these conditions. However, if we increase the discount rate to $r = .10$ then both the UNM and UW students prefer the nontraditional path to degree attainment. Note that this simplification of human capital assumes no uncertainty regarding the costs and benefits of college, and does not consider the possibility of binding credit constraints. However, the exercise demonstrates that students may rationally choose a longer, nontraditional path to degree attainment under reasonable assumptions.

3.3 Is delayed graduation punished in the labor market?

Existing literature does not answer the question of whether lengthened time to degree penalizes workers. Previous studies do not control for one confounding factor or another: none control for institutional characteristics, which likely impact both time to degree and earnings after graduation (Groot and Oosterbeek, 1994; Brodaty *et al.*, 2009; Flores-Lagunes and Light, 2010; Aina and Pastore, 2012). Only Groot and Oosterbeek (1994) include a proxy for student ability. Without adequately controlling (or instrumenting) for such factors, it is impossible to isolate the effect of time to degree in the earnings function. For example, low ability students generally earn less money than their high ability counterparts, but will still increase their earnings by obtaining a degree. Moreover, students that attend lower quality institutions may earn less than those attending higher quality institutions, however it may be the characteristics of the institution that are contributing to the wage penalty, not necessarily the time it takes the student to complete. It is thus unlikely that previous studies are able to fully separate student and institutional characteristics from time to degree.

To operationalize our investigation of wages and time to degree, we present a linear model of wages for baccalaureate degree earners closely following that of Brodaty *et al.* (2009). Let d be the graduate's time to degree in months. Subscripts are omitted for ease of exposition. Graduation delay, D, is defined as the individual's time to degree less "normal time" to degree, defined as 45 months, so that $D \equiv d - 45$.[26] We assume that an individual's productivity is given by q and takes the form

(4) $$ln(q) = a_0 S + b_0 D + \boldsymbol{X} c_0 + \theta_1 + \theta_2$$

where $\boldsymbol{X}$ is a vector of controls (including potential experience and its square, race, ethnicity, and gender) observed by both the researcher and the employer. Graduate degree attainment is given by S, indicating receipt of a master's or doctoral degree at the time of follow-up. The direct productivity effect of schooling is given by a_0, which we expect to be nonnegative. Similarly the direct productivity effect of graduation delay is b_0, expected to be nonnegative since it can be viewed as a measure of age or maturity. The terms θ_1 and θ_2 measure student ability, the former being observed only by the employer, the latter unobserved by both the researcher and the employer. Ability measures are assumed to be normally distributed with mean zero, finite variances, and nonnegative covariances. Employers observe graduation delay, given by

(5) $$D = \boldsymbol{X}c_1 + \boldsymbol{Z}g_1 + f_1\theta_0 + \xi_1$$

where θ_0 is another unobserved measure of student ability that is likely positively correlated with θ_1 and θ_2. The vector $\boldsymbol{Z}$ captures exogenous sources of variation in graduation delay not observed by the employer, assumed to be uncorrelated with θ_j for all j. These instruments affect graduation delay but are not reported in job application materials nor are they otherwise observable by the potential employer.[27] To identify a causal link between time to degree and subsequent wages, graduation delay is instrumented using the ratio of six- to four-year graduation rates at the student's degree granting university.

If employers set wages equal to the expected productivity of workers, conditional on their information set, wages can be expressed as

(6) $w = E[q|\boldsymbol{X}, S, D, \theta_1]$.

We assume that wages are distributed as log-normal. Since ability and error terms are normally distributed, $ln(q)$ conditional on $(\boldsymbol{X}, S, D, \theta_1)$ is also normally distributed. Thus, we may write

(7) $w = exp\{E[ln(q)|\boldsymbol{X}, S, D, \theta_1] + \frac{1}{2}Var[ln(q)|\boldsymbol{X}, S, D, \theta_1]\}$.

Normal vectors exhibit the property that the conditional variance $Var[ln(q)|\boldsymbol{X}, S, D, \theta_1]$ does not depend on $(\boldsymbol{X}, S, D, \theta_1)$, meaning we can treat it as a constant and include it in $\boldsymbol{X}$. This gives us the equation

(8) $ln(w) = a_0S + b_0D + \boldsymbol{X}c_0 + \theta_1 + E[\theta_2|\boldsymbol{X}, S, D, \theta_1]$.

Having assumed that ability is normally distributed, conditional expectations are linear and given by

(9) $E[\theta_2|\boldsymbol{X}, S, D, \theta_1] = a_3S + b_3D + \boldsymbol{X}c_3 + f_3\theta_1$

which can be substituted in equation (8) to arrive at a modified version of Mincer's (1974) human capital earnings function

(10) $ln(w) = (a_0 + a_3)S + (b_0 + b_3)D + \boldsymbol{X}(c_0 + c_3) + (1 + f_3)\theta_1$.

Note that graduation delay appears in the earnings function as it is a signal conveying information about unobserved productivity, θ_2. The coefficient on graduation delay is composed into two components: b_0, the direct productivity effect of graduation delay and b_3, the signaling effect of graduation delay, each which are not identified individually. We infer whether results support human capital or the screening hypothesis based on the sign of $(b_0 + b_3)$. As mentioned, we expect b_0 to be nonnegative in sign. Thus, if the coefficient on graduation delay is negative, this constitutes evidence of time to degree communicating a negative productivity signal. A nonnegative coefficient on graduation delay is considered support for human capital theory.

3.4. NCES data

We use ELS:2002 data for our analysis, which allows us to observe students' secondary, postsecondary, and subsequent labor market outcomes for the 2004 graduating high school cohort. The ELS:2002 documentation explicitly notes the survey was intended to inform policymakers of the "rate of progress through postsecondary curriculum" and the "social and economic rate of return on education to both the individual and society" (Ingalls *et al.* 2014, pp. 10). The dependent variable in our analysis is the natural log of hourly wages in 2011 dollars. This measure is obtained from sample members at the fourth and final follow-up. The final follow-up takes place eight years after students' high school graduation cohort date, defined as June 1, 2004. We limit the sample to college students who earned a high school diploma and enrolled in college within two years. This includes students who did not graduate high school in normal time. Because the survey only follows students eight years after their expected high school graduation, it does not permit analysis of nontraditional students such as those that matriculate in their late 20s or later.

Bound *et al.* (2012) find that trends in time to degree across 1972 and 1992 graduating high school cohorts vary significantly according to the student's first institution type. The authors classify students' first institutions into five categories: non-top 50 public colleges, top 50 public colleges, less selective private colleges, highly selective private colleges, and community colleges. This categorization is based on 2005

U.S. News & World Report college rankings. Highly selective private colleges include the top 50 ranked private colleges, the top 65 ranked liberal arts colleges, and four U.S. Armed Services Academies: the U.S. Military Academy at Westpoint, the U.S. Naval Academy, the U.S. Coast Guard Academy, and the U.S. Air Force Academy. We use the same categorization scheme as Bound *et al.* (2012), with the exception of excluding students that started at community colleges since they do not have selectivity or Integrated Postsecondary Education Data System (IPEDS) data.[28]

Several variables are included in wage equations in order to help isolate the effect of time to degree on earnings. Standardized test scores, measured in terms of the composite ACT score, are included to capture observed student ability.[29] Higher standardized test score achievement is expected to be positively correlated with future earnings (Betts and Grogger, 2003). Potential work experience and its square are included, as workers earn more as they acquire additional labor market experience, but at a diminishing rate (Mincer, 1974; Heckman *et al.*, 2003; Lemieux, 2006). Variables capturing whether a post-baccalaureate degree has been earned are included, as additional credentials are expected to increase future earnings (Card, 2001). Also included are controls for respondent gender, race, and ethnicity. A set of state dummies account for heterogeneity across labor markets in the United States. Institution quality is proxied using the 2004 Barron's Admissions Competitiveness Index. We instrument the student's time to degree with the ratio of six- to four-year graduation rates at the student's first institution. These rates are reported in the IPEDS. This measure is intended to capture the prevalence of graduation beyond normal time at the institution-level.[30] See Table 3.2 for descriptive statistics of the instrument for various institution types.

Table 3.11 presents the cumulative distribution of time to degree as well as its mean for ELS:2002 data. We include data from Bound *et al.* (2012) in order to examine whether trends in time to baccalaureate degree have persisted using this most recent NCES survey. Table 3.1 suggests that the overall mean time to degree has not changed significantly across 1992 and 2004 graduating high school cohorts, standing most recently at 4.83 years. Average time to degree for those at non-top 50 public colleges held steady at 4.93 years across 1992 and 2004 cohorts. At top 50 public schools there was a marked decrease in time to degree from 4.66 to 4.42 years, an average difference of approximately three months. The percent of graduates completing in four years or less at these schools increased from approximately 40 to 57 percent. There were small increases (decreases) in time to degree for students starting at highly (less) selective private institutions. Overall it seems that the alarming increases in time to baccalaureate degree found by Bound *et al.* (2012) may have slowed if not stabilized.

Table 3.2 presents descriptive statistics by first institution type for variables included in the analysis. The average hourly wage for the full sample is $20.57, with wages highest for those starting at highly selective private institutions, and lowest for those starting at non-top 50 public institutions. Graduation delay varies widely according to first institution type. It averages just under one year for students at non-top 50 public universities but is only approximately 3.5 months for students at highly selective private schools. Figure 3.1 presents a histogram of graduation delay, revealing that roughly 45 percent of students in the sample graduated within six months of normal time. The largest spike in delay is between zero and six months after the 45-month mark, where the distribution decays thereafter. The instrumental variable, time to degree ratio, averages roughly two for the full sample. This is interpreted as having twice as many students graduating within six years relative to four years. We see similar variation by institution type compared to the graduation delay variable. Approximately 24 percent of the full sample had a master's degree at the last follow-up, while six percent held a doctoral degree. ACT composite scores are lowest for students starting at non-top 50 public institutions, and highest for those starting at highly selective private institutions. The majority of the sample consists of white women.

3.5. Results

3.5.1 Hausman Test for Endogeneity

We first wish to examine whether it is even necessary to use 2SLS in estimating wage penalties from delayed baccalaureate graduation. This can be done through implementing Hausman's (1978) test for endogeneity. This test offers a formal way of examining whether the error term in the earnings function is correlated with our measure of graduation delay. To conduct this test, we estimate the reduced form equation for graduation delay and save the fitted residuals from this equation. We then estimate the earnings function including all of exogenous variables, graduation delay, and the fitted residual. Rejecting the null hypothesis that the coefficient on the fitted residual is not statistically different from zero suggests that graduation delay is endogenous, and 2SLS must be used. We estimate the coefficient on the fitted residual to be 0.1003 with a standard error of 0.0385, providing strong evidence that endogeneity of the time-to-degree variable in the wage equation is a problem that must be dealt with.

3.5.2 Instrument Relevance and Instrument Exogeneity

Valid estimation via 2SLS requires two conditions be met. First, the instrument should be highly correlated with the endogenous variable. This can be assessed using a simple *t*-test on the instrument in the first stage equation. Second, the instrument should only affect the dependent variable through the endogenous variable. Because the latter requires knowledge of the true model error, this requirement cannot be tested and instead must be maintained. Table 3.3 presents results for the full sample. We note that the first stage regression suggests strong positive correlation between time to degree ratio and graduation delay variables. A one-unit increase in the time to degree ratio increases graduation delay by just less than one month. We reject the null of the simple *t*-test at the one percent level, providing evidence of instrument relevance. An F-statistic of 11.71 in the first stage suggests the 2SLS model does not suffer from weak instruments.

Instrument exogeneity requires that time to degree at the student's institution only affects earnings through the student's time to degree. We suspect that time to degree at the student's institution affects the student's own time to degree through what could be considered a sort of "peer effect." If a large proportion of one's peers in college are planning on overshooting normal time, the student may be more likely to consider this a valid path to degree attainment.[31] There are many other reasons why this relationship may hold as well. It may reflect institution quality in some broad sense, the resources available to the student, or the additional tuition costs associated with delayed graduation.

This motivates us to include the admissions competitiveness index in the models in order to capture many broad measures of institution quality, helping to isolate the causal path from institution time to degree to wages through the student's own time to degree.

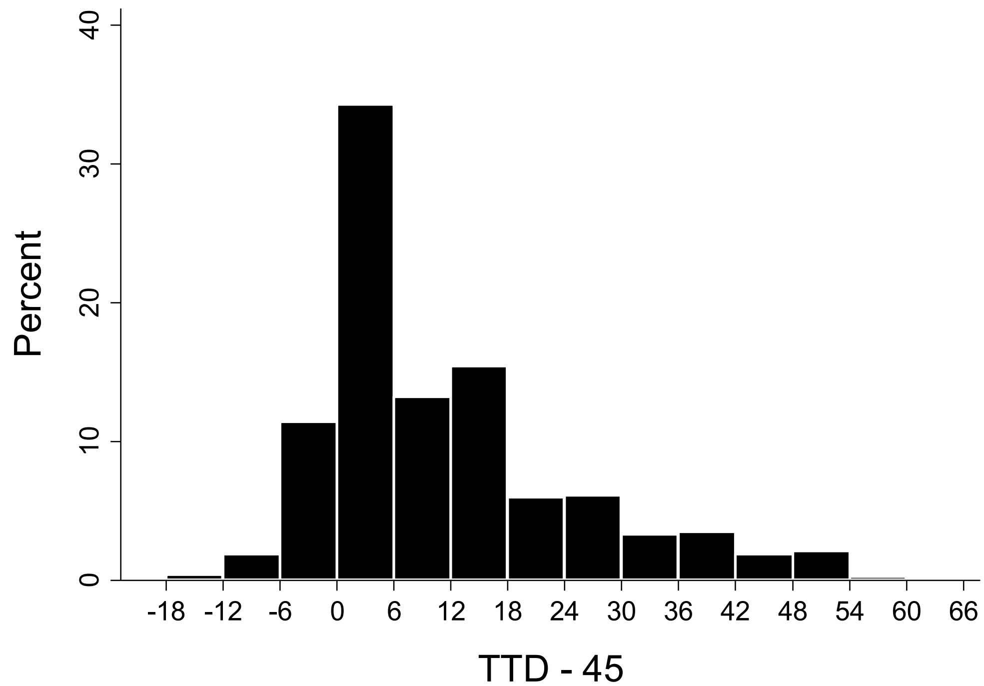
First stage results show that higher ability translates into shorter time to degree, consistent with Flores-Lagunes and Light (2010). Women were found to be less likely to exceed graduation in normal time, while Hispanic students were shown to have longer time to degree. We interpret admissions competitiveness dummies as relative to schools deemed as "Most Competitive." Coefficients on these measures suggest that the less competitive the school, the higher the time to degree. Wage equation estimates also bear many features that one would expect from an earnings function. We see large positive returns to obtaining a master's and doctoral degree. Wages are increasing in experience, but at a decreasing rate. Higher ability results in higher wages—a ten point increase in the composite ACT score results in a ten percent increase in wages. Results reveal a 5.4 percent wage penalty for women compared to men, a 9.4 percent for black workers compared to white. Students starting at less competitive colleges earn lower wages.

Most importantly, Table 3.3 reveals a pattern we find repeatedly—while OLS estimates find that a one month delay in graduation results in wage penalty of approximately 0.5 percent (approximately six percent for one year of delay), we find no evidence of any wage penalty after instrumenting for graduation delay. OLS estimates, of which previous studies report similar effect sizes, are clearly misleading. Table 3.4 provides evidence that results do not differ by the student's first institution type. Non-top-50 public schools, top-50 public schools, and less selective private schools again reveal one-month graduation delay penalties of 0.5-0.6 percent (6-7 percent per year), while 2SLS reveal no penalty. For highly selective private schools, both OLS and 2SLS results are statistically insignificant.

Overall, preliminary findings support the human capital hypothesis that earning the same degree over a varying length of time has no effect on the returns to a college degree. In other words, we find no evidence that time to degree serves as a productivity signal to prospective employers.

3.6 Conclusions

Time to degree may be costly for students, institutions, and society at large. Our results provide evidence that delaying one's graduation does not result in any sort of wage penalty. It is hoped this study will inform policymakers on the costs and benefits of lengthened time to degree, especially those at institutions currently considering or actively discouraging alternative paths to degree completion which take longer than normal time. Reducing time to degree, which appears to have been taking place at some institutions since the study of the 1992 high school cohort, may free up additional resources for new students. Our results provide evidence that utility-maximizing students may be better off pursuing a longer path to degree attainment. It is important to then consider the growing proportion of nontraditional students in higher education, and to promote policies accommodating their rational decision to work during school and take fewer credit hours per semester, in contrast to supporting policies which penalize students for not remaining on track to graduate in four years—policies which may ultimately hurt their very chances of completing college at all.



Source: Education Longitudinal Study of 2002. *TTD* is the student's time to degree in months. The histogram displays the distribution of graduation delay using six month bins. Approximately 45 percent of students in the sample graduated with six months of normal time.

Figure 3.1 Histogram of graduation delay, baccalaureate earners, ELS:2002

Table 3.1 Time to degree (TDD) distributions for all graduates by first institution type

	TTD Distribution				Mean
	4	5	6	7	TTD
Full Sample:					
NLS72	53.1	81.8	90.6	96.3	4.48
NELS:88	39.4	72.7	88.3	94.7	4.81
ELS:2002	42.3	72.1	85.7	93.5	4.83
Non-top 50 public:					
NLS72	49.7	82.3	91.1	96.3	4.49
NELS:88	29.1	68.8	87.8	95.1	4.93
ELS:2002	34.2	68.5	85.0	94.1	4.93
Top 50 public:					
NLS72	52.7	81.5	89.2	96.4	4.49
NELS:88	39.7	82.0	93.7	96.6	4.66
ELS:2002	56.7	85.2	95.2	98.1	4.42
Less selective private:					
NLS72	66.7	87.3	94.0	98.7	4.28
NELS:88	58.0	84.6	93.4	98.6	4.60
ELS:2002	56.1	83.4	92.5	96.1	4.51
Highly selective private:					
NLS72	65.2	88.2	93.8	96.8	4.31
NELS:88	73.1	91.9	98.1	99.8	4.20
ELS:2002	68.6	91.7	96.3	98.2	4.28
Community college:					
NLS72	36.5	67.8	83.0	92.6	4.90
NELS:88	15.5	44.2	70.8	83.6	5.58
ELS:2002	16.5	43.9	64.4	81.6	5.69

Note: NLS72 and NELS:88 figures reproduced from Bound, Lovenheim, and Turner (2012). ELS:2002 calculations were made using third follow up panel weights. In each survey the sample includes baccaluareate-earners enrolling at a postsecondary institution with two years of their high school cohort graduation month. High school cohort graduation month is assumed to be June 1972 for NLS72, June 1992 for NELS:88, and June 2004 for ELS:2002. Students were followed for eight years following their high school cohort graduation month.

Table 3.2 Sample characteristics of employed college graduates in the ELS:2002

	All Institutions		Non-Top 50 Public		Top 50 Public		Less Selective Private		Highly Selective Private	
Variable	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Hourly Wage	20.574	11.165	19.542	10.011	21.894	11.930	19.826	10.884	23.844	13.636
Graduation Delay	7.872	12.222	11.173	13.214	5.401	10.559	5.607	11.132	3.494	9.119
Time to Degree Ratio	2.043	1.472	2.593	1.031	1.76	0.351	1.532	2.451	1.343	0.633
Master's	0.239		0.204		0.280		0.271		0.245	
Doctorate	0.059		0.033		0.099		0.055		0.099	
Experience	3.241	1.186	3.368	1.069	3.098	1.327	3.163	1.177	3.147	1.327
ACT Composite Score	24.200	4.362	22.614	3.891	25.794	3.860	23.858	4.187	28.414	3.562
Female	0.564		0.563		0.544		0.593		0.554	
White	0.767		0.751		0.749		0.811		0.777	
Hispanic	0.063		0.057		0.064		0.071		0.070	
Black	0.077		0.113		0.046		0.060		0.023	
American Indian	0.004		0.002		0.005		0.003		0.006	
Asian	0.051		0.039		0.083		0.027		0.090	
Two or More Races	0.037		0.036		0.053		0.029		0.031	
Nat. HI/Pac. Islander	0.001		0.003		0.001		0.000		0.000	
Barron's - Most Competitive	0.080		0.005		0.119		0.000		0.477	
Barron's - Highly Competitive	0.129		0.031		0.296		0.094		0.293	
Barron's - Very Competitive	0.294		0.193		0.444		0.394		0.225	
Barron's - Competitive	0.392		0.602		0.141		0.391		0.005	
Barron's - Less Competitive	0.058		0.103		0.000		0.051		0.000	
Barron's - Noncompetitive	0.027		0.052		0.000		0.017		0.000	
Barron's - Special Designation	0.004		0.001		0.000		0.017		0.000	
Observations	3297		1369		713		780		435	

Source: Education Longitudinal Study of 2002. Descriptives are shown for the full sample and for each first institution type. Time to degree is measured in months and centered at 45, the time it takes to complete a bachelor's degree in "normal time." Third follow-up panel weights are used in calculations. Institutions were categorized using 2005 *US News and World Report* rankings. Institution characteristics are for the first institution attended. Experience is measured in years. ACT composite scores also include SAT composite scores that have converted by NCES.

Table 3.3 OLS and 2SLS wage models of graduation delay penalty, all institutions

		OLS	2SLS
VARIABLES	(1) Graduation Delay	(2) Wages	(3) Wages
Graduation Delay		-0.005***	0.006
		(0.001)	(0.009)
TTD Ratio	0.778***		
	(0.146)		
Master's		0.204***	0.258***
		(0.064)	(.0079)
Doctorate		0.537***	0.575***
		(0.130)	(0.138)
Experience	0.839	0.198**	0.224**
	(0.624)	(0.090)	(0.095)
Experience Squared	0.274**	-0.021*	-0.026**
	(0.123)	(0.012)	(0.013)
ACT Composite	-0.424***	0.005**	0.010**
	(0.055)	(0.002)	(0.005)
Female	-2.900***	-0.089***	-0.054*
	(0.406)	(0.018)	(0.033)
Hispanic	1.999**	-0.016	-0.037
	(0.841)	(0.037)	(0.042)
Black	0.735	-0.086**	-0.094**
	(0.792)	(0.035)	(0.037)

Table 3.3 OLS and 2SLS wage models of graduation delay penalty, all institutions (continued)

		OLS	2SLS
	(1)	(2)	(3)
VARIABLES	Graduation Delay	Wages	Wages
American Indian	-1.198	-0.049	-0.034
	(3.324)	(0.148)	(0.152)
Asian	1.042	-0.000	-0.012
	(0.918)	(0.041)	(0.043)
Two or More Races	-0.295	-0.050	-0.046
	(1.042)	(0.046)	(0.048)
Hawaiin or Pacific Islander	-3.030	0.152	0.179
	(5.235)	(0.232)	(0.240)
Highly Competitive	0.940	-0.003	-0.014
	(0.855)	(0.038)	(0.040)
Very Competitive	1.717**	-0.001	-0.023
	(0.774)	(0.034)	(0.039)
Competitive	2.883***	-0.047	-0.087*
	(0.802)	(0.035)	(0.048)
Less Competitive	7.943***	0.010	-0.085
	(1.147)	(0.051)	(0.091)
Non-Competitive	3.428**	-0.095	-0.144*
	(1.444)	(0.064)	(0.076)
Special Designation	0.894	-0.018	-0.030
	(3.030)	(0.135)	(0.139)
Observations	3,297		
F-statistic	11.71	5.10	4.35
Adjusted R-squared	0.198	0.100	0.044

Source: Education Longitudinal Study of 2002. Robust standard errors are reported below estimated coefficients. The dependent variable is the natural log of hourly wages in 2011, at the third follow-up. Time to degree is in months and centered at 45, the time it takes to complete a bachelor's degree in "normal time." All specifications include state fixed effects. In 2SLS specifications, the student's time to degree is instrumented by the ratio of six- to four-year graduation rates at their university.

Table 3.4 Estimates of graduation delay penalty by first institution type

	Non-Top 50 Public		Top 50 Public		Less Selective Private		Highly Selective Private	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	OLS	2SLS	OLS	2SLS	OLS	2SLS	OLS	2SLS
Time-to-Degree	-0.005	0.005	-0.006	-0.001	-0.006	-0.009	0.002	0.012
	0.001	0.012	0.002	0.010	0.002	0.045	0.004	0.015
Observations		1369		713		780		435
Adjusted R-Squared	0.136	0.090	0.221	0.208	0.174	0.170	0.241	0.227

Source: Education Longitudinal Study of 2002. Robust standard errors are reported below estimated coefficients. The dependent variable is the natural log of hourly wages in 2011, at the third follow-up. Time to degree is in months and centered at 45, the time it takes to complete a bachelor's degree in "normal time." All specifications include state fixed effects. In 2SLS specifications, the student's time to degree is instrumented by the ratio of six- to four-year graduation rates at their university.

Notes

[22] See, for example, the "15 to Finish" policy promoted by Complete College America and other nonprofits. Online at http://completecollege.org/docs/GPS_Summary_FINAL.pdf, accessed 13 November 2016. As of 2013, five statewide higher education systems and at institutions in fifteen states had adopted 15 to Finish. This information is online at http://www.completecollege.org/news.html, accessed 13 November 2016.

[23] U.S. Department of Education, Fact Sheet: Helping More American Complete College: New Proposals for Success, released 19 January 2016, online at http://www.ed.gov/news/press-releases/fact-sheet-helping-more-americans-complete-college-new-proposals-success, accessed 13 November 2016.

[24] See Appendix 4.A for the derivation of equation (3).

[25] Unemployment rates and earnings by educational attainment, 2016. Bureau of Labor Statistics, online at https://www.bls.gov/emp/ep_chart_001.htm (accessed 22 March 2018).

[26] Forty-five months was chosen as it represents a "four-year" stay from fall in year one to spring in year four. A histogram of time to degree in months is presented below which appears to support this choice.

[27] As demonstrated in Brodaty *et al.* (2009), this assumption may be relaxed without loss of generality.

[28] See Appendix 4.B for a detailed list of which colleges fall in each category.

[29] The National Center for Education Statistics provides a standardized test score variable in the ELS:2002 which includes all composite ACT scores and includes an equivalent score in terms of composite ACT for students that chose to instead take the SAT.

[30] This measure was chosen because average time to degree is not available at the institution-level in the IPEDS.

[31]As one respondent told Bowen, Chingos, and McPherson (2009), graduating in four years is like “leaving the party at 10:30pm.”

Chapter 4: Merit Aid Scholarships and Human Capital Production in STEM: Evidence from New Mexico

The New Mexico Legislative Lottery Scholarship is a broad, "low-bar," state lottery-funded scholarship designed to increase access to higher education on behalf of New Mexico residents. The natural experiment of a state lottery scholarship is used to measure the effect of generous financial aid on major choice at New Mexico's flagship public university. A potential unintended consequence of state merit aid scholarships is to discourage the production of human capital in science, technology, engineering, and mathematics (STEM) fields. This may occur if students avoid more rigorous majors in order to increase the likelihood of scholarship retention. I find no evidence that the scholarship decreased the overall likelihood that a student first declares a STEM major or earns a STEM degree. There are significant effects when disaggregating by academic preparation: less-academically prepared entering freshmen are 6.8 percentage points (40 percent) less likely to initially declare a STEM major, while more-academically prepared entering freshmen are 12.1 percentage points (44.3 percent) more likely to initially declare a STEM major. No significant effects are found when examining whether a STEM degree was earned. Evidence suggests these effects are at least in-part due to compositional changes in the student body before and after the advent of the lottery scholarship.

4.1 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. I analyze the effect of the New Mexico Legislative Lottery Scholarship (NMLLS), a uniquely "low-bar" merit-based scholarship, on student major choice. Specifically, this paper is interested in two related research questions. First, do generous, low-bar merit scholarships discourage students from choosing majors in science, technology, engineering, and mathematics (STEM)? Second, do such scholarships affect the number of STEM degrees produced?

The major focus of this paper is on the first research question. Since merit-based scholarships require students to maintain a set level of academic achievement in order to continue to receive aid, there are potential unintended consequences that may occur, including dissuading students from studying more difficult subjects, including those categorized as STEM. The consequences of this outcome may be significant to economic interests at both the state and national levels, as STEM occupations are often seen as major drivers of innovation, and well as key to economic growth.

Since 1993, 27 states have implemented merit-based scholarships, the first and most studied being Georgia's Helping Outstanding Pupils Educationally, or HOPE, scholarship program.[32] HOPE marked the beginning of what has been a major restructuring of the financial aid landscape in America. 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 2013-2014 academic year, New Mexico was one of 13 states where this percentage was below 40 percent.[33]

We know more about the relationship between financial aid, enrollment, and degree completion than financial aid and major choice. Different types of financial aid have varying effects on college enrollment. Loans tend to have little to no effect, while grants have a positive and significant effect on student enrollment (Linsenmeier *et al.* 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 are sensitive to financial aid offers. Several studies show a significant and positive relationship between grant aid and student enrollment (Seftor and Turner 2002; Kane 2003; Heller 2009) 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 four percent. Also studying HOPE, Cornwell *et al.* (2006) find the program increased student enrollment by 6 percent. In New Mexico, Binder and Ganderton (2002, 2004) find that while the NMLLS boosted enrollment at four-year colleges in New Mexico, the effect appears to be driven by additional enrollment of students that otherwise would have attended college out-of-state. The effect of merit aid on college completion has also been studied.

Analyzing statewide educational attainment data, Sjoquist and Winters (2012, 2015b) found no difference in college attainment for those exposed to lottery scholarship programs. Using a similar methodology, Jia (2017) 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 cut-off for West Virginia's lottery-funded PROMISE scholarship program, compared to students just below. Using similar strategies, Bruce and Carruthers (2014) and Welch (2014) found no program effect for Tennessee's lottery scholarship. Erwin and Binder (see Chapter 2) found no overall effect of generous, low-bar merit aid on college completion. Divergent effects appeared when disaggregating the sample by academic preparation. Less-academically prepared students appeared to exhibit lower completion rates as a result of the scholarship while more-academically prepared students exhibited higher completion rates, two significant effects similar in magnitude but opposite in sign. The authors argue that changes in student composition are potentially driving results.

I examine how the NMLLS affects STEM engagement at the University of New Mexico (UNM) by exploring changes in 1) the likelihood of initially declaring a STEM major and 2) the likelihood of earning a baccalaureate degree in a STEM field 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 NMLLS on declaring a STEM major or earning a STEM degree. However, there are large and statistically significant completion effects after disaggregating by academic preparation. Academically less-prepared eligible freshmen are 6.8 percentage points (40 percent) less likely to first declare a STEM major, while academically more-prepared freshmen are 12.1 percentage points (44.3 percent) more likely to first declare a STEM major, compared to ineligible peers with similar high school GPAs. In addition, there is evidence that some program effects may be a result of the NMLLS inducing compositional changes in the student body.

The paper proceeds as follows: Section 2 discusses existing literature regarding merit-aid and major choice, and introduces the NMLLS; Section 3 presents a theoretical model of major choice; 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 found in the results; and Section 8 concludes.

4.2 Merit aid and major choice

The natural experiment of lottery-financed merit-based aid programs provides a promising avenue for determining the relationship between aid and major choice. Several studies have analyzed how students sort into different majors. An early study of this behavior can be found in Berger (1988). Berger uses a life cycle approach that assumes students choose majors based on the expected discounted stream of future earnings rather than beginning wages following graduation. The author provides evidence to support this approach using data from the National Longitudinal Survey of Young Men. Montmarquette *et al.* (2002) relax two assumptions common in previous literature, including Berger (1988): uniform probabilities of success across majors and constant earnings streams across majors. Using the National Longitudinal Survey of Youth, the authors estimate the probability of success across different majors for all students in the sample. These data are combined with estimates of predicted future earnings in all majors from Rumberger and Thomas's (1993) analysis of the 1987 Survey of Recent College Graduates, which allows the construction of a multinomial logit model of major choice. Results suggest that one's expected earnings stream is the most significant factor influencing major choice, yet the probability of success is an important factor as well. Arcidiacono *et al.* (2012) argues that both expected earnings and students' perceived abilities across majors are important determinants of major choice.

Arcidiacono (2004) estimates a dynamic model of college major choice, finding that even after controlling for selection, large earnings premiums and ability differences still exist for some majors. Differences in monetary returns explain little of the ability sorting across majors. Instead, Arcidiacono (2004) provides evidence that virtually all ability sorting is due to differences in preferences for taking particular majors in college and workplace preferences for jobs likely to be obtained after graduation, the former being more influential than the latter. Similarly, Beffy *et al.* (2010) find a small, but statistically significant, positive earnings elasticity of major choice, suggesting that nonpecuniary factors are a large part of major choice (e.g., preferences for workload, workplace conditions, opportunities field research, *et cetera*.).

Focusing on STEM fields, Wang (2013) finds that choosing a STEM major is positively related to high school performance, as well as initial college performance/experiences. Similarly, Griffith (2010) finds that differences in academic preparation and educational experiences drive differences in persistence rates in STEM majors. Wiswall and Zafar (2015) find that while expected earnings and perceived ability play a major role in choosing STEM, unobserved tastes are the largest factor in major choice. Henry and Rubenstein (2002) argue that merit aid may result in greater effort on behalf of high school students, thus better preparing students for difficult majors such as those included in STEM.

Four studies directly examine the relationship between merit aid and major choice. Analyzing Georgia's HOPE Scholarship, Cornwell *et al.* (2006) use administrative data to compare qualified residents and nonqualified nonresident students in a difference-in-differences framework. The authors find that HOPE resulted in a small 1.2 percentage point increase in the likelihood that residents chose education majors, relative to their nonresident counterparts. Cornwell *et al.* (2006) do not find any meaningful change in the likelihood that students chose STEM majors due to the advent of HOPE. Both Dynarski (2000) and Cornwell *et al.* (2006) find evidence that state merit-based scholarships increase the likelihood that highly-academically prepared students stay in-state for college, and thus affect the type and quality of institutions attended. This implies that crowding out of moderately-academically prepared students may occur as competition increases within more difficult majors.

Using Integrated Postsecondary Education Data System files, Zhang (2011) examines whether Georgia's HOPE Scholarship and Florida's Bright Futures Scholarship affected the likelihood that students embarked on a course of study within STEM fields. Zhang uses differences-in-differences estimation for aggregate state data, and finds a statistically significant 1.6 percentage point (11.4 percent) increase in the proportion of degrees classified as STEM at private institutions in Florida, but no broader effect of merit aid on STEM degree completion in either Florida or Georgia. Two significant problems should be noted with this approach. Since the unveiling of such programs affects how students sort into institutions, it is difficult to distinguish compositional change from real program impact (see Chapter 2, for example). Also, asymptotic refinement should be applied in situations where there are relatively few treated units or policy changes in a difference-in-differences framework (Conley and Taber, 2011).

Stater (2011) uses administrative data from three large public universities to examine the relationship between tuition and financial aid on the first major a student declares. He finds that larger net tuition results in students being more likely to choose professional fields such as architecture, business, or law and less likely to declare majors in humanities and sciences. Merit aid was shown to increase the likelihood of declaring majors in humanities and sciences, while having a negative effect on social sciences. It is difficult to view these estimates as causal, however, since Stater does not address the endogeneity of merit aid: students that receive merit aid are better academically prepared for college. Thus, recipients may be more likely to choose STEM majors for reasons other than merit aid.

A recent paper regarding the relationship between merit aid and major choice comes from Sjoquist and Winters (2015a). Their analysis relies on a difference-in-differences strategy using American Community Survey (ACS) microdata. They assign treatment status to individuals that were 18 years of age in a state where a merit aid program was in place, with all others assigned to the control group. Sjoquist and Winters divide the 27 adopting states into "strong" and "weak" merit aid state categories, based on their judgement of how broad-based programs are and how much funding they provide. New Mexico is defined as a strong merit aid state. Findings suggest that state merit-based scholarships reduce the rates of STEM completion. Overall, strong merit aid programs (from 9 states) were found to reduce the number of male STEM graduates by 8 percent, with no meaningful impact on women in the sample. The overall impact of merit aid on the production of STEM degrees is estimated to be -6.5 percent. The authors argue that men may be more willing to switch majors in order to retain a merit-based scholarship. Weak merit aid programs were not found to have any effect on STEM degree completion. There are several notable weaknesses in Sjoquist and Winters (2015a). The authors are also not able to control for student-level characteristics, which is important as merit aid may result in changes in student composition. Also, as noted in Jia (2017), program features matter, and vary considerably across programs. With this in mind, approaches which treat all state merit-aid programs as homogeneous are problematic.

Literature on the relationship between merit aid and major choice is not in agreement, but the most dependable studies suggest either null or negative effects on STEM degree completion. In this study, I employ a rich administrative data set from New Mexico's flagship university to revisit this question and others. The main contribution to the literature is that I control for, and disaggregate by, student-level characteristics, which allows for more detailed insight into the effects of merit aid on subpopulations. Cornwell *at al.* (2006) control for high school GPA, but do not split the sample as I do, so it's difficult to interpret how academic preparation impacts major choice. I also consider how compositional changes in academic preparation of the student body play an important role in interpreting results.

4.2.1 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 equivalency in New Mexico and enroll at a public postsecondary institution in the first regular fall or spring semester following 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 16 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 maintained a cumulative 2.5 GPA. Only 58 percent of first semester students over 1994-1999 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 1993-1994 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 the NMLLS not only supplemented rather than supplanted student aid, but drastically changed the student aid landscape throughout the state.

Compared to 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.[34] 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.[35] 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.[36]

If financial constraints are binding for students, then the NMLLS should have the desired effect of increasing access to higher education and boosting resident enrollment. However, due to low-bar initial and ongoing scholarship qualifications of the NMLLS, much of the increase in resident enrollment may be on behalf of less-academically prepared students 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).[37] Because the NMLLS is structured so that students lose the scholarship permanently if they fail to meet renewal requirements in any semester, some students may respond to merit aid by choosing easier majors which improve their chances of scholarship retention. In this case, the NMLLS could have the unintended consequence of decreasing the proportion of students choosing and ultimately completing degrees in STEM fields. However, as discussed above, renewal requirements for the NMLLS are relatively low. If students expect their probabilities of success in STEM majors are sufficiently to satisfy eligibility criteria then students may not avoid pursuing majors in STEM. The direction of any estimated program effects relies heavily on the academic preparation of resident students before and after the introduction of the NMLLS, and is ultimately an empirical question. Composition of the student body at UNM is discussed in detail below.

4.3 Merit aid and major choice

Students' choice of college major are modeled using a modified life-cycle approach developed by Montmarquette *et al.* (2002). This approach holds that students choose college majors so as to maximize lifetime utility, which depends on expected earnings and the likelihood of merit scholarship retainment. For simplicity, assume that students who are able to retain merit aid ultimately graduate with a bachelor's degree. Let p_{ij} be the likelihood of scholarship retainment for student i in major j. The expected lifetime utility for student i choosing major j, $E(U_{ij})$, is a function of predicted future earnings so that:

$$(1) \quad E(U_{ij}) = p_{ij}(\boldsymbol{X})e_{ij}(\boldsymbol{Z}) + \left(1 - p_{ij}(\boldsymbol{X})\right)e_{i0}(\boldsymbol{Z}), \;\; i = 1, \dots, N; j = 1, \dots, m,$$

where **X** includes factors influencing the probability of retaining the scholarship, including academic preparation. The vector **Z** includes factors affecting earnings after college. e_{ij} are the discounted value of lifetime earnings after completing a degree in major j and e_{i0} are discounted value of lifetime earnings after losing the scholarship and dropping out of college without a degree. Students will choose major j over major k whenever $E(U_{ij}) \geq E(U_{ik})$ for all $k \neq j$, or whenever,

$$(2) \quad p_{ij}(\boldsymbol{X})\left[e_{ij}(\boldsymbol{Z}) - e_{ik}(\boldsymbol{Z})\right] + \left[p_{ij}(\boldsymbol{X}) - p_{ik}(\boldsymbol{X})\right]\left[e_{ik}(\boldsymbol{Z}) - e_{i0}(\boldsymbol{Z})\right] \geq 0.$$

According to equation (2), if the likelihoods of retaining merit aid differ substantially across majors, and lifetime earnings differences across majors are relatively small, then success probabilities will play a major role in major choice. If likelihoods of scholarship retainment are approximately the same, then expected earnings will be the major driver in the choice of major. Thus, one would expect highly-academically prepared students, whose likelihoods of retaining the merit scholarship are high across all majors, to be more likely to choose majors based on which has the highest expected return (i.e., STEM).[38] For less-academically prepared students, I assume the likelihood of retaining the merit scholarship is lower for some majors relative to others, thus these students choose majors primarily on the basis of success probabilities, and choose majors which are less difficult. Importantly, this simplified model does not account for tastes and preferences of students, which the literature has indicated plays an important role in major choice (Arcidiacono, 2004; Beffy *et al.*, 2010; Wiswall and Zafar, 2015).

In the context of whether broad, low-bar merit scholarships such as the NMLLS affect student major, the theoretical framework above suggests that more-academically prepared students will tend to embark on more difficult, higher-paying majors such as those in STEM fields, while less-academically prepared students will tend to avoid such majors in favor of less-difficult majors, such as those within education and the liberal arts, for example.

4.4 Data set

The analysis uses administrative data for all first-time, full-time entering freshmen at UNM before and after the implementation of the NMLLS to estimate effects on major choice. 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. Data include socio-demographic 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 outcomes by semester (credits earned, declared major, college GPA, date of graduation). Majors are categorized into five areas using ACS definitions: STEM, liberal arts, education, business, social science, and health-related.[39] I also consider alternative definitions provided by UNM as a robustness check. Data are complete with the exception of family income and high school GPA. The data set only contains family income for FAFSA-filers, constituting 51 percent of the sample. For those that did not submit a FAFSA, it is assumed 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 a 1995-1996 Federal Pell Grant End-of-Year Report showing that less than two percent of Pell recipients had family income in excess of $40,000.[40] This assumption is not perfect. King (2004) estimates that in 2000 over ten percent of all Pell-eligible students did not file a FAFSA.[41] If the analysis in King (2004) holds for our data set, then there would exist systematic measurement bias in the family income variable—some lower income students would be incorrectly placed in the higher income category. High school GPA is missing for home-schooled students, a small portion of matriculating students at UNM. For these students, they are assigned the mean high school GPA of 3.28.

Models concentrate on the years 1994 to 1999, bounding the policy change by three years before and after implementation. These years encompass the largest economic expansion in the U.S. since World War II. During this period labor market conditions in New Mexico were gradually tightening but remained relatively stable, so one needs not to worry much that broad economic conditions are driving results. To my knowledge, there were no concurrent policy changes at the high school or postsecondary level in New Mexico over the 1994-1999 period which would have differentially impacted enrollment and/or major choice for residents and nonresidents.

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

Table 4.2 compares summary statistics for resident and nonresident students before and after the implementation of the NMLLS. It appears the composition of 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 to years following the implementation of the scholarship. 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 4.2 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 improved following implementation of the scholarship, according to HSGPA and composite ACT scores. Also note the statistically significant decline in resident students initially declaring a STEM major—a decline not seen in the nonresident group. Table 4.3 presents descriptive statistics for those earning a degree at UNM during the study period. Note there is less evidence of a compositional change in resident students, with only a small decline in high school GPA.

For degree earning residents, there is no descriptive evidence of a decline in STEM degree production after the initiation of the NMLLS.

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 threaten the validity of our difference-in-differences model of STEM engagement 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 treatment and control groups. As a visual check of this identifying assumption, Figure 4.1 presents pre-treatment trends in the likelihood of declaring a first major in STEM for residents and nonresidents between 1994 and 1999. Visual inspection supports the validity of a difference-in-differences identification strategy examining six-year graduation rates. Figure 4.2 presents pre-treatment trends in the likelihood of earning a STEM degree for residents and nonresidents over the same time period. Because completion rates at UNM are relatively low, there are far fewer observations for this group and consequently the graph is quite noisy, especially for nonresidents who are greatly outnumbered by resident students at UNM (by nearly 11 times over). Although Figure 4.1 seems reasonably comparable before the NMLLS was launched in 1997, Figure 4.2 does not pass visual inspection. An empirical test of the common trends assumption is conducted following Autor (2003). Autor suggests estimating flexible difference-in-differences models 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

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

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 which the treatment started (k = 1997 in this case). X_{ist} contains controls for race, ethnicity, gender, family income, remedial coursework in college, high school GPA, and standardized test scores. Models report robust standard errors. In equation (3), m and q are the number of leads and lags of the treatment effect included. Two leads and three lags are included in the test, defining 1999 as the reference cohort.

Testing the common trends assumption using (3) requires examining whether

$$(4)\ \beta_j = 0\ \forall j < 0.$$

In other words, the common trends assumption holds when the coefficients on all leads of the treatment are zero. This specification can also have 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 for the two STEM outcomes using ordinary least squares and results are presented in Appendix A. Results provide evidence that the common trends assumption holds for all specifications, as estimated coefficients on all leads are not statistically different from zero.

Data include 10,022 resident students, 6,307 of which enrolled during the post-NMLLS period and were eligible for the Bridge to Success Scholarship. Of these, 2,664 met cumulative GPA and credit attainment requirements to begin the NMLLS in their second semester. Table 4.4 documents the number of students that maintain the scholarship in the second through ninth semester. It is apparent scholarship loss was quite common. Of the 2,664 students that qualified for the NMLLS, approximately 30 percent were still eligible for the NMLLS going into their third year.

4.5 Empirical model

Difference-in-differences matching estimation on the propensity score is conducted to mitigate any observable differences between resident and nonresident students. The approach uses kernel matching, a one-to-many matching technique assigning larger weights to control units closer in propensity score. The general form of the matching estimator is given by

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

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. Major choice outcomes 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

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

where K is the Epanechnikov kernel function and $\hat{l}(\cdot) \equiv \ln\left(\frac{\hat{p}(\cdot)}{1-\hat{p}(\cdot)}\right)$ is the fitted linearized propensity score from a logistic regression model estimated by maximum likelihood. Linearized propensity scores are used 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.[42] Results of the propensity score model are presented in Table 4.5. It is important to note that while 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 the estimates. In order to increase the precision of estimated treatment effects, and to avoid imposing functional form where possible, kernel density matching is chosen.[43] This method has the advantage of lower variance since more information is used. On the other hand, it may result in an increase in bias due 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, matching is limited to those individuals whose propensity scores lie in the common support region, which is over 99.5 percent of the original sample. No observations are trimmed from the analysis. As a sensitivity analysis, effects are estimated using various fixed bandwidths, h, in 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 the appendices to chapter 2.

In addition to estimating the overall effect of the NMLLS, I am also interested whether program effects differ depending on academic preparation. This is explored by estimating separate models on students above and below the mean high school GPA.[44] Robustness checks using various STEM definitions, cohorts, and smoothing parameters are discussed in Section 6.1.

While difference-in-differences models hinge on the comparability of pre-treatment trends in outcomes across residents and nonresidents, combining difference-in-differences methods with propensity score matching controls for compositional changes in groups over time (Stuart *et al.* 2014). It is also worth noting that regressions control for high school achievement and standardized test scores, the main indication of compositional change. Also, because UNM is a *de facto* open enrollment institution, changes in selectivity are not likely to confound the analysis (Binder and Ganderton, 2004). It is clear that compositional change in the student body occurred, yet this does not diminish the validity of treatment effects estimated.

4.6 Results

Means and normalized differences after kernel matching are presented in Table 4.6. 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 (-.122) far below one-quarter of a standard deviation unit in absolute value. These statistics are produced by academic preparation as well, revealing 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 4.7 presents results of the difference-in-differences kernel matching estimation. Results provide no evidence of an overall effect on either first declaring a STEM major or earning a STEM degree. Furthermore, there is no evidence suggesting the NMLLS had an impact on earned STEM degrees when the sample is disaggregated by academic preparation. Considering students' decisions to first declare a major in STEM, there appears to be a divergent effect: students with below average academic preparation are 6.8 percentage points (40 percent) less likely to declare their first major to be in a STEM field, while those with above-average academic preparation are 12.1 percentage points (44.3 percent) more likely to declare a first major in STEM. Effects are significant at ten and one percent-levels, respectively. These divergent effects mask any overall program effect of the NMLLS on declaring a first major in STEM.

In summary, results reveal no meaningful impact on first declaring a STEM major or earning a STEM degree in the aggregate. In terms of declaring an initial major in STEM, I find that less-academically prepared students are averse to doing so. Conversely, I find that more-academically prepared students declare initial majors in STEM at higher rates compared to their nonresident counterparts as a result of the scholarship.

4.6.1 Alternative STEM definitions, smoothing parameters, and freshmen cohorts

Robustness checks are conducted to examine the sensitivity of results to various assumptions. Appendix 4.C offers three different definitions of STEM based on the student's major. Table 4.C1 presents STEM majors from the ACS, our preferred categorization scheme. We prefer this set of STEM majors as it was developed by the U.S. Census Bureau, is sufficiently narrow in scope, and is the most comprehensive list that can be found. Further, it is employed by previous literature which we are keen to compare our results to (Sjoquist and Winters, 2015a). Tables 4.C2 and 4.C3 present alternative lists of STEM majors compiled by the STEM Collaborative Center (SCC) at UNM. Table C2 presents the "broad" list of STEM majors compiled by SCC while 4.C3 presents the "narrow" version. The broad list is problematic because it includes many majors which one may not agree qualify as being designated as STEM, including anthropology, economics, geography, and nursing. The narrow list should be a subset of the broad list put out by SCC, yet it is not. For example, the narrow list includes statistics while the broad list does not. I nonetheless run models of STEM major declaration and STEM degree completion using broad and narrow lists from SCC. Appendix 4.D displays results of these regressions.

Table 4.D1 presents estimates using the narrow STEM definition provided by the SCC. Aggregate results and those disaggregated by academic preparation are shown. In general, point estimates are similar to our preferred results using the ACS definition, but are attenuated in both magnitude and statistical significance. Using the narrow definition, the point estimate for first majoring in STEM for less-academically prepared students remains negative, but is no longer precisely estimated. The point estimate for first

majoring in STEM for more-academically prepared students is still positive, yet the magnitude is smaller and it achieves statistical significance at a lower level. Table 4.D3 is structured just as other results tables, but employs the broad list from SCC. One would expect the broader scope of this definition to result in further attenuation in terms of magnitude and statistical significance, which it does with one exception. Results using the broad definition estimate a large statistically significant decline in STEM major declaration for the most-academically prepared entering freshmen, although the point estimate is significant only at the ten percent-level.

In addition to examining the sensitivity of results to various definitions of STEM, it is also imperative to examine whether results are sensitive to the choice of smoothing parameter used in the kernel matching procedure. Appendix 4.E presents such sensitivity tests. According to test performed in Appendix 4.B, the matching procedure performed remarkably well. This is further evidenced by Table 4.E1 where one notes that only a few additional observations are included when increasing the bandwidth from 0.1 to 0.3. Point estimates using bandwidths of $h = \{0.1, 0.2, 0.3\}$ are remarkably close in magnitude and statistical significance. There are no sign changes when varying the bandwidth across these values. This provides evidence that bandwidth choice is not a significant driver of our main results presented in Table 4.7.

Appendix 4.F presents results using different sets of freshmen cohorts. Although a bit noisier than robustness checks using alternative bandwidths, we see a similar pattern of completion rates emerge as compared to our preferred specification. Some coefficients become imprecisely estimated when including either the 1993 cohort, the 2000 cohort, or both.

4.6.2 Compositional effects

Key results from Section 6 are not entirely in agreement with the most thorough treatment of this subject to date. Estimates in this paper reveal no meaningful effect of the NMLLS on the likelihood that students earn degrees in STEM fields, in contrast to Sjoquist and Winters (2015a). However, results also provide no evidence that merit aid decreases students' likelihoods of majoring in STEM, in agreement with Cornwell *et al.* (2006). Further, estimates provide evidence of negative STEM degree effects for men, with no statistically meaningful effects for women in the sample, in-line with findings in Sjoquist and Winters (2015a).[45] It is valuable to entertain compositional effects as an alternative hypothesis for the results obtained.

The NMLLS was designed to increase access to higher education for resident students, which it certainly did. According to Table 4.2, the post-NMLLS period of the sample showed a resident population increase of 70 percent (with a much smaller 11 percent increase in nonresidents). After the NMLLS was introduced, however, resident high school GPAs and standardized test scores fell significantly, and resident students were required to take more remedial courses at UNM. This apparent change in student composition is likely key to interpreting much of the results found in Table 4.7. According to the theoretical model presented in Section 4.3, academically marginally prepared students are likely to respond to merit aid by choosing majors for which their probability of success is higher. This may explain why results show that less-academically prepared students majored in STEM significantly less in response to the NMLLS. On the other hand, theory predicts that more-academically prepared students have high probabilities of success in all majors, and so are likely to choose majors with higher expected lifetime earnings, such as STEM.

4.7 Conclusions

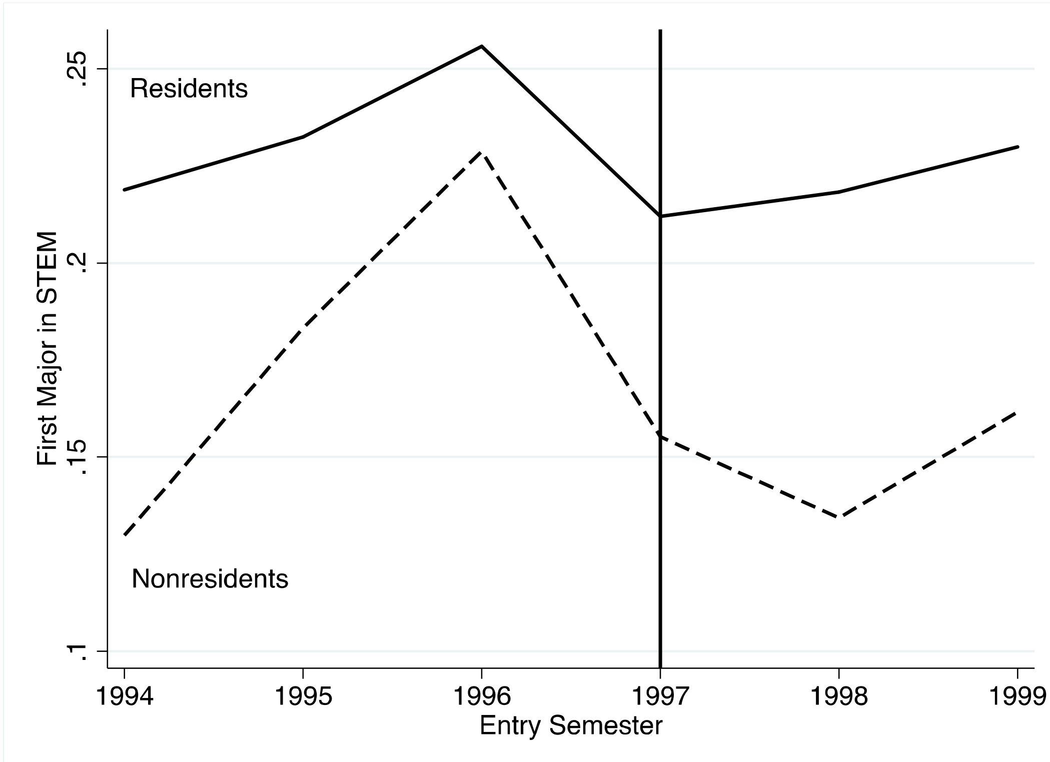
I examine the effect of an exceptionally generous and low-bar merit-based scholarship on intitially declaring a major in STEM and ultimately earning a degree in STEM. Variants of the difference-in-differences model are estimated using qualified resident students as the treatment group and a matched sample of ineligible nonresident students as the control group. The common trends assumption is supported empirically. The sample is stratified by academic preparation and gender to see which, if any, subgroups are driving completion effects. Kernel matching is conducted and its success is examined through rigorous statistical 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. Additionally, I use alternative definitions of STEM, finding similar patterns in results that are attenauted in magnitude and significance-level.

Results reveal find no meaningful program effects in terms of declaring a STEM major or earning a STEM degree in the aggregate. As per declaring an initial major in STEM, less-academically prepared students are more likely to declare a non-STEM major, an effect that appears to be driven by women. Conversely, I find that more-academically prepared students declare initial majors in STEM at higher rates compared to their nonresident counterparts as a result of the scholarship, an effect that is again driven by women at UNM. These effects are similar in magnitude but opposite in sign, masking any program effect in the aggregate.

In motivating the paper, two main research questions were proposed. First, do generous, low-bar merit scholarships discourage students from choosing majors in STEM? Results suggest the answer is "no" in the aggregate, but "yes" on behalf of less-academically prepared students. Moreover, such programs may actually increase interest in STEM majors on behalf of well-academically prepared students. Second, do scholarships such as the NMLLS affect the number of STEM degrees produced? The answer is a resounding "no" according to my results.

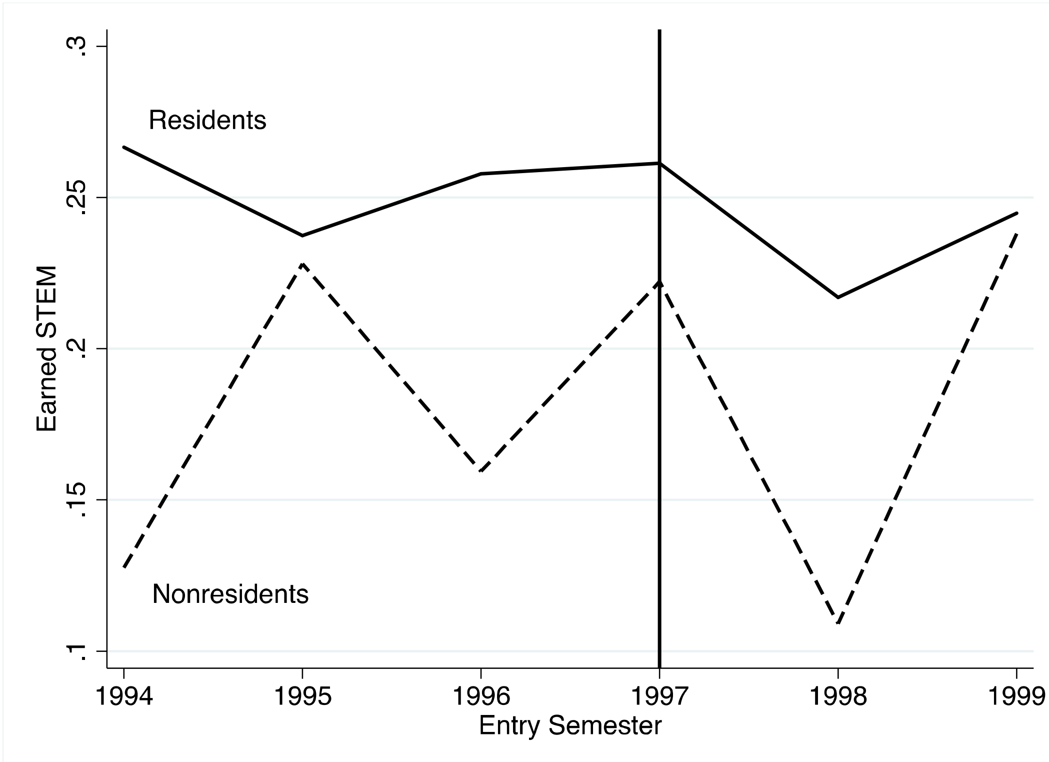
The main conclusion we can draw from the analysis is that low-bar merit-based scholarships neither increase nor decrease the production of STEM degrees. I find little evidence that merit aid eligibility requirements result in students pursuing easier, non-STEM course of study. Although overall production of STEM degrees is not affected by such scholarships, they may alter the composition of who majors in STEM and who eventually completes a STEM degree. To my knowledge, no other studies have looked at merit aid and STEM degree production by high- and low-achieving students. We find a divergent effect of the NMLLS on major choice, in accordance with the theoretical model posed by Montmarquette *et al.* (2002): more-academically prepared students are more likely to declare a major in STEM, while less-academically prepared students are less likely to do so.

Since its inception in 1997, the NMLLS has seen significant changes. Starting in the 2014-2015 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 2017-2018 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 choice of major at UNM.



Note: The plot above show the likelihood of declaring the first major in STEM for incoming cohorts over the period 1994 to 1999. Solid lines represent resident students while dashed lines represent nonresident students. The vertical bars at 1997 mark the implementation of the New Mexico Legislative Lottery Scholarship.

Figure 4.1 Pre-post trends in the likelihood of declaring first major in STEM, by residency



Note: The plot above show the likelihood of declaring the first major in STEM for incoming cohorts over the period 1994 to 1999. Solid lines represent resident students while dashed lines represent nonresident students. The vertical bars at 1997 mark the implementation of the New Mexico Legislative Lottery Scholarship.

Figure 4.2 Pre-post trends in the likelihood of earning a degree in STEM, by residency

Table 4.1 Full-time resident tuition at all NMLLS-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
Mesalands Community College	2	1,990
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 Institute	2	730

Source: Institution financial aid department websites. Accessed 28 March 2017. Figures present tuition and fees for one academic year taking fifteen credit hours per semester. For two-year schools it is assumed the student is within the community college district, where applicable.

Table 4.2 Student characteristics before and after initiation of the NMLLS program, first major declared, ACS major codes

	Residents			Nonresidents		
Variable	Before	After	Diff.	Before	After	Diff.
First Major Declared:						
STEM	.236	0.221	-.015*	.179	0.151	-0.026
Liberal Arts	0.158	0.184	.026***	0.206	0.237	0.031
Education	.074	0.101	.027***	.061	0.088	.027*
Business	0.075	0.094	.019***	0.065	0.071	0.006
Social Science	.110	0.101	-0.009	.112	0.122	0.01
Health-Related	0.133	0.114	-.019***	0.121	0.083	-.038**
Never Declared	0.214	0.183	-.031***	0.256	0.248	-0.008
HSGPA	3.312	3.273	-.038***	3.233	3.3	.067**
	-0.502	-0.471		-0.532	-0.503	
ACT	22.53	22.176	-.354***	22.317	22.861	.544**
	-3.834	-3.887		-4.109	-4.096	
Remedial	0.264	0.29	.026***	0.164	0.227	.063***
Income < $40K	0.23	0.205	-.025***	0.155	0.162	0.007
Female	0.571	0.565	-0.006	0.526	0.545	0.019
Hispanic	0.386	0.375	-0.011	0.147	0.166	0.02
Native	0.043	0.045	0.002	0.041	0.051	0.01
Asian	0.047	0.037	-.010**	0.034	0.026	-0.008
Black	0.021	0.022	0.002	0.082	0.08	-0.002
Observations	3,715	6,307		587	649	

Source: Freshmen Tracking System, Office of Institutional Analytics, UNM. ***, **, and * represent statistical significance at the 1, 5, and 10 percent-levels, respectively. Standard deviations are in parentheses.

Table 4.3 Student characteristics before and after initiation of the NMLLS program, degree type earned, ACS major codes

	Residents			Nonresidents		
Variable	Before	After	Diff.	Before	After	Diff.
Degree Type Earned:						
STEM	.253	0.24	-0.013	.173	0.195	0.022
Liberal Arts	0.262	0.249	-0.013	0.341	0.326	-0.015
Education	.116	0.081	-.035***	.121	0.026	-.095***
Business	0.157	0.19	.033***	0.185	0.163	-0.022
Social Science	.167	0.177	0.01	.145	0.237	.092**
Health-Related	0.044	0.064	.020***	0.035	0.053	0.018
HSGPA	3.479	3.455	-.024*	3.483	3.473	-0.01
	-0.467	-0.439		-0.46	-0.442	
ACT	23.268	23.085	-0.183	23.526	23.807	0.281
	-3.761	-3.784		-3.865	-3.888	
Remedial	0.192	0.196	0.004	0.138	0.15	0.012
Income < $40K	0.194	0.173	-.021*	0.128	0.14	0.012
Female	0.616	0.616	0	0.622	0.609	-0.013
Hispanic	0.366	0.358	-0.008	0.097	0.159	.062*
Native	0.022	0.023	0.001	0.01	0.039	.029*
Asian	0.052	0.041	-.011*	0.041	0.019	-0.021
Black	0.016	0.018	0.002	0.102	0.058	-0.044
Observations	1,547	2,543		173	190	

Source: Freshmen Tracking System, Office of Institutional Analytics, UNM. ***, **, and * represent statistical significance at the 1, 5, and 10 percent-levels, respectively. Standard deviations are in parentheses.

Table 4.4 NMLLS student attrition, 1994-1999

Semester	Residents Eligible	Percent Remaining
2	2,664	100.0%
3	2,249	84.4%
4	2,017	75.7%
5	1,863	69.9%
6	1,734	65.1%
7	1,629	61.1%
8	1,568	58.9%
9	1,510	56.7%

Source: Office of Institutional Analytics, University of New Mexico. We consider the sample of resident students that met cumulative GPA and credit requirements in their first semester to qualify for the NMLLS.

Table 4.5 Estimated parameters for propensity score model of NMLLS data, 1994-1999

Variable	Estimate	Std. Error
HSGPA	1.729**	.724
ACT	.498***	.090
Remedial	.891***	.118
Income < 20K	.268*	.158
Income < 40K	.160	.108
Female	1.670***	.367
Hispanic	1.865***	.550
Native American	1.884**	.923
Asian	.032	.207
Black	-5.729***	1.155
Declined to state race-ethnicity	-.108	.282
ACT^2	-.013***	.002
ACT*Black	.141***	.045
Female*White	-.571***	.146
$HSGPA^2$	-.461***	.116
ACT*Female	-.053***	.016
ACT*HSGPA	.059***	.020
Remedial*Asian	1.147**	.505
GPA*Black	.546	.339
ACT*Native	-.082**	.041
Female*Native	-.608*	.317
HSGPA*Hispanic	-.312*	.165
Constant	-7.711***	1.600
Observations		11,258

Standard errors are in parentheses. ***, **, and * correspond to statistical significance at the 1, 5, and 10 percent levels, respectively. Propensity scores are estimated using a logistic model. Forty-nine observations were dropped following estimation of the propensity score to ensure overlap, leaving 11,209 observations. The variable *Declined to state race-ethnicity* is equal to one if the student declined to state their race-ethnicity, and zero otherwise.

Table 4.6: Means and normalized differences after kernel matching, full sample, 1994-1999

	Pre-NMLLS			Post-NMLLS		
Variable	Res.	Nonres.	ND	Res.	Nonres.	ND
HS GPA	3.31	3.27	0.088	3.27	3.33	-0.122
Composite ACT	22.56	22.37	0.047	22.19	22.58	-0.099
Remedial	0.26	0.24	0.032	0.29	0.28	0.012
Income < $40,000	0.22	0.21	0.04	0.2	0.21	-0.032
Female	0.57	0.58	-0.009	0.56	0.59	-0.063
Hispanic	0.39	0.39	-0.019	0.37	0.36	0.025
Native	0.04	0.04	0.001	0.05	0.05	-0.03
Asian	0.04	0.03	0.058	0.04	0.03	0.019
Black	0.02	0.02	-0.023	0.02	0.02	0.018

Means are from Epanechnikov kernel matching performed with a bandwidth of $h = .2$. Normalized differences (ND) are calculated by taking the difference average covariate values by residency status and dividing by a measure of standard deviation.

Table 4.7 NMLLS and major choice by academic preparation, American Community Survey definition, 1994-1999

Group	Obs.	First Declared STEM	Obs.	Majored in STEM
Full Sample	11,209	.026	4,438	-.012
		(.030)		(.057)
$\bar{Y}$		.221		.240
HSGPA ≤ 3.28	5,473	-.068*	1,507	.147
		(.040)		(.093)
$\bar{Y}$		.170		.145
HSGPA > 3.28	5,734	.121***	2,930	-.051
		(.046)		(.073)
$\bar{Y}$		.273		.291
HSGPA > 3.78	2,105	-.063	1,271	-.061
		(.067)		(.119)
$\bar{Y}$		.334		.386

Robust standard errors are reported in parentheses. *, **, and *** denote statistical significance at the 10, 5, and 1 percent-level, respectively. Estimates are from difference-in-differences kernel matching performed with a bandwidth of h = .2 using the Epanechnikov kernel function. We report estimates for students with below average or average high school 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 the baseline rate of STEM major choice by academic preparation.

Notes

---

[32] See Sjoquist and Winters (2015a) for a complete list.

[33] The College Board, Trends in Student Aid 2015, Figure 28A and Figure 28B. Retrieved January 29, 2016 from http://trends.collegeboard.org/sites/default/files/trends-student-aid-web-final-508-2.pdf.

[34]Georgia Student Finance Commission, GACollege411, Georgia Hope Scholarship Program Overview. Retrieved May 29, 2013 from https://secure.gacollege411.org/Financial_Aid_Planning/HOPE_Program/Georgia_s_HOPE_Scholarship_Program_Overview.aspx.

[35]Tennessee Student Assistance Corporation, Tennessee Hope Scholarship. Retrieved May 29, 2013 from http://www.tn.gov/collegepays/mon_college/hope_scholar.htm.

[36]Florida Department of Education, Office of Student Financial Aid, Florida Student Scholarship and Grant Programs, Chart of Eligibility and Award Criteria. Retrieved May 29, 2013 from http://www.floridastudentfinancialaid.org/ssfad/PDF/BFEligibilityAwardChart.pdf.

[37]Consider full-time tuition at all 16 participating public institutions in New Mexico as depicted in Table 4.1. A student better matched at Santa Fe Community 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.

[38]Sjoquist and Winters (2015a) calculate mean earnings for persons aged 40 – 49

using 2009 – 2011 ACS data, finding that those majoring in STEM fields earned $95,389; those with business degrees earned $78,122; those in social science earned $67,735; those with health-related degrees earned $58,937; those with liberal arts degrees earned $58,823; and those with degrees in education earned $46,169. They choose this age range because 1) these respondents are too old to be affected by state merit aid programs and 2) according to Berger (1988) mid-career earnings are likely to be more relevant than early-career earnings.

[39] Majors are categorized into these bins according to the U.S. Census Bureau, found online at https://www2.census.gov/programs-surveys/acs/tech_docs/code_lists/2016_ACS_Code_Lists.pdf (accessed 19 Feb 2019).

[40]1995-1996 Federal Pell Grant Program End-of-Year Report, U.S. Department of Education, online at https://www2.ed.gov/finaid/prof/resources/data/pell-historical/pell-eoy-1995-96.pdf (accessed 26 March 2017).

[41] King, Jacqueline E. "Missed Opportunities: Students who do not Apply for Financial Aid," American Council on Education Issue Brief, 2004. Online at http://www.soe.vt.edu/highered/files/Perspectives_PolicyNews/10-04/2004FAFSA.pdf (accessed 1 April 2017).

[42]A sequential search for quadratic terms to include in the propensity score model was conducted. The first step involved estimating logistic models including all terms in levels and one of all possible quadratic terms. I then calculate the likelihood ratio statistic for the null hypothesis that the most recently added quadratic term has a coefficient of zero. The quadratic term with the highest test statistic over 2.71, corresponding to a z-statistic of 1.645, is selected for inclusion. This covariate is then

added to the "baseline" model and the process repeated until all remaining likelihood ratio statistics are below the threshold of 2.71.

[43] There are 9,979 resident students and only 1,233 nonresident students in the sample. One-to-many matching allows us to proceed without a significant loss in information. For example, if I was to conduct a simple nearest neighbor matching procedure, estimates would (at most) be based on 1,233 matches, or 2,466 observations, which constitutes approximately 22 percent of the sample.

[44] Results are similar when the sample is split around the median high school GPA.

[45]Although we estimate regressions splitting the sample by academic preparation *and* gender, we do not report these as the number of control units is problematically small when disaggregating the sample in this way.

Chapter 5: Conclusions: How students respond to merit aid and how employers react to lengthened time to degree

The work in this dissertation focuses on two major issues germane to the market for higher education: the changing structure of financial aid in the United States and the changing paradigm of time to degree. The advent of state merit-based scholarships in the U.S. has changed students' college-going decisions, as well as their choices while in college. While good intentions belie these programs, do such scholarships help or hinder a student's ability to complete college? Do they affect the academic trajectory of the student, ultimately affecting their career path? Chapters 2 and 4 address this issue directly. The longstanding trend of taking longer to complete a baccalaureate degree is also at question. How long is too long? Do employers entertain this variable when making job offers? Chapter 3 addresses this issue. I summarize the main findings, limitations, and policy implications of each essay in turn.

In chapter two, "Does Broad-Based Merit Aid Improve College Completion? Evidence from New Mexico's Lottery Scholarship," we investigate whether broad-based merit aid results in any meaningful change in college completion rates. Surprisingly, we find that merit aid, as it is structured in New Mexico, results in some students being less likely to graduate, with others being more likely to graduate. This divide hinges on academic preparation. Although completion rates are unaffected in the aggregate, we provide evidence that such scholarships result in meaningful changes in student composition. With low initial eligibility criteria, students appear to utilize the scholarship at the university providing the highest expected return on to degree, often corresponding to the flagship university or the university with the highest cost of attendance. Low-bar scholarships generous in funding may promote overmatching, which occurs when a student attends an institution for which they are academically underprepared. We find that low-achieving high school students that acquire such scholarships are less likely to complete their studies. Conversely, students that perform well in high school are more likely to complete. What is the best solution to increasing access to higher education at the state-level? Not surprising to economists, it appears that broad-based merit-aid scholarships have both costs and benefits to recipients. The trade-off focuses is one between access and completion: broad-based merit scholarships significantly increase access to higher education, yet they distort the choice of where to go, and may harm the marginally prepared student seeking out the highest return. Disaggregating results by family income suggests that program effects are likely driven by students from low-income households. Because the NMLLS has had many difficulties regarding solvency over the years, we recommend that a need-based component to the scholarship be considered. Additionally, researchers would like to have data on the entire postsecondary system in New Mexico, which would allow for a richer analysis of compositional change as a result of such scholarships.

In chapter three, "Wage Effects of Baccalaureate Time to Degree in the United States," we examine whether how long an undergraduate student spends obtaining a degree matters to employers after college. Using a nationally-representative longitudinal study of high school students, we develop a test of whether longer time to degree serves as a negative productivity signal. Previous literature estimates that each additional year beyond the four-year mark results in up to an eight percent wage penalty. Being skeptics, and assuming that college students are rational actors in the economy, we test this hypothesis ourselves. The major problem with previous literature is that time to degree is endogenous in the earnings equation. That is, there are several factors which may both affect time to degree and earnings, such as student ability, college quality, for example, and unobservable factors that impact how long a student typically graduates at a given college. We confront this endogeneity by controlling for student ability (vis-à-vis standardized test scores), institution quality (vis-à-vis Barron's Admissions Competitiveness data), and instrumenting the student's time to degree by the average at their institution. The instrument appears to be relevant and exogenous. Without using instrumental variables techniques, we are able to mimic the large and significant wage penalties found in previous studies. However, after controlling for the above and using instrumental variables, results suggest that time to degree is not taken as a productivity signal, and there is no wage penalty associated with lengthened time to degree. Results provide fodder to arguments that punishing prolonged time to degree is a waste of resources. Indeed, we offer theoretical and empirical evidence showing that rational, utility-maximizing students may prefer to earn a degree and work part-time simultaneously over six years, rather than foregoing work and completing a degree in four years.

In chapter four, "Merit Aid Scholarships and Human Capital Production in STEM: Evidence from New Mexico," I examine whether merit based aid dissuades students from studying more difficult subjects, such as STEM. The advent of state merit aid scholarships begs the question: do students respond by avoiding more difficult majors, such as those in STEM? I again utilize the rich administrative data set provided by UNM (same as chapter two). Using the same matching algorithm, which appears to have been successful, I estimate whether students are more- or less-STEM averse after receiving the scholarship. Using two different outcomes related to studying STEM (declaring a first major in STEM and earning a degree in STEM), I find that there is no meaningful impact of the NMLLS in the aggregate. However, there does appear to be a divergent effect when disaggregating by academic preparation, as proxied by one's high school performance. High achieving students seem to study more STEM in response to merit aid, while low-achieving student have an opposite reaction. This is in line with the theoretical literature. Results suggest that adopting or killing state merit aid scholarships will not affect degree production in STEM, although it may change the composition of those earning it in terms of academic preparation and gender.

This work is by no means a comprehensive study relating to the relationship between state merit aid student outcomes. In fact, it only scratches the surface. We provide evidence of the costs and benefits of such aid, but implore universities and state governments to make more data available, so that we may make more informed decisions regarding the trade-off between accessibility to higher education and student success.

Appendices to Chapter 2

Appendix 2.A Assessing properties of the propensity score

In order to examine the effectiveness of our matching procedure, we first assess overlap in the propensity score both before and after the NMLLS is in place. Note that tests conducted in this section use only information concerning covariates and residency classification, and do not consider completion rates, therefore cannot intentionally introduce bias in subsequent analyses. For a thorough treatment of these tests, see Imbens and Rubin (2015).

Figure 2.A1 presents histogram estimates of the distribution of linearized propensity scores before and after the implementation of the NMLLS, by residency. First inspection reveals substantial overlap in the linearized propensity score across residents and nonresidents, both before and after the NMLLS was launched. As a more formal check, we calculate the percent of observations where there exists an observation of the opposite treatment status with a difference in linearized propensity score less than 10 percent. These measures are presented in Table 2.A1. For residents, approximately 99 percent of students had at least one closely matching nonresident student in terms of linearized propensity score both before and after the launch of the NMLLS. For nonresidents, this percentage was approximately 97 percent. This suggests we should be able to credibly estimate causal effects of the NMLLS on student graduation under the assumption of unconfoundedness.

We next perform two tests assessing the balancing property of the propensity score, which asserts that conditional on the propensity score, treatment assignment and student characteristics are independent of one another. We perform these tests both before and after the NMLLS is launched. If results of these tests are favorable, this constitutes evidence supporting the assumption of unconfoundedness, although it cannot be directly tested. The balancing property can be formally represented as:

$$W_i \perp\!\!\!\perp X_i \mid l(X_i) \qquad (2.A1)$$

where W_i is a binary treatment indicator equal to one if student i is a New Mexico resident, and zero otherwise, X_i is a vector of covariates, and $l(X_i)$ is the true linearized propensity score. Because we do not know the true linearized propensity score, we approximate this test by instead using its estimated counterpart, $\hat{l}(X_i)$. Our strategy is to stratify the sample into J blocks, $B_i(1), \dots, B_i(J)$, so there will be no significant difference between linearized propensity scores within each block. This way, (2.A1) becomes

$$W_i \perp\!\!\!\perp X_i \mid B_i(1), \dots, B_i(J). \qquad (2.A2)$$

Equation (2.A2) can be examined by testing whether residency classification and covariates are uncorrelated within each of the J blocks, so that

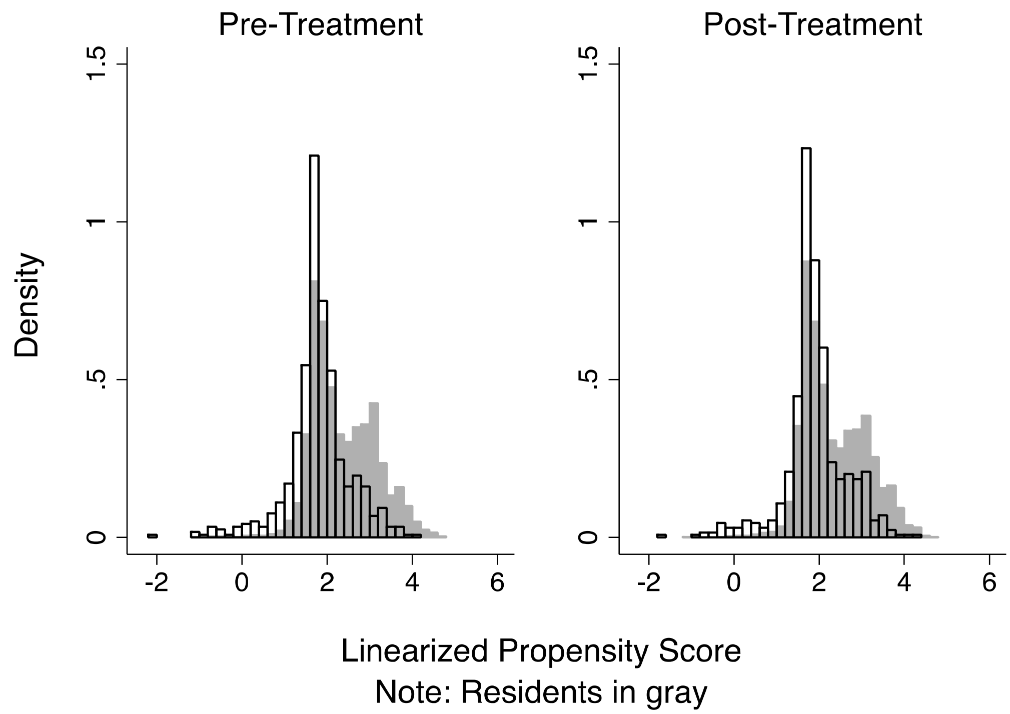
$$E[X_i | W_i = 1, B_i(j) = 1] = E[X_i | W_i = 0, B_i(j) = 1] \qquad (2.A3)$$

for all blocks, $j = 1, \dots, J$. Tables 2.A2 and 2.A3 present the results of this stratification procedure. For the pre-NMLLS period, we split the sample into 11 blocks using the linearized propensity score. Near the upper end of the propensity score distribution, we were not able to further split blocks 10 and 11 due to a small number of nonresident students relative to the number of resident students.[46] We also encountered this issue when stratifying the sample in the post-NMLLS period, although at the opposite end of the propensity score distribution. We nonetheless consider the stratification successful, as only two of the 25 blocks created were left with propensity scores that were significantly different across resident and nonresident groups at the five percent level.

With pre- and post-NMLLS samples stratified, we then assess covariate balance within blocks. These tests can be thought of "pseudo treatment effects" as they examine the effect of treatment on pre-treatment covariates, where the effects are a priori known to be zero. Confirmation that pseudo treatment effects are zero constitutes evidence that equation (2.A3) holds, supporting the assumption of unconfoundedness. We conduct two different tests. First, we test separately, by each covariate, whether within-block differences between residents and nonresidents are equal to zero. Second, for each covariate we test whether the weighted average of within-block differences between residents and nonresidents are equal to zero. Results of these tests for pre- and post-NMLLS cohorts are reported in Tables 2.A4 and 2.A5, respectively.

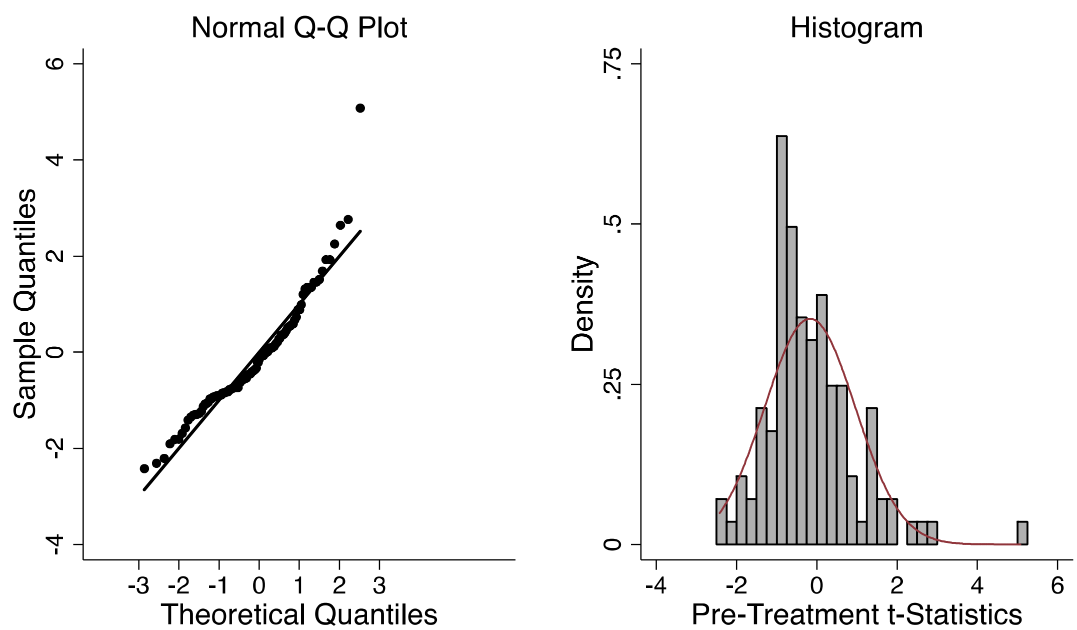
We analyze the results of these tests as if data arose from a stratified random experiment. The first approach for assessing covariate balance focuses on one covariate-block dyad at a time. We calculate z-statistics testing the null hypothesis that the difference between residents and nonresidents in the dyad is equal to zero. These tests produce a large amount of information, however they are not very informative when examined individually. Of the 113 pre-NMLLS tests, only seven (six percent) had z-statistics above two. Similarly, of the 148 post-NMLLS tests, only twelve (eight percent) exceeded two in absolute value. It is informative to present these statistics in Q-Q plots, where z-values are compared against their expected values under independent draws from a standard normal distribution. If the distributions of z-values closely follow the 45 degree lines in these plots, it is evidence that the propensity score was effective in balancing covariates as if treatment was randomly assigned within blocks. Q-Q plots are presented in Figures 2.A2 and 2.A3. Both appear to follow the normal distribution reasonably well, although they are slightly skewed to the right (especially for pre-treatment cohorts). One major outlier deserves attention in Figure 2.A2—it is due to the incomparability of black resident and nonresident students at UNM at a particular region of the propensity score distribution.[47]

The column labeled as the overall t-statistic tests the null hypothesis that the block-adjusted weighted average of within block differences is equal to zero. Finding z-values larger in absolute value than we would expect if they were drawn independently from a standard normal distribution is evidence that the stratification does not adequately balance covariates, suggesting that the propensity score model is not satisfactory and may need to be refined. According to Table 2.A4, there do not appear to be any significant balance issues for pre-NMLLS cohorts. For these cohorts, the largest t-statistic we find is 1.73, suggesting excellent balance. Table 2.A5 reveals that there may exist some imbalance in the high school GPA and ACT composite score covariates for post-NMLLS cohorts. The z-statistics for these covariates are 2.17 and 2.14 in absolute value, respectively, indicating that we can reject the null hypothesis that the weighted averages of within-block differences are zero for both of these variables at the five percent level. Analyzing normalized differences between residents and nonresidents for these covariates after matching is performed provides an additional check as to whether this imbalance requires estimating a more flexible propensity score or perhaps trimming the sample. Although the propensity score model did not perform as well as would randomization into treatment within blocks, overall we feel it worked adequately in balancing covariates across resident and nonresident students.



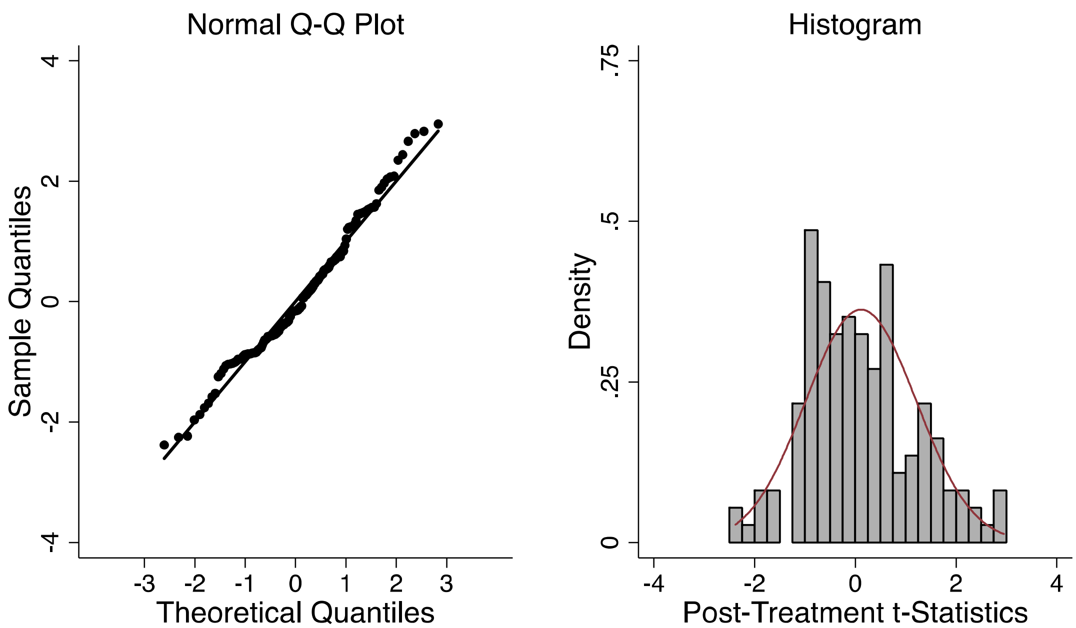
Note: The left and right panels overlap linearized propensity scores for residents and nonresidents before and after the implementation of the NMLLS, respectively, allowing for visual inspection of sufficient overlap, a critical requirement for successful propensity score matching. Both figures indicate there is sufficient overlap of residents and nonresidents.

Figure 2.A1 Linearized propensity scores, by residency and NMLLS implementation



Note: The left panel presents a normal Q-Q plot of t-statistics from within-block tests before implementation of the NMLLS. Normal Q-Q plots graph actual percentiles against theoretical percentages from a normal distribution with the same mean and standard deviation. Normality is evidenced by a straight line of plotted values. Above right is a histogram of the same t-statistics with a fitted normal curve. Both plots provide visual evidence of slight positive skew.

Figure 2.A2 Visual check of normality of within-block differences across resident status, pre-NMLLS



Note: The left panel presents a normal Q-Q plot of t-statistics from within-block tests after implementation of the NMLLS. Normal Q-Q plots graph actual percentiles against theoretical percentages from a normal distribution with the same mean and standard deviation. Normality is evidenced by a straight line of plotted values. Above right is a histogram of the same t-statistics with a fitted normal curve. Again, both plots provide visual evidence of slight positive skew.

Figure 2.A3 Visual check of normality of within-block differences across resident status, post-NMLLS

Table 2.A1 Proportion of units with match discrepancy in terms of linearized propensity score less than 10 percent

Measure	Pre-NMLLS	Post-NMLLS
$q_{nonresident}$	.971	.968
$q_{resident}$	.989	.995

Table 2.A2 Propensity score blocks and their boundaries, pre-NMLLS

Block	Lower	Upper	Width	Nonresidents	Residents	t-Statistic
1	.238	.688	.449	37	29	-.854
2	.688	.757	.070	24	43	-.796
3	.757	.800	.043	47	86	-.598
4	.800	.830	.030	57	209	-.057
5	.830	.843	.013	47	219	.054
6	.843	.851	.009	62	205	.154
7	.851	.888	.036	167	899	.783
8	.888	.920	.032	62	471	.304
9	.920	.945	.025	38	496	.065
10	.945	.961	.015	23	510	-3.274
11	.961	.985	.024	21	513	-1.519

Above presents results of an attempt to stratify the sample on the linearized propensity score. *t*-statistics are for the null hypothesis of equality in linearized propensity scores between resident and nonresident students. Blocks 10 and 11 could not be split further as there would be an insufficient number of members in new groups for subsequent hypothesis testing.

Table 2.A3. Propensity score blocks and their boundaries, post-NMLLS

Block	Lower	Upper	Width	Nonresidents	Residents	t-Statistic
1	.249	.717	.468	40	68	-3.393
2	.717	.780	.063	25	83	-1.314
3	.780	.813	.033	28	188	-1.493
4	.813	.832	.019	49	383	-.835
5	.832	.852	.020	110	758	.419
6	.852	.869	.017	109	755	-.516
7	.869	.888	.020	89	779	-.557
8	.889	.896	.007	36	180	-.292
9	.896	.904	.008	18	199	-.877
10	.904	.922	.017	26	407	-.971
11	.922	.937	.015	31	402	.328
12	.937	.946	.010	22	411	-1.339
13	.946	.962	.015	45	821	-.569
14	.962	.987	.025	20	846	-.390

Above presents results of an attempt to stratify the sample on the linearized propensity score. *t*-statistics are for the null hypothesis of equality in linearized propensity scores between resident and nonresident students. Block 1 could not be split further as there would be an insufficient number of members in new groups for subsequent hypothesis testing.

Table 2.A4 Tests for balance conditional on propensity score, pre-NMLLS

Covariate	Within Blocks											Overall
	1	2	3	4	5	6	7	8	9	10	11	t-Statistic
High School GPA	0.37	-0.37	-1.81	**-2.21**	1.51	-0.15	-0.74	-1.27	-0.78	-0.03	-0.22	1.73
Composite ACT	-0.94	**-2.31**	-1.58	-1.82	-0.75	-0.60	0.09	-1.35	-0.56	**2.25**	0.09	0.85
Remedial	**-2.43**	0.22	-0.85	0.67	1.46	-0.40	0.01	0.45	-0.97	-0.92	-0.83	0.83
Income < 20K	0.27	-0.76	-0.07	0.99	1.20	0.59	-0.83	-1.04	-0.36	0.40	0.35	0.59
Income < 40K	0.19	-0.85	-0.85	-0.90	-0.45	1.46	-0.47	0.88	-0.97	-1.08	0.29	1.17
Female	-0.83	-1.07	0.05	-0.42	0.58	-0.90	0.50	-1.91	1.69	-0.74	0.16	0.27
Hispanic	-1.13	1.35	-0.74	1.92	-	-0.78	-0.08	0.73	-0.58	0.12	1.31	-0.76
Native	-	-0.74	-1.69	-0.54	**2.76**	-0.07	-0.39	-0.33	0.01	0.53	-	-0.10
Asian	0.88	1.35	-0.07	1.93	-0.93	0.08	-0.02	-1.23	**2.64**	-0.64	-1.31	-0.03
Black	-1.41	-0.63	-1.29	-0.90	0.36	-0.55	0.54	**5.08**	-0.55	-	-	-1.08
Declined	-	1.35	-0.74	-0.91	-0.46	-	-1.30	0.11	-0.48	-0.21	-	1.22

z-statistics test the null hypothesis of equality of means within blocks for resident and nonresident students. Overall t-statistics test the null hypothesis that the weighted average of differences across blocks is equal to zero. *Declined* is equal to zero if the student declined to state their race-ethnicity, and zero otherwise.

Table 2.A5: Tests for balance conditional on propensity score, post-NMLLS

Covariate	**Within Blocks**														**Overall**
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	t-Statistic
High School GPA	-1.88	**-2.38**	-1.76	1.49	1.56	**2.07**	1.20	-0.19	0.67	0.94	1.50	0.82	0.51	0.06	**-2.17**
Composite ACT	**-2.24**	-0.63	0.42	**2.35**	1.90	-0.88	**2.44**	0.34	1.85	0.53	-0.56	0.57	-0.15	0.54	**-2.14**
Remedial	**2.79**	1.45	-0.43	0.25	0.11	1.23	-1.97	0.74	**-2.25**	-0.99	1.63	-1.03	1.46	-0.80	0.28
Income < 20K	-0.11	-0.38	-0.87	1.45	0.13	1.24	-0.54	0.17	0.66	-0.11	-1.06	0.22	-0.61	0.66	-0.33
Income < 40K	1.28	-0.16	-1.25	0.35	-0.85	0.62	-0.14	-0.91	**2.95**	-0.87	-0.87	-0.15	-1.69	**2.09**	-0.69
Female	-0.96	-0.82	-0.08	1.34	0.06	1.97	-0.07	-0.27	-1.02	0.75	0.84	1.25	-0.58	0.10	-1.14
Hispanic	-0.77	-0.55	-0.67	-0.36	0.74	-0.85	0.45	-1.52	-0.44	0.53	-1.04	-0.64	-1.19	0.38	0.86
Native	-	-0.17	-0.49	-0.50	-0.35	0.69	-1.12	**2.04**	-0.16	0.16	0.83	0.46	0.19	-0.15	-0.45
Asian	-	-0.55	1.57	1.04	-0.85	-1.58	-0.33	-0.25	0.20	-0.32	0.74	1.53	1.55	-0.36	-0.06
Black	1.47	0.55	-0.58	-0.95	0.70	0.30	0.35	-0.78	-0.52	-0.72	-0.39	-0.57	-	-	0.38
Declined	-	**2.66**	-0.96	-0.88	0.29	0.43	1.24	-1.01	0.10	**2.83**	-0.84	-0.57	-0.41	-	-1.23

z-statistics test the null hypothesis of equality of means within blocks for resident and nonresident students. Overall t-statistics test the null hypothesis that the weighted average of differences across blocks is equal to zero. Declined is equal to zero if the student declined to state their race-ethnicity, and zero otherwise.

Notes

[46] In order to perform subsequent hypothesis testing, we are not able to further split blocks where new blocks would have fewer members than $K + 2$, where K is the number of covariates.

[47] Although this outlier is visually striking, it is driven by the relatively small number of black students at UNM (less than 3 percent of the sample).

Appendix 2.B Alternative bandwidths

Table 2.B1 NMLLS kernel matching estimates with bandwidths of .1, .2, and .3; 1994-1999

Group	Obs.	Graduation Rates by Years since First Enrollment			
		4 Years	4 ½ Years	5 Years	6 Years
Full Sample					
$h = 0.1$	11,207	-.035	-.027	-.020	-.013
$h = 0.2$	11,209	-.035	-.030	-.024	-.019
$h = 0.3$	11,210	-.035	-.030	-.024	-.019
GPA ≤ 3.28					
$h = 0.1$	5,470	-.014	-.037	.064	-.082*
$h = 0.2$	5,473	-.015	-.035	-.069*	-.087**
$h = 0.3$	5,474	-.016	-.034	-.070*	-.089**
GPA > 3.28					
$h = 0.1$	5,732	-.026	.016	.066	.089*
$h = 0.2$	5,734	-.022	.016	.069	.094*
$h = 0.3$	5,735	-.023	.013	.067	.096*
GPA > 3.78					
$h = 0.1$	2,103	.041	.095	.098	.113
$h = 0.2$	2,105	.031	.082	.093	.107
$h = 0.3$	2,105	.034	.086	.103	.115

Robust standard errors are reported in parentheses. *, **, and *** denote statistical significance at the 10, 5, and 1 percent-level, respectively. Estimates are from difference-in-differences kernel matching using an Epanechnikov kernel function with various bandwidth parameters, *h*. We report estimates for students with below average or average high school GPAs (≤ 3.28), above average high school GPAs (> 3.28), and high school GPAs greater than one standard deviation above the mean (> 3.78).

Table 2.B2 NMLLS kernel matching estimates with bandwidths of 0.1, 0.2, and 0.3; family income < $40,000; 1994-1999

		Graduation Rates by Years since First Enrollment			
Group	Obs.	4 Years	4 ½ Years	5 Years	6 Years
Full Sample					
$h = 0.1$	2,291	-.010	-.014	-.041	-.016
$h = 0.2$	2,296	-.011	-.012	-.037	-.020
$h = 0.3$	2,297	-.013	-.015	-.040	-.025
GPA ≤ 3.28					
$h = 0.1$	1,128	.003	-.013	-.206**	-.204**
$h = 0.2$	1,131	.003	-.018	-.197**	-.202**
$h = 0.3$	1,133	.004	-.019	-.183*	-.198**
GPA > 3.28					
$h = 0.1$	1,160	.047	.054	.180*	.226*
$h = 0.2$	1,162	.022	.037	.161	.200*
$h = 0.3$	1,163	.017	.033	.156	.205*
GPA > 3.78					
$h = 0.1$	403	-.086	-.036	.148	.054
$h = 0.2$	404	-.091	-.032	.151	.054
$h = 0.3$	404	-.071	-.012	.159	.060

Robust standard errors are reported in parentheses. *, **, and *** denote statistical significance at the 10, 5, and 1 percent-level, respectively. Estimates are from difference-in-differences kernel matching using an Epanechnikov kernel function with various bandwidth parameters, h. We report estimates for students with below average or average high school GPAs (≤ 3.28), above average high school GPAs (> 3.28), and high school GPAs greater than one standard deviation above the mean (> 3.78).

Table 2.B3 NMLLS kernel matching estimates with bandwidths of 0.1, 0.2, and 0.3; family income ≥ \$40,000; 1994-1999

		Graduation Rates by Years since First Enrollment			
Group	Obs.	4 Years	4 ½ Years	5 Years	6 Years
Full Sample					
$h = 0.1$	8,904	-.037	-.028	-.018	-.011
$h = 0.2$	8,908	-.035	-.028	-.015	-.010
$h = 0.3$	8,909	-.035	-.027	-.012	-.008
GPA ≤ 3.28					
$h = 0.1$	4,333	-.014	-.038	-.042	-.061
$h = 0.2$	4,335	-.018	-.040	-.035	-.055
$h = 0.3$	4,336	-.019	-.038	-.030	-.052
GPA > 3.28					
$h = 0.1$	4,569	-.030	.014	.029	.049
$h = 0.2$	4,571	-.027	.011	.033	.058
$h = 0.3$	4,571	-.032	.005	.030	.056
GPA > 3.78					
$h = 0.1$	1,699	.052	.110	.089	.123
$h = 0.2$	1,701	.055	.109	.096	.130
$h = 0.3$	1,701	.055	.109	.106	.139*

Robust standard errors are reported in parentheses. *, **, and *** denote statistical significance at the 10, 5, and 1 percent-level, respectively. Estimates are from difference-in-differences kernel matching using an Epanechnikov kernel function with various bandwidth parameters, h. We report estimates for students with below average or average high school GPAs (≤ 3.28), above average high school GPAs (> 3.28), and high school GPAs greater than one standard deviation above the mean (> 3.78).

Appendix 2.C Alternative sets of cohorts

Table 2.C1 NMLLS kernel matching estimates with alternative cohort sets

		Graduation Rates by Years since First Enrollment			
Group	Obs.	4 Years	4 ½ Years	5 Years	6 Years
Full Sample					
1993 – 1999	12,755	-.022	-.011	.010	.024
1993 – 2000	15,208	-.037*	-.032	-.002	.018
1994 – 1999	11,209	-.035	-.030	-.024	-.019
1994 – 2000	13,715	-.047**	-.047*	-.030	-.015
GPA ≤ 3.28					
1993 – 1999	6,309	-.015	-.027	-.031	-.033
1993 – 2000	7,510	-.012	-.041	-.045	-.038
1994 – 1999	5,473	-.015	-.035	-.069*	-.087**
1994 – 2000	6,706	-.016	-.052**	-.076**	-.084**
GPA > 3.28					
1993 – 1999	6,441	-.011	.024	.082*	.107**
1993 – 2000	7,696	-.038	.003	.067	.096**
1994 – 1999	5,734	-.022	.016	.069	.094*
1994 – 2000	7,009	-.046	-.003	.063	.098**
GPA > 3.78					
1993 – 1999	2,359	-.004	.080	.117	.103
1993 – 2000	2,838	-.061	.018	.084	.080
1994 – 1999	2,105	.031	.082	.093	.107
1994 – 2000	2,583	-.019	-.031	.073	.094

Robust standard errors are reported in parentheses. *, **, and *** denote statistical significance at the 10, 5, and 1 percent-level, respectively. Estimates are from difference-in-differences kernel matching using an Epanechnikov kernel function with various bandwidth parameters, *h*. We report estimates for students with below average or average high school GPAs (≤ 3.28), above average high school GPAs (> 3.28), and high school GPAs greater than one standard deviation above the mean (> 3.78).

Table 2.C2 NMLLS kernel matching estimates with alternative cohort sets, family income < $40,000

Group	Obs.	Graduation Rates by Years since First Enrollment			
		4 Years	4 ½ Years	5 Years	6 Years
Full Sample					
1993 – 1999	2,611	.045	.053	.062	.047
1993 – 2000	3,009	.008	-.016	-.001	.005
1994 – 1999	2,296	-.011	-.012	-.037	-.020
1994 – 2000	2,749	-.052	-.078*	-.094	-.074
GPA ≤ 3.28					
1993 – 1999	1,291	.007	.012	-.105	-.131
1993 – 2000	1,505	.002	-.045	-.145*	-.145*
1994 – 1999	1,131	.003	-.018	-.197**	-.202**
1994 – 2000	1,379	-.002	-.053	-.190***	-.171**
GPA > 3.28					
1993 – 1999	1,318	.116	.127	.269***	.229*
1993 – 2000	1,503	.003	.008	.150	.173
1994 – 1999	1,162	.022	.037	.161	.200*
1994 – 2000	1,369	-.083	-.069	.061	.097
GPA > 3.78					
1993 – 1999	456	-.031	-.0002	.199	-.114
1993 – 2000	518	-.160	-.129	.145	.049
1994 – 1999	404	-.091	-.032	.151	.054
1994 – 2000	467	-.178	-.102	.174	.124

Robust standard errors are reported in parentheses. *, **, and *** denote statistical significance at the 10, 5, and 1 percent-level, respectively. Estimates are from difference-in-differences kernel matching using an Epanechnikov kernel function with various bandwidth parameters, *h*. We report estimates for students with below average or average high school GPAs (≤ 3.28), above average high school GPAs (> 3.28), and high school GPAs greater than one standard deviation above the mean (> 3.78).

Table 2.C3 NMLLS kernel matching estimates with alternative cohort sets, family income ≥ $40,000

		Graduation Rates by Years since First Enrollment			
Group	Obs.	4 Years	4 ½ Years	5 Years	6 Years
Full Sample					
1993 – 1999	10,140	-.030	-.019	.005	.022
1993 – 2000	12,195	-.046*	-.036	-.004	.017
1994 – 1999	8,908	-.035	-.028	-.015	-.010
1994 – 2000	10,955	-.048*	-.042	-.014	-.001
GPA ≤ 3.28					
1993 – 1999	5,015	-.020	-.039	-.014	-.010
1993 – 2000	6,001	-.016	-.046	-.022	-.011
1994 – 1999	4,335	-.018	-.040	-.035	-.055
1994 – 2000	5,316	-.017	-.049	-.026	-.038
GPA > 3.28					
1993 – 1999	5,121	-.017	.024	.055	.078*
1993 – 2000	6,192	-.044	.006	.050	.075
1994 – 1999	4,571	-.027	.011	.033	.058
1994 – 2000	5,639	-.043	.002	.045	.075
GPA > 3.78					
1993 – 1999	1,902	-.0001	.102	.113	.141*
1993 – 2000	2,320	-.042	.052	.083	.101
1994 – 1999	1,701	.055	.109	.096	.130
1994 – 2000	2,116	-.012	.041	.059	.100

Robust standard errors are reported in parentheses. *, **, and *** denote statistical significance at the 10, 5, and 1 percent-level, respectively. Estimates are from difference-in-differences kernel matching using an Epanechnikov kernel function with various bandwidth parameters, *h*. We report estimates for students with below average or average high school GPAs (≤ 3.28), above average high school GPAs (> 3.28), and high school GPAs greater than one standard deviation above the mean (> 3.78).

Appendix 2.D Accumulation of credits models

Table 2.D1 Cumulative credit-taking by year since first enrollment, difference-in-differences kernel matching, 1994-1999

		Cumulative Credits by Year Since First Enrollment					
Group	Obs.	1	2	3	4	5	6
Full Sample	11,209	-.374 (.755)	-.633 (1.673)	.094 (2.610)	.359 (3.443)	1.160 (3.899)	1.903 (4.098)
$\bar{Y}$		21.834	40.315	56.372	71.446	81.615	86.387
GPA ≤ 3.28	5,473	-1.480 (1.016)	-3.361 (2.279)	-3.453 (3.611)	-5.793 (4.586)	-7.465 (5.291)	-7.693 (5.591)
$\bar{Y}$		19.040	33.752	46.131	57.592	66.780	57.591
GPA > 3.28	5,734	1.777** (.902)	4.123* (2.104)	6.757** (3.268)	10.809** (4.517)	13.957*** (5.172)	15.355*** (5.395)
$\bar{Y}$		24.679	46.700	66.804	88.558	96.748	101.866
GPA > 3.78	2,105	-.594 (1.199)	1.099 (3.128)	5.319 (5.289)	10.470 (7.472)	12.735 (8.310)	14.802* (8.557)
$\bar{Y}$		27.376	53.024	76.187	98.428	109.629	114.472

Robust standard errors are reported in parentheses. *, **, and *** denote statistical significance at the 10, 5, and 1 percent-level, respectively. Estimates are from difference-in-differences kernel matching performed with a bandwidth of $h = .2$ using the Epanechnikov kernel function. We report estimates for students with below average or average high school 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 baseline cumulative credits earned by high school performance and years since first enrollment for resident students.

Table 2.D2 Cumulative credit-taking by year since first enrollment, difference-in-differences kernel matching, family income < $40,000, 1994-1999

Group	Obs.	Cumulative Credits by Year Since First Enrollment 1 Year	2 Years	3 Years	4 Years	5 Years	6 Years
Full Sample	2,296	-1.723	-0.245	1.637	2.808	3.098	3.617
		-1.962	-3.82	-5.846	-7.665	-8.803	-9.281
$\bar{Y}$		20.75	38.104	53.051	67.034	77.005	82.442
GPA ≤ 3.28	1,131	-3.626	-5.867	-8.928	-11.115	-15.468	-16.037
		-2.232	-4.294	-6.495	-7.923	-9.924	-10.591
$\bar{Y}$		18.047	31.834	43.347	53.54	53.54	66.242
GPA > 3.28	1,162	2.935	13.304***	24.526***	33.005***	38.627***	39.994***
		-2.734	-4.983	-7.47	-10.001	-11.452	-12.084
$\bar{Y}$		23.492	44.463	62.893	80.721	92.731	98.874
GPA > 3.78	404	-4.381*	-0.628	4.091	13.061	13.061	18.191
		-2.556	-7.089	-13.148	-18.96	-18.96	-21.774
$\bar{Y}$		26.365	49.749	70.301	90.726	90.726	108.808

Robust standard errors are reported in parentheses. *, **, and *** denote statistical significance at the 10, 5, and 1 percent-level, respectively. Estimates are from difference-in-differences kernel matching performed with a bandwidth of $h = .2$ using the Epanechnikov kernel function. We report estimates for students with below average or average high school 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 baseline cumulative credits earned by high school performance and years since first enrollment for resident students.

Table 2.D3 Cumulative credit-taking by year since first enrollment, difference-in-differences kernel matching, family income ≥ $40,000, 1994-1999

		Cumulative Credits by Year Since First Enrollment					
Group	Obs.	1 Year	2 Years	3 Years	4 Years	5 Years	6 Years
Full Sample	8,908	-.301 (.795)	-.980 (1.848)	-.523 (2.853)	-.227 (3.846)	.831 (4.344)	1.697 (34.532)
$\bar{Y}$		22.107	40.872	57.210	72.559	82.778	87.382
GPA ≤ 3.28	4,335	-.848 (1.168)	-3.147 (2.596)	-2.789 (4.022)	-5.251 (5.250)	-6.2137 (6.009)	-6.337 (6.377)
$\bar{Y}$		19.290	34.235	46.831	58.612	62.083	72.437
GPA > 3.28	4,571	.860 (.880)	1.810 (2.169)	2.973 (3.414)	6.085 (4.774)	8.551 (5.428)	10.097* (5.630)
$\bar{Y}$		24.979	47.641	67.793	86.781	97.964	102.623
GPA > 3.78	1,701	-.395 (1.311)	1.913 (3.311)	6.943 (5.510)	12.085 (7.756)	14.463 (8.659)	16.734* (8.976)
$\bar{Y}$		27.613	53.793	77.568	100.236	111.264	115.802

Robust standard errors are reported in parentheses. *, **, and *** denote statistical significance at the 10, 5, and 1 percent-level, respectively. Estimates are from difference-in-differences kernel matching performed with a bandwidth of h = .2 using the Epanechnikov kernel function. We report estimates for students with below average or average high school 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 baseline cumulative credits earned by high school performance and years since first enrollment for resident students.

Appendices to Chapter 3

Appendix 3.A Mathematical derivations

Recall r is the discount rate, Y_H and Y_C are earnings with a high school diploma and a baccalaureate degree, respectively, and F is direct full-time schooling costs per year. Students prefer a six-year, .75 FTE (i.e., 30 hour per week) employment approach to traditional baccalaureate degree attainment (i.e., .25 FTE, four-year path) when

$$(3.A1)\ \frac{3}{4}\sum_{t=1}^{6}\frac{Y_H}{(1+r)^t}+\sum_{t=7}^{T}\frac{Y_C}{(1+r)^t}-\sum_{t=1}^{6}\frac{F}{(1+r)^t}>\frac{1}{4}\sum_{t=1}^{6}\frac{Y_H}{(1+r)^t}+\sum_{t=5}^{T}\frac{Y_C}{(1+r)^t}-\sum_{t=1}^{4}\frac{F}{(1+r)^t}.$$

To simplify this expression, we pull out constant terms from the summation operators, combine like terms, and apply the rule of finite geometric series which states that

$$(3.A2)\ \sum_{k=1}^{n}a^k=\frac{a(1-a^n)}{1-a},$$

giving the following equation:

$$(3.A3)\ \frac{Y_H}{4}\left[\frac{\frac{3}{1+r}\left[1-\left(\frac{1}{1+r}\right)^6\right]}{1-\frac{1}{1+r}}-\frac{\frac{1}{1+r}\left[1-\left(\frac{1}{1+r}\right)^4\right]}{1-\frac{1}{1+r}}\right]+Y_C\left[\sum_{t=7}^{T}\frac{1}{(1+r)^t}-\sum_{t=5}^{T}\frac{1}{(1+r)^t}\right]+F\left[\sum_{t=1}^{4}\frac{1}{(1+r)^t}-\sum_{t=1}^{6}\frac{1}{(1+r)^t}\right]>0.$$

Removing summation notation for the second and third terms allows the expression to be reduced to:

$$(3.A4)\ \frac{Y_H}{4}\left[\frac{\frac{3}{1+r}\left[1-\left(\frac{1}{1+r}\right)^6\right]}{1-\frac{1}{1+r}}-\frac{\frac{1}{1+r}\left[1-\left(\frac{1}{1+r}\right)^4\right]}{1-\frac{1}{1+r}}\right]-(Y_C+F)\left[\frac{1}{(1+r)^5}+\frac{1}{(1+r)^6}\right]>0$$

Simplifying the second term and adding it to both sides produces:

$$(3.A5)\ \frac{Y_H}{4}\left[\frac{\frac{3}{1+r}\left[1-\left(\frac{1}{1+r}\right)^6\right]}{1-\frac{1}{1+r}}-\frac{\frac{1}{1+r}\left[1-\left(\frac{1}{1+r}\right)^4\right]}{1-\frac{1}{1+r}}\right]>(Y_C+F)\left[\frac{r+2}{(1+r)^6}\right]$$

Moving all the discount rate parameters to the left-hand side and simplifying yields the solution in equation (2):

$$(3.A6)\ \frac{2(1+r)^6+(1+r)^2-3}{r(r+2)}>\frac{4[Y_C+F]}{Y_H}.$$

Appendix 3.B U.S. News & World Report Classification

Table 3.B1 U.S. News & World Report College Rankings, by institution type, 2005

	Highly Selective Private Schools	
Top 50 Public Schools	Top 65 Private Schools	Top 50 Liberal Arts Schools
University of California–Berkeley	Harvard University	Amherst College
University of Virginia	Princeton University	Williams College
University of Michigan–Ann Arbor	Yale University	Swarthmore College
University of California–Los Angeles	University of Pennsylvania	Wellesley College
University of North Carolina–Chapel Hill	Duke University	Carleton College
College of William and Mary	MIT	Middlebury College
University of Wisconsin–Madison	Stanford University	Pomona College
University of California–San Diego	California Institute of Tech.	Bowdoin College
University of Illinois	Columbia University	Davidson College
Georgia Institute of Technology	Dartmouth College	Haverford College
University of California–Davis	Northwestern University	Claremont-McKenna
University of California–Irvine	Washington Univ. of St. Louis	Wesleyan University
University of California–Santa Barbara	Brown University	Grinnell College
University of Texas–Austin	Cornell University	Vassar College
University of Washington	Johns Hopkins University	Harvey Mudd College
Pennsylvania State University	University of Chicago	Washington and Lee
University of Florida	Rice University	Smith College
University of Maryland–College Park	Notre Dame University	Hamilton College
Rutgers University–New Brunswick	Vanderbilt University	Colgate University
University of Georgia	Emory University	Oberlin College
University of Iowa	Carnegie Mellon University	Colby College
Miami University (Ohio)	Georgetown University	Bates College
Ohio State University	Wake Forest University	Bryn Mawr College
Purdue University	Tufts University	Colorado College
Texas A&M–College Station	Univ. of Southern California	Macalester College
University of Connecticut	Brandeis University	Scripps College
University of Delaware	New York University	Mt. Holyoke College
University of Minnesota–Twin Cities	Case Western Reserve	Barnard College
Indiana University	Lehigh University	Kenyon College
Michigan State University	Univ. of Rochester	College of the Holy Cross
Clemson University	Tulane University	Trinity College

Note: Adopted from Bound *et al.* (2012) and the 2005 U.S. News & World Report College Rankings. Highly selective private colleges also include the four U.S. Armed Services Academies: the U.S. Military Academy at Westpoint, the U.S. Naval Academy, the U.S. Coast Guard Academy, and the U.S. Air Force Academy.

Table 3.B1 U.S. News & World Report College Rankings, by Institution Type, 2005 (continued)

	Highly Selective Private Schools	
Top 50 Public Schools	Top 65 Private Schools	Top 50 Liberal Arts Schools
SUNY at Binghamton	Rensselaer Polytechnic	Lafayette College
University of California–Santa Cruz	Yeshiva University	Occidental College
University of Colorado–Boulder	George Washington Univ.	Bard College
Virginia Tech.	Pepperdine University	Furman University
University of California–Riverside	Syracuse University	Whitman College
Iowa State University	Worcester Polytechnic	Union College
North Carolina State University	Boston University	Franklin and Marshall
University of Alabama	University of Miami	Sewanee College
University of Missouri–Columbia	Fordham University	University of Richmond
Auburn University	Southern Methodist Univ.	Connecticut College
University of Kansas	Brigham Young University	Centre College
University of Tennessee–Knoxville	Clark University	Dickinson College
University of Vermont	Stevens Inst. of Technology	Skidmore College
Ohio University	St. Louis University	Gettysburg College
University of Arizona	Baylor University	Pitzer College
University of Massachusetts–Amherst	American University	DePauw University
University of Nebraska–Lincoln	Howard University	Rhodes College
University of New Hampshire	Marquette University	Reed College
	University of Denver	
	University of Tulsa	
	Texas Christian University	
	University of Dayton	
	Drexel University	
	Illinois Institute of Technology	
	University of San Diego	
	Catholic University	
	Loyola University	
	Univ. of San Francisco	
	University of the Pacific	
	New School	
	Northeastern University	
	Seton Hall University	
	University of St. Thomas	

Note: Adopted from Bound *et al.* (2012) and the 2005 U.S. News & World Report College Rankings. Highly selective private colleges also include the four U.S. Armed Services Academies: the U.S. Military Academy at Westpoint, the U.S. Naval Academy, the U.S. Coast Guard Academy, and the U.S. Air Force Academy.

Appendices to Chapter 4

Appendix 4.A Flexible difference-in-differences results

Table 4.A1 Common trends assumption test, American Community Survey definition, 1994-1999

Leads/Lags	First Declared	Degree Earned
NMLLS $_{t-3}$	.026	.064
	(.037)	(.075)
NMLLS $_{t-2}$	-.052	-.018
	(.040)	(.081)
NMLLS $_{t-1}$	-.058	.050
	(.042)	(.073)
NMLLS $_{t0}$	.026	-.003
	(.037)	(.074)
NMLLS $_{t+1}$	.009	.070
	(.036)	(.070)
R^2	.0609	.1249
Prob > F	.464	.666
Observations	11,258	4,453

Robust standard errors are reported in parentheses. Ordinary least squares estimates for all students entering UNM between 1994 – 1999 given. Reported coefficients are on interactions between cohort years and a resident dummy variable. Models include resident and cohort dummies as well as controls for race, ethnicity, standardized test scores, high school GPA, gender, and family income. The period t_0 is 1997, the year the NMLLS was implemented. *, **, and *** denote statistical significance at the 10, 5, and 1 percent-level, respectively. NMLLS$_{t+2}$ (1999) serves as the base year. Prob > F displays the p-value of the null hypothesis that estimated coefficients on leading periods are jointly different from zero.

Appendix 4.B. Categorizing of STEM majors

Table 4.B1 Majors classified as STEM according to the American Community Survey

ACS Code	ACS Code Description
2402	Biological engineering
2403	Architectural engineering
2404	Biomedical engineering
2405	Chemical engineering
2406	Civil engineering
2407	Computer engineering
2408	Electrical engineering
2409	Engineering mechanics, physics, and science
2410	Environmental engineering
2411	Geological and geophysical engineering
2412	Industrial and manufacturing engineering
2413	Materials engineering and materials science
2414	Mechanical engineering
2415	Metallurgical engineering
2416	Mining and mineral engineering
2417	Naval architecture and marine engineering
2418	Nuclear engineering
2419	Petroleum engineering
2499	Miscellaneous engineering
2500	Engineering technologies
2501	Engineering and industrial management
2502	Electrical engineering technology
2503	Industrial production technologies
2504	Mechanical engineering related technologies
2599	Miscellaneous engineering technologies
3600	Biology
3601	Biochemical sciences
3602	Botany
3603	Molecular biology
3604	Ecology
3605	Genetics
3606	Microbiology
3607	Pharmacology
3608	Physiology
3609	Zoology
3611	Neuroscience
3699	Miscellaneous biology

Table 4.B1 (continued)

ACS Code	ACS Code Description
3700	Mathematics
3701	Applied mathematics
3702	Statistics and decision science
3801	Military technologies
4002	Nutrition sciences
4003	Neuroscience
4005	Mathematics and computer science
4006	Cognitive science and biopsychology
5000	Physical sciences
5001	Astronomy and astrophysics
5002	Atmospheric sciences and meteorology
5003	Chemistry
5004	Geology and earth science
5005	Geosciences
5006	Oceanography
5007	Physics
5008	Materials science
5098	Multi-disciplinary or general science
5102	Nuclear, industrial radiology, and biological technologies
5901	Transportation sciences and technologies
6106	Health and medical preparatory programs
6108	Pharmacy, pharmaceutical sciences, and administration
6202	Actuarial science
6212	Miscellaneous information systems and statistics

The code list from the American Community Survey was referenced 22 Jan 2018 and can be found online at https://www2.census.gov/programssurveys/acs/tech_docs/code_lists/2016_ACS_Code_Lists.pdf. See Sjoquist and Winters (2015a) for a more exhaustive list that categorizes majors into other categories including liberal arts, health-related, social sciences, education, and business.

Table 4.B2 Majors classified as STEM according to the STEM Collaborative Center at the University of New Mexico, Broad Definition

Major Code(s)	Major Description
5, ANTH	Anthropology
6, ARCH	Architecture
249, BIOC	Biochemistry
12, BIOL	Biology
15, CHE	Chemical Engineering
16, CHEM	Chemistry
17, CE	Civil Engineering
171, CPE	Computer Engineering
109, 168, ACS, CS	Computer Science
262, CONE	Construction Engineering
263, 474, CMGT, CONM	Construction Management
22, 23, DEHY, DHYG	Dental Hygiene
340, EPS	Earth and Planetary Sciences
27, ECON	Economics
173, EE	Electrical Engineering
379, EMS	Emergency Medical Services
438, ENSC	Environmental Science
371, GENG	General Engineering
46, GEOG	Geography
481, HMHV	Health, Medicine and Human Values
INGV	Integrative Studies
110	Management Information Systems
64, MATH	Mathematics
65, ME	Mechanical Engineering
353, MEDL	Medical Laboratory Sciences
76, NE	Nuclear Engineering
77, 456, NUR, NURS	Nursing
24, NDIT	Nutrition/Dietetics
81, PHYC	Physics
405, PAP	Physics and Astrophysics
FANT	Pre Anthropology
FBIC	Pre Biochemistry
FBIO	Pre Biology
FCH	Pre Chemical Engineering
FCHM	Pre Chemistry
FCE	Pre Civil Engineering
FCP	Pre Computer Engineering
FCS	Pre Computer Science
FEPS	Pre Earth and Planetary Sciences
FECO	Pre Economics

Table 4.B2 (continued)

Major Code(s)	Major Description
FEE	Pre Electrical Engineering
FESC	Pre Environmental Science

The code list was provided by the Office of Institutional Analytics at the University of New Mexico. STEM-designated majors are according to the University of New Mexico STEM Collaborative Center and can be found online at https://stem.unm.edu/tools-for-faculty-and-staff/5517-broad-data.pdf (accessed 24 Jan 2018). This is considered the “broad” list of STEM majors at the University of New Mexico.

Table 4.B3 Majors classified as STEM according to the STEM Collaborative Center at the University of New Mexico, Narrow Definition

Major Code(s)	Major Description
249, BIOC	Biochemistry
12, BIOL	Biology
15, CHE	Chemical Engineering
16, CHEM	Chemistry
17, CE	Civil Engineering
171, CPE	Computer Engineering
109, 168, ACS, CS	Computer Science
262, CONE	Construction Engineering
263, 474, CMGT, CONM	Construction Management
340, EPS	Earth and Planetary Sciences
173, EE	Electrical Engineering
438, ENSC	Environmental Science
371, GENG	General Engineering
64, MATH	Mathematics
65, ME	Mechanical Engineering
76, NE	Nuclear Engineering
81, PHYC	Physics
405, PAP	Physics and Astrophysics
FANT	Pre Anthropology
FBIC	Pre Biochemistry
FBIO	Pre Biology
FCH	Pre Chemical Engineering
FCHM	Pre Chemistry
FCE	Pre Civil Engineering
FCP	Pre Computer Engineering
FCS	Pre Computer Science
FCOE	Pre Construction Engineering
FCON	Pre Construction Management
FEPS	Pre Earth and Planetary Science
FEE	Pre Electrical Engineering
FESC	Pre Environmental Science
FMAT	Pre Mathematics
FME	Pre Mechanical Engineering

Table 4.B3 (continued)

FNE	Pre Nuclear Engineering
FPHY	Pre Physics
FSTA	Pre Statistics
STAT	Statistics

The code list was provided by the University of New Mexico. STEM-designated majors are according to the University of New Mexico STEM Collaborative Center and can be found online at https://stem.unm.edu/common/pdfs/17-benchmarking-narrow.pdf (accessed 5 Feb 2018). This is considered the "narrow" list of STEM majors at the University of New Mexico.

Appendix 4.C Alternative STEM definitions

Table 4.C1 NMLLS and major choice by academic preparation, UNM narrow STEM definition, 1994-1999

Group	Obs.	First Declared STEM	Obs.	Majored in STEM
Full Sample	11,209	.022 (.029)	4,692	.009 (.052)
$\bar{Y}$		.197		.194
HSGPA ≤ 3.28	5,473	-.022 (.031)	1,574	.121 (.086)
$\bar{Y}$		.153		.110
HSGPA > 3.28	5,734	.084* (.045)	3,117	-.011 (.068)
$\bar{Y}$		.312		.238
HSGPA > 3.78	2,105	-.099 (.063)	1,357	-.094 (.099)
$\bar{Y}$		.299		.332

Robust standard errors are reported in parentheses. *, **, and *** denote statistical significance at the 10, 5, and 1 percent-level, respectively. Estimates are from difference-in-differences kernel matching performed with a bandwidth of h = .2 using the Epanechnikov kernel function. We report estimates for students with below average or average high school 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 the baseline rate of STEM major choice by academic preparation.

Table 4.C2 NMLLS and major choice by academic preparation, UNM broad STEM definition, 1994-1999

Group	Obs.	First Declared STEM	Obs.	Majored in STEM
Full Sample	11,209	.025	4,692	-.033
		(.033)		(.057)
$\bar{Y}$		.264		.276
HSGPA ≤ 3.28	5,473	-.009	1,574	.065
		(.041)		(.096)
$\bar{Y}$		.217		.187
HSGPA > 3.28	5,734	.079*	3,117	-.054
		(.047)		(.072)
$\bar{Y}$		.312		.323
HSGPA > 3.78	2,105	-.126*	1,357	-.112
		(.067)		(.103)
$\bar{Y}$		.364		.400

Robust standard errors are reported in parentheses. *, **, and *** denote statistical significance at the 10, 5, and 1 percent-level, respectively. Estimates are from difference-in-differences kernel matching performed with a bandwidth of $h = .2$ using the Epanechnikov kernel function. We report estimates for students with below average or average high school 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 the baseline rate of STEM major choice by academic preparation.

Appendix 4.D Alternative bandwidths

Table 4.D1 NMLLS kernel matching estimates with bandwidths of 0.1, 0.2, and 0.3; 1994-1999

	STEM First Declared	
Group	Obs.	Estimate
Full Sample		
$h = 0.1$	11,207	.025
$h = 0.2$	11,209	.026
$h = 0.3$	11,210	.027
GPA ≤ 3.28		
$h = 0.1$	5,470	-.063*
$h = 0.2$	5,473	-.068*
$h = 0.3$	5,474	-.057
GPA > 3.28		
$h = 0.1$	5,732	.115**
$h = 0.2$	5,734	.121***
$h = 0.3$	5,735	.119***
GPA > 3.78		
$h = 0.1$	2,103	-.088
$h = 0.2$	2,105	-.063
$h = 0.3$	2,105	-.049

Robust standard errors are reported in parentheses. *, **, and *** denote statistical significance at the 10, 5, and 1 percent-level, respectively. Estimates are from difference-in-differences kernel matching using an Epanechnikov kernel function with various bandwidth parameters, *h*. We report estimates for students with below average or average high school GPAs (≤ 3.28), above average high school GPAs (> 3.28), and high school GPAs greater than one standard deviation above the mean (> 3.78).

Table 4.D2 NMLLS kernel matching estimates with bandwidths of 0.1, 0.2, and 0.3; 1994-1999

	STEM Degree Earned	
Group	Obs.	Estimate
Full Sample		
$h = 0.1$	4,437	-.024
$h = 0.2$	4,438	-.012
$h = 0.3$	4,439	-.016
GPA ≤ 3.28		
$h = 0.1$	1,506	.156*
$h = 0.2$	1,507	.147
$h = 0.3$	1,508	.134
GPA > 3.28		
$h = 0.1$	2,929	-.045
$h = 0.2$	2,930	-.051
$h = 0.3$	2,930	-.065
GPA > 3.78		
$h = 0.1$	1,271	-.028
$h = 0.2$	1,271	-.061
$h = 0.3$	1,271	-.118

Robust standard errors are reported in parentheses. *, **, and *** denote statistical significance at the 10, 5, and 1 percent-level, respectively. Estimates are from difference-in-differences kernel matching using an Epanechnikov kernel function with various bandwidth parameters, h. We report estimates for students with below average or average high school GPAs (≤ 3.28), above average high school GPAs (> 3.28), and high school GPAs greater than one standard deviation above the mean (> 3.78).

Appendix 4.E Alternative cohorts

Table 4.E1 NMLLS kernel matching estimates with alternative cohort sets

	First Declared STEM Major	
Group	Obs.	Estimate
Full Sample		
1993 – 1999	12,788	.027
1993 – 2000	15,308	.013
1994 – 1999	11,209	.026
1994 – 2000	13,756	.012
GPA ≤ 3.28		
1993 – 1999	6,335	-.036
1993 – 2000	7,564	-.072
1994 – 1999	5,473	-.068*
1994 – 2000	6,725	-.097**
GPA > 3.28		
1993 – 1999	6,451	.091**
1993 – 2000	7,742	.079**
1994 – 1999	5,734	.121***
1994 – 2000	7,029	.112**
GPA > 3.78		
1993 – 1999	2,364	-.025
1993 – 2000	2,860	-.018
1994 – 1999	2,105	-.063
1994 – 2000	2,601	-.054

Robust standard errors are reported in parentheses. *, **, and *** denote statistical significance at the 10, 5, and 1 percent-level, respectively. Estimates are from difference-in-differences kernel matching using an Epanechnikov kernel function with various freshmen cohorts included. We report estimates for students with below average or average high school GPAs (≤ 3.28), above average high school GPAs (> 3.28), and high school GPAs greater than one standard deviation above the mean (> 3.78).

Table 4.E2 NMLLS kernel matching estimates with alternative cohort sets

	Earned STEM Degree	
Group	Obs.	Estimate
Full Sample		
1993 – 1999	4,932	-.024
1993 – 2000	5,953	.001
1994 – 1999	4,438	-.012
1994 – 2000	5,466	.002
GPA ≤ 3.28		
1993 – 1999	1,712	.165*
1993 – 2000	2,037	.165**
1994 – 1999	1,507	.147
1994 – 2000	1,836	.155*
GPA > 3.28		
1993 – 1999	3,219	-.068
1993 – 2000	3,915	-.040
1994 – 1999	2,930	-.051
1994 – 2000	3,629	-.034
GPA > 3.78		
1993 – 1999	1,400	-.104
1993 – 2000	1,711	-.083
1994 – 1999	1,271	-.061
1994 – 2000	1,581	-.062

Robust standard errors are reported in parentheses. *, **, and *** denote statistical significance at the 10, 5, and 1 percent-level, respectively. Estimates are from difference-in-differences kernel matching using an Epanechnikov kernel function with various freshmen cohorts included. We report estimates for students with below average or average high school GPAs (≤ 3.28), above average high school GPAs (> 3.28), and high school GPAs greater than one standard deviation above the mean (> 3.78).

References

Aina, C., & Pastore, F. (2012). *Delayed graduation and overeducation: A test of the human capital model versus the screening hypothesis* (No. 6413). Institute for the Study of Labor (IZA).

Angrist, J. D. and A. B. Krueger. (1991). Does compulsory school attendance affect schooling and earnings. Quarterly Journal of Economics, 106, 979-1014.

Arcidiacono, P. (2004). Ability sorting and the returns to college major. Journal of Econometrics, 121(1-2), 343-375.

Arcidiacono, P., Hotz, V. J., & Kang, S. (2012). Modeling college major choices using elicited measures of expectations and counterfactuals. Journal of Econometrics, 166(1), 3-16.

Arcidiacono, P., Aucejo, E. M., & Hotz, V. J. (2016). University differences in the graduation of minorities in STEM fields: evidence from California. The American Economic Review, 106(3), 525-562.

Arrow, K. J. (1973). Higher education as a filter. Journal of Public Economics, 2(3), 193-216.

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

Beffy, M., Fougere, D., & Maurel, A. (2012). Choosing the field of study in postsecondary education: Do expected earnings matter?. Review of Economics and Statistics, 94(1), 334-347.

Berger, M. C. (1988). Predicted Future Earnings and Choice of College Major. Industrial and Labor Relations Review, 41(3), 418.

Bettinger, E. (2004). How financial aid affects persistence. In C.M. Hoxby (Ed.), College choices: The economics of where to go, when to go, and how to pay for it (pp. 207-238). Chicago, IL: University of Chicago Press.

Betts, J. R., & Grogger, J. (2003). The impact of grading standards on student achievement, educational attainment, and entry-level earnings. *Economics of Education Review*, *22*(4), 343-352.

Binder, M., & Ganderton. P. T. (2002). Incentive effects of New Mexico's merit-based state scholarship program: Who responds and how? In D. E. Heller & P. Marin. (Eds), Who should we help? The negative social consequences of merit scholarships (pp. 41-56). Cambridge, MA: The Civil Rights Project, Harvard University.

________. (2004). The New Mexico lottery scholarship: Does it help minority and low-income students? In D. E. Heller & P. Marin. (Eds), State merit scholarships and racial inequality (pp. 101-122). Cambridge, MA: The Civil Rights Project, Harvard University.

Bound, J., Lovenheim, M. F., & Turner, S. (2010). Increasing time to baccalaureate degree in the United States (No. w15892). National Bureau of Economic Research.

Bound, J., Lovenheim, M. F., & Turner, S.. (2012). Increasing time to baccalaureate degree in the United States. *Education Finance and Policy*, *7*(4), 375-424.

Bowen, W. G., Chingos, M. M., & McPherson, M. S. (2009). *Crossing the finish line: Completing college at America's public universities*. Princeton University Press.

Brodaty, Thomas, Robert Jean Michel Gary-Bobo, and Ana Prieto. (2009). Does speed signal ability? A test of Spence's theory. Working Paper No. 2009-02. Centre de Recherche en Economie et Statistique.

Bruce, D. J., and Carruthers, C. K. (2014). Jackpot? The impact of lottery scholarships on enrollment in tennessee. Journal of Urban Economics, 81, 30-44.

Card, D. (1995). "Using geographic variation in college proximity to estimate the return to schooling", in: Louis N. Christofides, E. Kenneth Grant and Robert Swidinsky (Eds.), Aspects of labour market behaviour: essays in honour of John Vanderkamp (University of Toronto Press, Toronto, Canada) pp. 201-222.

________. (1999). The causal effect of education on earnings. In O. Ashenfelter and D. Card (Eds.) Handbook of Labor Economics, Volume 3, Part A, 1801-1863.

________. (2001). Estimating the return to schooling: Progress on some persistent econometric problems. *Econometrica*, *69*(5), 1127-1160.

Castleman, B. L., and Long, B. T. (2013). Looking beyond enrollment: the causal effect of need-based grants on college access, persistence, and graduation (No. w19306). National Bureau of Economic Research.

Cattaneo, M. D., Jansson, M., and X. Ma. (2017). Simple local polynomial density estimators. Working paper, University of Michigan.

Chabé-Ferret, S. (2015). Analysis of the bias of matching and difference-in-difference under alternative earnings and selection processes. Journal of Econometrics, 185, 110-123.

Cohodes, S. R., and 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, 251-285.

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., & David, B. (2006). Mustard. The Effects of State-Sponsored Merit Scholarships on Course Selection and Major Choice in College. Institute for the Study of Labor (IZA) Discussion Paper Series, (IZA DP No.1953).

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.

Cuñado, J., & de Gracia, F. P. (2012). Does education affect happiness? Evidence for Spain. Social Indicators Research, 108(1), 185-196.

Dillon, E. W., & Smith, J. A. (2013). The determinants of mismatch between students and colleges (No. w19286). National Bureau of Economic Research.

Dynarski, S. (2000). Hope for whom? Financial aid for the middle class and its impact on college attendance. National Tax Journal, 53(3), pt. 2: 629-61.

Flores-Lagunes, A., & Light, A. (2010). Interpreting degree effects in the returns to education. *Journal of Human Resources*, *45*(2), 439-467.

Groot, W., & Oosterbeek, H. (1994). Earnings effects of different components of schooling; Human capital versus screening. *The Review of Economics and Statistics*, *76*(2), 317-321.

Hausman, J. A. (1978). Specification tests in econometrics. *Econometrica: Journal of the Econometric Society*, *46(3)*, 1251-1271.

Heckman, J. J., Lochner, L. J., & Todd, P. E. (2003). *Fifty years of Mincer earnings regressions* (No. w9732). National Bureau of Economic Research.

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

Henry, G. T., & Rubenstein, R. (2002). Paying for grades: Impact of merit-based financial aid on educational quality. Journal of Policy Analysis and Management, 21(1), 93-109.

Imbens, G.W., and D. B. Rubin. (2015). Causal inference for statistics, social, and biomedical sciences: An introduction. New York, NY: Cambridge University Press, 2015.

Imbens, G. W. (2015). Matching methods in practice: Three examples. Journal of Human Resources, 50, 373-419.

Ingalls, Steven J., Daniel J. Pratt, Christopher P. Alexander, Donna M. Jewell, Erich Lauff, Tiffany L. Mattox, and David Wilson. (2014). Education Longitudinal Study of 2002: Third Follow-up Data File Documentation (NCES 2014-364). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Jaeger, D. A., & Page, M. E. (1996). Degrees matter: New evidence on sheepskin effects in the returns to education. The Review of Economics and Statistics, 733-740.

Jia, Ning. (2017). Heterogeneous effects of state merit scholarship programs: Do program features matter? Working Paper. China Center for Human Capital and Labor Market Research, Central University of Finance and Economics.

Kane, T. J. (2003). A quasi-experimental estimate of the impact of financial aid on college-going (No. w9703). National Bureau of Economic Research.

Kane, T. J. and C. E. Rouse. (1993). Labor market returns to two- and four-year colleges: is a credit a credit and do degrees matter? Working paper no. 4268 (NBER, Cambridge, MA).

Kinsley, P. & Goldrick-Rab, S. (2015). "Making the Grade: The Academic Side of College Life Among Financial Aid Recipients." Working Paper. Wisconsin Hope Lab.

LaLonde, R. J. (1986). Evaluating the econometric evaluations of training programs with experimental data. The American Economic Review, 76, 604-620.

Lemieux, Thomas. (2006). The "Mincer equation" thirty years after *schooling, experience, and earnings*. In *Jacob Mincer: A Pioneer of Modern Labor Economics* (pp. 127-145), edited by Shoshana Grossbard, New York: Springer.

Linsenmeier, D. M., Rosen, H. S., & Rouse, C. E. (2006). Financial aid packages and college enrollment decisions: An econometric case study. The Review of Economics and Statistics, 88(1), 126-145.

McCrary, J. 2008. Manipulation of the running variable in the regression discontinuity design: A density test. Journal of Econometrics 142(2): 698-714.

McPherson, M. S., & Schapiro, M. O. (1991). Does student aid affect college enrollment? New evidence on a persistent controversy. The American Economic Review, 81(1), 309-318.

Mincer, J. A. (1974). Schooling and earnings. In *Schooling, experience, and earnings* (pp. 41-63). National Bureau of Economic Research.

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

Montmarquette, C., Cannings, K., & Mahseredjian, S. (2002). How do young people choose college majors? Economics of Education Review, 21(6), 543-556.

Ness, E. C. (2008). Merit aid and the politics of education. In P. G. Altbach (Ed.), Studies in higher education. New York, NY: Routledge Press.

Rumberger, R. W., & Thomas, S. L. (1993). The economic returns to college major, quality and performance: A multilevel analysis of recent graduates. Economics of Education Review, 12(1), 1-19.

Scott-Clayton, J. (2011). On money and motivation a quasi-experimental analysis of financial incentives for college achievement. Journal of Human Resources, 46, 614-646.

Seftor, N. S., & Turner, S. E. (2002). Back to school: Federal student aid policy and adult college enrollment. Journal of Human Resources, 336-352.

Sjoquist, D. L., and Winters, J. V. (2012). Building the stock of college-educated labor revisited. Journal of Human Resources, 47, 270-285.

________. (2015a). State merit aid programs and college major: A focus on STEM. Journal of Labor Economics, 33(4), 973-1006.

________. (2015b). State Merit-Based Financial Aid Programs and College Attainment. Journal of Regional Science, 55, 364-390.

Staiger, D. and J. H. Stock. (1997). Instrumental variables regression with weak instruments. Econometrica, 65, 557-586.

Stater, M. (2011). Financial aid, student background, and the choice of first-year college major. Eastern Economic Journal, 37(3), 321-343.

Stuart, E. A, Huskamp, H. A., Duckworth, K., Simmons, J., Song, Z., Chernew, M. E., and C.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, 166-182.

Turner, S. (2004). Going to college and finishing college. In C. Hoxby (Ed.), College choices: The economics of where to go, when to go, and how to pay for it. Chicago, IL: University of Chicago Press (pp. 27-33).

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

Wang, X. (2013). Why students choose STEM majors: Motivation, high school learning, and postsecondary context of support. American Educational Research Journal, 50(5), 1081-1121.

Welch, J. G. (2014). HOPE for community college students: The impact of merit aid on persistence, graduation, and earnings. Economics of Education Review, 43, 1-20.

Wiswall, M., & Zafar, B. (2014). Determinants of college major choice: Identification using an information experiment. The Review of Economic Studies, 82(2), 791-824.

Xue, Y., & Larson, R. (2015). "STEM crisis or STEM surplus? Yes and yes," Monthly Labor Review, U.S. Bureau of Labor Statistics, https://doi.org/10.21916/mlr.2015.14.

Zhang, L. (2011). Does merit-based aid affect degree production in STEM fields? Evidence from Georgia and Florida. The Journal of Higher Education, 82(4), 389-415.