2014

Determinants of Undergraduate GPA and Persistence at Connecticut College

Patrick Russo
Connecticut College, prusso@conncoll.edu

Determinants of Undergraduate GPA and Persistence at Connecticut College

Author: Patrick Russo

Thesis Advisor: Professor Terry-Ann Craigie, Ph.D.

AN HONORS THESIS PRESENTED TO THE ECONOMICS

DEPARTMENT IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR HONORS IN THE MAJOR FIELD

Abstract

This thesis uses a unique dataset of five class-years to determine the factors that affect the academic experience of students at Connecticut College, particularly their cumulative grade point averages (CGPA) and persistence into sophomore year. Chapter I examines how changes in high school inputs affect students' performance at Connecticut College. This chapter finds that high school inputs (*viz.*, pupil-teacher ratio, total expenditure per pupil, and expenditure on teacher salaries per pupil) have a significant effect on college GPA even after controlling for demographics and prior academic performance. Chapter II examines the factors affecting persistence at Connecticut College. It finds that various measures of fit between a student and his or her peers have a strong impact on persistence into sophomore year.

CONNECTICUT COLLEGE

NEW LONDON, CONNECTICUT

MAY 2014

Acknowledgements

I would like to thank my advisor, Professor Terry-Ann Craigie, who has provided invaluable assistance and support. Professor Craigie has offered great counsel, guidance and feedback throughout the year. This thesis utilized the methods learned in Professor Craigie's Advanced Econometrics class.

I would like to thank Professor John Nugent, Director of Institutional Research, for his assistance throughout this year. Professor Nugent assisted with data collection, provided thoughtful feedback, and gave several helpful suggestions.

I would also like to thank Professor Candace Howes for her valuable guidance and feedback. Professor Howes's comments and suggestions helped pull together parts of this thesis. Professor Howes' Econometrics I class was the foundation of all econometric methods used.

I would also like to thank Jessica Ricker, Director of Admission Information & Systems, who assisted in data collection. Jessica Ricker (née Dietz) completed her 2006 master's thesis *Qualitative and Quantitative Predictors and Outcome Measures of Success at Small, Residential Liberal Arts Colleges* on a similar topic and is cited in this thesis.

I would like to thank all others who assisted in data collection and those who have discussed this thesis with me in any capacity. I have had fruitful conversations with countless people that helped shape this thesis.

Introduction

Previous research has yielded conflicting results on the effect of school resources on student performance. The first chapter seeks to determine if resources in a student's high school (such as measured by the pupil-teacher ratio or expenditures per pupil) have an impact on a student's collegiate academic performance. By studying five-class years of students at Connecticut College, the study explores the effects of these high school inputs on collegiate GPA). This chapter finds that the pupil-teacher ratio has a significant negative effect on collegiate GPA and that teacher salary expenditure has a significant positive effect. This chapter also finds that total expenditure per pupil has a negative effect on collegiate GPA, which is the opposite of the hypothesized direction.

The second chapter sets out to investigate factors that affect the likelihood of a student persisting at Connecticut College, including how similarities and differences between students and their peers can affect student persistence. By examining different measures of matching between a student and his or her peers, this chapter finds that female persistence decisions are strongly affected by demographics and freshman fall grades, while male persistence decisions are strongly affected by their academic performance relative to their classmates. This chapter also finds that females are more likely to persist when assigned a roommate with similar high school performance, yet males are more likely to persist when assigned a roommate with dissimilar high school performance. The implications of all these findings are discussed in detail.

Table of Contents

Contents

THE EFFECTS OF HIGH SCHOOL INPUTS ON COLLEGIATE ACADEMIC PERFORMANCE

Abstract

Previous research has yielded conflicting results on the effect of school resources on student performance. The first chapter seeks to determine if resources in a student's high school (such as measured by the pupil-teacher ratio or expenditures per pupil) have an impact on a student's collegiate academic performance. By studying five-class years of students at Connecticut College, the study explores the effects of these high school inputs on collegiate GPA). This chapter finds that the pupil-teacher ratio has a significant negative effect on collegiate GPA and that teacher salary expenditure has a significant positive effect. This chapter also finds that total expenditure per pupil has a negative effect on collegiate GPA, which is the opposite of the hypothesized direction.

I. Introduction

Economists argue that an individual's career prospects, future earnings, and standard of living improve when they earn a bachelor's degree. Besides degree attainment, one important measure of achievement in college is a student's collegiate grade point average. A student's undergraduate GPA can be a determining factor as to whether a graduate is hired by an employer, awarded a fellowship, or admitted to a graduate school. Researchers have found that undergraduate GPA has a positive impact on earnings. For example, Jones and Jackson (1990) found that an increase in GPA of one grade point causes an increase in annual earnings of over 8% (for individuals five years after college graduation). Due to the strong effect that undergraduate GPA can have on earnings and employment, it is important to study factors that affect a student's undergraduate GPA.

Of particular question to this report is how high school inputs (*viz.*, the pupil-teacher ratio, expenditure per pupil, and expenditure on teacher salaries per pupil) affect collegiate GPA. There has been relatively little research on whether the effects of these factors have an impact on college performance, yet this is question that is likely of interest to high school educators and college educators alike. In a college classroom, some students may be better prepared for college than others. Can this difference be mostly explained by prior high school performance or is it in part due to the resources (staffing or spending) in a student's high school? In other words, do the resources of a high school continue to affect a student's performance once he or she is in college (even after controlling for the student's prior academic performance)?

This study examines Connecticut College students graduating between 2009 and 2013 and estimates the effects that students' demographics, prior performance, and the resources of students' high schools have on cumulative GPA. This paper combines two active areas of study, those estimating the long-term effects of school inputs and those estimating the determinants of GPA, by using a unique data set consisting of Connecticut College students combined with data from the Census Bureau and National Center for Education Statistics. To the author's knowledge, Betts and Morell (1999) are the only authors to previously connect these two areas of research; they performed a study similar to that in this chapter at the University of California San Diego. This study uses regression analysis to determine how high school inputs (such as the pupil-teacher-ratio, total expenditure per pupil, and expenditure on teacher salaries per pupil) may impact a student's cumulative GPA (CGPA) at Connecticut College. This study finds that pupil-teacher ratio has a negative effect on GPA, total expenditure per pupil has a negative effect on GPA, and that teacher salary per pupil has a positive effect on GPA. More specifically, a 10-student decrease in the pupil-teacher ratio at a student's high school is expected to increase his or

her GPA at Connecticut College by .057 points, holding all else constant. A $1,000 increase in total expenditure per pupil at a student's high school is expected to decrease his or her GPA at Connecticut College by .016, but a $1,000 increase in teacher salary expenditure per pupil at a student's high school is expected to increase his or her GPA at Connecticut College by .053 (holding all else constant).

This chapter contains a literature review discussing previous research on the effects of school inputs and determinants of undergraduate GPA. Next, section III of this chapter discusses the data that is used and the methods used to collect, compile, and prepare it for analysis. Section IV, data problems, discusses issues related to this data set, such as why a composite of prior student performance had to be used as opposed to high school GPA. Next, the empirical method section, section V, discusses the regression equation. The problems with omitted variable bias section, section VI, discusses how a latent variable, academic aptitude, might impact regression estimates. The results section, section VII, provides summaries of the data, presents regression results, and presents the interpretation of these results. This is followed by section VIII, the discussion which provides more in depth interpretations and explains how omitted variable bias may impact the interpretations. The final section, the conclusion, summarizes findings, implications, and proposes future research.

II. Literature Review

Several authors (such as Mosteller, 1995 and Fredriksson, Öckert, and Oosterbeek, 2013) have studied how school inputs (such as the pupil-teacher ratio or expenditure per pupil) affect student test scores, years of schooling, and career earnings. Scholars disagree about whether increasing school inputs (more teachers or more spending) has an impact on student's

performance (Hanushek, 1996; Betts, 1996). Some authors (Stater, 2009; Cohn, Cohn, Balch, & Bradley, 2004; Geiser & Studley, 2002) have found that factors such as prior academic performance (*e.g.*, SAT scores), demographics (*e.g.*, gender and median income in a student's home ZIP code), and financial aid affect students' grade point averages in college.

Relatively few studies, however, have investigated how school inputs in a student's high school may affect his or her college performance. This thesis sets out to bridge the gap between authors studying the effects of high school inputs on test scores and the authors studying the effects of test scores and other predictors of college GPA. To do this, I first examine the research on the effects of school inputs on student test performance and on career. Then I examine the research on factors affecting student's college GPA's.

High School Inputs

Many educators in high schools strive not just for students to succeed in high school, but after high school as well. While some studies have linked school inputs (such as pupil-teacher ratios and expenditure per pupil) to higher earnings and higher educational achievement, most prior research studying the impacts of the pupil-teacher ratio has focused on the effect of the pupil-teacher ratio in lower grades, not in high school. The pupil-teacher ratio is defined as the number of full-time students in a school divided by the number of full-time teachers. The larger the pupil-teacher ratio, the larger the class sizes, and the less individualized attention, the less tailored the learning is to individual students' needs. Therefore, pupil-teacher ratio is expected to have a negative impact on learning and development.

Studies of pupil-teacher ratio and school expenditures can either be short-term (*e.g.*, studying how class size affects current academic performance or academic performance the year after being in a small class) or long-term (*e.g.*, studying how small classes affected a student's

performance several years later). Several studies have found short-term relationships between the pupil-teacher ratio and academic success (such as Mosteller, 1995) or expenditure and academic success (such as Papke, 2004); however, other short-term studies find mixed results (such as described in Hanushek, 1996 and 1999). Several studies examine the longer-term effects of school inputs; however these often find mixed results.

Several studies (such as Nye, Hedges, and Konstantopoulos, 1999 and Mosteller, 1995) have performed short-term and long-term analyses of Project STAR (Student-Teacher Achievement Ratio), a natural experiment in Tennessee in which elementary school students were randomly assigned to classes of different sizes for three consecutive years (*viz*., a student was either in large class (22-25 students) or a small class (13-17 students) for each of first through third grade). The short-term impacts of smaller class sizes included higher scores on both standardized and curriculum based tests; students who had been in smaller classes from first through third grade continued to show better exam scores through seventh grade relative to students who had been in larger classes Mosteller (1995).

Fredriksson, Öckert, and Oosterbeek (2013) find long-term effects of elementary school class sizes—students in smaller classes were more likely to have higher cognitive ability by 13, higher achievement by 16, and wages that were 4.4% higher as adults. Krueger and Whitmore (2001) found that students in the smaller classes were over 9% more likely to take the ACT or SAT. Those who take the ACT or SAT would be more likely to attend college (taking the ACT or SAT is often a requirement for admission into a college or university). Fredriksson, Öckert, and Oosterbeek (2013) find that smaller class sizes in elementary school "increase completed education level, wages, and earnings at age 27 to 42."

By linking project STAR data with federal tax returns, Chetty et al. (2011) found that small class sizes increased the likelihood that a student would attend college. While they did find that teacher quality was positively related to the earnings of experiment subjects at age 27, they did not find a significant effect of smaller elementary school classes on earnings.

Betts (1996) reviews articles trying to link school inputs such as teacher quality, teacher education, pupil-teacher ratio, and spending per pupil to earnings. Some studies reviewed (*e.g.*, Card & Kreuger, 1992) have found significant negative relationships between the pupil-teacher ratio and earnings, while others (*e.g.*, Grogger, 1996) found significant positive relationships between the pupil-teacher ratio and earnings, or no statistically significant relationship (*e.g.*, Wachtel, 1975). Betts (1996) considers the literature "unsettled" on any consensus of school inputs on earnings. He notes that the effects authors find often depend on the level of analysis (such as at the state level versus the district level). Studies that use spending per pupil on the state level (such as by using the *Biennial Survey of Education)* often find a positive significant effect of spending on earnings (Rizzuto & Wachtel (1980) and Nechyba (1990)). Studies that use spending per pupil on a school level do not find consistent effects of spending per pupil on an individual's earnings. Rizzuto and Wachtel (1980) and Wachtel (1975) also find that teacher salary has a significant positive effect on earnings.

Betts and Morell (1999) seem to be the first large study examining how high school inputs, such as the pupil-teacher ratio, affect college GPA for individual students. Betts and Morell (1999) find a negative but insignificant relationship between pupil-teacher ratio and college GPA and a positive significant relationship between teacher experience and college GPA. Betts and Morell's study is limited to students at the University of California San Diego who came from public high schools in California.

Betts and Morell (1999), Dustmann, Rajah, and Soest (2003), Chetty et al. (2011), and Fredriksson, Öckert, and Oosterbeek (2013), are some of the most recent studies of the long-term effects of the pupil-teacher ratio yet they do not even have a consensus. Fredriksson, Öckert, and Oosterbeek (2013) point this out, saying "while there is a large literature estimating the short-term effects of class size, credible estimates of long-term effects of class size are sparse." Dustmann, Rajah, and Soest (2003) propose that the mechanism through which class sizes affect earnings is by affecting a student's decision to stay in school (or, not drop out) and to pursue a higher degree.[1]

This thesis sets out to provide another estimate for the effect of the pupil-teacher ratio and other school inputs on college GPA, while controlling for a student's demographics and performance in high school. While the GPA at one college or university may not be directly comparable to that at another, the factors driving students' GPAs are expected to be similar. Students and their backgrounds do vary between schools (for example, the students at the University of San Diego (that were included in the Betts and Morell (1999) study) may be different from students at Connecticut College[2]). At any college or university, one's cumulative GPA (or GPA at the time of graduation) has important implications for post-graduate employment and post-graduate study.

College GPA

Authors have studied several factors that impact college performance, but almost all studies examine or control for the effect of some measure of prior academic performance (such

[1] This study will not be able to test Dustmann, Rajah, and Soest's hypothesized mechanism of action since all students at Connecticut College will have chosen to continue education beyond high school and to pursue an advanced degree.

[2] According to *Forbes Top Colleges 2013*, UC San Diego has over 23,000 undergraduate students, while Connecticut College has just shy of 1900 students. The student-faculty ratio at UC San Diego is 19, whereas at Connecticut College, it is 9. Forbes ranks Connecticut College as the 102nd ranked college and UC San Diego as the 114th ranked college.

as high school GPA, SAT scores, or a composite measure of performance). Several studies have demonstrated that a student's prior academic performance has strong positive predictive power on a student's undergraduate success (such as Cohn, Cohn, Balch, & Bradley, 2004; Geiser & Studley, 2002; Zwick & Sklar, 2005; Dietz, 2006). These can be measures of, or proxies for, student characteristics such as motivation, intelligence, and ability. Standardized tests, such as the SAT are "intended to assess students' capacity for future learning" (Geiser & Studley, 2002).

In a study of University of California students (Geiser & Studley, 2002), SAT II scores were the best predictor of freshman year GPA, followed by high school GPA (HGPA), followed by SAT I. Freshman year GPA, in turn, is one of the strongest predictors of cumulative undergraduate GPA, or cumulative GPA (CGPA).[3]

Several studies (including Cohn, Cohn, Balch, & Bradley, 2004; Geiser & Studley, 2002; Betts & Morell (1999)) have made an effort to determine factors that have a significant effect on a student's undergraduate GPA. These studies estimate the effect of HSGPA, SAT I Scores, and class rank on undergraduate GPA (and generally find these results to be positive and significant). Generally, these studies control for gender, race, and other demographic variables. Demographics can be so important, in fact, that Geiser and Studley (2002) find that after controlling for family income and parent's education, SAT I scores no longer help explain freshman year GPA. Cohn et al. (2004) find that being a white female has a significant positive effect on college GPA.

[3] Butler and McCauley (1987) find correlations of .89 and above between a student's freshman year GPA and a student's GPA for another year for students at the United States Military Academy. They also find that SAT scores and high school rank are just as correlated with senior-year GPA as freshman year GPA. Estimates of these correlations at civilian colleges were considerably less—Humphreys (1968) found a decline in the correlation between undergraduate first quarter GPA and last quarter GPA (with the correlation between the first and last quarters as low as .35). Similarly, Humphreys also found rapidly declining correlations between high school rank and each semesters GPA. In a previous study of Connecticut College students, Dietz (2006) finds a correlation of .85 between freshman year GPA and CGPA.

Stater (2009) adds the demographics of a student's home ZIP code (*e.g.*, median income, percent of adults who have a bachelor's degree, and percent of home ZIP code urbanized) to the student's personal demographics which were significant in at least some models. Information about a student's ZIP code can help us learn more about the socioeconomic environment of where they live. A high population of educated adults may have a positive impact on local schools and the community as a whole.

Not all studies examining predictors of undergraduate GPA follow the same basic blueprint; for example, Winston and Zimmerman (2002) included SAT scores and background demographics, but not HSGPA. Winston and Zimmerman were most interested in how a student's roommate's SAT scores can affect his or her CGPA, and thus only included the SAT scores of a student and their roommate. Baron and Norman use class rank as an explanatory variable in lieu of HGPA. Most studies use some sort of control for prior academic ability, although the measure used varies from study to study. Instead of any individual measures, Sacerdote (2001) uses a composite index of HSGPA, SAT I scores, and SAT II scores as a measure of high school achievement (each weighted as 1/3 of the index). This index was created by the Admission Office at Dartmouth College. Sacerdote finds that after including the index in his regression model he does not greatly increase his explanatory power of GPA by adding any of the three covariates into the model. Section IV of this chapter discusses the reasons that this chapter uses a composite measure of performance; the wide range of measures of academic performance used in the literature do not suggest that using a composite would be problematic.

Studies have had conflicting results on the effects of school inputs on student outcomes—primarily long-term outcomes. Betts and Morell (1999) are the first, to my knowledge, to examine to direct effects of high school inputs on collegiate GPA. Their study demonstrates that

high school inputs can impact student performance years later. While Betts and Morell test the effects of class sizes in public schools in California, this chapter is not limited to a single state or to public schools (in the case of class sizes). The chapter is also able to test additional high school inputs, two measures of expenditure, for students from public high schools.

III. Data Description and Methods

Individual student-level data were collected at Connecticut College with identifying information removed except for student ID numbers, which were used to merge data from different sources. On-campus sources of data included the Office of the Institutional Researcher, the Office of Admission, the Financial Aid Office and the Registrar's Office. The Office of Admission data include demographic (such as gender, race/ethnicity, home state and ZIP code), previous performance (SAT scores, HGPA, etc.), and information about a student's high school (such as its name, zip code).

Demographic Data

Previous research has shown that males, on average, have lower CGPAs than females. Research has also shown that blacks or African Americans, and Hispanics or Latinos tend to have lower CGPAs than whites. First generation college students, students whose parents did not attend college, have also been shown to have lower CGPAS than students whose had at least patent who attended college.

Median Income

Data from the Census Bureau were merged with the student-level data set by a student's home ZIP code to calculate the median income in a the ZIP code in which a student lives.

Median income of an area can impact socioeconomic status, and is expected to have a positive effect on CGPA.

Percent of Adults in Zip Code with a Bachelor's Degree

Data from the Census Bureau were merged with the student-level data set by a student's home ZIP code to determine the percent of adults residing in a student's home zip code with a bachelor's degree or higher. This is a measure of the socioeconomic status where a student lives and is expected to have a positive impact on CGPA.

Aid Ratio

The amount of need based financial aid was calculated as the *aid ratio*, the dollar amount of the award a student's freshman year divided by the cost of tuition at Connecticut College. This was calculated as a ratio because the tuition at Connecticut College rose each year of study.

High School Inputs

In addition to student level data, this study utilizes publicly available aggregate data such from the National Center for Education Statistics. By knowing the name, ZIP code, and College Entrance Examination Board ID (CEEB) of a student's high school, it is possible to merge this information with data on all high schools to learn more about the high school a student attended. By merging the student level data set with data (the Common Core of Data and the Private School Survey) from the National Center for Education Statistics (part of the U.S. Department of Education) the pupil-teacher ratio from a student's high school could be acquired.[4] By this method it was also possible to determine whether the high school a student attended was a public school or a private school. By linking these data with information on a district, it was possible to

[4] This merging is being done by string, so the merge was originally imprecise and needed to be done in an iterative process. It was not possible to find the matching data on a student's high school for 123 students.

determine the total expenditure per pupil, and the spending on teacher salaries per pupil in the district.

Pupil-Teacher Ratio

The pupil-teacher ratio was created as the number of full time teachers in a school divided by the number of full-time students. This was available for both public and private school students. The pupil-teacher ratio is expected to have a negative relationship with CGPA.

Total Expenditure per Pupil

The total expenditure per pupil was gathered from the NCES data at the district level. This is not as accurate as a measure of total expenditure per pupil in a school would be, but such a measure is unavailable. Prior research has often had to suffice for using high school inputs at the district or state wide level. This measure was only available for students from public high schools. Total expenditure per pupil is expected to have a positive effect on CGPA.

Expenditure on Teacher Salaries Per Pupil

Expenditure on teacher salaries per pupil was also gathered from the NCES data at the district level. This measure was only available for students from public high schools. This is a measure of the amount of money that is spent on teacher salaries. Districts with higher on teacher salaries per pupil may pay teachers higher salaries or may have more teachers per student. This is expected to have a positive effect on CGPA.

IV. Data Problems

HGPA, while used frequently by researchers studying factors affecting GPA in college, could not accurately be used in this study as a measure of past academic performance. High schools use different GPA scales, and some do not use GPA at all. Those that have a GPA may

be on a scale of 4 points, 5 points, 16 points, 100 points, or some other system.[5] In fact, the highest high school GPA was 377.07. An effort was made to estimate the scale that a GPA might have been on, but this was a relatively crude method that did not guarantee accuracy. After reviewing literature, I chose to use an alternative measure similar to that used by Sacerdote (2001), discussed above. Like the Office of Admission at Dartmouth, the Office of Admission at Connecticut College also generates a composite measure of past performance called the Academic Reader Rating (ARR). The ARR is a composite measure of high school GPA and class rank, standardized test scores, and the rigor of the classes a student took and ranges from 1 to 7 (Dietz, 2006).[6] An ARR of 7 is the best and an ARR of 1 is the worst (so a student with an ARR of 7 would be the strongest academically and a student with an ARR of 1 would be the weakest academically). The Office of Admission uses this as a way to compare students on different GPA scales, who take different course-loads, and who take different standardized tests, on a single scale. This study will use ARR as the measure of previous performance in lieu of high school GPA.[7]

Study population and sample.

Data was gathered on the 2475 students who started their undergraduate career at Connecticut College between 2005 and 2009 (this excludes students who transferred to Connecticut College, but includes students who transferred from or dropped out of Connecticut College). Of these students, the 1989 who graduated from Connecticut College in 4 years or less are the basis of this study. Out of the 1989 students in the study, 1904 identified as having

[5] The mean high school GPA was 21.83 and the median was 3.8.
[6] ARR was inverted so that a high ARR indicated stronger high school performance.
[7] ARR regressed on HGPA (after attempting to correct for different scales) and SAT or HSGPA and ACT scores yields an R^2 of .43 or .51, respectively. This indicates that ARR captures the effects of both high school GPA and standardized test scores pretty well.

attended high school in the United States. In merging these 1904 students to the national high school data base, 123 students' high schools could not be identified. This could have been due to a high school changing name, a high school using different names in different places, a change in ZIP code, or a misprint of the ZIP code or high school name. Merging student data with the NCES data was first done by the CEEB ID[8] and subsequently by the high school name and ZIP code. If a match could not be found, this process was reiterated on various forms of the high school name (such as the first three letters of a school name or the first five letters of a school name). The inability to match some students brought our sample down to 1781 students.

Of the students who were matched with their high school, 92 were from high schools where the pupil-teacher ratio was not reported. In public schools the pupil-teacher ratio was taken to be the median of the pupil-teacher ratios from each of 2005 to 2008.[9] The only years within this time frame in which private high schools reported data were in 2006 and 2008. The pupil-teacher ratio in private schools was simply the mean of the two observations (which is the median of two numbers).

Almost all of the public school districts in this study reported total expenditure per pupil and salary expenditure per pupil (which can include salaries of teachers, administrators, support staff, etc.). For only 842 of the 1250 students from public high schools was data on expenditure on teacher salaries in a student's district available. Like with the pupil-teacher ratio, the median of the 5 years of data was used as the measure for each of total expenditure per pupil, and teacher salary per pupil.

[8] Which had to be cross-walked with the NCESID by using information from the Postsecondary Electronic Standards Council.

[9] This allowed for the final figure not to be skewed by grossly misreported data that occurred in a single year. For example, one school reported a pupil-teacher ratio of 254 in 2006, but of around 18 in other years.

V. Empirical Method

To estimate the effects of high school inputs on academic performance at Connecticut College an ordinary least squares regression model was built in much the same way as most researchers who were to studying CGPA (or freshman year GPA). Models included the ARR of students (an index of their prior performance), a vector of demographic controls (*e.g.*, gender and whether a student's parents attended college), and a vector of high school inputs (which differs between models). The OLS model also contains an error term, ε_i, and a latent variable θ_i, a student's academic aptitude. Academic aptitude will be discussed more below. The following OLS model was used:

$$CGPA = \beta_1 ARR_i + \boldsymbol{Demographics_i \beta_2} + \boldsymbol{High\ School\ Inputs_i \beta_3} + \theta_i + \varepsilon_i$$

The first ordinary least squares regression, column 1, in table B regresses CGPA on past performance (*i.e.* ARR) and demographic controls. Column 2 of table B adds a dummy variable for whether or not a student graduated from a private high school. Column 3 of table B regresses CGPA on past performance, demographic controls, the indicator for private high schools, and the pupil-teacher ratio from a student's high school.

Table C presents the results of a series of analyses on students who graduated from public high schools. Column 1 of table C is the same regression as column 3 of table B, but with a sample limited to students who graduated from public high schools. Columns 2 and 3 replace pupil-teacher ratio with total expenditure per pupil and teacher salary expenditure per pupil, respectively. Columns 4 and 5 investigate the inclusion of multiple measures of high school inputs in a regression model.

VI. Omitted Variable Bias

Assuming that a student's academic ability is a product of their aptitude (*i.e.*, their ability before high school) and the resources of their high school (such as the pupil-teacher ratio), a student whose high school had more resources (say, a lower pupil-teacher-ratio) will be better educated (or have higher academic ability) at the end of high school than a student of equal aptitude who attended a high school with fewer resources. Being a highly selective college, Connecticut College does not admit all applicants, only strong applicants (or those with higher academic ability). That being said, some colleges and universities are ranked higher (for example, by U.S. News and World Report or by Forbes) than Connecticut College. Higher ranking institutions are generally more selective than Connecticut College (in that they admit a smaller proportion of applicants) and may be more desirable.[10] With this as a starting point, it is reasonable to hypothesize that students with the highest academic ability level may elect to attend higher ranked colleges or universities and that those with lower academic ability may attend lower ranked colleges or universities. Therefore, aptitude and pupil-teacher ratios may be positively correlated for students enrolled at Connecticut College.

In other words, students with the highest academic aptitude who attended the best high schools (those with lower pupil-teacher ratios) may be more likely to attend an institution ranked higher than Connecticut College (but a student with a slightly lower academic aptitude from the same school might enroll in Connecticut College). On the other hand, a student with low academic aptitude who attended a low quality high school (such as with a high pupil-teacher ratio) might not have been admitted into Connecticut College (but a student with higher academic aptitude from the same high school might be accepted and enroll in Connecticut

[10] While the selectivity of a college is not the only factor students consider when choosing a college or university in which to enroll, it is often an important factor.

College). Through this process of selection (selection of which institution to attend and selection of which students to admit) the Connecticut College student body will include high aptitude students from lower quality schools (those with a higher pupil-teacher ratio) and lower aptitude students from higher quality schools (those with lower pupil-teacher ratios). As such, it is expected that there is a negative relationship between academic aptitude and high school quality for Connecticut College students. This hypothesis implies that pupil-teacher ratio and aptitude are positively correlated at Connecticut College[11]. It is expected that aptitude would have a positive impact on CGPA. Since aptitude is omitted from the regression model, the coefficient on the pupil-teacher ratio is expected to be biased upwards. Since the pupil-teacher ratio is expected to have a negative effect on CGPA, the estimated effect of the pupil-teacher ratio will be attenuated towards zero.

Following a similar line of reasoning, one can apply this to other high school inputs such as total expenditure per pupil and teacher salary expenditure per pupil. Total expenditure per pupil, salary expenditure per pupil, and teacher salary per pupil are expected to be negatively correlated with aptitude, which would cause coefficients on both total expenditure per pupil and teacher salary per pupil to be downward biased (which would attenuate any effect that was found).

VII. Results

Summary statistics for the students in this study can be seen in table A below. It presents the summary statistics for all students in the sample who graduated as well as summary statistics

[11] This argument could likely be generalized to describe students at other institutions, however this could not be used to discuss students pooled from several institutions; in a pooled sample, there would not necessarily be a selection process.

disaggregated by type of high school. The average college GPA in this analysis sample was 3.45 and 38% of the sample is male. In addition, 740 had attended private high schools and 1040 had attended public high schools. Connecticut College students coming from private schools tended to have lower ARR's, less financial aid, and lower pupil-teacher ratios in their high schools than students from public high schools.

[*Place table A here.*]

Analyses

The first portion of these analyses will be conducted on the entire sample and a smaller subset of these analyses (examining the effects of school expenditures) will be conducted on students coming from public high schools. Column 1 of table B regresses CGPA on ARR and demographic variables (male, race/ethnicity, first-generation college student, median income, percent with a bachelor's degree or higher, aid-ratio, and a dummy for class year).[12] As expected, the higher a student's ARR, the higher CGPA that student is expected to achieve at Connecticut College.[13] As several other authors have found, males are expected to have lower CGPA's than females (for example, see Betts and Morell, 1999 or Stater, 2009) and black and Hispanic students are expected to have lower GPA's than white students. Students who are first generation college students are estimated to have a CGPA 0.12 points lower than students who had at least one college educated parent. Unlike Stater (2009), I did not find any significant effect of need-based financial aid on a student's CGPA.[14] The regressions also indicate that

[12] Significance levels are reported using two-tailed t-tests.
[13] No attempt was made to disaggregate this effect into the effect of HGPA vs. the effect of SAT scores. Several authors (such as Geiser and Studley, 2002) have studied how HGPA and SAT scores each affect CGPA. ARR is used in this study because the scale of HGPA varied from high school to high school.
[14] This chapter uses a ratio for the financial aid figure, the percent of Connecticut College's tuition that is paid by need based financial aid in the student's freshman year.

CGPA varies significantly between class years. Students of the class of 2009 are the base category; the coefficient on class of 2010 in column 1 of table 2 means that a student from the class of 2010 is expected to have a GPA .058 points lower than an identical student from the class of 2009 and the coefficient on class of 2011 means that a student from the class of 2011 is expected to have a GPA .072 points lower than an identical student from the class of 2009. Clearly, CGPA varies between years.

Column 2 shows that students who attended a private high school have a lower CGPA, on average, after controlling for their background characteristics and their ARR's. Column 3 indicates that the pupil-teacher ratio in a student's high school has a significant negative effect on CGPA.[15] By the estimate in column 3, a 10 student increase in a student's high school class size is estimated to decrease a student's CGPA by .058. While this does not appear to be a large effect, it is significant (at α=.05) and may be attenuated towards zero as was argued above (so the true effect of the pupil-teacher ratio may actually be greater). Note that once pupil-teacher ratio is included in the regression we find that the coefficient on private high school is even more negative—this is because on average the pupil-teacher ratio is lower in private schools.

[Place Table B Here]

Public Schools

Table C below tests the effects of two additional high school inputs, district-level expenditure per pupil and teacher salary per pupil (for students from public high schools). A regression was run on each of the school inputs independently followed by regressions on combinations of them. After regressing CGPA on demographics and pupil-teacher ratio for

[15] Note that the interaction between private schools and pupil-teacher ratios was tested in the regression model and was found to be insignificant (not shown here), thus we cannot reject that school size has the same effect in private high schools and public high schools.

students from public high schools (column 4 in table C, we again find that the pupil-teacher ratio in public schools has a significant negative effect on CGPA.

The total expenditure per pupil in a district is found to have a significant negative effect on CGPA (column 5 in table C). For every additional $1,000 spent per pupil, CGPA is expected to decrease by .0059 points. Expenditure on teacher salary per pupil[16] is found to have a negative but insignificant effect on CGPA (column 6). Recall that both of these estimates may be downward biased—this could mean that total expenditure per pupil truly has no effect on CGPA, or the estimate could even have the wrong sign. Implications for this bias will be given below in the next section.

Regressing CGPA on teacher salary per pupil while controlling for students' backgrounds and past performance yielded an insignificant positive coefficient on teacher salary per pupil (column 6 of table C). However, when total expenditure per pupil was added back into the regression (column 7), the results show that both total expenditure per pupil and teacher salary per pupil are significant and working in opposite directions. A $1,000 increase in total expenditure per pupil is expected to decrease a student's CGPA by .0175 points, holding teacher salary per pupil constant (*i.e.*, if this added spending is used on something besides on increased spending on teachers). A $1,000 increase in teacher salary per pupil is expected to increase CGPA by .0705, while holding total expenditure per pupil constant (*viz.* if $1,000 per pupil is reallocated from spending not on teacher salaries to teacher salaries). A $1,000 increase in both total expenditure per pupil and expenditure on teacher salaries per pupil is expected to increase a

[16] Note that expenditure on salaries per pupil can increase by either increasing the pay of teachers or by increasing the number of teachers (more experienced and better educated teachers demand higher pay, so higher salaries per pupil could occur from hiring teachers with higher qualifications, simply raising pay, or by hiring more teachers).

student's CGPA by .053 points (which is significantly greater than 0 at the .01 level) (*i.e.*, raising teacher salary per pupil by $1,000 through brand new spending).

Next in column 8, the pupil-teacher ratio at a student's high school was added to the model in column 7 with total expenditure per pupil and teacher salary per pupil; the coefficient on total expenditure per pupil is significant and little changed from column 5. Teacher salary per pupil is slightly smaller but is still significant.

In this model, the coefficient on pupil-teacher ratio is not significantly different from 0 at the .05 level, so holding total expenditure per pupil and expenditure on teacher salaries per pupil constant, the pupil-teacher ratio has no significant effect on CGPA. Holding total expenditures per pupil and the number of teachers constant, an increase in teacher salary per pupil is expected to increase CGPA. That is to say, increasing teacher salaries (not the number of teachers) has a significant positive effect on CGPA at the .05 level. A $1000 increase in teacher salaries per pupil by redirecting spending from elsewhere in the budget while holding the pupil-teacher ratio constant yields a .053 point increase in CGPA.

In this model, the linear combination of the coefficients on total expenditures per pupil and teacher salary per pupil is significantly greater than 0 at the .05 level of significance. So, while holding the number of teachers constant, we find that raising total expenditure per pupil by raising expenditure on teacher salaries per pupil is expected to increase CGPA by .037 points. In a regression model that includes all three of these high school inputs, we see that redirecting resources from spending not related to salaries, to teacher salaries may be the most productive

method of school spending.[17] The fact that in column 8 pupil-teacher ratio is no longer significant may partially be due to the decreased sample size.

VIII. Discussion

The estimated effects of school inputs were mixed, which is similar to results in past literature, such as was noted by Betts (1996) and Hanushek (1996). Table B shows the finding that the pupil-teacher ratio in a student's high school has a significant negative effect on his or her CGPA if we do not control for school expenditures. From Column 3 of table B, a two standard deviation increase in high school class size (an 8.36 student increase per teacher) is expected decrease CGPA by .0456 points; while significant, this effect is smaller than the effect of simply being from a different class year.[18] A two standard deviation increase in teacher salary per pupil ($2194) will cause a .113 point increase in CGPA, holding all else constant.[19]

For the sample of students from public schools, total expenditure per pupil was hypothesized to be positive but with a downward omitted variable bias, yet it was significantly negative. While it is possible that the true effect of expenditure on CGPA is negative, it is not something we should accept given the downward bias. It may be more accurate to conclude that we cannot necessarily determine the direction of the true effect because it is downward biased. With aptitude omitted, it is possible that the true effect of total expenditure per pupil is positive, but close enough to zero that the bias switches the sign of the coefficient. Or, if total expenditure

[17] Productive in this instance is limited, in fact it simply means the best way to spend money to increase the CGPA of students who enroll in and attend a small selective liberal arts college. We cannot know how spending in high school would affect students enrolled at other institutions.
[18] The *Omitted Variable Bias* section above argues that this estimated effect may in fact have been attenuated towards zero, so it is possible that the pupil-teacher ratio may truly have a stronger effect.
[19] Likewise, this estimate may be attenuated towards zero due to a downward bias.

per pupil truly has no effect on CGPA, then a downward bias would explain the negative coefficient. In a study estimating the effects of high school inputs on wages, Card and Krueger (1992) argue, "We suspect that the quality of education is more directly linked to indexes of pupil teacher ratios and teacher salaries than to total expenditures per pupil, and indeed this is suggested by the results in Welch (1966)." These are the three high school inputs included in this study. The conflicting signs of high school inputs has been seen repeatedly in the literature (Beets, 1996; Hanushek 1996), so the fact that total expenditure per pupil has a negative coefficient is not necessarily alarming.

The analyses of the effect of one type of spending over another in the results section above are based upon the assumption that we have accurately estimated all coefficients (*i.e.,* if you believe the coefficient on total expenditure per pupil to truly have a negative effect). If you believe that total expenditure per pupil may have no effect or a slightly positive effect, we would have one of the two following implications.

- Expenditure per pupil has no effect on CGPA (while holding teacher salaries per pupil constant): In this instance, it would not matter where the money came from to pay teacher salaries—taking money from elsewhere in the budget and adding new expenditure would each have the same impact on CGPA.
- Expenditure per pupil has a positive effect on CGPA (while holding teacher salaries per pupil constant): An increase in expenditure on teacher salary per pupil would still have a positive effect on CGPA (regardless of where the money came from). Expenditure on teacher salary per pupil (via new expenditures) would be the most effective spending (in regards to raising predicted CGPA).

Even if the true effect of total expenditure per pupil were negative, policymakers can still improve long-term student outcomes by increasing expenditure on teacher salaries either through new expenditure of by reallocating funds. Increased spending on teacher salaries can increase the number of teachers and/or increase the quality of the teachers which could lead to long-term benefits not just in high schools, but in Colleges as well. While these results were found at Connecticut College, further research could indicate whether teacher expenditure has positive impacts on students at other institutions as well. This researching would be worth expanding to other institutions as well, and could better inform policy makers on the long-term effects of school inputs on students.

IX. Conclusion

Studying five class years of Connecticut College students revealed that students' collegiate GPAs are significantly influenced by the resources provided by their high schools (even after controlling for a composite of students' high school academic performance). Teacher salary expenditure per pupil and pupil-teacher ratio had significant effects on CGPA as was predicted. Total expenditure per pupil has a significant negative effect on CGPA, which is the opposite direction than expected. The estimates suggest that increasing expenditure on teacher salaries (via hiring more experienced teachers, hiring more teachers, or offering higher pay) may improve students' GPA's in college.

The results of this research could likely be extended to similar colleges (particularly highly selective small liberal arts colleges); however, they may not be appropriate to describe students nationally. In extrapolating these results, one might not know the point at which a student might have a high enough academic ability to select a higher ranked institution than

Connecticut College—likewise, one might not know when a student might not have a high enough academic ability to be accepted into Connecticut College. This first course of follow up study may be to estimate the effect that high school inputs have on the likelihood that a student would attend Connecticut College. This would allow a researcher to determine what kind of selection occurred and how to correct for it.

On the topic of a national study, the theoretical argument for biases on school inputs would not hold as the process of selection would no longer occur. In fact, a study of students at all colleges (or a representative sample of colleges) would remove these biases and may allow for more accurate estimates; on the other hand one might need to allow for fixed effects at each college. While the literature remains unsettled, Betts (1999) and this chapter both provide evidence that some high school inputs may have persistent effects on college performance for students in highly selective colleges. Resources provided to high school students can clearly impact those students for several years.

Works Cited

Betts, J. R. (1996). Is There a Link between School Inputs and Earnings? Fresh Scrutiny of an Old Literature. *Does Money Matter?: The Effect of School Resources on Student Achievement and Adult Success*, 1141.

Belts, J., & Morrell, D. (1999). The Determinants of Undergraduate Grade Point Average. *The Journal of Human Resources*, *34*, 268-293.

Butler, R. P., & McCauley, C. (1987). Extraordinary stability and ordinary predictability of academic success at the United States Military Academy.*Journal of educational psychology*, *79*(1), 83.

Card, D., & Krueger, A. B. (1992). Does School Quality Matter? Returns to Education and the Characteristics of Public Schools in the United States.*The Journal of Political Economy*, *100*(1), 1-40.

Chetty, R., Friedman, J. N., Hilger, N., Saez, E., Schanzenbach, D. W., & Yagan, D. (2011). How does your kindergarten classroom affect your earnings? Evidence from Project STAR. *The Quarterly Journal of Economics*, *126*(4), 1593-1660.

Cohn, E., Cohn, S., Balch, D. C., & Bradley Jr, J. (2004). Determinants of undergraduate GPAs: SAT scores, high-school GPA and high-school rank.*Economics of Education Review*, *23*(6), 577-586.

Cyrenne, P., & Chan, A. (2012). High school grades and university performance: A case study. *Economics of Education Review*, *31*(5), 524-542.

Dietz, J. (2006) *Qualitative and Quantitative Predictors and Outcome Measures of Success at Small, Residential Liberal Arts Colleges.* Master's thesis, Connecticut College, New London, Connecticut.

Dustmann, C., Rajah, N., & Soest, A. (2003). Class Size, Education, and Wages*. *The Economic Journal*, *113*(485), F99-F120.Fredriksson, Peter, Björn Öckert, and Hessel Oosterbeek. "Long-term effects of class size." *The Quarterly Journal of Economics* 128.1 (2013): 249-285.

Geiser, S., & Studley, W. R. (2002). UC and the SAT: Predictive validity and differential impact of the SAT I and SAT II at the University of California.*Educational Assessment*, *8*(1), 1-26.

Grogger, J. (1996). School expenditures and post-schooling earnings: evidence from high school and beyond. *The Review of Economics and Statistics*, 78(4), 628-37.

Hanushek, E. A. (1996). School resources and student performance. *Does money matter*, 43-73.

Hanushek, E. A. (1999). Some findings from an independent investigation of the Tennessee STAR experiment and from other investigations of class size effects. *Educational Evaluation and Policy Analysis*, *21*(2), 143-163.

Humphreys, L. G. (1968). The fleeting nature of the prediction of college academic success. *Journal of Educational Psychology*, *59*(5), 375.Jones, Ethel B., and John D. Jackson. "College grades and labor market rewards." *The Journal of Human Resources* 25.2 (1990): 253-266.

Krueger, A. B., & Whitmore, D. M. (2001). The effect of attending a small class in the early grades on college-test taking and middle school test results: Evidence from Project STAR. *The Economic Journal*, *111*(468), 1-28.

Mosteller, F. (1995). The Tennessee study of class size in the early school grades. *Future of children*, *5*, 113-127.

Nechyba, T. J. (1990). The southern wage gap, human capital and the quality of education. *Southern Economic Journal*, 308-322.

Nye, B., Hedges, L. V., & Konstantopoulos, S. (1999). The long-term effects of small classes: A five-year follow-up of the Tennessee class size experiment. *Educational Evaluation and Policy Analysis*, *21*(2), 127-142.

Papke, L. E. (2005). The effects of spending on test pass rates: evidence from Michigan. *Journal of Public Economics*, *89*(5), 821-839.

Rizzuto, R., & Wachtel, P. (1980). Further evidence on the returns to school quality. *Journal of Human Resources*, 240-254.

Sacerdote, B. (2001). Peer effects with random assignment: Results for Dartmouth roommates. *The Quarterly Journal of Economics*, *116*(2), 681-704.

Stater, M. (2009). The impact of financial aid on college GPA at three flagship public institutions. *American Educational Research Journal*, *46*(3), 782-815.

Vincenty, T. (1975). Direct and inverse solutions of geodesics on the ellipsoid with application of nested equations. *Survey review*, *23*(176), 88-93.

Winston, G., & Zimmerman, D. (2004). Peer effects in higher education. In*College choices: The economics of where to go, when to go, and how to pay for it* (pp. 395-424). University of Chicago Press.

Zwick, R., & Sklar, J. C. (2005). Predicting college grades and degree completion using high school grades and SAT scores: The role of student ethnicity and first language. *American Educational Research Journal*, *42*(3), 439-464.

Table A

Pooled Sample

Variable	Mean	Standard Deviation	Observations
CGPA	3.452	0.314	1989
ARR	4.657	1.043	1989
Male	0.383	0.486	1989
Asian	0.065	0.246	1901
Black or African American	0.046	0.209	1901
Hispanic or Latino	0.063	0.242	1901
Other	0.002	0.040	1901
White	0.825	0.380	1901
First Generation College Student	0.117	0.321	1989
Median Income(000's)	91.865	38.179	1867
Pct Bachelor's or Higher	0.205	0.135	1867
Aid Ratio	0.272	0.373	1989
Private High School	0.415	0.493	1781
Total Expenditure (000's)	14.906	4.194	1030
Salary Expenditure (000's)	3.997	1.091	694
Pupil Teacher Ratio	11.972	4.202	1689

	Public High School Student			Private High School Students		
Variable	Mean	Standard Deviation	Observations	Mean	Standard Deviation	Observations
CGPA	3.487	0.319	1041	3.399	0.305	740
ARR	4.795	0.988	1041	4.394	1.049	740
Male	0.340	0.474	1041	0.427	0.495	740
Asian	0.052	0.222	1001	0.040	0.196	698
Black or African American	0.040	0.196	1001	0.054	0.227	698
Hispanic or Latino	0.073	0.260	1001	0.040	0.196	698
Other	0.001	0.032	1001	0.003	0.053	698
White	0.834	0.372	1001	0.862	0.345	698
First Generation College Student	0.136	0.343	1041	0.089	0.285	740
Median Income(000's)	94.266	40.104	1027	90.048	35.232	709
Pct Bachelor's or Higher	0.195	0.125	1027	0.217	0.143	709
Aid Ratio	0.305	0.377	1041	0.204	0.339	740
Private High School	0.000	0.000	1041	1.000	0.000	740
Total Expenditure (000's)	14.906	4.194	1030			0
Salary Expenditure (000's)	3.997	1.091	694			0
Pupil Teacher Ratio	14.407	2.982	1005	8.395	3.007	684

Table B

VARIABLES	(1) cgpa	(2) cgpa	(3) cgpa
ARR	0.124***	0.120***	0.121***
	(0.006)	(0.007)	(0.007)
Male	-0.117***	-0.118***	-0.119***
	(0.014)	(0.014)	(0.014)
Asian	-0.045	-0.078*	-0.067+
	(0.031)	(0.035)	(0.034)
Black or African American	-0.191***	-0.197***	-0.178***
	(0.045)	(0.046)	(0.048)
Hispanic or Latino	-0.167***	-0.162***	-0.169***
	(0.034)	(0.035)	(0.039)
Other	0.026	0.042	0.040
	(0.205)	(0.212)	(0.223)
First Generation College Student	-0.086**	-0.089**	-0.079**
	(0.027)	(0.028)	(0.029)
Median Income	0.000	0.000	-0.000
	(0.000)	(0.000)	(0.000)
Pct Bachelor's or Higher	0.093*	0.123*	0.142**
	(0.047)	(0.049)	(0.049)
Financial Aid Ratio	-0.005	-0.012	-0.022
	(0.022)	(0.023)	(0.023)
Class of 2010	-0.058**	-0.058**	-0.057**
	(0.020)	(0.021)	(0.021)
Class of 2011	-0.072***	-0.066**	-0.066**
	(0.020)	(0.021)	(0.021)
Class of 2012	-0.098***	-0.101***	-0.101***
	(0.021)	(0.022)	(0.022)
Class of 2013	-0.069***	-0.064**	-0.064**
	(0.020)	(0.021)	(0.020)
Private High School		-0.042**	-0.077***
		(0.014)	(0.020)
Pupil-Teacher Ratio			-0.006*
			(0.002)
Constant	3.108***	3.142***	3.237***
	(0.033)	(0.037)	(0.051)
Observations	1,785	1,657	1,573
R-squared	0.302	0.307	0.313

Robust standard errors in parentheses
*** p<0.001, ** p<0.01, * p<0.05, + p<0.1

Median income was divided by 1000.

Table C

VARIABLES	(4) cgpa	(5) cgpa	(6) cgpa	(7) cgpa	(8) cgpa
ARR	0.119***	0.119***	0.127***	0.126***	0.123***
	(0.009)	(0.009)	(0.010)	(0.010)	(0.010)
Male	-0.114***	-0.119***	-0.151***	-0.153***	-0.153***
	(0.020)	(0.020)	(0.025)	(0.025)	(0.024)
Asian	-0.054	-0.094*	-0.081	-0.076	-0.048
	(0.043)	(0.045)	(0.053)	(0.052)	(0.050)
Black or African American	-0.191**	-0.160*	-0.183*	-0.146+	-0.122
	(0.068)	(0.064)	(0.072)	(0.076)	(0.077)
Hispanic or Latino	-0.158**	-0.143**	-0.178***	-0.148**	-0.157**
	(0.050)	(0.044)	(0.047)	(0.047)	(0.054)
Other	-0.438***	-0.412***	-0.430***	-0.357***	-0.399***
	(0.044)	(0.038)	(0.045)	(0.050)	(0.061)
First Generation College Student	-0.104**	-0.091**	-0.082*	-0.071*	-0.075*
	(0.034)	(0.033)	(0.037)	(0.036)	(0.037)
Median Income	0.000	0.000	-0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Pct. Bachelor's or Higher	0.230***	0.177*	0.124	0.149+	0.225**
	(0.067)	(0.069)	(0.090)	(0.083)	(0.078)
Class of 2010	-0.064*	-0.053*	-0.044	-0.050	-0.063+
	(0.027)	(0.026)	(0.034)	(0.033)	(0.033)
Class of 2011	-0.075**	-0.060*	-0.070*	-0.076*	-0.098**
	(0.028)	(0.028)	(0.033)	(0.032)	(0.032)
Class of 2012	-0.120***	-0.104***	-0.126***	-0.134***	-0.149***
	(0.028)	(0.029)	(0.036)	(0.036)	(0.035)
Class of 2013	-0.073**	-0.064*	-0.076*	-0.092**	-0.105***
	(0.026)	(0.026)	(0.032)	(0.031)	(0.031)
Pupil-Teacher Ratio	-0.006*				-0.006
	(0.003)				(0.004)
Total Expenditure Per Pupil		-0.006**		-0.017***	-0.016**
		(0.002)		(0.005)	(0.005)
Salary Expenditure Per Pupil			0.015	0.071***	0.053*
			(0.011)	(0.018)	(0.021)
Constant	3.224***	3.188***	3.088***	3.139***	3.296***
	(0.070)	(0.053)	(0.063)	(0.062)	(0.108)
Observations	955	978	658	657	626
R-squared	0.302	0.297	0.341	0.357	0.355

Robust standard errors in parentheses
*** p<0.001, ** p<0.01, * p<0.05, + p<0.1

Median income, total expenditure per pupil, and teacher salary were all divided by 1000.

PEER AND CLASSMATE EFFECTS ON PERSISTENCE

Abstract

This chapter sets out to investigate factors that affect the likelihood of a student persisting at Connecticut College, including how similarities and differences between students and their peers can affect student persistence. By examining different measures of matching between a student and his or her peers, this chapter finds that female persistence decisions are strongly affected by demographics and freshman fall grades, while male persistence decisions are strongly affected by their academic performance relative to their classmates. This chapter also finds that females are more likely to persist when assigned a roommate with similar high school performance, yet males are more likely to persist when assigned a roommate with dissimilar high school performance. The implications of all these findings are discussed in detail.

I. Introduction

This chapter sets out to investigate factors that affect the likelihood of a student persisting at Connecticut College, including how similarities and differences between students and their peers can affect student persistence. Unique to this study is not just the data set, but also the ability to compare students to their peers in their dorms and in their classes. College persistence and completion is important to students, educators, parents and administrators (Astin & Oseguera, 2012). It is important that students graduate from college because an undergraduate degree can impact career opportunities and career earnings (Day & Newburger, 2002). An individual's graduation from college is important for both students and parents because of the economic cost to attend college. Lastly, retention is important to college administrators because of the lost revenue when students leave, lower retention rates (which can hurt college rankings

(*e.g.*, Forbes and U.S. News & World Report use the retention rate as a factor in determining college rankings)), and perceptions on campus. College administrators want their students to succeed and to remain at their college, for the benefit of all those involved.

Despite the importance of college retention and degree college degree, there has been little improvement in retention and graduation rates across the U.S. over time. About half of those who enroll in a 4-year institution typically receive a bachelor's degree within the next six years. After 6 years, only 56% of those who enrolled in a four-year institution earned a bachelor's degree. Only 50% of students at four-year institutions received a bachelor's degree from the first four-year institution at which they enrolled (Radford, Berkner, Wheeless, & Shepherd, 2010).

Tinto (1987) theorizes that a major factor in the decision to persist at an institution is the "match" between a student and that institution. This chapter will examine several possible measures of matching and peer effects to determine what may have an impact on the likelihood of persistence. Specifically, this chapter defines persistence as enrollment at Connecticut College for more than one year.[20]

This chapter tests several unique measures of environmental matching between a student and their roommate and a student in their classes—to my knowledge, several of these measures of matching have not previously been examined in the literature. This chapter finds that various covariates and dimensions of matching may have different effects on the likelihood of persistence for males and females at Connecticut College and compares these possible differences. Females who had high school academic performance similar to their roommates are more likely to persist than females who had different high school performance than their

[20] A student who returns for sophomore year is defined as persisting, as is a student who takes a gap year between freshman year and sophomore year.

roommates, yet males who had high school academic performance similar to their roommates are less likely to persist than males who had different high school performance than their roommates. Males who performed substantially better or substantially worse than others in their freshman fall classes were less likely to persist than males who performed similarly to their classmates.

Section II of this chapter provides a literature review discussing the prominent theory on student departure, and factors that have been found to be predictive of college persistence. Section III presents and discusses the empirical model. Section IV describes the data used in this study. Section V will focus on the idea of "matching" between a student and a college and will propose several dimensions of matching that will be tested in this thesis chapter. There can be several dimensions of matching, such as the difference between a student's SAT scores and the institution's average SAT scores (Cragg, 2009). Section VI presents the interpretation of results followed by a discussion of their implications in section VII. Section VIII provides concluding remarks and summarizes the results of this study.

II. Literature Review

There are two distinct branches of literature on college retention and completion. Some studies examine the factors affecting the likelihood that a student will graduate from a given institution (such as the first institution in which a student enrolls) (Light and Strayer, 2000; Cragg, 2009) and other studies examine the factors affecting the likelihood that a student will graduate from any four-year institution (Attewell, Heil, & Reisel, 2011; Adelman, 1999). Many students (approximately 25%) who enroll in a four-year institution change colleges at least once

(Radford, Berkner, Wheeless, & Shepherd, 2010). The factors affecting the likelihood that a student will graduate from the institution they started at may be different from the factors affecting the likelihood that a student graduates from any college.[21] After examining factors that affect the likelihood of a student obtaining a bachelor's degree, this section will discuss factors that affect persistence and graduation within a specific institution. Tinto (1987) developed one of the main theories of college retention and is cited by Attewell, Heil, and Reisel (2011), and several others. Tinto argues that a strong academic and social match between a student and an institution will increase the likelihood of a student remaining at, and graduating from, an institution.

Attainment of a bachelor's degree

Several factors influence a student's likelihood of completing college, including demographics (such as gender and race), family characteristics (such as parental education and family income), high school preparation and performance, employment status, and institutional factors (public versus private institution, selectivity of the institution, etc.).[22]

Several studies find that the strength of a student's high school academics as measured by the highest level of mathematics taken in high school (Adelman, 1999; Attewell, Heil, & Reisel, 2011) and various other measures of high school academic success such as standardized test scores and high school GPA (Astin & Oseguera, [23] 2012; Attewell, Heil, & Reisel, 2011; Ishitani, 2006; Adelman, 1999) are predictive of the attainment of a four-year degree. Others, however,

[21] Factors that are predictive of a student not attaining a degree from any institution are often predictive of a student not earning a degree at a particular institution as well.

[22] For the purpose of this study, institutional factors and employment status are not of great relevance; this study takes place within an institution, not across institutions. While data on employment of students is not available, cases of full-time employment by students at Connecticut College would be extremely rare, if there was any.

[23] Moreover, Astin and Oseguera (2012) find that standardized test scores and high school GPA are predictive of graduation from the first college a student attended.

have found that the ability of SAT scores to predict success in college varies by race and ethnicity. Fleming and Garcia (1998), in their meta-analysis, find that while a white student's SAT scores are consistently highly predictive of his or her collegiate GPA, a black student's SAT scores do not consistently predict success across studies. This chapter expands the idea of factors having different effects on different subgroups and explores whether various factors can have different effects on persistence of males and females.

Attewell, Heil, and Reisel (2011) find that several factors, such as high school grades and SAT scores, are predictive of degree attainment regardless of the selectivity of an institution. They find that taking a high-level math class (such as Calculus) in high school was predictive of a student completing a degree at a highly selective institution, but having taken a high-level high school math classes did not affect a student's persistence at a less selective institution. They estimate a separate regression for students enrolled in least selective (lowest mean SAT scores), moderately selective, and highly selective (highest mean SAT scores) four year colleges. Connecticut College, with a median SAT score between 1250-1390[24] would fall in the highly selective category.

For students in highly selective colleges, "race and gender, academic preparation, financial aid and work hours are each significant predictors of graduation" (Attewell, Heil, & Reisel, 2011). Specifically, being female, having a parent with a master's degree or higher, taking upper-level math classes, and high school GPA have a positive impact on the likelihood of graduating from a highly selective institution. While socioeconomic status, when measured as a composite of factors, was a highly predictive factor of non-completion in least selective and moderately selective institutions, it did not predict completion for highly selective institutions

[24] Using the middle 50% of 2006 SAT scores at Connecticut College. These were found through the National Center for Education Statistics.

(Attewell, Heil, & Reisel, 2011). These results indicate that in studying persistence at a highly selective college one should control for academic and demographic variables, but that some socioeconomic factors may not be as significant as they might be in less selective institutions. These results also indicate that results of this study may not necessarily be applicable to less selective institutions.

While Attewell, Heil, and Reisel (2011) did not find socioeconomic status to be predictive of degree attainment at a highly selective institution, Adelman (1999, 2006) finds that a composite measure of socioeconomic status is positively related to degree attainment when the sample isn't limited by institutional selectivity. Several studies find that both parental education and parental income have a positive relationship to college completion (Astin & Oseguera, 2012; Attewell, Heil, & Reisel, 2011; Ishitani, 2006). Adelman (1999), however, did not find self-reported measures of the level of parental education to be predictive.

While Adelman (1999, 2006) finds that first-year college grades are positively related to degree attainment, it is possible that the effect of first-year college grades varies between institutions, or is even non-linear when looking at how first-year college grades affect the likelihood of graduating from the first institution a student attends. For instance, while first-year college grades are positively related to attaining a college degree, it is possible that students could leverage high first-year college grades to transfer to a more highly selective institution. Succeeding the first semester or first year at a college (in terms of grades) could increase the chances that a student be accepted to transfer to another institution. On the other hand, low grades could increase the likelihood of transferring to a less selective institution or to stop attending college altogether. It follows that freshman year GPA may have quadratic or opposing effects on student persistence within an institution. For these reasons it is important to

investigate factors that affect a student remaining at, and graduating from, the institution they begin at.

Retention, Matching, and Peer Effects

In studying retention at a college, it is important to look not only at students who fail to earn a college degree, but also at students who transfer out of an institution. For this reason, we must also look at factors that affect the likelihood that a student earns a degree from the first college or institution that he or she attends. For example, Light and Strayer (2000) find that students with SAT scores more than 200 points above the institutional mean are less likely to graduate from that institution.

Several studies show that the selectivity of an institution is positively related to a student persisting and obtaining a degree at that institution (Melguizo, 2008; Adelman, 1999; Bowen, Chingos & McPherson, 2009; Oseguera & Rhee, 2009). Oseguera and Rhee (2009) find that the average high school GPA of students attending a college is positively related to the likelihood of a student graduating from the institution they start at, and Cragg (2009) finds a positive relationship between mean SAT scores and the likelihood of a student graduating from the institution he or she starts at.

Tinto (1987) states that incongruence (defined as a "mismatch or lack of fit between the needs, interests and preferences of the individual and those of the institution") can be one of the reasons a student may leave an institution. This mismatch can stem from differences in "abilities, skills, and interests" of the student and of others at the institution. Incongruence can be academic or social. Academic mismatches are easier to quantify, as measures could include a difference in SAT scores or a difference in scores on the Armed Forces Qualification Test

(AFQT). Academic factors need not be separate from social factors though. The difference academically between a student and his or her peer(s) could cause social incongruence as well.

Light and Strayer (2000) separate four-year institutions into quartiles based upon their median SAT scores and separate students who were part of the National Longitudinal Survey of Youth into quartiles based upon their scores on the AFQT. They find that individuals in the lowest quartile of AFQT are more likely to graduate from the first college in which they enroll if they attend a less selective college rather than a more selective college. Similarly, students in the top quartile who attend highly selective colleges are more likely to graduate than top quartile students who attend less selective colleges. These findings help confirm the matching theory presented by Tinto (1987) in regards to academic matching. Not only is AFQT is positively related to the likelihood that a student will graduate from the first college he or she attends, being at a school with similar students can have a strong impact as well.

Cragg (2009) tests the effect that a student's match with the institution he or she is attending has on his or her likelihood of graduating from that institution. Cragg defines the academic match as the difference between that student's SAT score and the average student's SAT score at that institution. The financial match is defined as the difference between a student's Expected Family Contribution (*i.e.*, what that student is expected to be able to pay for college) and the cost of enrollment after financial aid (*i.e.*, what the institution is charging the student); the financial match is the unmet financial need of a student. Cragg then breaks these differences into categorical variables.[25]

[25] SAT matching was broken into four categories, having an SAT score more than 200 points above the institution's average, having an SAT score between 51 and 199 points higher than the institutional average, being within 50 points of the institutional average, having an SAT score between 51 and 199 points lower than the institutional average, and having an SAT score more than 200 points less than the institutional average. Being within 50 points of the institution's average was the base case and other

Light and Strayer include an interaction between AFQT quartile and the quartile of college selectivity to determine if there is a "match" effect, while Cragg uses the categorized difference between a student and the institution's mean. Both of these approaches are categorical and their results suggest that the effect of "matching" with a school might not be linear. For instance, Cragg (2009) finds that students more than 200 points above or below the average SAT score at an institution are more than 5% more likely to leave that institution compared to students with SAT scores within 50 points of an institution's mean SAT score. To account for this nonlinearity, this chapter includes differences and squared differences (an alternative to partitioning the data and using categorical dummy variables) when looking at "matching." The dimensions of matching and peer effects are discussed after the data section.

Clearly, the prior academic performance of a student's peers can have a large impact on that student's persistence and degree completion. Perceptions on a college campus can even affect persistence; Oseguera and Rhee (2009) find that the average self-reported likelihood of dropping out of an institution[26] is significantly and negatively related to the probability of a student graduating from that college, after controlling for an individual's own response. Attewell, Heil, and Reisel (2011) conclude that while "there is no single dominant factor that is associated with better chances of graduation." The literature revealed that there are several factors that affect student degree attainment and student persistence at an institution. This chapter is an attempt to understand how some of the most commonly noted factors affect persistence at Connecticut College, and to explore factors that have been relatively unexplored by the literature.

---

categories were included as dummy variables. These dummy variables allow for non-linearity between categories. Financial match was constructed in an analogous manner (Cragg, 2009).

[26] Based on students' self-indicated intentions to drop out or to transfer. These intentions are given on a scale of how likely they believe it is that they will drop out or transfer.

III. Empirical Model

This study uses logistic regression to estimate the effects that covariates have on persistence. This chapter uses persistence beyond the first year at Connecticut College as the outcome. While some students may leave during or after their second year, they would still be counted as persisting. This narrow definition of persistence was chosen to simplify the model.[27]

The logistic regression models estimated includes a composite measure of a student's high school performance (the academic reader rating (ARR)), a vector containing demographic variables, a student's freshman fall GPA and squared freshman fall GPA, a vector containing variables describing a student's housing, and a vector consisting of various levels of matching between the student and his or her peers. The logistic regression model also contains an error term, ε_i.

The logistic regression form of this empirical model is expressed as:

$$Ln\left(\frac{Prob(Persist = 1)}{Prob(Persist = 0)} \mid \boldsymbol{X}\right)$$
$$= \beta_1 ARR_i + \boldsymbol{Demographics_i \beta_2} + \beta_3 FreshmanFallGPA_i$$
$$+ \beta_4 FreshmanFallGPA_i^2 + \boldsymbol{Housing_i \beta_5} + \boldsymbol{Matching_i \beta_6} + \varepsilon_i.$$

The results from the logistic regression are presented using odds ratios. Odds (which are not quite the same as probabilities) give the relative likelihood of a student persisting versus not persisting at Connecticut College. An estimated odds ratio on an explanatory variable would be the estimated effect of that variable on persistence. An odds ratio of 1 would mean the variable

[27] In studying the effects of freshman year grades and roommates the most closely related outcome is persistence into sophomore year.

has no effect on persistence, an odds ratio between 1 and 0 would mean that the variable has a negative effect on persistence and an odds ratio greater than 1 would mean that the variable has a positive effect on persistence.[28]

While freshman fall GPA is expected to be explanatory of persistence (Adelman, 1999, 2006) and is thus important to include in the logistic regression model, 27 students did not have freshman fall GPA's—so using freshman fall GPA as a variable would omit these students, leading to selection bias. Logistic regressions will be estimated twice, once omitting freshman fall GPA and freshman fall GPA squared, and once including both; the first estimation will not have sample selection bias, the second one will. The estimates to the logistic regression including freshman fall GPA, then, will be the effect of a variable on the likelihood that a student who has completed his or her first semester at Connecticut College will persist into a second year.

Preliminary results suggested that the effects of various covariates may differ by gender (several factors were significant for females, but not for males). Some covariates were significant when the sample was divided by gender but not in the pooled sample. To further investigate this, the models below will be estimated separately for males and females. Like chapter one, this chapter will use two-tailed t-tests to determine the significance of odds ratios, and will use a significance level of α=.05 unless otherwise noted.

[28] More specifically, an odds ratio of 1.25 would mean that a one unit increase in that variable increases the odds of persistence by 25%, and an odds ratio of .75 would mean that a one unit increase in that variable would decrease the odds of persistence by 25%.

IV. Data

This chapter uses student-level data from Connecticut College merged with data on students' home ZIP codes obtained from the Census Bureau. Data from Connecticut College includes data from the Office of the Registrar, the Office of Financial Aid Services, the Office of Admission, the Office of Residential Education and Living, and the Office of Institutional Research. The variables used to examine the research question are briefly described below.

Financial Aid Ratio

The amount of financial aid that a student received (including grants, loans, and work study) his or her freshman year is divided by the cost of tuition that year; this is done as a way of standardizing financial aid between years.

ARR

As detailed in chapter one, it is not possible to use high school GPA directly in this study.[29] Therefore, as a measure of prior academic success, this study uses the ARR which ranges from 1 to 7 and is a composite measure of high school GPA and class rank, standardized test scores, and the rigor of the classes a student took (Dietz, 2006).[30] An ARR of 7 is the strongest, and an ARR of 1 is the weakest; so a student with stronger high school academics would have a higher ARR. It is expected that ARR will be positively related with persistence.

Roommate's ARR

Using housing data, each student was matched with his or her freshman year roommate(s). By doing so it is possible to determine a student's roommate's ARR(s). These

[29] High school GPA is not necessarily on a constant scale between schools
[30] ARR was inverted so that a high ARR indicated higher high school performance.

data are used to create the difference in ARR between a student and his or her roommate(s) and this difference is discussed further in section IV below.

Number of Roommates

With these housing data and the matching method used to calculate the difference in ARR, it was also possible to count the number of roommates that a student had his or her freshman year. Most students had between 1 and 3 roommates (although a few did not have any).

Median GPA in Dorm

The median GPA in a dorm was calculated by combining housing records and grade records. For students in a given dorm in the fall semester, the median of cumulative GPAs (from the previous spring) was calculated. The cumulative GPA was this omits freshmen, since they would not have had a GPA from the prior spring). Thus, the median GPA in a dorm is the median GPA of all non-freshmen in the dorm.

Freshman Fall GPA

Using grade data, a freshman's GPA in his or her fall semester classes was calculated. Students who didn't complete their fall semester had a freshman fall GPA of 0. These students were omitted from any analysis that includes freshman fall GPA because the GPA of 0 would not be accurate; this is discussed further under the subheading *sample selection*. Freshman fall GPA is expected to have a non-linear effect on persistence with a negative second derivative.

Median Grade in a Class

Using the Registrar's records of classes and grades, it was possible to calculate the median grade in each class since 2005. This grade is measured by GPA points on a 4.0 scale.[31]. Section IV below discusses the median grade in each class and how to is used to compare a student performance relative to the median in that class.

Distance from Connecticut College

The latitude and longitude of a student's home ZIP code (as determined by the Census Bureau) was used to calculate that ZIP code's distance from Connecticut College.[32] The distance between Connecticut College and a student's home is approximated by the distance between Connecticut College and latitude and longitude of the student's ZIP code. For this study, the natural logarithm of this distance is used. For some institutions, Ziskin, Gross, and Hossler (2006) found that the distance that a student lives from the institution has a negative influence on a student's intent to persist at that institution. The distance that a student lives from Connecticut College is expected to be negatively related to persistence.

Population Density

Also using Census Bureau data, the population density of a ZIP code is calculated as the population of a zip code divided by the land area (square miles) of that ZIP code as determined by the 2010 Census. For the empirical model, the natural logarithm of population density is used. Connecticut College is located in a small city, New London, about two hours away from any major city (New York or Boston). Because of this, curiosity about those who came to

[31] Scoring an A in a class is worth 4 points, scoring an A- is worth 3.7 points, scoring a B+ is worth 3.3 points, scoring a B is worth 3 points, scoring a B- is worth 2.7 points, scoring a C+ is worth 2.3 points, scoring a C is worth 2 points, scoring a C- is worth 1.7 points, scoring a D+ is worth 1.3 points, scoring a D is worth 1 point, scoring a D- is worth 0.7 points and a failing grade is worth 0 points

[32] To calculate this distance, one has to account for the curvature of the Earth's surface. The *vincenty* command in STATA uses the formula developed by Vincenty (1975) to account for not just the Earth's curvature by assuming it's a sphere, but the precise ellipsoidal shape of the Earth.

Connecticut College from cities arose. To test whether those from cities or those from more rural environments were more likely to persist, the population density of a student's home ZIP code is included in the empirical model.

Sample Selection Bias

Note that two of the variables included were created using data from the Census Bureau distance and population density. The inclusion of these variables means that any student whose place of residence is outside of the United States is omitted from any regression estimates—thus we are estimating the effects that various factors have on domestic student persistence. Because this study is in several cases looking at the effects that a student's roommates might have on persistence, students who had no roommates (45 students) were omitted.[33] In addition, 54 students did not have values for the academic reader rating and were excluded from this study. These restrictions, albeit necessary, account for sample selection which may bias the findings of this study.

V. Dimensions of Matching

According to Tinto's theory (1987), academic and social factors significantly influence a student's decision of persistence. A student's freshman-year roommate for example, may have a profound impact on the student's first-year college outcomes. Sacerdote (2001) showed that there is a positive relationship between a student's freshman-year GPA and his or her roommate's freshman-year GPA (where students have been randomly assigned roommates).

[33] Since freshmen are not generally placed in singles, these individuals could be outliers. The experience of living alone could be significantly different than living with roommates. Furthermore, these observations could not be included in in estimations that included the difference between a student's ARR and his or her roommate's ARR(s).

Winston and Zimmerman (2004) found that a student's GPA may be affected by his or her roommate's SAT scores. The average perceptions on campus about the likelihood of degree completion can have an impact on a student's degree completion (Oseguera & Rhee, 2009).This gives credence to the importance of the match effect on college retention and completion. To empirically assess this, several measures of matching are constructed via housing and class data.

Due to the strong affects that roommates can have on a student (Sacerdote, 2001; Winston & Zimmerman, 2004) the first matching strategy was the match between a student and his or her roommate(s). The difference between a student's ARR and his or her roommate's ARR(s) were calculated; these differences were averaged together for each student to calculate the average difference between a student and his or her roommate(s) (if a student had only one roommate, then no averaging was needed). This measure is referred to as *roommate difference in ARR.* A student's decision to stay or leave Connecticut College may be more strongly related to the magnitude of the roommate difference in ARR (rather than the direction of the distance), so squared difference in ARR is created as the squared term of the difference in ARR. While I do not have a prediction of the effect of the difference on persistence, I expect that the squared difference is negatively related to persistence (or that the more different a student is academically than his or her roommate, the less likely he or she would be to persist)..

The median collegiate GPA of all students in a dorm (excluding freshmen) was calculated using students' cumulative GPAs from the prior spring as was discussed in section III.[34] Oseguera and Rhee (2009) found that the average high school GPA at a college has a positive effect on degree completion; I predict that a similar effect holds using the median cumulative college GPA across dorms. Both high school GPA and cumulative GPA are

[34] Although this isn't explicitly "matching" like the other terms discussed here, it is another quantitative descriptor of a dorm a student lives in.

measures of prior academic success, which could contribute to future academic success, such as persistence. While there could be several reasons for this effect, Carnevale and Rose (2003) give one possible reason that students might be more likely to graduate when attending highly ranked colleges "Perhaps peer interactions and high expectation about performance at top-tier colleges create an atmosphere in which students work harder and graduate." It is possible that the attitudes that other students in a dorm have towards academics (as would be indicated by cumulative GPA) could influence the decisions of freshmen to persist.

The next dimension of matching is that of class performance. A student's average grade relative to classmates was calculated as the average of the differences between a student's grade in a class and the median grade in that class.[35] For a student in four classes, the calculation would be:

$$Average\ Grade\ Relative\ to\ the\ Median\ =$$

$$\frac{1}{4}(GradeClass1 - MedianGradeClass1 + GradeClass2 - MedianGradeClass2 +$$

$$GradeClass3 - MedianGradeClass3 + GradeClass4 - MedianGradeClass4\ )$$

It was predicted that the difference between a student's grades and the median grades in his or her classes (hereafter referred to as *relative performance*) may affect his or her likelihood of persisting. Assuming that a grade reflects a student's understanding of the material in a class, those who do not understand or learn the material as well as their classmates may be disheartened and may transfer or otherwise withdraw. On the other hand, those who do very well in a class (if they score above the median) may feel that they have mastered the material and

[35] For example, a student who scored .5 GPA points above the median of two classes but 0.5 below the median of two other classes would have an average relative grade of 0, while a student who scored 0.5 points below the median in all 4 classes would have an average relative GPA of -0.5)

may feel like they are not being challenged academically.[36] Those who excel in their classes may want more challenging academics and thus be more likely to transfer.

VI. Results

Descriptive Statistics

The sample of 2,466 Connecticut College students included in this study enrolled in Connecticut College as freshmen between Fall 2005 and Fall 2009. Students included had between one and three roommates (the average was 1.73). On average, 91.0% the freshman year students returned in the fall of their sophomore year or at some later point.[37] On average, males were 0.1 percentage points more likely to persist than females. The average ARR was 4.39; females had a higher average ARR (4.77) than males (4.39). Females had a lower average freshman fall GPA of 3.41, compared to males' of 3.19. The summary statistics can be seen in table A below, first for all students and then disaggregated by gender.

[Place table A here.]

The Effects of Demographics on Persistence

Persistence was regressed on previous performance and demographics for all students. This can be seen in column 1 of table B and omits the variable freshman fall GPA (and freshman

[36] Note that the median grade also includes the grades of upper classmen who may be enrolled in the class. A grade above the median would also (potentially) means performing better than several upperclassmen.

[37] Students who miss sophomore fall but return to Connecticut College any time after that are considered to have persisted. Occasionally, students occasionally take a semester or a year off for personal reasons, but return. These students should not be considered the same as those who withdrew or transferred.

fall GPA squared).[38] As expected, financial aid had a positive effect on persistence (at α=.05). The number of roommates that a student had and the log of the population density of a student's home ZIP code were both positively related to persistence (at α=.05).

Further, the natural log of the distance of a student's home from Connecticut College had a negative effect on student persistence. The results are reported as odds ratios, so coefficient on logged distance means that a 100% increase in the distance from Connecticut College decreases the odds of persistence by 18.2%, all else held constant. On average, a student has a probability of persisting of 91.0%. This can be expressed in as odds as being 10.11 more likely to persist than not (odds= $\frac{Persist}{\sim Persist}$ =91/9). To ease the interpretation of odds ratios, this results section will occasionally use an odds ratio to determine how a change in a variable would affect a student with the mean likelihood of persisting, 91.0%.[39]

If one compared a student with the mean odds of persisting (10.11) with an identical student who lived twice as far away (a 100% increase in distance), the student who lived farther away would have odds of persisting of 8.27 (the odds are multiplied by the coefficient on logged distance, which gives 10.11*81.8%). So this student who is living further away has a probability

[38] One might expect that the size of classes that students are in their freshman fall may impact their persistence. Class size was constructed as the number of students in a section of a class which is accurate unless a class also has an accompanying lab (e.g. Biology 106), where the different sections are defined as the lab sessions. The class size for a student in classes with labs would be skewed downward, and would not be accurate (a student in a lecture with over 100 students may only have a class size of 14, the size of a lab section in Biology 106). Many classes are offered at multiple points during the week—and to separate students in those classes, they needed to be divided by sections. This would need to be revisited using more detailed data from the Registrar before any conclusions can be drawn about the effect of class sizes.

[39] For ease of reading, it may be helpful to suppose an individual with a probability of graduating of 91.0% and to use the following:
If this individual's odds of persisting increased (or decreased) by 10%, his or her probability of graduating would increase (or decrease) by 0.9 percentage points (pp).
If this individual's odds of persisting increased (or decreased) by 20%, his or her probability of graduating would increase (or decrease) by 2.0pp.
If this individual's odds of persisting increased (or decreased) by 50%, his or her probability of graduating would increase (or decrease) by 7.5pp.

of graduating of 89.2% (calculated as 8.27/(1+8.27)). While this student's odds of persisting are 18.2% less, his or her probability of graduating is only 1.8 percentage points (pp) less.[40]

[Place table B here.]

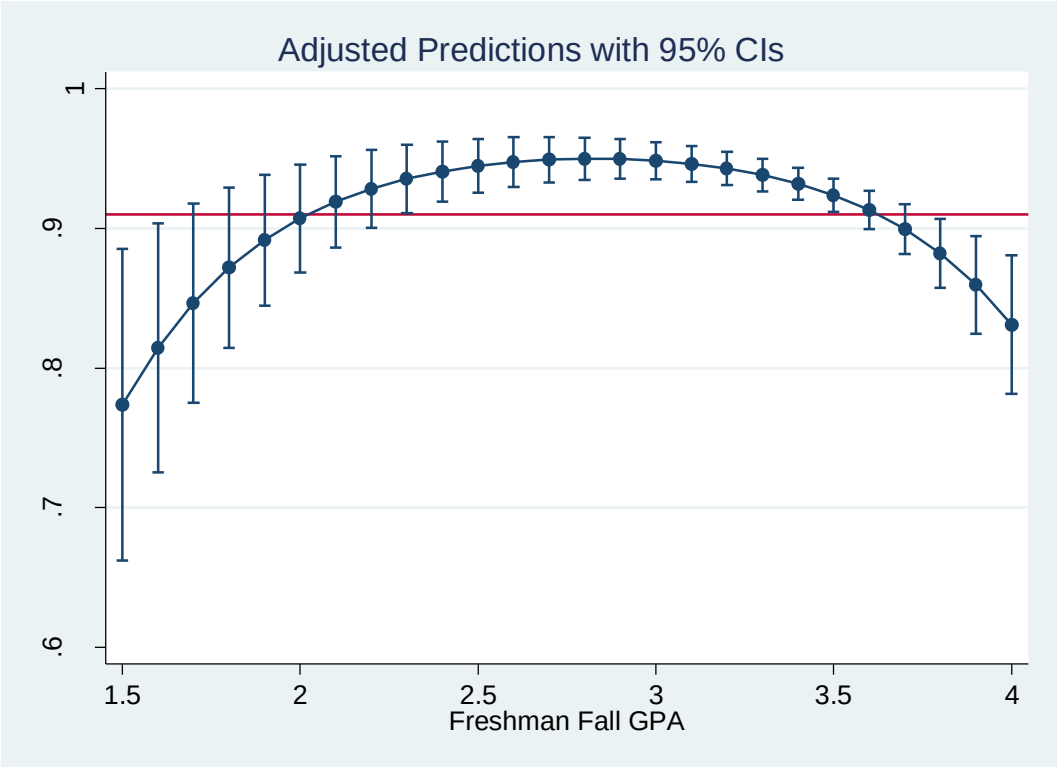
The Effects of Performance on Persistence

The results in column 2 show that freshman fall GPA and freshman fall GPA squared have a significant impact on student persistence. The interpretation of these is slightly more difficult, but can be seen graphically in the predicted margins plot in figure 1 below. A student with a freshman fall GPA either greater than or less than 2.82 is less likely to persist than an identical student with a freshman fall GPA of 2.82, all else held constant.[41] On the Y-axis is the probability of persistence and on the X-axis is freshman fall GPA; this figure has a reference line at 91% probability of persistence, the mean for the sample. Striking is that for students whose demographics and prior performance are set to the means, those with a freshman fall GPA of 3.8 or more are significantly less likely to persist than the average student at Connecticut College; students with freshman fall GPAs of 3.8, 3.9, and 4.0 have expected probabilities of persisting of 88%, 86%, and 83%, respectively. Likewise, those with a freshman fall GPAs less than 1.6 are significantly less likely to persist than the average student. On average, those with a freshman fall GPA of 2.82 have a 95.0% chance of persisting. Clearly, freshman fall GPA has a strong, non-linear impact on persistence.

[40] A coefficient may not always increase by units of 1 (*e.g.*, log distance). If distance increased by 10%, then the odds of graduating would decrease by $\left(1 - 0.818^{\frac{1}{10}}\right) * 100\% = 2.0\%$, and the likelihood of graduating would decrease by 0.17%.

[41] By logging the odds ratios, one can calculate the GPA at which the probability of persistence is maximized (the negative second derivative indicates that the curve is an upside down parabola). The maximum point is calculated as $\frac{-\log(246.345)}{2*\log(.377)}$=2.82.

Figure 1. The effect of freshman fall GPA on the probability of persistence (column 2)

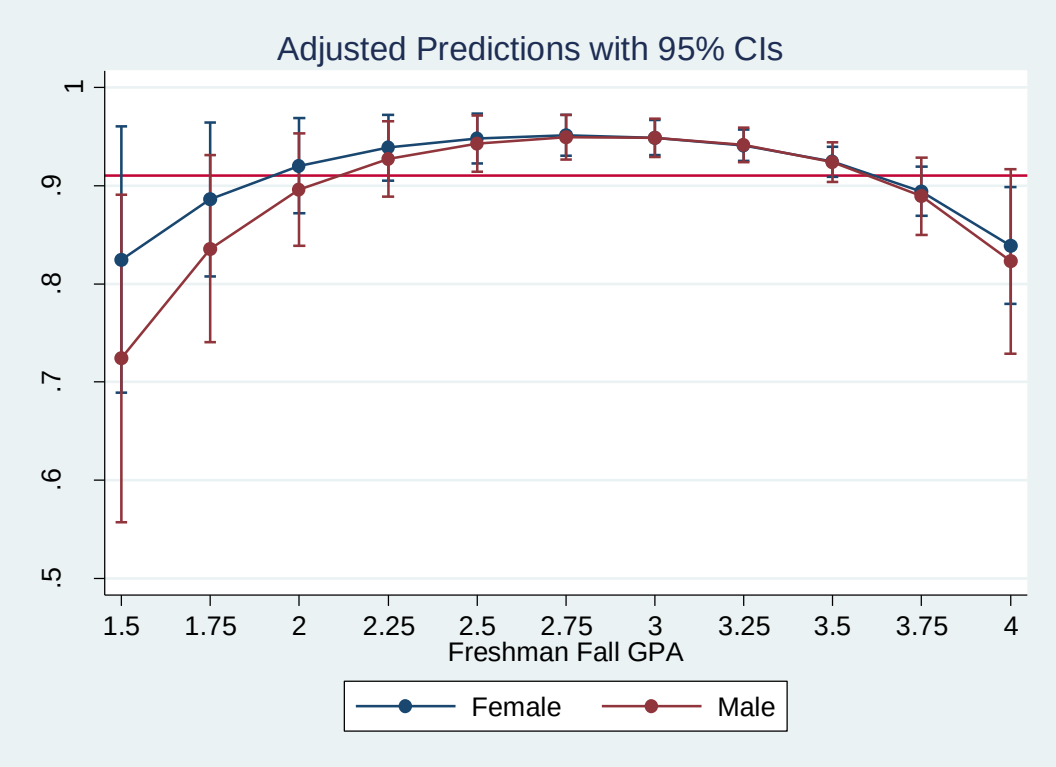


The logistic regression in column 1 was separated into two logistic regressions, one for males and one for females. The estimates for males and females can be seen in columns 3 and 5 in table B, respectively. These display the estimated effects of demographics and prior performance on persistence when we do not control for freshman fall GPA. Of note is that none of the explanatory variables are significant predictors of persistence for males in column 3. In column 5, the financial aid ratio and logged population density both have a significant positive impact on a female's persistence (α=.05). The logged distance has a significant negative effect on female persistence. ARR and the number of roommates only has a significant effect on persistence for females only at the .10 level.

Columns 4 and 6 of table B estimate the effects of freshman fall GPA on persistence for males and females, respectively. The effect of freshman fall GPA and its squared term are

significant for both genders. While the odds ratios on freshman fall GPA look quite different for males than females, the estimated effects of freshman fall GPA are not; these effects can be seen below in figure 2, which plots predicted persistence for males and females at different levels of freshman fall GPA with all other variables set to their means. The vertical lines in figure 2 give confidence intervals for predicted persistence at each level of freshman fall GPA. The horizontal line at 91% is the average probability of persistence. Note that females with a freshman fall GPA of 3.9 or 4.0 are significantly less likely to persist than the average student, and males with a freshman fall GPA of 1.5 or less are significantly less likely to persist than the average student.

Figure 2. The effect of freshman fall GPA on the probability of persistence for males and females (columns 4 and 6, respectively)



The next analyses analyze the effects the difference in roommate ARR and the median GPA in a dorm on persistence. Because of the difference in significance between predictors of persistence between males and females, further analyses estimate logistic regressions separately by gender. Table C includes three explanatory variables regarding a student's housing, roommate difference in ARR, roommate difference in ARR squared, and median GPA in a dorm. In column 7 of table C, we find that the estimated odds ratio on roommate difference in ARR squared is significantly greater than 1. This was unexpected and implies that the larger the difference in ARR between a male and his roommate(s), the more likely he is to persist (regardless of whether he or his roommate has the higher ARR). This is the opposite of the effect found for females. For females (column 9), the odds ratio on roommate difference in ARR squared is less than 1. This indicates that the larger the difference in ARR between a female and her roommate(s), the less likely she is to persist (regardless of whether she or her roommate has the higher ARR); this was the expected direction of the effect. As before, the financial aid ratio, logged distance from Connecticut College, and population density have significant effects on female persistence.

For both males and females (columns 8 and 10, respectively), we find that controlling for freshman fall GPA does not greatly impact the odds ratios on roommate difference in ARR squared. Calculating the estimated effects of roommate difference in ARR for males in column 8, we find that a 1-point ARR increase in the difference between a male and his roommate(s) increases the likelihood of persisting by 0.44pp, and a 3-point ARR increase in the difference between a male and his roommate(s) increases his likelihood of persisting by 5.55pp. Calculating the estimated effects of roommate difference in ARR for females in column 10, we find that a 1-point ARR increase in the difference between a female and her roommate(s)

decreases the likelihood of her persisting by 0.84pp, and a 3-point ARR increase in the difference between a female and her roommate(s) decreases her likelihood of persisting by 7.45pp.

[Place table C here.]

A student's relative performance was predicted to factor affecting persistence. Table D estimates the logistic regressions that include relative performance and relative performance squared. To estimate the effect of relative performance, it is also important to control for freshman fall GPA; by controlling for freshman fall GPA the specific hypothesis we are investigating is whether a student's performance in direct comparison to his or her classmates affects his or her decision to persist while holding his or her grades constant. This measure may be the effect that a student's perception of his or her grades or intelligence has on persistence (as opposed to the level grade itself, which would be freshman fall GPA).

[Place table D here.]

For males (column 11) the odds ratio on relative performance squared is significantly less than one. This indicates that students performing either better than or worse than their classmates are less likely to persist than those performing the same as their classmates. Relative performance has a parabolic effect on persistence for males, with males having the highest expected persistence when performing -0.26 grade points worse than their classmates. For males, the odds ratios on freshman fall GPA and freshman fall GPA squared are no longer significant after controlling for relative performance. For females, while neither term of relative performance has a significant effect on persistence, freshman fall GPA and freshman fall GPA squared are still significant (column 12). The effect of relative performance on persistence is graphed below in figure 3, disaggregated by gender. For males (column 11), we find that the

only significant predictors of persistence are roommate difference in ARR squared and relative performance squared.

Figure 3. The effect of average grades relative to classmates on the probability of persistence, by gender.

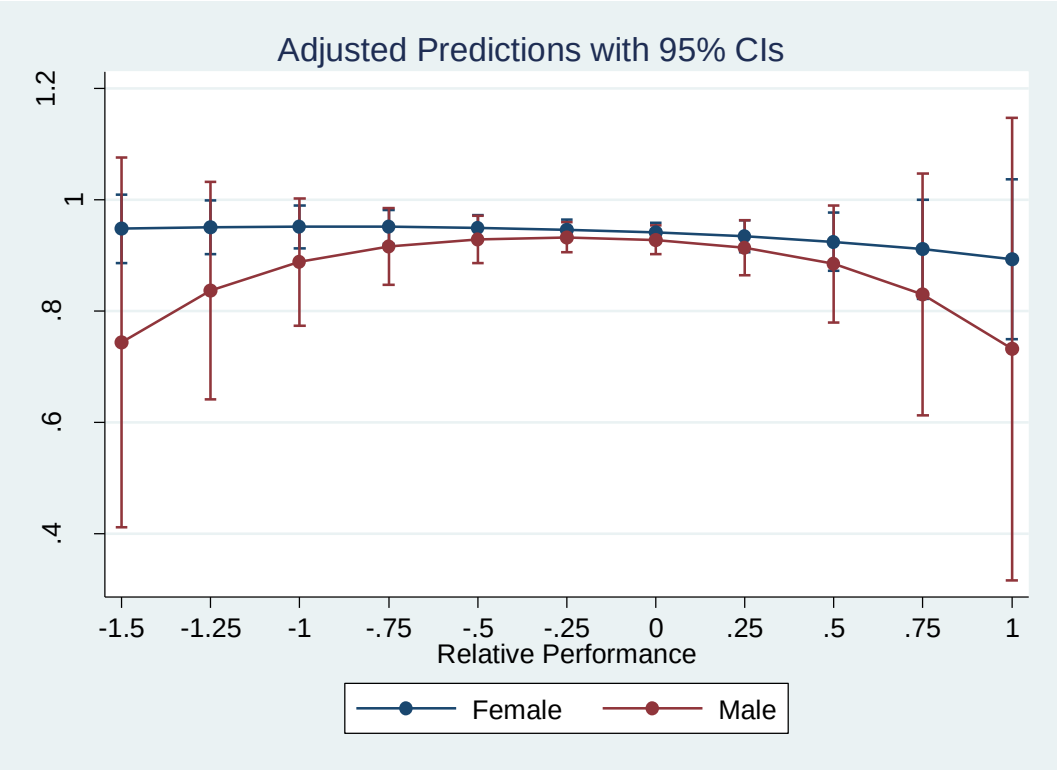


Figure 3 above clearly shows that relative performance can have a strong impact on persistence for males and that this effect is much stronger than the impact on females (as indicated by the steeper parabola for males). A male who scores 0.5 points above the median in his class is expected to persist 88.5% of the time, but a male who scores 0.5 points below the median in his classes is expected to persist 92.9% of the time. Males who score 1 point or 1.5 points below the median in their classes are expected to persist 88.9% and 74.4% of the time respectively.

VII. Discussion

The logistic regression estimates above provide some insight into student persistence decisions. The factors affecting persistence that are significant for males are very different from those that are significant for females. In fact, comparing columns 11 (males) and 12 (females) of table D we see that the only factor that is predictive of persistence for both genders is squared difference in ARR between roommates—which has opposite effects on males and females.

Section IV above stated that there is sample selection bias in regressions including freshman fall GPA. Comparing the estimated odds ratios of models including freshman fall GPA with those of models not including freshman fall GPA, one notes that there is little change in the estimates. Moreover, there is only one instance in which a variable that is significant in one model is not significant when freshman fall GPA is added or removed (roommates is not significant in model 2 when freshman fall GPA is added). While including freshman fall GPA does not seem to cause major sample selection bias, biases could still exist from the other sample restrictions that were necessary.

The median GPA of students in a dorm was not found to have a significant effect on student persistence at Connecticut College. While Oseguera and Rhee (2009) found that the average high school GPA of students at a college has a significant impact on persistence, a similar effect does not hold across dorms within Connecticut College.

Several factors are found to be predictive of persistence for females. Among them, distance from Connecticut College consistently had negative effects on persistence, while population density and financial aid consistently had significant positive effects on persistence. Freshman fall GPA and freshman fall GPA squared were also predictive of female persistence.

While the odds ratios on freshman fall GPA and freshman fall GPA squared were significant for males in early specifications (namely columns 4 and 8), they were not significant when relative performance and relative performance squared were included in the regression model (column 11).

The squared difference in ARR between a female and her roommate consistently has a significant negative effect on persistence, while the squared difference in ARR between a male and his roommate consistently has a significant positive effect on persistence. Figure 4 below was created to help visualize the effect that roommate difference in ARR has on persistence, disaggregated by gender. This graph was created using the estimates from columns 11 and 12 of table D.

Figure 4. The effect of the difference in ARR between a student and his or her roommate on persistence

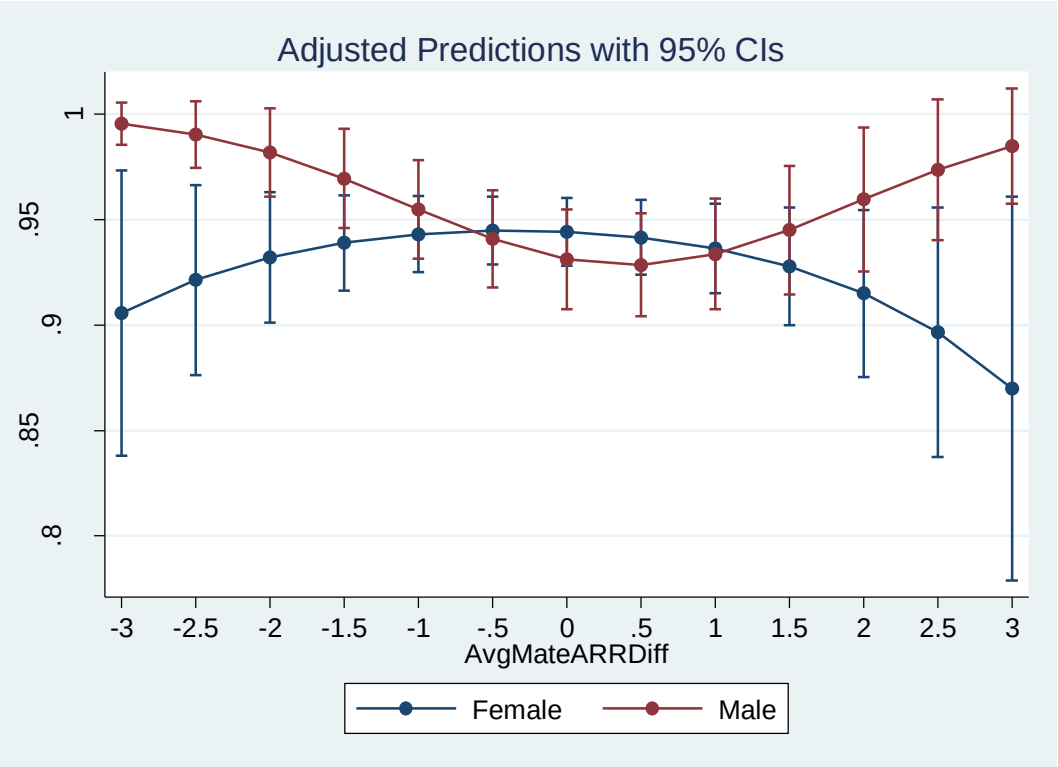


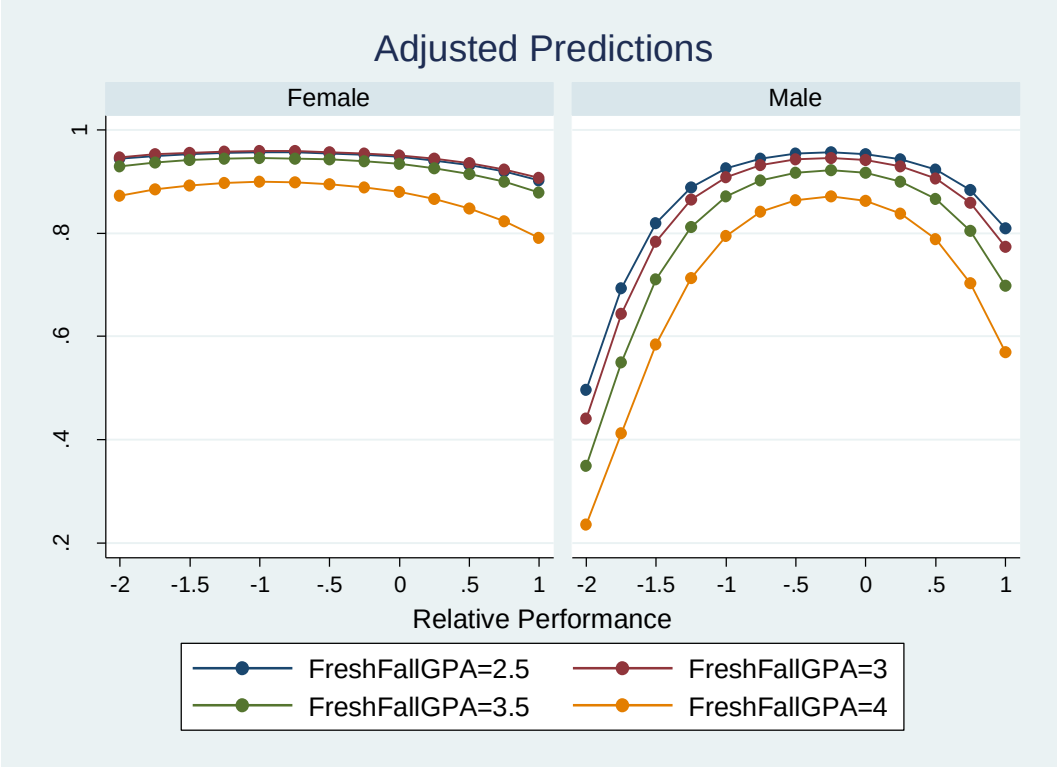
Figure 4 demonstrates that the roommate difference in ARR has substantially different effects on males than on females. These findings suggest that male students have more success roommates who are academically different than them, while female students have more success with roommates who are academically similar to them. Male students may generally prefer to associate with individuals who are similar in non-academic regards, while females may generally prefer to associate with a roommate who is similar to them academically. Qualitative research should be conducted in this area to provide empirical explanations for these patterns. Surveys about roommate interactions and satisfaction would be one way to study the reasons behind these trends.

The results suggest that females who are assigned a roommate with a similar ARR may be more likely to persist than females who are assigned a roommate with an ARR different from their own. If one was to reassign female students roommates with the exact same ARR as themselves, the retention rate of females would only increase (the average predicted values of persistence) by 1.2pp (this is the change in average probability of persisting for the classes of 2009-2013 if you set the difference in ARR to 0). This would increase the mean persistence of females from 90.6% to 91.8%. Given that roommates are not randomly assigned, but matched to some degree, matching females with other females with the exact same ARR may cause other differences to occur between roommates. Therefore, ARR should be one measure considered while matching females, but should not necessarily a decisive factor in matching roommates.

The regressions in table D revealed that while a male's relative performance has a significant impact on his persistence, a female's relative performance does not have a significant impact on her persistence but freshman fall GPA does. Males might be more concerned about how they do relative to those in their classes rather than their overall freshman fall GPA. Below

are two separate plots to provide a better picture of the effects of relative grades and freshman fall GPA on persistence. In the first chart, figure 5, we see the effects that relative grades have on persistence for various freshman fall GPA levels for males and females.

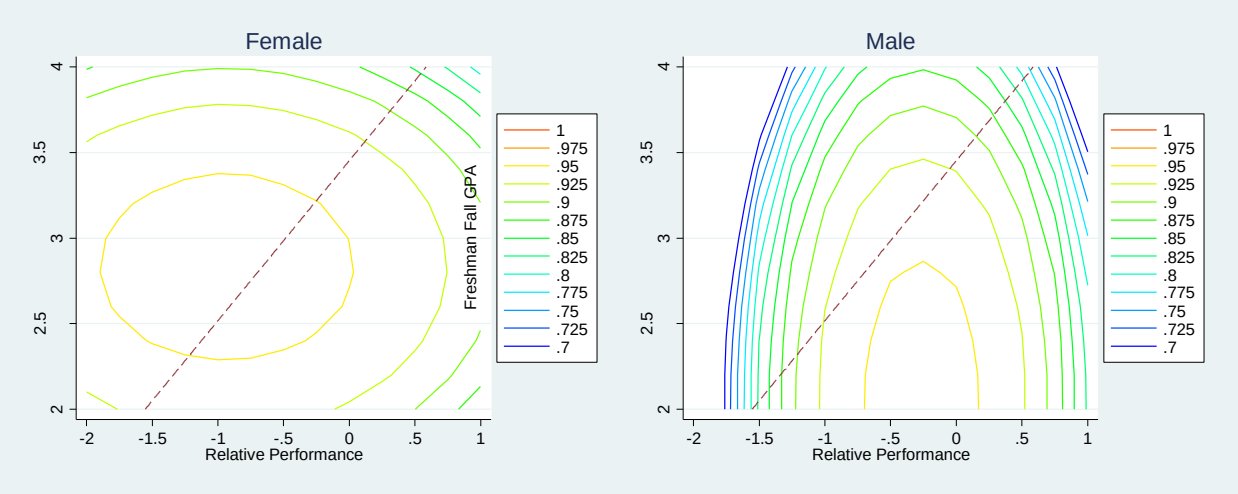
Figure 5. The effect of relative performance on persistence, by gender and by freshman fall GPA



From these graphs, it is clear that relative performance has a much stronger effect on males than on females (as was also seen by the estimates). Also, note that the parabolas with the lowest predicted persistence are for a freshman fall GPA of 4.0 for both males and females. Note that a student who has a positive relative performance would likely have a high GPA. The graph above shows that a male with a 4.0 who has a high relative performance would have a lower probability of persisting than most other students. Because the lines for various levels of

freshman fall GPA are close together, a contour plot was created and included below in figure 6 to show how relative performance and freshman fall GPA combined effect persistence.

Figure 6. The effect of relative performance and freshman fall GPA on persistence, by gender



These contour plots show the predicted probabilities of persistence graphed against freshman fall GPA and relative performance for both females and males. The dotted line is the bivariate regression line of freshman fall GPA on relative performance[42]; this gives the expected freshman fall GPA for a given level of relative performance.[43] Interpreting the graph by using the dotted line as a guide, it is expected that females with a freshman fall GPA of 4.0 would have a relative performance of 0.59, and a probability of persistence of between 82.5% and 85%; males with a freshman fall GPA of 4.0 would be expected to have a relative performance of 0.59, and a probability of persistence of between 75% and 77.5%. Unlike previous interpretations, the use of contour plots allows both freshman fall GPA and relative performance to vary at once (as

[42] The estimated regression line is *Freshman Fall GPA* = .935**Relative Performance* +3.452.

[43] For example the line demonstrates that a student with a freshman fall GPA of 3.0 has a relative performance, on average, of -0.5. It also demonstrates that at a freshman fall GPA of 4.0 relative performance is, on average, about 0.59. This makes sense and prevents one from considering a negative relative performance when freshman fall GPA is 4.0.

opposed to looking at one while holding the other constant). Now consider a female with a freshman fall GPA of 4.0 who has a relative performance of 0.75 (a little to the right of the dotted line); she would have an expected probability of persistence 82.5%. A male with a freshman fall GPA of 4.0 who has a relative performance of 0.75 would have an expected probability of persistence 70%.

Figure 6 illustrates that for males with a low freshman fall GPA, having a lower than expected relative performance (being to the left of the dotted line) decreases the likelihood of persisting, while for males with a high freshman fall GPA having a higher than expected relative performance (being to the right of the dotted line) decreases the likelihood of persisting. This effect is much smaller for females.

As was previously seen from the estimates, relative performance is expected to play a larger role in the persistence decisions of males than of females. The contour plots shows that both freshman fall GPA and relative performance may strongly impact a males persistence decisions; there is a much greater variability in the probabilities of persistence for males than of females in figure 6. This means that while demographics and roommate difference in ARR may explain a large portion of the persistence decision of females, freshman fall GPA and relative performance may explain a small portion of the persistence decision of females. On the other hand, it appears that both freshman fall GPA and relative performance may explain a large portion of the persistence decision of males.

It is also important to consider the accuracy of predicted probabilities of persistence. Using the estimates in columns 11 and 12, those with a predicted probability of persisting greater than 90% persisted 93.5% of the time (1776 students), those with a predicted probability to persist between 80%-90% persisted 86.2% of the time (587 students), those with a predicted

probability to persist between 70%-80% persisted 76.0% of the time (75 students), and those with a predicted probability to persist less than 70% persisted 50% of the time (28 students). These predictions of the probability of persistence are fairly accurate, however, are not precise; the factors considered do affect the probability of persistence, yet a large portion of the decision to persist is still unexplained. It is possible that these models could be used to help identify some students with a higher risk of not persisting so as to allocate resources to them. While it would not be possible to determine who exactly would withdraw or transfer, it may possible to target some students with a lower probability of persisting with some sort of mentoring or initiative.

VIII. Conclusion

The series of logistic regressions indicate that it may be advantageous to study or to model the factors that affect persistence of males and females separately. While demographic and background factors were predictive of female persistence, they were not predictive of male persistence. Roommate difference in ARR squared was found to have opposing effects for males and females. While it was expected to have a negative effect on persistence, it was only found to have a negative effect on female persistence; it had a positive effect on male persistence. This is a result that requires more detailed study (such as survey analysis) to explore the mechanisms behind these effects. This is a measure that could be used to improve roommate matches it were further studied.

Freshman fall GPA has a quadratic effect on persistence of females; females with very high and very low freshman fall GPAs experience negative effect from freshman fall GPA. Relative performance has a strong effect on persistence of males; males who have a high

freshman GPA and a high relative performance are less likely to persist and males with a low freshman fall GPA and a low relative performance are less likely to persist. This is a factor that administration could monitor as an early warning indicator that a student may not persist so as to provide mentoring or assistance to those male students who are performing so well or so poorly that they might have a lower probability of persisting.

None of the specific factors that were tested stood out as being the main reason for driving student transfers or withdrawals. The decision to transfer or withdraw can be complex, and we can model only some of the factors. Tinto (1987) theorized that the decision to persist or not is based upon both social and academic integration. With the roommate differences and the relative grades we were able to test some possible academic differences, but more academic differences may exist. Social differences may be more difficult to quantify, but the difference in ARR between a student and his or her roommate(s) could explain a small piece of the social fit that a student experiences. A study of students' social behavior relative to their roommates and those in their dorms may help shed some light on this matter—this study only explains a sliver of the many complex reasons that impact persistence.

Works Cited

Adelman, C. (2006). The Toolbox Revisited: Paths to Degree Completion From High School Through College. US Department of Education.

Adelman, C. (1999). Answers in the Tool Box. Academic Intensity, Attendance Patterns, and Bachelor's Degree Attainment. US Department of Education.

Astin, A., & Oseguera, L. (2012). Pre-college and institutional influences on degree attainment. *College student retention: Formula for student success*, 245-276.

Attewell, P., Heil, S., & Reisel, L. (2011). Competing explanations of undergraduate noncompletion. American Educational Research Journal, 48(3), 536-559.

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

Carnevale, A. P., & Rose, S. J. (2003). *Socioeconomic status, race/ethnicity, and selective college admissions*. New York: Century Foundation.

Cragg, K. M. (2009). Influencing the probability for graduation at four-year institutions: A multi-model analysis. Research in Higher Education, 50(4), 394-413.

Dietz, J. (2006) *Qualitative and Quantitative Predictors and Outcome Measures of Success at Small, Residential Liberal Arts Colleges.* Master's thesis, Connecticut College, New London, Connecticut.

Fleming, J., & Garcia, N. (1998). Are standardized tests fair to African Americans?: Predictive validity of the SAT in black and white institutions. Journal of Higher Education, 471-495.

Ishitani, T. T. (2006). Studying attrition and degree completion behavior among first-generation college students in the United States. Journal of Higher Education, 861-885.

Light, A., & Strayer, W. (2000). Determinants of college completion: School quality or student ability?. Journal of Human Resources, 299-332.

Melguizo, T. (2008). Quality matters: Assessing the impact of attending more selective institutions on college completion rates of minorities. Research in Higher Education, 49(3), 214-236.

Oseguera, L., & Rhee, B. S. (2009). The influence of institutional retention climates on student persistence to degree completion: A multilevel approach. Research in Higher Education, 50(6), 546-569.

Radford, A. W., Berkner, L., Wheeless, S. C., & Shepherd, B. (2010). Persistence and Attainment of 2003-04 Beginning Postsecondary Students: After 6 Years. First Look. NCES 2011-151. National Center for Education Statistics.)

Sacerdote, B. (2001). Peer effects with random assignment: Results for Dartmouth roommates. The Quarterly Journal of Economics, 116(2), 681-704.

Tinto, V. (1987). Leaving college: Rethinking the causes and cures of student attrition.

Ziskin, M., Gross, J., & Hossler, D. (2006). Institutional practices and student persistence: Extending the empirical record. In *annual meeting of the Association for the Study of Higher Education, Anaheim, CA*.

Table A

Variable	Mean	Standard Deviation	Observations
Persistence	0.910	0.287	2421
ARR	4.62	1.04	2387
Male	0.39	0.49	2421
Aid Ratio	0.26	0.37	2421
Ln(Distance)	4.93	1.55	2242
Ln(Population Density)	7.57	1.76	2242
Roommates	1.73	0.69	2421
First Gen. College Student	0.11	0.31	2421
Freshman Fall GPA	3.34	0.48	2391
Roommate difference in ARR	0.00	1.27	2421
Roommate difference in ARR^2	1.62	2.38	2421
Median GPA in Dorm	3.41	0.08	2410
Relative Performance	-0.12	0.46	2380
Relative $Performance^2$	0.23	0.69	2380

	Male			Female		
Variable	Mean	Standard Deviation	Observations	Mean	Standard Deviation	Observations
Persistence	0.915	0.278	946	0.906	0.292	1475
ARR	4.39	1.04	929	4.77	1.02	1458
Male	1	0	946	0	0	1475
Aid Ratio	0.25	0.37	946	0.26	0.37	1475
Ln(Distance)	4.87	1.53	865	4.98	1.56	1377
Ln(Population Density)	7.52	1.76	865	7.6	1.75	1377
Roommates	1.76	0.7	946	1.7	0.69	1475
First Gen. College Student	0.1	0.3	946	0.12	0.32	1475
Freshman Fall GPA	3.22	0.52	932	3.42	0.43	1459
Roommate difference in ARR	0	1.28	946	0	1.27	1475
Roommate difference in ARR^2	1.64	2.44	946	1.61	2.34	1475
Median GPA in Dorm	3.41	0.07	944	3.4	0.08	1466
Relative Performance	-0.21	0.51	926	-0.07	0.42	1454
Relative $Performance^2$	0.3	0.82	926	0.18	0.6	1454

Table B

VARIABLES	(1) Combined	(2) Combined	(3) Male	(4) Male	(5) Female	(6) Female
ARR	0.907	1.029	1.042	1.066	0.842+	1.002
	(0.065)	(0.092)	(0.113)	(0.133)	(0.082)	(0.121)
Aid Ratio	2.206**	2.139**	1.708	1.842	2.586**	2.346*
	(0.590)	(0.591)	(0.754)	(0.857)	(0.857)	(0.792)
Ln(Distance)	0.818***	0.809***	0.848+	0.853	0.804**	0.788***
	(0.045)	(0.046)	(0.080)	(0.085)	(0.053)	(0.055)
Ln(Population Density)	1.103*	1.104*	1.058	1.064	1.131*	1.129*
	(0.047)	(0.048)	(0.076)	(0.077)	(0.059)	(0.062)
Roommates	1.256*	1.230+	1.177	1.153	1.303+	1.270+
	(0.136)	(0.140)	(0.198)	(0.212)	(0.185)	(0.184)
First Gen. College Student	0.870	0.959	1.092	1.131	0.785	0.883
	(0.258)	(0.298)	(0.595)	(0.701)	(0.276)	(0.308)
Freshman Fall GPA		246.345***		467.426***		139.389***
		(195.790)		(589.571)		(132.462)
Freshman Fall GPA^2		0.377***		0.341***		0.411***
		(0.054)		(0.080)		(0.071)
Constant	12.009***	0.005***	8.725**	0.002***	14.327***	0.014**
	(6.892)	(0.007)	(7.305)	(0.004)	(11.423)	(0.024)
Observations	2,213	2,186	851	840	1,362	1,350
Pseudo R-squared	0.0249	0.0562	0.0140	0.0624	0.0347	0.0749

Robust standard errors in parentheses
*** p<0.001, ** p<0.01, * p<0.05, + p<0.1

Table C

VARIABLES	(7) Male Dorm	(8) Male Dorm	(9) Female Dorm	(10) Female Dorm
ARR	1.113	1.184	0.856	1.057
	(0.167)	(0.201)	(0.111)	(0.154)
Aid Ratio	1.494	1.578	2.362**	2.292**
	(0.604)	(0.668)	(0.712)	(0.711)
Ln(Distance)	0.842+	0.835+	0.803**	0.785***
	(0.082)	(0.085)	(0.055)	(0.056)
Ln(Population Density)	1.049	1.050	1.134*	1.132*
	(0.077)	(0.077)	(0.060)	(0.063)
Roommates	1.136	1.107	1.272+	1.229
	(0.196)	(0.211)	(0.183)	(0.180)
Freshman Fall GPA		575.896***		139.525***
		(729.079)		(139.074)
Freshman Fall GPA^2		0.329***		0.411***
		(0.078)		(0.074)
Roommate difference in ARR	0.959	0.890	0.981	0.941
	(0.134)	(0.138)	(0.100)	(0.096)
Roommate difference in ARR^2	1.173*	1.201*	0.918*	0.921*
	(0.090)	(0.108)	(0.031)	(0.031)
Median GPA in Dorm	20.650+	14.816	0.994	1.288
	(36.791)	(28.191)	(1.246)	(1.678)
Constant	0.000	0.000*	15.784	0.006
	(0.001)	(0.000)	(66.340)	(0.028)
Observations	845	835	1,347	1,344
Pseudo R-squared	0.0299	0.0810	0.0400	0.0813

Robust standard errors in parentheses

*** p<0.001, ** p<0.01, * p<0.05, + p<0.1

Table D

VARIABLES	(11) Male	(12) Female
ARR	1.199	1.071
	(0.222)	(0.159)
Aid Ratio	1.684	2.273**
	(0.744)	(0.706)
Ln(Distance)	0.841+	0.787***
	(0.087)	(0.056)
Ln(Population Density)	1.101	1.130*
	(0.085)	(0.062)
Roommates	1.085	1.219
	(0.202)	(0.180)
Freshman Fall GPA	3.956	56.440**
	(7.827)	(83.635)
Freshman Fall GPA^2	0.718	0.490**
	(0.251)	(0.122)
Roommate difference in ARR	0.815	0.942
	(0.143)	(0.097)
Roommate difference in ARR^2	1.274*	0.920*
	(0.136)	(0.031)
Median GPA in Dorm	32.024+	1.209
	(64.397)	(1.582)
Relative Performance	0.586	0.650
	(0.449)	(0.357)
Relative Performance^2	0.360**	0.795
	(0.128)	(0.195)
Constant	0.000	0.019
	(0.000)	(0.103)
Observations	829	1,332
Pseudo R-squared	0.113	0.0617

Robust standard errors in parentheses
*** p<0.001, ** p<0.01, * p<0.05, + p<0.1