

Three Studies on the Value and Risk of Higher Education

by

Katharina Ley Best

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Industrial and Operations Engineering)
in The University of Michigan
2013

Doctoral Committee:

Associate Professor Jussi S. Keppo, Chair
Professor Romesh Saigal, Chair
Professor Edward D. Rothman
Professor Lawrence M. Seiford

© Katharina Ley Best 2013

All Rights Reserved

For Mama, Papa, and Christopher.

ACKNOWLEDGEMENTS

I would like to thank my advisor, Professor Jussi Keppo, without whom this dissertation could not have been completed. It is thanks to his guidance, encouragement, and dedication to these projects that I was able to complete this work and grow into a successful independent researcher. Even across continents and time zones, Professor Keppo's interest in my work and progress was unmatched throughout my time at the University of Michigan. Many thanks are also due to my committee co-chair, Professor Romesh Saigal, for providing a new outlook on my research work and guiding me through the final stages of the dissertation writing process. Thanks also to my committee members, Professor Larry Seiford and Professor Ed Rothman, who helped develop my interest in the application of financial engineering and operations research techniques to new areas.

The third chapter of this dissertation is based on a paper published in "Education Economics" (Best and Keppo, 2012). Thanks to Taylor & Francis for allowing me to include this work here. The work presented in the fourth chapter of this dissertation was generously supported by the American Association of University Women (AAUW) through the award of an American Fellowship. Thank you to the AAUW for providing an entire year of research funding and living support, as well as to AAUW staff for all of their efforts.

This research was made possible through the amazing support of the staff of the Industrial and Operations Engineering, especially Tina Picano Blay, Christopher Konrad, and Mary Winter. Thanks also to Wanda Dobberstein for her guidance

throughout my time as a Graduate Student Instructor. Teaching is one of the most rewarding and challenging things I had the opportunity to do while at the University of Michigan, and without her help I would have been much less successful and much more terrified. Portions of the work presented in this dissertation required access to restricted data from the U.S. Department of Education and the Bureau of Labor statistics. I would like to thank the staff at both of these organizations.

Finally, I want to thank my wonderful friends and family. Thanks to Kathryn Schumacher for being an unbeatable office mate for over four years, to my parents for their constant encouragement and pride, to my husband for making me feel so loved, and to our tiny “belly monster” for sending little pains and flutters to remind me of the most important things in life throughout the writing process.

TABLE OF CONTENTS

DEDICATION	ii
ACKNOWLEDGEMENTS	iii
LIST OF FIGURES	viii
LIST OF TABLES	ix
LIST OF APPENDICES	xi
CHAPTER	
I. Introduction	1
II. For what it's worth: Higher education spending in North America and Western Europe	3
2.1 Introduction	3
2.2 Background	4
2.3 Data	6
2.4 Model	7
2.5 Conclusion	13
III. The credits that count: The effect of credit growth and fi- nancial aid on college tuition and fees	15
3.1 Introduction	15
3.2 Literature review	17
3.3 Data collection	22
3.4 Model	25
3.5 Results	29
3.6 Conclusion	36

IV. A major choice: An examination of educational choices, student characteristics, and ability-adjusted income	39
4.1 Introduction	39
4.2 Background and literature review	43
4.3 Data	48
4.3.1 Data sources	48
4.3.2 Sample selection	50
4.4 Debiasing: Creating university and student groups	52
4.4.1 A new approach for addressing selection bias in school choice	52
4.4.2 Potential criticisms and shortcomings of the NLSY 1997 data	54
4.5 Modeling income	56
4.5.1 Post-college earnings	56
4.5.2 Effect of college attendance and other factors	59
4.5.3 Lifetime income	60
4.6 Drivers of college choice	65
4.6.1 Choice of attendance: The first stage binary logit model	66
4.6.2 Choice between colleges: The second stage conditional logit model	70
4.7 Conclusions	73
4.8 Future work	75
V. Conclusion	77
APPENDICES	80
A.1 Starting point for alternate models	81
B.1 Final model residual plots	83
B.2 Demand model with offerings change tuition component	85
B.3 Main dataset	86
B.4 Pre-estimation test	91
B.5 Preliminary model residual plots	93
B.6 Comparison of parameter estimates in data subsets	94
B.7 Ljung-Box test of residual autocorrelation	97
C.1 Effect of selectivity of school attended on post-college income, unconditional and conditional on selectivity of best school admitted to	98
C.2 Variables necessary for building discrete choice model	100
C.3 Projecting lifetime income	105
C.3.1 Functional form of the income curve	105
C.3.2 Parameter estimation	106
C.3.3 Income by educational attainment	109

C.4	Conditional post-college income effects using categorical selectivity data	112
C.5	Effect of attending college at all	114
C.6	Logistic choice models	115
C.7	Comparison to models in Long (2004b)	118
C.8	Collinearity and model selection	121
BIBLIOGRAPHY		122

LIST OF FIGURES

Figure

2.1	Gini coefficient by country.	11
2.2	Country fixed effect versus Gini coefficient.	12
3.1	Example of variable detrending. Detrended net tuition per student and actual net tuition per student.	27
3.2	Model enrollment and actual enrollment, detrended.	31
3.3	Detrended net tuition and credit variables.	34
3.4	Detrended net tuition and expenditures per student.	37
4.1	Average income over the first four years after expected college graduation by selectivity of top school admitted to, with effects of attendance and major decisions.	44
4.2	Average post-college earnings by selectivity of top school admitted to.	58
4.3	Best estimates of post-college income by selectivity of the best school admitted to and attendance decision.	59
B.1	Residuals over time, final models in tables 3.4 and 3.5.	84
B.2	Residuals over time, preliminary models.	93
C.1	Average income over the first four years after expected college graduation by selectivity of the school attended.	99
C.2	Third (curvature) term in equation (C.2), $\frac{1-e^{-t/\tau}}{t/\tau} - e^{-t/\tau}$, over τ and age (t).	107
C.3	Census synthetic lifetime income earnings curve fits based on equation (C.2) with fixed $\tau = \tau^*$ only.	108
C.4	Census synthetic lifetime income earnings curve fits based on equation (C.2) for different estimation procedures.	111

LIST OF TABLES

Table

2.1	Countries included in analysis.	7
2.2	Summary of data and sources.	8
2.3	Effect of GDP on new higher education enrollment.	9
2.4	Effect of GDP on new enrollment, including country fixed effects.	11
2.5	Effect of GDP on new enrollment, including fixed effects for public/total spending ratio categories.	12
2.6	Reform categories: Higher education reform in Europe.	14
3.1	Dataset summary.	25
3.2	Summary of first principal components.	26
3.3	Instrumental variables for the first stage regression of net tuition, equations (3.2)–(3.4).	28
3.4	Supply model (enrollment per high school graduate, equation (3.2)).	30
3.5	Demand model (enrollment per high school graduate, equation (3.3)).	30
4.1	Number of students in sample, by attendance and admission groups.	51
4.2	Predicted post-college income by selectivity of top school admitted to and school attended.	57
4.3	Effect of student characteristics on post-college income.	61
4.4	Number of students in lifecycle income projection sample, by attendance and admission groups.	63
4.5	Percentage of selected NLSY 1997 sample used in analysis, by attendance and admission groups.	64
4.6	First stage model: Binary logit model of whether to attend college or not.	68
4.7	Second stage model: Conditional logit model of college choice given college attendance.	73
A.1	Effect of population correlated and uncorrelated with GDP on new enrollment.	82
A.2	Effect of spending on tertiary education correlated and uncorrelated with GDP on new enrollment.	82
B.1	Demand model, offerings change only (enrollment per high school graduate).	85
B.2	Main dataset, part 1	87

B.3	Main dataset, part 2	88
B.4	Main dataset, part 3	89
B.5	Main dataset, part 4	90
B.6	Pre-estimation test results.	92
B.7	Parameter estimates in data subsets, supply model.	95
B.8	Parameter estimates in data subsets, demand model.	96
B.9	Ljung-Box test for autocorrelation of residuals, p-values with and without AR(1) adjustment.	97
C.1	Candidate variables used in model construction.	100
C.2	Separating effects of top school admitted to and school attended selectivity on Nelson-Siegel income parameter using interval selectivity data.	110
C.3	Separating effects of top school admitted to and school attended selectivity on post-college income, categorical selectivity data.	113
C.4	Effect of binary decision to attend college on post-college income.	114
C.5	Comparison model for first stage logit regression model of decision to attend college at all.	119
C.6	Comparison model for second stage conditional logit regression model of choice between candidate schools.	120
C.7	Overview of model selection process.	121

LIST OF APPENDICES

Appendix

- A. Appendices for Chapter II: Higher education funding in the Western world 81
- B. Appendices for Chapter III: The effect of credit growth and financial aid on college tuition and fees 83
- C. Appendices for Chapter IV: An examination of educational choices, student characteristics, and ability-adjusted income 98

CHAPTER I

Introduction

College in the United States is very expensive. This statement has been true for at least a generation, but costs have reached new heights in recent years. List tuition prices have grown at a rate of about 5% above inflation over the last ten years, and the financial aid budgets of both the US government and individual institutions are expanding to keep up.¹ The studies presented in this dissertation tackle some of the finance-related questions that arise in discussions of the higher education market in the United States. We examine college funding and affordability, aid policies, and college choice. This work may help to inform governments, institutions, families, and students about the risks and rewards of higher education.

In chapter II, we examine the role of public and private funding for higher education in countries throughout North America and Western Europe. We compare and contrast the various systems of higher education funding that are found in wealthy Western nations, and examine the impact of recent reforms meant to overhaul some of these systems. We focus on drivers of overall enrollment in higher education, and study the impact of spending, social inequality, and GDP. This work provides context for further study of the US system. In general, the US higher education system relies more heavily on private funding and financial aid than its European counterparts,

¹See e.g. The College Board (2012b) and The College Board (2012c) for more on college pricing and financial aid trends.

though recent European reforms tend to move away from traditional publicly funded systems.

Chapter III focuses on the US higher education market. By developing a two-stage least squares model of supply and demand in higher education, we build a better understanding of the effect of aid mechanisms on school and student decisions. We study the effect of financial factors such as tuition and aid amounts, expenditures, and schools' revenues on enrollment levels. Additionally, we compile an extensive dataset of macroeconomic variables that capture changes in the higher education market and financial aid market over a thirty-year period, and develop supply-side and demand-side models. A primary motivator for this work is the hypothesis that increased aid availability may actually be helping to fuel higher tuition prices.

Chapter IV studies the impact of financial factors on higher education at an even more granular level. A unique individual-level dataset on college choice decisions allows us to create models of the decision to attend college and the choice between institutions. We study the relationship between post-college income, individual and school characteristics, and the student side of the college decision process. We examine drivers of income differences across students after adjusting for ability. This work introduces a new method of adjusting for bias introduced through differences in student ability across schooling decisions. Unique survey data that captures all stages of students' application processes allows us to condition on the selectivity level of the top school to which a student was admitted, capturing valuable information about that students' intrinsic ability. We focus on the effect of student and family characteristics as well as school and major decisions made by individual students after adjusting for ability. This chapter also outlines important areas of future research that will enhance understanding of the relationship between educational decisions and future income levels.

CHAPTER II

For what it's worth: Higher education spending in North America and Western Europe

2.1 Introduction

Developed Western countries have taken very different approaches to financing higher education. On one extreme, students attending private four-year colleges and universities in the United States now pay an average of \$36,993 per year to attend school (College Board Advocacy and Policy Center, 2010). On the other hand, European students in Finland, Norway, Austria, and other countries pay no official tuition, though fees and housing costs may apply (EDALO Education Promotional Services S.L., 2011). Some governments have instituted complicated financial aid systems through which federal, state, and local governments can assist students, while other countries rely on private lenders. There is an extensive literature examining the similarities and differences, failures and successes of these different approaches to tertiary education. Overall, universities in North America and Western Europe provide some of the best education in the world, but experts and the media have warned in recent years that Western educational institutions may be losing ground (see e.g. Economist (2010) and Harris and Beschloss (2011)). In order for universities to remain competitive, affordable, and well-funded, governments must maximize the impact of aid and

subsidies.

This chapter examines differential effects of GDP, population, and tertiary education spending on enrollment across North America and Western Europe. Our goals are to study the effect of public versus private spending on enrollment, to identify economic drivers of enrollment, and to examine differences in enrollment levels caused by the introduction of certain reforms. New enrollment captures the initial decision to attend school and gives an indication of access to tertiary education. Improved access to higher education is a common goal for policymakers. The strongest predictor of new enrollment in tertiary education is GDP in the country in question, not spending on tertiary education alone. We find a positive relationship between GDP and enrollment. Although this primarily captures the fact that larger countries have greater enrollment, we investigate differences in enrollment based on the enactment of reforms and public to private spending ratios.

We find that the effect of GDP on enrollment varies between countries with different levels of private spending and tuition. Specifically, for the same GDP levels, enrollment is higher in countries with a high level of private spending on tertiary education. This may indicate that better economic conditions have a more positive effect in these countries. This result may also stem from the fact that private spending on education is more highly correlated with GDP than public spending, indicating that GDP increases raise enrollment both directly and indirectly through higher private spending levels. We find no significant differences in baseline enrollment or effect of GDP over time. We also do not see an effect of education and funding reforms.

2.2 Background

A rise in earnings inequality in the 20th century brought renewed relevance to the question of return on investment in higher education. During this time, influential papers like Becker and Chiswick (1966) and Mincer (1974) first applied sophisticated

human capital models to the question of returns on education. Those papers set the stage for future returns estimates based on ordinary least squares regression, various instrumental variables approaches, semi-log earnings functions, and natural experiments based on twin data and policy changes (see e.g. Psacharopoulos and Mattson (1998), Behrman and Rosenzweig (1999), Heinrich and Hildebrand (2005), and Hämäläinen and Uusitalo (2008)). Private returns are calculated based on out-of-pocket education costs and after-tax earnings differences between those with a tertiary degree and those without. Social returns are based on public expenditures and productivity differentials. In general, private returns exceed social returns, and higher levels of economic development and previous education reduce the impact of further education (see Psacharopoulos (1972)).

In this study, we focus on highly developed economies in Europe and North America. According to Psacharopoulos (1972), the richest countries in these regions have among the lowest returns on education in the world. Returns are higher in Eastern European countries like the Czech Republic, Poland, and Hungary, but are still below international average returns. Returns on education in these areas are fairly consistent over time, even in countries of the former Eastern Bloc that have recently transitioned to market economies (see Flabbi et al. (2011)). Each country strikes a different balance between public and private funding of tertiary education. In the United States, close to 40% of the contribution to tertiary education comes from households. In much of Europe, this number is well below 10% (OECD, 1998). In the last several years, many European countries have struggled to reduce government spending, and educational expenditures are a popular place for cuts. Costs can be shared among governments and taxpayers, students and their families, and philanthropic organizations. Mechanisms for such cost-sharing include public and private grants, student tuition and fees, government sponsorship of schools (tuition-free education), government subsidies for schools, government subsidies for students including

subsidized loans, indirect government support such as child allowances, and tax credits (for more on these see Johnstone (2005)). This work does not seek to create a new estimate of returns on investment in tertiary education. Instead, we investigate differences in the effect of educational spending and GDP on new enrollment in tertiary education.

2.3 Data

Lack of consistent, reliable data is a primary barrier to research on higher education spending. It is difficult to get relevant spending and earnings data in each country, and reporting differences between countries make it hard to compare results. Additionally, a large number of control variables must be accounted for, such as differences in age, intelligence, motivation, and other personal information (see e.g. Psacharopoulos (1972)). Past estimates of returns of higher education use data reported by individual governments, international groups like the European Union, UNESCO, or OECD, and individual survey results from studies like the European Community Household Panel (see, for example, Heinrich and Hildebrand (2005)).

For this work, we use data from the OECD, UNESCO, and Eurostat to build a panel dataset of enrollment, population, GDP, and expenditures. We also include a variety of adjustment variables which allow us to conduct analysis on a per student basis. These results are not provided here but support conclusions similar to those from the aggregate analysis that is presented. This dataset is supplemented with information about recent tertiary education reforms in Europe from a variety of sources, including Centro de Estudios en Gestión de la Educación Superior (2007), CHELPS, IOE London, Technopolis Group (2010), Johnstone (2005), OECD (1998), EDALO Education Promotional Services S.L. (2011), and ICHEFAP (2011). The dataset spans an eleven-year period from 1997 to 2008 and includes twelve countries, as shown in table 2.1. We consider new enrollment in tertiary education, graduation

Table 2.1: Countries included in analysis.

Europe	North America
Austria	United States
Finland	
Hungary	
Ireland	
Italy	
Netherlands	
Norway	
Poland	
Spain	
Sweden	
United Kingdom	

rates, unemployment rates, and relative earnings of tertiary graduates versus non-graduates as metrics of educational success in each country. Only the new enrollment metric yields significant results, probably since relative earnings and graduation rates depend increasingly on the actual quality of education received. On the spending side, we examine GDP, total expenditures on tertiary institutions, public expenditures on tertiary institutions, a ratio of public over total spending, and population as candidate predictors. As described in the next section, we employ a fixed effects approach to account for country differences, allowing us to limit the number of predictor variables in the model. A list of main variables used is provided in table 2.2.

2.4 Model

We use simple ordinary least squares regression to model the relationship between spending metrics and outcomes. The panel data compiled for this study are non-stationary as determined by inspection and unit root tests.¹ We therefore take

¹Fail to reject presence of unit roots in time series for new enrollment (p-value = 0.99 at lag 1).

Table 2.2: Summary of data and sources.

Category	Summary	Data source
Outcome metrics		
New enrollment	New entrants into tertiary (college-level, non-professional, ISCED 5A ^a) education, per population	OECD (2011)
Graduation rate	Graduates per enrollment in tertiary (ISCED 5 & 6) education	OECD (2011)
Employment rate	Relative employment rate for tertiary (ISCED 5 & 6) graduates versus upper secondary (ISCED 3 & 4) graduates	OECD (2011)
Earnings difference	Relative earnings for tertiary (ISCED 5 & 6) graduates versus upper secondary (ISCED 3 & 4) graduates	OECD (2011)
Spending and economic metrics		
Total expenditures	Expenditure on education institutions (total, public and private sources), tertiary (5,6), \$BN	OECD (2011)
Public expenditures	Expenditure on education institutions and administration, public sources, tertiary, \$BN	OECD (2011)
Ratio public/total expenditures	Public expenditures/total expenditures	OECD (2011)
GDP	Gross domestic product, \$BN	World Bank (2011)
Population	Population of each country studied	OECD (2011), European Commission (2011)
Adjustment factors		
Enrollment	Total enrollment in tertiary education (ISCED 5 & 6)	UNESCO (2011)

^aISCED is the International Standard Classification of Education developed by UNESCO. For more information, see UNESCO (2013).

ordinary differences of all data. Log differences were also considered and yielded similar results. The strongest relationship among the spending and outcome variables described in the previous section exists between GDP and new enrollment in tertiary education. Graduation rate, employment rate, and earnings difference are metrics of educational success rather than simply attendance. There may be too many other factors affecting quality of education to allow us to identify the relationship between these variables and per student spending. We therefore use only GDP and new enrollment for our analysis. Table 2.3 shows the simplest possible model demonstrating the relationship between GDP and new enrollment. We see that a \$1 million increase in GDP leads to an enrollment increase of 0.19 students. This relationship captures primarily the fact that larger economies enroll more students in tertiary education. We use this as a baseline to compare differences not explained by country size.

Table 2.3: Effect of GDP on new higher education enrollment. $\text{newEnrollment}_{i,t} = \beta_0 + \beta_1 \text{GDP}_{i,t} + \epsilon_{i,t}$.^a

Ordinary least squares regression			
Adjusted R ²	0.2757		
Variable ^b	Coefficient value	St. error	p-value
(Intercept)	2,402.52	2,678.64	
GDP (\$BN)	190.46	26.74	6.47x10 ⁻¹¹

^aData include yearly GDP and enrollment figures for countries listed in table 2.1 from 1997 to 2008.

^bFor detailed variable descriptions see table 2.2.

We investigate many alternative models to the one presented in table 2.3. A similar model using log-differenced data yields consistent results. Models where all data are normalized to account for country size (i.e. enrollment is given as a percentage of population and per capita GDP is used) also provide similar results. We also experiment with using population and spending on tertiary education instead of GDP. We find that the portion of these variables directly correlated with GDP provide most of the predictive power. Results regarding enrollment differences are confirmed. These models are provided in appendix A.1.

We use the model given in table 2.3 to test for differences between countries. We introduce country and time fixed effects in order to reduce possible omitted variable bias. Time fixed effects account for unobserved shifters of enrollment that change over time and affect all countries. Country fixed effects account for unobserved shifters of enrollment that may affect a particular country but are constant over time. We find no significant time fixed effects. However, we do see that, for the same GDP levels, enrollment is higher in the United States, the United Kingdom, and Poland, as shown in table 2.4. Investigating the causes of this difference is beyond the scope of this chapter, but we suggest it may be due to a larger private return on education; these countries have higher levels of social inequality than other countries studied, as shown in figure 2.1. This figure plots Gini coefficients for the countries in our sample. Gini coefficients measure social inequality, with a larger index suggesting greater inequality.² Figure 2.2 plots country fixed effects from the model in table 2.4 for each country against the corresponding Gini coefficient, showing a strong trend of greater enrollment for the same GDP level as Gini coefficient increases. Countries with higher Gini coefficients and higher enrollment relative to GDP also have a higher ratio of private to total spending, implying that more education funding comes from tuition than in other countries.

We can study the effect of private versus public spending by categorizing countries according to their private to total spending ratios. Table 2.5 shows this model, revealing higher enrollment levels for countries with higher private to total spending ratios when GDP and total spending are held constant. A possible explanation for this phenomenon is that GDP growth may translate more directly into more education spending in countries with high private spending. Private individuals can choose to take on loans or grants for educational purposes in response to a growing economy more quickly than governments can increase spending. Additionally, GDP growth is

²Gini coefficient data compiled from Central Intelligence Agency (2011).

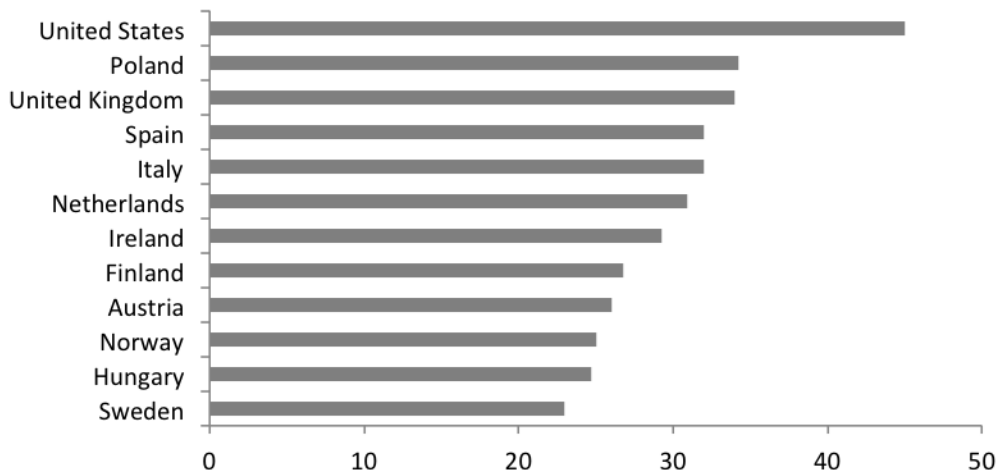


Figure 2.1: Gini coefficient by country.^a

^aGini coefficient data compiled from Central Intelligence Agency (2011). Gini coefficient measures social inequality, with a larger index suggesting greater inequality.

more highly correlated with private spending on education than public spending, so increases in GDP raise enrollment levels directly as well as through higher private spending levels.

Table 2.4: Effect of GDP on new enrollment, including country fixed effects. $\text{newEnrollment}_{i,t} = \beta_0 + \beta_1 \text{GDP}_{i,t} + \gamma_i \text{countryDummy}_i + \epsilon_{i,t}$.^{a,b}

Fixed effects regression			
Adjusted R ²	0.6000		
Variable ^c	Coefficient value	St. error	p-value
GDP	-2,157.66	42.19	2.87x10 ⁻⁴
Poland	15,400.86	6,609.50	0.02
United Kingdom	18,483.89	6,958.13	8.98x10 ⁻³
United States	138,237.45	14,327.32	2.00x10 ⁻¹⁶

^aOnly coefficients significantly different from zero are presented.

^bData include yearly GDP and enrollment figures for countries listed in table 2.1 from 1997 to 2008.

^cFor detailed variable descriptions see table 2.2.

We further investigate interaction effects between country dummies as well as ratio categories and GDP. These interaction effects test for differences in the effect of GDP changes across countries or different private to total spending ratios. The interaction terms in the regression equation allow the impact of GDP on enrollment

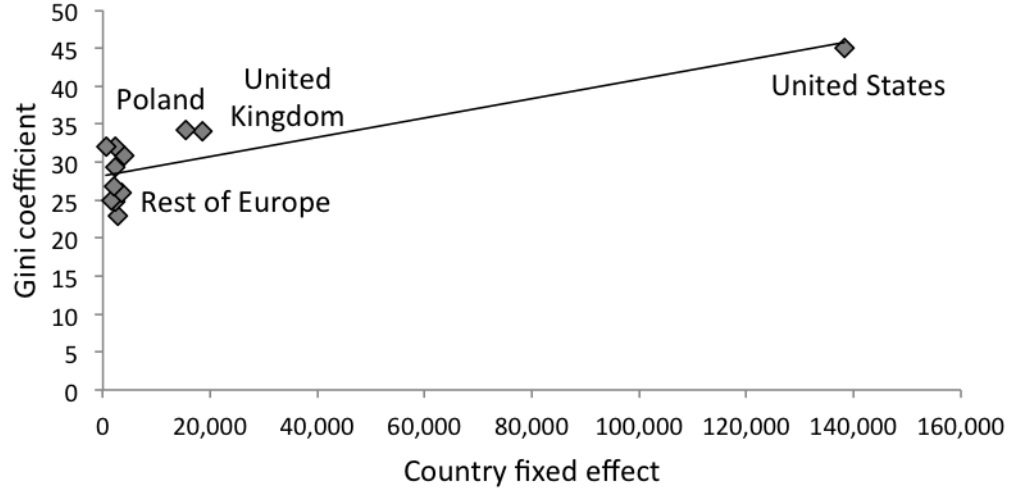


Figure 2.2: Country fixed effect versus Gini coefficient.^{a,b,c}

^aGini coefficient data compiled from Central Intelligence Agency (2011). Gini coefficient measures social inequality, with a larger index suggesting greater inequality.

^bFit line equation given by $\text{giniCoefficient} = 0.0001 * \text{countryDummy}_i + 28.193$.

^cCountries plotted are those listed in table 2.1.

Table 2.5: Effect of GDP on new enrollment, including fixed effects for public/total spending ratio categories. $\text{newEnrollment}_{i,t} = \beta_1 \text{GDP}_{i,t} + \delta_i \text{ratio}_{i,t} + \epsilon_{i,t}$.^a

Fixed effects regression			
Adjusted R ²	0.4002		
Variable ^b	Coefficient value	St. error	p-value
GDP (\$BN)	74.01	37.36	0.05
Public/total spending < 0.6 (high relative tuition)	45,881.63	10,759.5	3.86x10 ⁻⁵

^aSpending ratio is calculated as public educational spending/total educational spending for each country. Fixed effects are estimated for each ratio category. Categories are those countries with a ratio > 0.9 (low relative tuition), countries with a ratio between 0.6 and 0.9 (medium tuition), and countries with a ratio < 0.6 (high relative tuition). Only coefficients significantly different from zero are presented. Data include yearly GDP, enrollment, and spending figures for countries listed in table 2.1 from 1997 to 2008.

^bFor detailed variable descriptions see table 2.2.

(i.e. the slope of the regression line) to vary between countries or ratio categories, as shown in equation (2.1).

$$\text{newEnrollment}_{i,t} = \beta_0 + \beta_1 \text{GDP}_{i,t} + \gamma_i (\text{countryDummy}_i * \text{GDP}_{i,t}) + \epsilon_{i,t} \quad (2.1)$$

where β_0 is the intercept, β_1 is the coefficient of GDP, γ_i is the estimated country

fixed effect for country i , and $\epsilon_{i,t}$ is a white noise error term. We find no significant interaction terms for the regression of enrollment on GDP. This suggests that GDP changes have a similar effect across countries.

We also look for evidence of regression discontinuity caused by recent reforms in European countries. In the past few decades, countries like Austria, Finland, and Hungary have eliminated or severely reduced tuition. Others, like Portugal and the United Kingdom, have introduced or increased tuition and fees. Still others have handed university administration over to non-government bodies for the first time, increasing institutional autonomy. Poland and Hungary, among others, introduced student loan and grant systems to help families pay for increasingly expensive tertiary education. For the current analysis, reforms are categorized according to a taxonomy suggested by ICHEFAP (2011), as seen in table 2.6. We test for differences in the effect of GDP on enrollment before and after reforms in several of these categories, including increased school autonomy, more government oversight of higher education, tuition bans, changes to loan programs, and changes to grant programs. We find no significant differences in the effect of GDP on enrollment based on these reform categories.

2.5 Conclusion

Our results indicate that GDP growth increases enrollment, with GDP being the main driver of enrollment changes. Although this result may simply demonstrate that larger countries have higher enrollment in tertiary education, the differences between countries provide interesting insights. Enrollment is higher for the same GDP and spending levels in countries where private spending contributes significantly to education. This may be because rising GDP levels in these countries fuel enrollment both directly and by improving families' ability to afford higher education. We find no significant differences in the effects of GDP based on recent European higher

Table 2.6: Reform categories: Higher education reform in Europe.

Reform category	Specific reform	Countries where reforms implemented
Cost sharing	Introduction of tuition or significant fees	Austria, Finland, Hungary, Italy, Poland, United Kingdom
	Tuition or significant fees banned	Austria, Finland, Hungary, Ireland
	Introduction of a second pay-track for non-government supported students	Hungary
	Tuition increase	United Kingdom
	Increase in the amount of money received from governments, or how it is allocated	Austria, Ireland, Italy, Spain
Autonomy	Increase in university autonomy	Austria, Ireland, Italy, Netherlands, Spain, Sweden, United Kingdom
Financial aid	Loans introduced or expanded	Austria, Hungary, Netherlands, Poland, United Kingdom
	Grants introduced or expanded	Austria, Hungary, Poland
	Introduction of loan limits	United States

education reforms, suggesting that these reforms have not had a significant impact on the effectiveness of subsidies to education, at least in terms of attendance and access. This is an important result for policymakers as they strive to improve access to higher education. Future work should seek to confirm the existence of enrollment differences based on public versus private spending differentials. A more complete dataset over a longer period of time would be required. Additionally, future research should attempt to adjust for the effects of significant shifts in the demographics of many Western European nations. An ageing population and changes in ethnic composition due to immigration could be significant drivers of higher education enrollment which are not studied here.

CHAPTER III

The credits that count: The effect of credit growth and financial aid on college tuition and fees

3.1 Introduction

Fifty-eight percent of US college students take out loans to help pay for tuition and fees. According to *The Wall Street Journal*, their average debt load upon graduation is \$23,186 and rising (Chaker, 2009). This comes as no surprise; college tuition prices increased by a staggering 326% between 1987 and 2007 (5.8% annually). To provide some comparison, the rise in medical costs during the same period was just 186% (Bureau of Labor Statistics, 2009). Rising tuition rates make it more and more difficult for families to pay for a college education. Student loans contribute to high levels of both personal and public debt, which are exacerbated by high tuition prices and easy credit from government student loan programs, respectively. These trends may be unsustainable.

Rising tuition rates have become a popular topic in the media and a priority for policymakers. Federal and state governments try to improve access to higher education for Americans through expanding financial aid programs. These programs allow thousands of Americans to attend school, but prices continue to rise, prompting governments to offer more and more financial aid. As early as 1987, Secretary of Edu-

cation William Bennett, Jr. suggested that readily available student loans and grants may actually be fueling the increase in tuition prices (Bennett, 1987). The current economic environment has infused renewed relevance into investigating a revised version of the “Bennett Hypothesis”: If public funding artificially inflates prices, they may eventually collapse, forcing schools to close, shattering perceptions of the worth of higher education, and destabilizing the education lending system through tighter lending standards and higher interest rates. In the meantime, prices may become too high for many Americans to afford without plunging into debt. Additionally, there is growing concern that US institutions may be losing ground in relation to international peers (see e.g. Harris and Beschloss (2011)), and that top tier elite institutions no longer carry their former prestige and value (see e.g. Kim et al. (2009)). Increasingly easy access to educational materials and skilled labor around the world is rapidly changing the US higher education market, creating uncertainty about the feasibility of high tuition prices in the future.

The main objective of this work is to conduct a macro-level analysis of the factors affecting supply and demand of higher education. By creating a single picture of supply and demand dynamics, we provide a context for previous research that studies the importance of individual shifters of supply and demand. We confirm that tuition, financial aid, and credit have a significant effect on US higher education supply and demand and build a better understanding of their directional impact within a holistic model. We use a two-stage least squares regression with a first order autoregressive error term to estimate supply and demand models for undergraduate education. Supply and demand are measured by enrollment quantity (undergraduate enrollment per high school graduate). We find that the benefit of a college education, household debt, and student loans shift demand, while cost of operations for schools, government aid to schools, and tuition and non-tuition revenue affect supply. Our model supports the theory that schools benefit from spending more on their students and increasing

tuition prices. However, our results indicate that credit constraints are a concern for students and their families, driving down demand and possibly limiting the ability of schools to continue on their trend of improving facilities, better services, and higher costs. Tuition and debt are highly correlated, suggesting that people use loan aid to cover greater schooling costs.

3.2 Literature review

A primary challenge in postsecondary education research is the availability of data. In the US, the federal government is the main data provider, and maintains sources of both aggregate and individual student data over time. Federal aggregate data are compiled in the Digest of Education Statistics and the Integrated Postsecondary Education Data System, while individual student data are provided through a series of longitudinal surveys (National Center for Education Statistics, 2008, 2009). Several private institutions such as *US News & World Report* and The College Board also keep data on rankings, tuition price, financial aid, and student bodies.

A second major challenge in this field is that the benefits of education are difficult to quantify. There are signs that the benefit of college has begun to decline in recent years, with more people with bachelor's degrees finding it difficult to secure employment in a field that takes advantage of their qualifications. For example, according to *The Chronicle of Higher Education*, 13% of parking lot attendants currently hold bachelor's degrees or higher (Vedder, 2010). However, recent return-on-investment studies (see e.g. Heckman et al. (2008)) claim that returns on a college education are steady or even increasing. These studies do not fully account for inherent differences between those that attend college and those that do not, or adjust for the possibility that college is unfordable for some. The benefit of college varies widely with the quality of school attended, and depends on each student's individual goals, financial situation, and location. Dale and Krueger (1999) show that the best predictor of a

college education's worth is tuition charged at a particular university, with a more expensive education consistently yielding better pay, even after adjusting for the differing quality of student bodies at various institutions. However, a more recent study by Dale and Krueger (2011) shows that controlling for inherent student ability based on SAT scores almost eliminates the wage increase effect. Higher tuition price may also lead to increased public perception of prestige, especially at elite schools (John, 1992). Hsing and Chang (1996) show that college enrollment has become more and more sensitive to tuition and related costs, with enrollment falling as costs increase. All else being equal, higher tuition prices for the same education reduce the net value of a college education. There may also be significant peripheral benefits to education that allow educated people to participate more effectively in consumer markets. For example, Grinblatt et al. (2009) show that education and intelligence may improve performance in mutual fund markets. Education may also have a signaling value that goes beyond the improvement to earnings potential actually created by education itself. Hämäläinen and Uusitalo (2008), for example, use recent reforms in the Finnish Polytechnic education system to show that there is a signaling effect on earnings difference between Polytechnic and vocational school graduates. This type of signaling may allow more prestigious schools to charge more for the same services, since admission to a top tier school signals ability to employers. In this study, we use a combination of the factors identified in past literature as a measure of college education benefit, namely unemployment rate, the earnings difference between college and high school graduates, expenditures by schools, and disposable income.¹

Past research on drivers of demand for higher education provides mixed results. Campbell and Siegel (1967) study historical demand levels from 1919 to 1964 and find that enrollment rises with income and falls with tuition prices. Hight (1975)

¹Disposable income is included here because it is highly correlated with the other proxies for benefit of college, meaning its inclusion as an individual variable leads to cross-correlation and biased estimators in the final model.

studies the period from 1927 to 1972, uncovering an increasingly complicated demand structure. Hight finds that the divergence in demand for public versus private education is fueled by tuition gaps and differing effects of income levels. With the growth of federal, state, and private financial aid programs, demand structures are becoming even more complicated. Carneiro et al. (2010) use an instrumental variables approach to estimate the marginal effect of policy changes, including tuition prices, on demand. They find that the level of effectiveness in encouraging college enrollment varies greatly with the type of policy change made. It also differs from student to student, depending on their specific situation and especially their prior probability of attending college.

Several groups have specifically examined the effect of financial aid and student loans on demand and/or tuition price. Conclusions on the institutional effect of financial aid vary. Due to difficulties in data compilation and the large number of possible factors affecting supply and demand, most studies focus on a particular program, locality, or event. Long (2004a), for example, looks at the impact of the Georgia HOPE Scholarship program, and concludes that scholarships are causing schools to increase tuition and fees for all students. Dynarski (2004) shows that Georgia's subsidies also increase college attendance, as do similar programs in other states. In a larger sample of four-year schools, however, Long (2003) shows that there is no significant relationship between federal tax credits and tuition price. Further, Long (2006) suggests that aid is not a significant factor in raising tuition prices, and that the link of aid and tuition must be studied by examining net tuition prices actually paid by students instead of list prices. We make this adjustment in the current study. Singell and Stone (2007) evaluate the validity of the "Bennett Hypothesis" with regard to Pell grants and conclude that these grants do not raise public universities' in-state tuition. At private institutions, however, they find that tuition growth parallels Pell grant increases. Curs et al. (2005) conclude that the institutional impact of Pell grants varies

greatly with school selectivity. Rizzo and Ehrenberg (2004) suggest that schools do not increase tuition in response to state or federal financial aid to students. However, they do not examine aid directly to schools.

College supply dynamics are less well studied. A focus on cutting edge research has allowed America's universities to rise to the top internationally, but is also tremendously expensive and may be detracting from the teaching that students pay so much for (Schumpeter, 2010). Peña (2006) finds that rising tuition costs are highly correlated with increased wealth among colleges, at least at four-year private institutions.² According to Peña, the growing benefits of college and increase in wealth are the main drivers of tuition increases in recent years. Wealthier families are able to pay for more expensive schooling, allowing institutions to raise tuition prices without reducing their applicant pool. If schools compete (and spend) to be top tier, high-tuition institutions, Americans looking for a decent education at a reasonable price will have fewer and fewer options. Rankings are, at least in part, driven by spending levels (U.S. News and World Report, 2012). However, it is not clear that a better ranking necessarily means students receive a better education. Most schools are non-profit entities whose prestige increases with spending more, which is problematic for accessible education pricing.

It is difficult to distinguish cause from effect in this analysis of financial aid and tuition prices. However, something is clearly driving prices upwards, and student loan debt levels are causing severe financial problems for many graduates. In 2009, total student loan debt surpassed total credit card debt in the United States, with close to \$850 billion in outstanding student loan debt (Pilon, 2010). Additionally, default rates are on the rise. Private for-profit institutions such as the University of Phoenix, which are quickly gaining market share, post an average 11.6% default rate among their students (McCluskey, 2010). We cannot say that financial aid causes tuition

²Wealth is measured by endowment values and the gap between tuition revenue and expenditures.

prices to rise, but it may be that current pricing mechanisms are somehow unbalanced, leading to higher and higher prices rather than to a more efficient, accessible education system. A report by the Goldwater Institute finds that the number of administrators per 100 students at leading US universities has increased by almost 40% since the early 1990s (Greene, 2010). Meanwhile, the number of researchers and teachers has grown by less than 20%. Increasing government subsidies help universities cover ever larger administrative costs, reducing schools' need to increase efficiency. Top tier schools, such as Yale and MIT, spend more on administration costs per student than many other schools spend in total. Yale, for example, spends about \$60,000 per student on administration, while the University of North Texas, among the most frugal, spends under \$10,000 per student in total (Greene, 2010). Schools spend vast amounts of money on professional-grade athletic facilities, sports programs, and clubs, and pay the most money to tenure-track professors who teach the fewest students (Hacker and Dreifus, 2010). These professors may be extracting high economic rents.

Instead of studying causal effects of specific reforms or policies, this chapter provides a comprehensive picture of the supply and demand market designed as a backdrop for understanding other findings. We are able to use aggregate data for individuals and schools across the United States to understand the market as a whole. We consider the whole college and university system, not e.g. private and public schools separately or competition among colleges. Our results provide support for many theories developed through study of natural experiments or more restrictive datasets, supporting the generalizability of these results. We see that consumers desire the most expensive education possible, but are hampered by credit constraints. Schools can thus use prices as a signal of quality, raising prices. In light of these higher prices, governments and private lenders offer loans to help student to achieve their educational goals, possibly allowing schools to raise prices even further. Our model allows us to build a picture of overall supply and demand dynamics, revealing

conditions that are exactly right for the existence and perpetuation of such a cycle of higher prices and higher debt.

3.3 Data collection

Due to the empirical nature of this study, all of our analysis rests on the compilation of a complete and accurate dataset. We compile a dataset of proxy variables for drivers of higher education supply and demand as well as a metric of quantity (enrollment). Candidate drivers cover spending, costs, financial aid, and college benefit. We focus on average behavior across four-year post-secondary institutions in the United States. Data include public, private, for-profit, and non-profit schools during the period 1976-2007. We select this period based on data availability and because it is recent enough to support an understanding of the current situation. In order to account for changes in the overall industry size, most data are reported per student, allowing for comparisons over the thirty-year period examined. Similarly, we measure enrollment quantity divided by high school graduates in that year. All analysis was also conducted without this normalization, modeling the effect of country-wide spending, cost, aid, and college benefit on total enrollment in number of students. This model yields similar results and consistent coefficient signs, alleviating concern that our results are based solely on the chosen normalization.

We define the quantity of both supply and demand as enrollment (in number of students) per high school graduate. The normalization factor is included here to account for the growth trend in the college-eligible population over the period under study. Using these data we construct two separate linear regression models, one for supply and one for demand. We identify candidate drivers of supply and demand from past literature and conversations with faculty, parents, students, and administrators. It is impossible to include every factor that may shift supply and demand in a given year. However, the goal here is to include main drivers identified in past literature

and identify internally consistent relationships between them. We focus especially on drivers related to financial aid and credit. We include those variables that we believe should have the largest impact as candidate drivers. We then use data availability to identify viable proxies. Candidate drivers of supply are the cost of operations for schools, financial aid given directly to schools, non-tuition revenue, and net tuition price. Candidate drivers of demand are the benefit of a college education, household financial situation, financial aid directly to students, and tuition price. For a list of candidate drivers see table 3.1. We separate aid given directly to schools from aid given through individual students in order to examine whether the mechanisms of aid-giving plays a role. In order to capture the price actually paid by students and received by universities, we use a net tuition estimate, defined as the list tuition price minus institutional aid (aid given directly to students by schools). Most students receive additional aid from other sources, thus further lowering the price they actually pay. For consistency between the supply and demand models, however, we include these adjustments in financial aid variables rather than in the tuition price itself. Additional financial aid sources provide aid to students, but do not affect institutional revenue.

The candidate drivers of supply capture schools' expenses as well as revenue from aid, tuition, and other sources. These are the main financial drivers of colleges' decisions, representing major cash flows. Non-financial drivers, such as rankings, are excluded here because they are relative metrics with unclear aggregate interpretations.³

Candidate demand drivers are designed to capture the factors individuals use when making decisions about whether and where to attend school, covering both the desire and the ability to attend. Estimates of the college education benefit capture the actual value of the education received. Tuition, aid, and credit variables measure

³It is possible that some sort of "ranking inflation" exists, meaning that higher rankings overall may indirectly affect other metrics. This is not explicitly treated in our model. Relative metrics could be included in future studies at the individual school level.

the price paid for an education and the ability to afford it. Of course, many personal factors, such as preferences regarding location, school size, non-financial career goals, and comfort with indebtedness, also influence enrollment decisions. These factors are again difficult to capture in an aggregate dataset, and are not explicitly included here.

We include a total of thirteen proxy variables for the underlying factors in the dataset, with seven variables aimed at explaining demand, five at explaining supply, and one endogenous variable (net tuition price) believed to shift both supply and demand. Other proxy variables were considered in the original dataset and are provided in appendix B.3. These other variables were discarded in preliminary analysis as inferior proxies for underlying variables. Table 3.1 gives a summary of the candidate proxy variables. The majority of data are from the Digest of Education Statistics maintained by the National Center for Education Statistics. A complete dataset for the period 1976–2007 is constructed by extrapolation based on the available data points. For the complete dataset and calculation details see appendix B.3.

Table 3.1: Dataset summary.

Candidate driver	Proxy variables	Model	Variable description
Quantity	Enrollment		Full-time equivalent enrollment at degree-granting institutions per high school graduate
Non-tuition revenue per student ^a	Endowment value per student	Supply	Average Harvard/Yale endowment market value, per student ^b
	Endowment distribution per student	Supply	Yale endowment spending, per student
	Donations per student	Supply	Donations, per student (voluntary support from persons and companies)
Cost of operations	Higher Education Price Index	Supply	HEPI inflation index ^c
Financial aid to schools per student	Aid to schools per student	Supply	Total government support directly to schools, per student (federal, state, and local)
Benefit of college	Earnings difference	Demand	Earnings difference (bachelor's degree or higher average - high school degree average)
	Expenditures by schools	Demand	Expenditures by schools, per student
	Unemployment	Demand	US unemployment rate
	Disposable income	Demand	US per capita disposable income
Grant aid to students per student	Grants per student	Demand	Average federal grant aid amount per full-time-equivalent student
Credit effects per student	Student loans per student	Demand	Average federal loan amount per full-time-equivalent student
	Household debt per student	Demand	US household debt, per student
Net tuition per student	Net tuition per student	Both	Average tuition and fees - institutional aid, per student

^aNon-tuition revenue per student, benefit of college, and credit effects are each summarized in a single proxy variable which represents the first principal component of the underlying proxies listed here. These principal component proxies appear in the final models.

^bHarvard and Yale endowment values are chosen as proxies since their investing strategies serve as a model for other schools and because data are publicly available.

^cThe Higher Education Price Index (HEPI) is an inflation index which tracks price changes for the basket of good commonly purchased by institutions of higher education, making it more specific than e.g. Consumer Price Index. It is computed annually and distributed free of charge by the Commonfund Institute (Commonfund Institute, 2009). The Index is used by schools to assist in the budgeting process. Categories included in the HEPI index are salaries for faculty and other employees, fringe benefits, utilities, supplies and materials, and miscellaneous services. HEPI index is adjusted for inflation during the time period studied here.

3.4 Model

The dataset described above creates several problems for regression modeling. Since multiple proxy variables for the same underlying candidate supply/demand driver are considered, there is strong correlation among the variables. In order to

reduce this correlation, we use principal component analysis to identify a single proxy variable for some candidate drivers. Specifically, we develop single proxy variables for non-tuition revenue in the supply model as well as for the benefit of a college education and credit in the demand model (see table 3.1 for proxy variables and candidate driver categories). We do not use principal component analysis across all candidate drivers since this would make coefficient signs difficult to interpret in the final models. Similarly, we avoid using principal components for variables which are of particular interest. A summary of the first principal components for non-tuition revenue, benefit of college, and credit are provided in table 3.2.

Table 3.2: Summary of first principal components.^a

Supply		Demand	
<u>Non-tuition revenue per student</u>		<u>Credit factors</u>	
Proportion of variance	0.916	Proportion of variance	0.633
Standard deviation	1.654	Standard deviation	1.126
<i>Loadings:</i>		<i>Loadings:</i>	
Endowment value per student	0.706	Household debt per student	0.707
Endowment distribution per student	0.708	Student loans per student	0.707
Donations per student	-0.018		
		<u>Benefit of college</u>	
		Proportion of variance	0.53
		Standard deviation	1.46
		<i>Loadings:</i>	
		Earnings difference	-0.507
		Expenditures by schools	-0.361
		Unemployment	0.459
		Disposable income	-0.633

^aFor detailed variable descriptions see table 3.1.

A second problem with the dataset is that variation is often driven by a strong time trend. To reduce concern that model fit is strong only because we are able to capture growth patterns, all variables are inflation adjusted. Additionally, we remove a linear trend as in equation (3.1) from all data:

$$x(t)_{\text{detranded}} = x(t)_{\text{actual}} - b - a(t - 1976) \quad (3.1)$$

where $x(t)$ is the variable in year t , with linear growth rate a and intercept b . This adjustment eliminates correlation among residuals in the supply model and reduces correlation in the demand model (for details see appendix B.7). A linear trend is chosen because this provides a better fit than an exponential curve for most variables. A sample plot of the actual and detrended net tuition variable is provided in figure 3.1. We continue to see heteroskedasticity and cyclical residuals even after this adjustment (see appendix B.5). To account for this, we introduce a first order lagged autoregressive error term (see e.g. Chatfield (2009)).

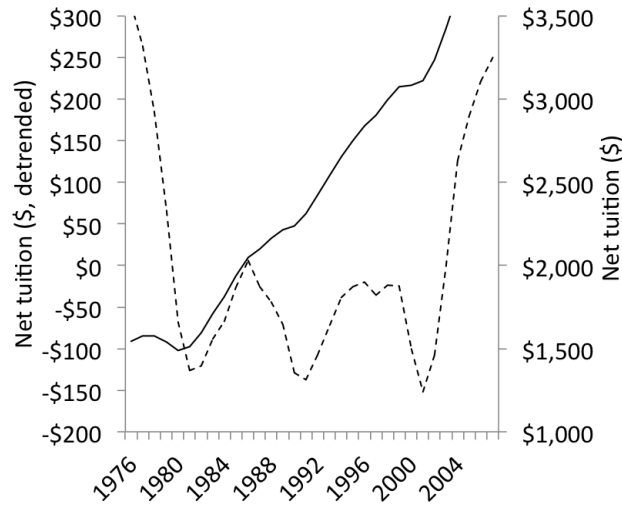


Figure 3.1: Example of variable detrending. Detrended net tuition per student (left axis, dotted) and actual net tuition per student (right axis, line). For detailed variable descriptions see table 3.1.

A third issue we must address is the endogeneity of the price variable, net tuition. Net tuition price shifts both demand and supply, violating ordinary least squares orthogonality assumptions. The two-stage least squares approach provides a method for adjusting a least squares model to account for the presence of endogenous variables (see e.g. Hayashi (2000)). Table 3.3 provides a list of candidate instrumental variables for the first stage estimation in each model.

Using the two-stage least squares approach, we develop the basic structure of the two models, which are provided in equations (3.2)–(3.4). These are not the

Table 3.3: Instrumental variables for the first stage regression of net tuition, equations (3.2)–(3.4).

Model	Instrumental variables used^{a, b}
Supply model	College benefit Credit effects per student
Demand model	HEPI Non-tuition revenue per student Aid to schools per student

^aInstrumental variables in the supply model are factors that shift net tuition but not supply. Instrumental variables in the demand model are factors that shift net tuition but not demand.

^bFor detailed variable descriptions see table 3.1.

final supply and demand models, but rather starting points for model identification. We include the first order autoregressive error term to account for heteroskedastic error terms caused by the time series nature of the input data (Chatfield, 2009). In this model, we use exogenous variables from the demand model as the instrumental variables in the supply regression (see e.g. Angrist and Krueger (2010)) and vice versa, yielding the same first stage regression for supply and demand. The first stage equation is presented in equation (3.4). All model equations are developed at a particular time (particular year) t for simplicity.

Supply model:

$$\begin{aligned}
 \text{enrollment}_t &= a_0 & (3.2) \\
 &+ a_{ex}[\text{costOfOperations}, \text{nonTuitionRevenue}, \text{aidToSchools}]_t \\
 &+ a_{en}[\text{netTuition}]_t + Y_t \\
 Y_t &= b_0 Y_{t-1} + e_{S,t}
 \end{aligned}$$

Demand model:

$$\begin{aligned} \text{enrollment}_t &= c_0 & (3.3) \\ &+ c_{ex}[\text{benefitOfCollege}, \text{grantAidToStudents}, \text{creditEffects}]_t \\ &+ c_{en}[\text{netTuition}]_t + W_t \\ W_t &= d_0 W_{t-1} + e_{D_t} \end{aligned}$$

First stage:

$$\begin{aligned} \text{netTuition}_t &= f_0 & (3.4) \\ &+ f_S[\text{costOfOperations}, \text{nonTuitionRevenue}, \text{aidToSchools}]_t \\ &+ f_D[\text{benefitOfCollege}, \text{grantAidToStudents}, \text{creditEffects}]_t + e_{1_t} \end{aligned}$$

We estimate the two-stage least squares models in equations (3.2)–(3.4) in a single step using R’s built-in two-stage least squares function⁴. We transform the original input variables in equations (3.2)–(3.4) to remove autocorrelation of error terms by first estimating b_0 and c_0 in equations (3.2) and (3.3). Upon model estimation, all results are transformed back into the original data space for reporting. All statistics and standard errors reported in the following sections assume that b_0 and c_0 are estimated without error. For more about the estimation of models with autocorrelated error terms, see e.g. Pandit and Wu (1990).

3.5 Results

Since we reduce the number of candidate variables and eliminate sources of collinearity prior to model fitting, almost all variables in the initial models are significant drivers of enrollment and are therefore present in the final models. There are two

⁴Fox, John. *Package ‘sem’*. CRAN. 2010.

notable exceptions in the demand model. Firstly, grant aid to students (per student) falls out due to strong correlation with student loans (per student), which is included in the credit effects principal component. Secondly, tuition is excluded from the final model here. We test a model where tuition in the demand model is adjusted for grant aid, i.e. net tuition = list tuition - institutional aid - grant aid. Signs in this adjusted model are consistent with the model in table 3.4, but correlation between grants and student loan aid makes it difficult to interpret credit effects. Tuition price has a positive sign when included in the model.⁵

Table 3.4: Supply model (enrollment per high school graduate, equation (3.2)).

Two-stage least squares regression			
b ₀ (residual autocorrelation)	0.838		
Correlation (model estimates vs. actual) ^a	0.73		
Exclusion restrictions	Test statistic	p-value	
Sargan	1.9837	0.1590	
Basmann	1.7183	0.1899	
Variable ^b	Coefficient value	St. error	p-value
(Intercept)	-0.0069	0.0140	
Cost of operations (HEPI)	0.0434	0.0101	0.0002
Non-tuition revenue per student	-0.2423	0.0546	0.0002
Financial aid to schools per student	0.0002	0.0001	0.0320
Net tuition per student	-0.0015	0.0006	0.0152

^aModel fit is captured by the correlation between the model estimates and actual values of supply and demand.

^bFor detailed variable descriptions see table 3.1.

Table 3.5: Demand model (enrollment per high school graduate, equation (3.3)).

Ordinary least squares regression			
d ₀ (residual autocorrelation)	0.653		
Correlation (model estimates vs. actual) ^a	0.73		
Variable ^b	Coefficient value	St. error	p-value
(Intercept)	-0.0027	0.0119	
Benefit of college	0.0319	0.0148	0.0399
Credit effects per student	-0.0790	0.0241	0.0028

^aModel fit is captured by the correlation between the model estimates and actual values of supply and demand

^bFor detailed variable descriptions see table 3.1.

⁵Note that since tuition falls out of demand model, we use ordinary least squares with a first-order autocorrelation term to estimate the demand model.

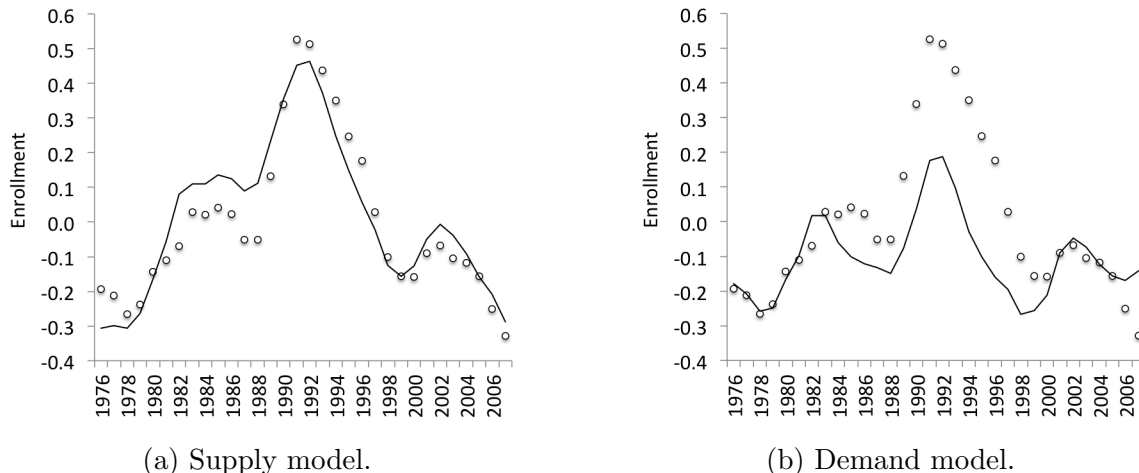


Figure 3.2: Model enrollment (line) and actual enrollment (circles), detrended.^a

^aModels are presented in tables 3.4 and 3.5. Models are based on equations (3.2)-(3.3).

Model robustness is difficult to test explicitly since small sample size precludes out-of-sample testing. When using time series data as individual observations, we must account for endogenous differences between different time periods. However, it is not possible to include, or even observe, all factors that may change over time (see e.g. Deming and Dynarski (2009)). Since we are using time points as individual observations, traditional fixed effects techniques (see discussion in e.g. Park (2009)) cannot be applied here. Alternatively, the autocorrelation term accounts for changing effects, creating a dynamic model which adjusts for changes in unobserved variables over time. This is similar to a dynamic approach for panel data that uses lagged observations to account for time trends (see e.g. Dhiring and Mordonu (2007)). Additionally, we compare model estimates using different types of standard errors, noting that standard and robust errors yield similar results (see e.g. Knutsen (2012)).

Since we are primarily interested in determining accurate coefficient signs, we focus model testing around these signs. We compare model signs with those from a pre-estimation test, in which each proxy variable is used to individually predict enrollment (for more detail regarding the pre-estimation test see appendix B.4). Signs are generally consistent between the final model and the pre-estimation test, except

as noted in appendix B.4. We further test model robustness by examining coefficient signs across two sub-periods, from 1976 to 1990 and from 1991 to 2007. We find that coefficients are generally stable between these two periods. Detailed results of this analysis are provided in appendix B.6. Finally, we test the validity and strength of the instrumental variables used in the first stage regression of the supply model. Sargan's and Basman's chi-squared tests reveal that the instrumental variables used for tuition in the supply model are valid, as shown in tables 3.4 and 3.5 (see e.g. Hipp et al. (2011) for more on these tests). The first stage regression yields an F-statistic of only 3.97 and a p-value of 0.01. This is a weakness in the specification, but limited data availability and consistency prevents us from adding additional instruments. However, model fit appears good given the small sample size, with an adjusted R^2 of 0.33. We also find consistent coefficient signs when using models that use ordinary differences instead of log differences, and when using aggregate rather than normalized per student data. This demonstrates that coefficient signs are not driven by the normalization factors.

The final models shown in tables 3.4 and 3.5 indicate that price and credit factors play a significant role in determining supply and demand for higher education. On the demand side, preliminary models which include tuition suggest that demand rises with tuition price, consistent with findings such as Dale and Krueger (1999) and John (1992). Tuition seems to provide some indication of quality or benefit, leading consumers to respond this way to increasing prices.⁶ On the other hand, the principal component variable of credit effects (per student), composed of the underlying proxy variables household debt (per student) and student loans (per student), has a negative effect on demand. Inside the principal component, both household debt (per student)

⁶Note that net tuition price used in these models, though adjusted for institutional aid, is not adjusted for grants, meaning this variable does not necessarily represent the price actually paid by students. Grant aid is highly correlated with net tuition (correlation is 76%), and therefore falls out of the final model. Creation of a grant-adjusted net tuition variable (i.e. tuition - institutional aid - grant aid) produces a model similar to that provided in table 3.5.

and student loans (per student) have a positive sign, implying that increased student loans and increased debt reduce enrollment demand. A possible explanation for this, which is supported by current high debt and default rates, is that families are not willing to take on additional debt to pay for an education. Students must find a balance between their desire to attend expensive schooling and their aversion to debt.

We find further support for this hypothesis by splitting the tuition variable into two components. The first component, expense change, is the portion of net tuition that is perfectly correlated with the cost of operations, i.e. the HEPI index. Expense change measures the rise in tuition caused by an increase in the cost for the same basket of goods purchase by schools. The second component, offerings change, represents the portion of net tuition price change that is not correlated with the HEPI index. This component represents changes in student offerings, such as better facilities and more academic and extracurricular resources. Inserting these two variables into the demand model in table 3.5, we see that offerings changes are significant and lead to demand growth, producing an effect counter to the traditional economic effect of decreasing demand as prices rise. Expense changes do not appear to be significant, suggesting price-inelastic demand in the traditional sense; demand for education does not fall as price increases when other factors remain constant. This degree of price-inelasticity is likely made possible by government loans and other subsidies. As explained above, credit constraints may allow rising tuition to reduce demand in an indirect way. This effect would likely be amplified if government loans to credit-constrained individuals were reduced. Details of the separated demand models are provided in appendix B.2.⁷ Credit effects are the main drivers behind changes in demand, appearing as highly significant in all models tested.

We also see that credit variables and net tuition price are highly correlated. Figure 3.3 provides a visual representation of the relationship between these variables

⁷Note that we exclude expenditures from the principal component estimate of college benefit for these models, as expenditures is now included as a separate variable.

over time. Even after removing the time trend, net tuition price (per student) and household debt (per student) exhibit a correlation of 91.5%, implying that families are responding to rising prices by borrowing. This relationship, though intuitive, has not been pointed out in past literature to our knowledge. We conclude that people cannot afford increasingly expensive schools and are willing to take on only limited debt, reducing the demand in times when credit is readily available, debt levels are already high, and net tuition is also high.

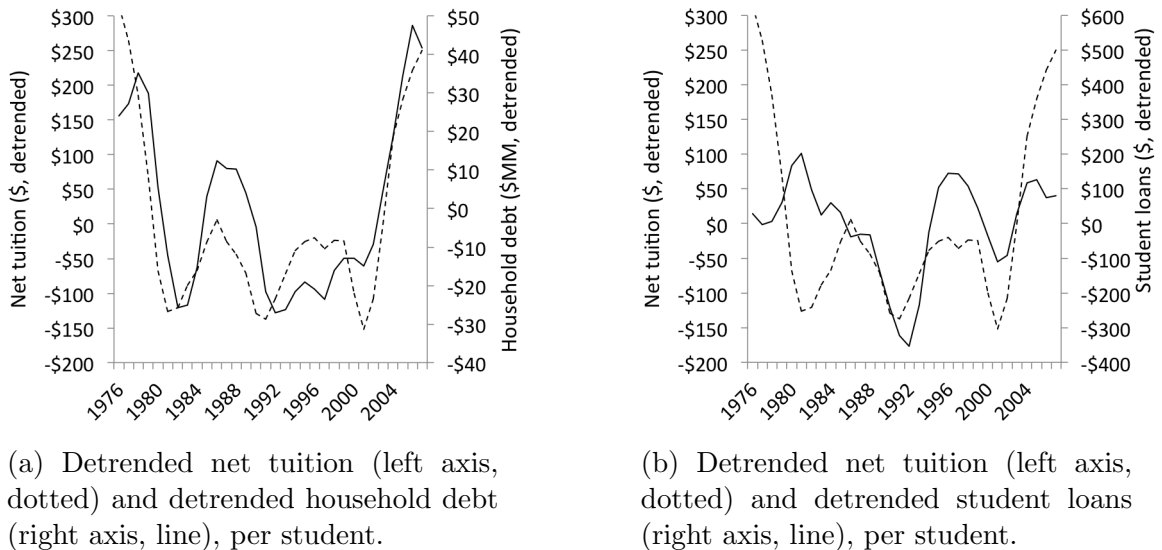


Figure 3.3: Detrended net tuition and credit variables. Detrended net tuition is the same in both plots above. For detailed variable descriptions see table 3.1.

The final significant variable in the demand model in table 3.5 is benefit of a college education. Due to the difficulty of measuring this benefit, we can say only that metrics of the benefits of college are important in predicting demand. Proxies used here include earnings difference between high school and bachelor’s graduates, spending by schools as a metric of services and facilities offered, unemployment rate, and per capita disposable income. These variables are selected because they have been identified as drivers in past research and because they provide basic metrics of the health of the overall economy (Fortin, 2006).

In the supply model shown in table 3.4, cost of operations, non-tuition revenue

(per student), aid to schools (per student), and net tuition price (per student) are significant. Cost of operations is measured via Commonfund's Higher Education Price Index, which is an inflation index for the basket of goods that institutions of higher education generally purchase. An increase in this index implies that the cost for the same offering goes up, rather than that schools increase their offerings to students by adding new facilities, programs or staff. Since supply rises with cost of operations, we conclude that schools enroll more students partly in order to cover increasing costs. A similar trend is suggested by the coefficient of non-tuition revenue (per student). As shown in table 3.1, this driver includes endowment value (per student), endowment distribution (the portion of the endowment actually spent, per student), and voluntary donations (per student).⁸ Enrollment falls with non-tuition revenue (per student), implying that schools can afford to enroll fewer students when they receive revenue from non-tuition sources. This makes sense in light of the non-profit nature of most colleges. An exception to this is government aid, which may come with stipulations about services offered or be associated with a particular group of students. However, aid directly to schools alone is not a significant driver of enrollment in the pre-estimation test (see appendix B.4). Increases in overall spending certainly increase enrollment, but our findings suggest that funding from some particular sources may have the opposite effect.

The last significant variable in the supply model in table 3.4 is net tuition price, which is adjusted for institutional aid to reflect the payment schools actually receive from students. We cannot treat net tuition price as a true predictor in the supply model because consumer demand is relatively price inelastic. Since many students are awarded grant aid, they do not pay list tuition price, meaning that schools are fairly free to set tuition prices as they see fit. However, the sensitivity to debt levels seen in

⁸Note that the endowment variables drive the non-tuition revenue principal component. Endowment data include only Harvard and Yale because of data availability and because these schools serve as models for other endowment funds.

the demand model suggests that some price sensitivity is present. Net tuition price is highly correlated with expenditures (correlation is 88%), which we omit from the supply model for this reason and because expenditures are determined by revenue for many schools due to their non-profit status (see figure 3.4). This is similar to Peña (2006), who finds that institution-level tuition price is highly correlated with school's wealth.⁹ The sign of the net tuition variable in our model suggests that higher net tuition price decreases supply, which is not intuitive. Taking into account the correlation with wealth (as in Peña (2006)) and expenditures, we can interpret the coefficient on tuition price at least partially as the effect of increased spending on new services and facilities for students. Higher quality, more expensive schools can afford to admit fewer students at higher prices, allowing schools to shrink supply. Students pay not only for an education, but for high-end athletic facilities and hundreds of student organizations. High school graduates are starting to expect these services from institutions of higher education (see e.g. Hacker and Dreifus (2010)). Our model supports the theory that schools succeed by increasing offerings and becoming more high-end and more expensive. These trends might pose a problem to students and may have contributed to the rapid price increases of the past thirty years.

3.6 Conclusion

This work models the dynamics of supply and demand in college education to provide a macro-level picture of the market. Our findings show that credit, financial aid, and pricing have a large effect on enrollment in higher education. We see that there exists a strong relationship between price, spending and wealth of schools, tuition levels, student loans, and indebtedness. We show how these relationships may lead to an upward spiral of tuition prices. Consumers desire more prestigious, more expensive education, allowing tuition price to become a signal for quality. Rankings

⁹Wealth is measured by endowment values and the gap between tuition revenue and expenditures.

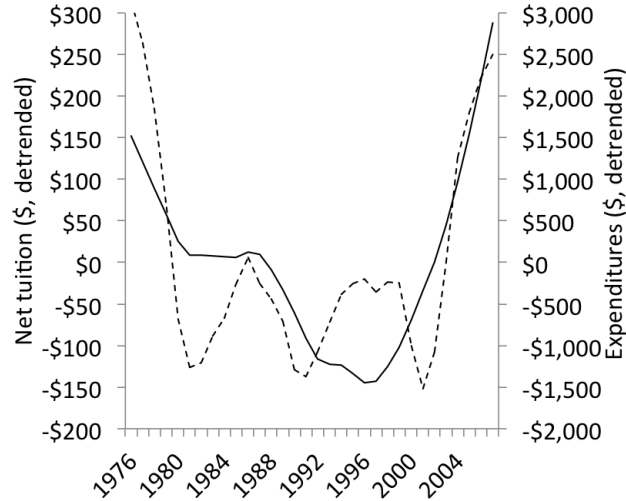


Figure 3.4: Detrended net tuition per student (left axis, dotted) and detrended expenditures per student (right axis, line). Detrended net tuition is the same as in figure 3.3. For detailed variable descriptions see table 3.1.

tied to tuition prices, facilities and services, and spending per student may amplify this effect. Based on the strong correlation between household debt levels and tuition price, we conclude that families turn to loans when faced with increasing prices. This leads to the need for a larger lending industry and, in turn, allows schools to charge more. On the other hand, credit constraints are a definite factor in demand side enrollment decisions, forcing consumers to battle with conflicting desires of attending a more expensive school and containing debt levels. The benefit of college, including spending by schools, student loan amounts, and tuition price drive demand, while school's operational costs, government aid to schools, tuition price, and non-tuition revenue shift supply. Several factors, such as increased resources and facilities for students, public perception of prestige, and/or ranking criteria linked to tuition prices, drive schools to become more expensive. Meanwhile, current financial aid structures fail to curb the resulting trend towards higher prices.

The models developed here succeed in identifying relevant drivers and their directional relationships. However, the causal relationships between highly correlated variables should be investigated further. Specifically, future research will help estab-

lish causal links between credit factors, aid, and tuition in the demand model as well as expenditures and tuition in the supply model. Changes to financial aid programs and incentive structures may allow us to decrease the risk of financial instability, ease financial strain for American families, increase the impact of government financial aid, and ensure that the US education system remains affordable, accessible, and internationally competitive.

CHAPTER IV

A major choice: An examination of educational choices, student characteristics, and ability-adjusted income

4.1 Introduction

We study post-college income across a representative sample of college applicants, and develop a model to study drivers of the decision to attend college and the choice between institutions. This analysis informs students, parents, and policymakers about the effects of educational choices on future earnings and the role these earnings play in the decision process. We introduce a new method to account for bias caused by endogenous ability differences in students. We find that post-college income levels are primarily driven by student ability and major choice, not by the quality of the school attended. However, school quality is a primary driver of students' attendance and institutional choice decisions.

Returns to schooling may depend on a variety of factors. Papers such as Brewer et al. (1999), Dale and Krueger (2011), and Cunha and Miller (2012) examine the effect of institutional characteristics on earnings outcomes. While Brewer et al. (1999) find that attending a highly selective school has a positive impact on future earnings, improved debiasing methods used by Dale and Krueger (2011) reveal that effects are

smaller than previously thought. A similar result is found by Cunha and Miller (2012), who show that large differences in earnings, graduation rate, and persistence across schools attended are greatly reduced after adjusting for student characteristics. Our analysis, using the new debiasing approach, supports the results of Dale and Krueger (2011) and Cunha and Miller (2012). A study by Hoekstra (2009) suggests that earnings gains from attending a state's flagship university over other schooling options may yield as much as a 20% increase in earnings for marginal students (those near the admissions cutoff for that school). However, this work does not include data on alternative schooling choices made by those who do not attend the flagship university or information on major choice. Our data include details about admissions success at several schools as well as major selection, capturing much more information on the ability and schooling decisions of each student. For example, we show that attending any school at all has a significant positive effect on income, so earnings differences in Hoekstra (2009) may be driven by differences in the percentage of students attending college at all. Aside from school choice, major choice is also a significant source of income disparity. Carnevale et al. (2013) show that income variation across majors is often larger than the variation across institutions attended. In this chapter, we show that major choice affects income even after adjusting for student ability differences. This is consistent with results in e.g. Arcidiacono (2004), who studies the effect of student ability on major sorting and finds that major choices are driven by personal preferences instead of the promise of monetary benefit. He also shows that major choice has an effect on earnings even after adjusting for ability sorting into different majors. Our work contributes to the school choice literature by demonstrating that major choice has a greater impact on future earnings than quality of the school attended, even though student decisions appear to be driven by quality concerns.

We construct a two-stage model of college choice, using a binary logit model for the attendance decision and an alternatives-based conditional logit model for the choice

between schools. We find that student ability and school quality of the most desirable school admitted to predict college attendance, while choice between schools is driven by financial concerns and school quality. The choice models developed in this chapter show that school quality and prestige are important factors in student's decisions about college attendance, suggesting that students and parents either believe that strong income differences exist based on quality of the school attended, or that they find quality inherently valuable. School quality drives school choice even though we do not find a significant impact of the quality of the school attended on earnings for most groups.

We use uniquely detailed data on college applications, acceptance, and attendance from the 1997 cohort of the National Longitudinal Survey of Youths (Bureau of Labor Statistics, 2013). Detailed individual-level data allow us to use information on the best school to which a student was admitted to adjust for intrinsic ability differences and identify drivers of post-college income after this adjustment. Our school quality metric is based on the selectivity of each school, meaning the percentage of applicants that are accepted for study. The selectivity metric uses data from Barron's selectivity index (Barron's, 2009). We use post-college income data reported in the National Longitudinal Survey to study income differences. Additionally, this dataset provides information on major, student and family characteristics including SAT scores, and financial aid offers which allow us to accurately capture the decisions facing each individual student.

Our model shows that student characteristics such as family background, mathematical ability, and major choice are much better predictor of future income than college attendance or school choice after removing sample bias. With the exception of attendance at an elite institution, quality of the school attended does not appear to have an effect on post-college earnings after adjusting for ability. Attending college at all is associated with an increase in earnings, but the benefit of college attendance

is not significantly different across students of different abilities. Future income is determined primarily by student ability, the decision to attend college at all, major choice, and attendance at an elite university.

We can see some of these results in raw post-college income data, which is presented in figure 4.1. Figure 4.1a portrays average income by the selectivity of the top school admitted to. We see a general downward trend in the whole sample (bar graph), with income declining as selectivity of the top school admitted to decreases. Income by the top school admitted to conditioned on the selectivity of the school attended is also displayed (lines). The effect of the attendance decision varies across groups:

- Conditional average income by school attended generally falls within the 95% confidence interval of unconditional (overall sample) average income (dotted lines).
- An exception occurs for student students who attend an elite university. For students who are admitted to a type 1 (top) school, attendance at this type 1 school appears to significantly increase post-college income (dark blue).
- Non-attenders consistently exhibit lower post-college income than attenders, with the exception of students accepted to only the lowest selectivity (type 5/6) schools (green).

Appendix C.1 provides an alternate look at this income information, displaying income levels by selectivity of the school attended, conditional on the selectivity of the top school admitted to.

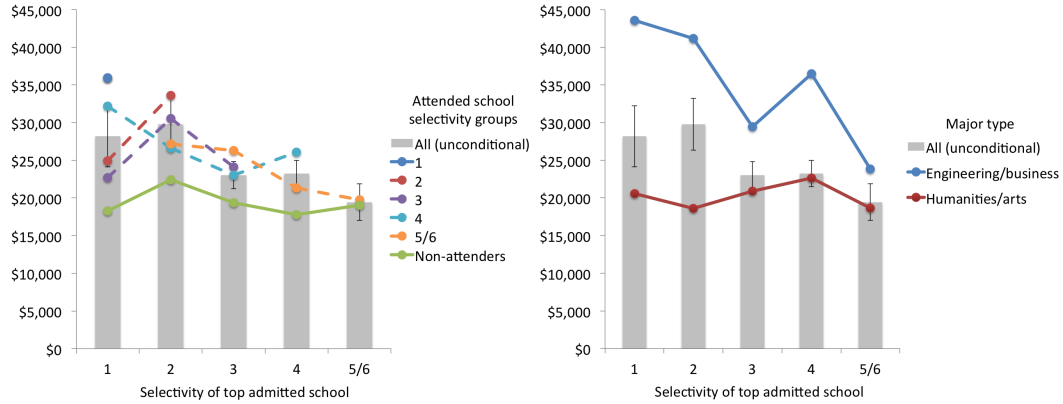
Figure 4.1b displays the same information on post-college income by top school admitted to. Lines indicate average post-college income for two major categories that exhibit above- and below-average incomes after adjusting for intrinsic ability and several other factors, as explained further in section 4.5.1. Those majoring in

business and especially engineering and computer related fields exhibit above-average post-college incomes across categories. Those majoring in social sciences, education, and fine arts have lower post-college salaries, especially among students admitted to top schools. This is true even after adjusting for student ability, suggesting that wage premiums for engineering majors are not caused only by self-selection of the best students into these programs. We continue to see this difference between majors after adjusting for mathematical ability (math SAT score), further suggesting that major choice affects salary beyond underlying student characteristics. We show in section 4.5.2 that major choice has a significant effect on income even for top students who attend elite universities. We also see similar results when using math SAT score as a proxy for student ability *instead of* the top school admitted to. However, significance levels are lower due to the fact that SAT data is only available for about one third of the students in our sample.

This chapter is divided into seven sections. Section 4.2 provides some context for this work. Section 4.3 familiarizes the reader with the input data. Sections 4.4 and 4.5 explain the new debiasing approach introduced in this chapter, the analysis of post-college income, and preliminary work on lifetime income projection. Section 4.6 presents the two-stage model of college choice. Sections 4.7 and 4.8 provide concluding remarks and future related work.

4.2 Background and literature review

Understanding the college attendance decision is important and currently relevant as college costs rise and the benefits of higher education become less obvious. Changes in the higher education and employment landscapes along with a new book by former Secretary of Education William Bennett have sparked a lively debate on the value and benefits of higher education in academic circles and the popular media (see e.g. Berger (2013), Selingo (2013), Carnevale et al. (2009)) (Bennett and Wilezol, 2013). Average



(a) Average income over the first four years after expected college graduation by selectivity of the best school admitted to (bars), with 95% confidence intervals. Lines indicate income by selectivity of school admitted to after conditioning on selectivity of school attended. Dotted lines indicate no differences from average post-college income. Attenders at elite universities have higher earnings (dark blue) while non-attenders have lower earnings (green). Type 1 indicates most selective schools.

(b) Average income over the first four years after expected college graduation by selectivity of the best school admitted to (bars), with 95% confidence intervals. Lines indicate income by major choice for business/engineering (blue, higher income) and social sciences/arts (red, lower income). Type 1 indicates most selective schools.

Figure 4.1: Average income over the first four years after expected college graduation by selectivity of top school admitted to, with effects of attendance and major decisions. ^{a, b, c, d}

^aExpected college graduation is considered to be four years after the beginning of a college education for those attending college, and four years after the start of the academic year for which students applied to college for those not attending.

^bSelectivity scores are based on Barron's selectivity index (Barron's, 2009). Types 5 and 6 are combined into a single category. Specialty schools (type 7) are omitted.

^cSelected sample is presented in table 4.1 and is derived from NLSY 1997 respondents (Bureau of Labor Statistics, 2013). Individuals included are those from the cross-sectional sample who completed high school, participated in the college choice section of the survey, were admitted to at least one four-year institution, and are not missing critical data.

^dMajor choice is based on the initial program decision made upon entering college. It does not necessarily reflect the final major in which a degree was attained, since some students change majors or do not complete their studies. Engineering/business majors include engineering, business, and computer-related fields. Social sciences/arts includes social sciences such as psychology, archeology, education, political science, sociology, and theology as well as fine arts studies.

list tuition and fees at four-year public colleges have been rising at an average annual rate of 5.2% above inflation throughout the last decade (The College Board, 2012b).

While private school tuition is growing less quickly, private institutions historically charge the highest prices. Students finance increasingly expensive education with private and federal loans; the student loan business disbursed a total of \$113.4 billion in such loans in the 2011-12 academic year (The College Board, 2012c). Private and government groups voice concerns about the quality and risk of student loan debt, and consumers are more and more reluctant to take on loans in a society that is deleveraging across the board (see e.g. Klein (2013)). With a general slow-down in the economy, salaries of bachelors degree recipients are stagnant, but degree-holders still earn considerably higher salaries than high-school graduates without a higher degree (The College Board, 2012a). Additionally, unemployment rates are much lower for those with a college or higher degree than those without, though unemployment has risen across the board. Wage increases from college attendance are especially large for minorities and women, who have historically had lower rates of college attendance (Becker and Murphy, 2007).

The college choice literature has historically focused on access to higher education, methods for increasing college attendance, and differences in college attendance rates across groups (Perna, 2006). College enrollment in the United States has increased steadily over the past twenty years, but attendance rates continue to be lower for students from less wealthy families (Mortenson, 2001). Bachelor's degree recipients earned on average 66% more in 2010 than those with only a high school degree, motivating studies focused on increased college access and better financial aid policies (The College Board, 2012a). As variance in earnings for those attending different institutions grows, researchers increasingly focus on the choice between schools in addition to the decision of whether or not to attend school at all. For example, Avery and Hoxby (2003) examine the effect of financial aid on choice between institutions, finding that students are highly sensitive to grant and loan packages. Similarly, Hurwitz (2012) takes advantage of exogenous differences between schools' financial aid

policies to study effects of institutional grant aid. Drewes and Michael (2006) find that institutional characteristics such as school size, services offered, class size, research activity, and rankings have an effect on a school's attractiveness to potential enrollees. Returns on a degree vary widely based on the school attended and major chosen, with aid-adjusted returns ranging from -15% to $+15\%$ across schools (PayScale.com, 2013). The benefits of a college education also depend on successful completion of a degree. Out of full-time college students tracked by the US government, only around 30% are recorded as graduating from the university they originally chose to attend, with many failing to complete any degree at all (The Chronicle of Higher Education, 2013). A college degree no longer guarantees a secure, well-paying job as it did a generation ago. A rise in the variance of earnings among college graduates makes college a much more risky investment than it used to be (Brown et al., 2012). Growth in the for-profit education sector, the rise of Massive Open Online Courses (MOOCs) and other alternative higher education models, and the general re-balancing of the economy threaten the status quo in the higher education market, and may change the way students make decisions about college attendance in the future.

Many methods are available for measuring the returns to schooling. Traditionally, papers on educational returns follow either the human capital or signaling schools of thought. In the human capital view, education benefits its participants by improving their skill set and making them more valuable to employers.¹ Under the signaling model, school admissions and attendance serves primarily as a signal of intrinsic student ability, making it easier for employers to identify high-value candidates. Because it is difficult to separate these effects, researchers often focus on controlled or natural experiments to distinguish between them. Hämäläinen and Uusitalo (2008), for example, take advantage of a policy change in Finland to investigate the effect of attending a more prestigious “polytechnic” instead of a vocational college. They find

¹For more on human capital models see Becker (1994) and related works.

evidence for the signaling model of education, suggesting that educational benefit on earnings is driven by the signal of ability instead of superior education taking place at the more prestigious schools. Manoli (2008) uses a change in compulsory schooling policy in the United Kingdom to compare signaling and human capital models. He finds that attainment of a diploma has a greater effect in terms of future earnings than the number of years of schooling received, again showing support for the signaling model. When natural experiments are not available, an alternative approach is to develop a sophisticated method for debiasing, adjusting for intrinsic student ability differences between groups before measuring educational returns. The current study contributes to this literature by introducing an improved debiasing method based on admissions decisions made by universities.

Some positive changes are starting to be implemented in the college decision environment. For example, there is a call for better information to enable improved decision-making and risk assessment. Government and private groups are pushing for increased emphasis on college accountability, oversight, and reporting of outcomes. One example of such a campaign is the Obama administration's introduction of the College Scorecard, which provides accepted students with information on costs, graduation rates, and default rates relative to other schools around the country (Department of Education: College Affordability and Transparency Center, 2013). User-friendly interfaces on government and private databases also provide information that may help in the college decision process (see e.g. National Center for Education Statistics (2009), Department of Education (2013), and PayScale.com (2013)). Several states, including Arkansas, Colorado, Tennessee, Texas, and Virginia, now work with College Measures, a group that compiles data on employment outcomes of graduates from different institutions. Students and parents can compare post-graduation earnings by school and major (College Measures, 2013).

The choice model presented in this chapter follows most directly from the work

of Long (2004b), who studies the changing factors influencing college attendance and choice decisions over time. Long uses an earlier cohort of the National Longitudinal Survey of Youth as well as the High School and Beyond dataset and the National Education Longitudinal Study 1988. Long presents a two-stage logistic choice model, modeling the attendance decision via a binary logit model and college choice via an alternative-based conditional logit model. She finds that college preferences are affected by price, distance to the college, and similarity to ability levels of peers at that college. Long finds that the importance of price has declined over time in the attendance decision, but that price continues to be a significant factor when deciding between schools. We compare our findings to Long’s work throughout this chapter, and find that our results are generally consistent with her conclusions. Although our dataset is too small to confirm all of Long’s results, we take advantage of richer information on each individual and potential school to test the importance of a larger number of factors. We expand upon past research and Long’s model by combining a college choice model with a focus on the effect of schooling decisions on income, by introducing a new method of adjusting for ability bias, and by focusing on drivers of income differences after making this adjustment.

4.3 Data

4.3.1 Data sources

Individual student data used in this project come from the 1997 cohort of the National Longitudinal Survey of Youth (NLSY 97) (Bureau of Labor Statistics, 2013). NLSY surveys are designed to follow a representative groups of students throughout their adult life, tracking employment, life events, health, and other areas of interest. The NLSY 1997 survey follows approximately 9,000 individuals who were born between the years of 1980 and 1984. They have been interviewed on an annual basis

since 1997, and data continue to be collected. The latest available data are from the 2010 survey. While older cohorts of the NLSY would provide more data points, especially for our evaluation of labor income, older cohorts do not include a set of questions focusing on college choice and are less relevant to today's high school graduates.

Our work draws on a subset of the NLSY 97 data, namely youths born in 1984. This subgroup of NLSY 1997 participants was selected to participate in a unique college choice portion of the survey, in which they were asked questions regarding all colleges they applied to, whether they were accepted, and whether aid was received. While all students, including those in older versions of the NLSY, are asked about the college(s) they chose to attend, the college choice dataset is unique in that it allows us to explicitly track college applications, admissions, and financial aid. Students in our cohort of interest turned eighteen years old in 2002 and are included in our sample if they applied to and were admitted to at least one college within two years of high school graduation. Demographic, residential, and other personal information about students is available within the survey data. Data that have the potential to uniquely identify students, such as geographical and school name information, were attained through a special license from the Bureau of Labor Statistics.

Data on individual students are supplemented by information on the colleges that survey respondents applied to and attended. While some school-specific information, such as financial aid awarded, is captured in the NLSY 97, most information pertaining directly to schools comes from the Department of Education's Integrated Postsecondary Education Data System (National Center for Education Statistics, 2009). Data include information on college finances, academic quality, location, and student body characteristics. To ensure consistency with information available to students at the time of their college decision, data from the 2002-2003 school year are used wherever possible.

Other sources of data include school selectivity information derived from Bar-

ron's Competitiveness Index and United States Postal Service zip code latitude and longitude information (Barron's, 2009). Barron's rankings were obtained through a restricted use license from the Department of Education. A list of variables used in model development is available in appendix C.2.

4.3.2 Sample selection

From the 9,000 individuals surveyed in the NLSY 1997, we select individuals that fit a variety of criteria making them suitable to our analysis. We are interested in the traditional post-high-school college decision process for those individuals that consider four-year institutions, and we require information on admissions for our debiasing work. Therefore, we reduce the sample as follows:

- Focus on cross-sectional sample that is designed to give a representative sample, excluding those from the Black and Hispanic oversample group. This leaves 6,748 individuals.²
- Select individuals from the 1984 cohort who completed high school. These individuals participated in the college choice section of the survey. There are 2,191 such individuals in the sample.
- Select individuals that applied and were accepted to at least one four-year institution, leaving 1,223 individuals.
- Remove individuals that apply to only international or specialty schools or where schools listed are not in IPEDS dataset. This leaves 1,121 applicants applying to a total of 914 colleges.
- Remove individuals where income data or critical regression data are missing.

We require sufficient post-college income data for the analysis outlined in section

²Oversample groups are included in the NLSY for researchers who wish to study impact of race and ethnicity. Oversampling ensures enough data points from individuals in minority groups for whom it would otherwise be difficult to establish sufficient sample sizes.

4.5. Regression inputs for key variables have very low missingness rates and do not have a significant impact on the number of usable data points.

- The final sample contains 1,053 individuals, including 796 four-year college attenders, as described in table 4.1. A small number of additional exclusions are necessary for some portions of the analysis presented here.

Table 4.1: Number of students in sample, by attendance and admission groups. Type 1 indicates most selective schools.^{a,b}

			Selectivity of school attended						
			More selective		Less selective		Non-attenders	Total	
			1	2	3	4			5,6
Best school admitted	More	1	39	21	27	18	0	16	121
	selective	2		56	44	29	7	19	155
		3			155	64	17	61	297
	Less	4				222	25	108	355
		selective	5,6					72	53
Total			39	77	226	333	121	257	1,053

^aSample is derived from NLSY 1997 respondents (Bureau of Labor Statistics, 2013). Individuals included are those from the cross-sectional sample who completed high school, participated in the college choice section of the survey, were admitted to at least one four-year institution, and are not missing critical data.

^bSelectivity scores are based on Barron's selectivity index, which measures the acceptance rate of students who have applied for study (Barron's, 2009). Types 5 and 6 are combined into a single category. Specialty schools (type 7) are omitted.

Table 4.1 gives some interesting insights into the college attendance decision. We see that it is quite common for students to attend schools below the level of the top school to which they were admitted. It is especially common to attend a school one type below one's top admitted option. Very few of the top students choose to attend schools in the least selective category. The top school admitted to is an especially common choice for those who were not admitted to the very highest types. We also see that most students fall into the middle admission categories, with a top admitted school in type 4 being the most common. Non-attendance is most common among students admitted only to lower-type schools, but occurs among all students.

4.4 Debiasing: Creating university and student groups

4.4.1 A new approach for addressing selection bias in school choice

One of the main contributions of this work is the introduction of a novel method for adjusting for endogenous student ability. Thanks to our unique data on the college application process, we can use an individual's success in college admissions to develop an understanding of his or her intrinsic ability. We condition on the Barron's selectivity index score of the best school to which a student was admitted, allowing us to reduce sample bias in our study of the effect of the attendance decision on income. School selectivity categories correspond to Barron's selectivity categories. Type 1 schools have acceptance rates of at most 33% while type 5 and 6 schools admit 100% of their applicants. We group type 5 and 6 schools (least selective) together due to the small number of observations for these school types.

Selectivity is used as the basis for forming school groups since it is common in the literature, it is available for our dataset, and it captures the detailed evaluation of student ability that is part of the college application process. An example of school groupings in the literature can be found in Fu (2012). In Fu's example, only four-year schools are addressed, which are subdivided into public and private groups. Within each of those groups, Fu creates a sub-group of selective schools and unselective (other) schools, yielding a total of four groups.³ Stratification based on public/private status and other metrics of school quality were considered in our work, but dismissed due to small sample size within each group. Adjustments of income expectations based on students' individual characteristics were not used for similar reasons.

This method of accounting for ability bias is superior to previously used methods such as debiasing by grades or standardized test scores because it captures more information and is more readily available for college choice applications. This approach

³Selective schools are those in the top thirty of US News & World Report's college rankings (U.S. News and World Report, 2012).

uses data that are common to all college applicants—information about where they applied and were admitted—instead of requiring participation in a particular exam or program that may not be offered to all potential college-goers. Additionally, college admissions staff are uniquely qualified to assess student ability given the large amount of information submitted with an application. We believe that the decisions of college admissions officers are likely to reflect intrinsic ability beyond differences in test scores or grades. Results suggesting that college attended has no effect on post-college income after adjusting for the top school admitted to are supported by alternative models where math SAT score is used as a metric of student quality.⁴ For example, figure 4.1 can be redrawn using math SAT score categories instead of top admitted school selectivity categories. We treat the acceptance decision as exogenous, assuming that students are generally admitted to schools which they are well qualified to attend. Previous work by e.g. Dillon and Smith (2013) suggests that this is a reasonable assumption. School grouping methods other than Barron’s selectivity index were also investigated and Barron’s scores are supplemented with SAT score percentiles and acceptance rates.⁵ We do not include individual characteristics such as high school grades or location of residence in our model since this reduces sample sizes within each group.⁶

There is some concern that school quality is endogenous, with school quality determined by the quality of the student body. Our metric of school quality is based on selectivity, directly linking school quality to the quality of students admitted. Such endogeneity is not problematic for our work since we model individual level decisions. Individual choices of attendance do not significantly affect the overall composition

⁴Math SAT score is a strong predictor of post-college income, as shown in table 4.3.

⁵In the sample of schools applied to by the students in this analysis, correlation between Barron’s selectivity index and the percentage of applicants admitted is 52%. Correlation between Barron’s and the student body’s 25th and 75th SAT score percentiles is -37% and -23% , respectively.

⁶Data on high school grades have high degrees of missingness. Regressions including location of residence, verbal SAT score, and top school admitted reveal that categorical location dummies and verbal SAT scores do not have a significant effect on post-college income after adjusting for top school admitted to (p-value > 0.05).

of the student body at a school. Additionally, our results support the hypothesis that school quality is based on student quality, since students admitted to selective schools perform well in terms of income even when they do not attend the selective institutions. This suggests that the education received at a selective school may not be superior to that at a different institution.

We do not explicitly model unemployment risk, graduation risk, or future tax burdens. Instead, we account for this type of uncertainty by including all individuals who begin education at a certain level in our sample, beginning with income earned in the year after an on-time graduation. Thus, unemployment and graduation risk are included since they affect the future income of each group. Currently available income data from the NLSY 97 does not capture all unemployment risk or future education decisions over the lifetime. As more data points are collected in the NLSY 97, we hope that more sophisticated income models can be developed, capturing income levels and variance over the whole lifetime.

4.4.2 Potential criticisms and shortcomings of the NLSY 1997 data

The main challenges in writing this chapter are due to shortcomings in the available data. The short time period for which NLSY 97 data are available is a concern. We are interested in studying the relationship between post-college income and schooling decisions, but we have only limited data on lifetime incomes. The latest NLSY 97 data are from 2010, and the 1984 cohort reached age eighteen (high school graduation) and age twenty-two (on-time college graduation) in 2002 and 2006, respectively. Therefore, we have only about four years of post-college income data available for those individuals who stayed in college for four years. As additional income data is collected in the NLSY 97, income estimates can be improved, allowing us to study income variance and the effect of individual characteristics more closely and over a longer period.

A further criticism may come from the fact that we examine only individuals from a single cohort. As pointed out by Davis and von Wachter (2011), individuals who graduate during a recession experience the repercussions of their inopportune entrance into the working world for the duration of their career, lowering salary expectations and utility of future income streams. Since we are primarily interested in the substitution effect between schools (e.g. what makes a person pick school A over school B), we believe that accurate modeling of the magnitude of future earnings is less important than the relationship between salaries for those attending different schools. It is not clear that graduation during a recession would significantly change the relative factors leading a student to choose one school over another. Previous cohorts of the NLSY did not include the college choice questions which provide the data used in our models. Therefore, comparative analysis cannot be performed using older cohorts.

We must make several assumptions in order to model the college choice portion of the application/acceptance/attendance game in isolation. We assume that students' application decisions reflect their ability as well as their desires about which school to attend. Throughout the model, we treat the application decision as exogenous and assume that students apply to the schools in which they are most interested. We also assume that high ability students apply to selective institutions, even though they may also apply to less selective schools, and that school's supply side decisions such as admissions and financial aid amounts are exogenous. While these assumptions may be violated in a small number of cases, they reflect the usual behavior of college applicants. Work by e.g. Dillon and Smith (2013) suggests that students generally apply to schools for which they are qualified, though less wealthy students may fail to apply to the highest quality schools given their ability levels. The number of schools applied to is not a significant predictor of post-college income in our models (p-value = 0.9626).

4.5 Modeling income

4.5.1 Post-college earnings

To understand the effect of educational choices on income, we compare raw post-college income data across students of different ability levels and educational decisions. Through this work, we build an understanding of the likely income level for a student of a particular ability who makes a specific schooling decision. Table 4.2 presents results of fixed-effects regressions estimating post-college income based on school attended and top school admitted to separately. Our results indicate that both being admitted to top schools and attending top schools is correlated with higher income levels. While school attended appears to be a more significant predictor of post-college income than school admitted to, categorization of individuals by the selectivity of school attended introduces sample bias since those students who are able to attend top schools must have been admitted to these institutions. The question we wish to answer is whether or not school attended has an effect on future income *after adjusting for intrinsic ability*, which we measure by top school admitted to. We test for the effect of admission within each attendance group separately as well as the effect of attendance within each admission group separately, but find that sample sizes in sub-groups are generally too small to draw definitive conclusions. The fixed effects factor models are presented in appendix C.4.

We are able to identify some significant pairwise differences in income levels by attendance decision within an admission category. Specifically, for students who are admitted to a type 1 school, attendance at that type 1 school yields significantly higher income than attendance at a type 2 or type 3 institution. Additionally, we show that attending college at all is better than attending no college for students whose top admitted school falls into categories 1, 3, and 4. These results are presented in appendix C.5. Salary for non-attenders is not significantly different across top

Table 4.2: Predicted post-college income by selectivity of top school admitted to and school attended.

(a) Predicted post-college income, by top school admitted to. Type 1 indicates most selective schools. $\text{avgIncome} = \alpha_0 + \alpha_1 \text{topAdmit} + \epsilon$.^{a, b, c}

Fixed effects regression model			
	Adjusted R^2 :		0.0764
	Estimate	St. Error	p-value
(Intercept)	39,457	2,122	
Type 2	1,217	2,792	0.6631
Type 3	-8,845	2,501	0.0004
Type 4	-9,898	2,41	4.59×10^{-5}
Types 5/6	-13,789	2,882	2.19×10^{-6}

^aBase case is type 1 (most selective).

^bSelectivity scores are based on Barron's selectivity index, which measures the acceptance rate of students who have applied for study (Barron's, 2009). Types 5 and 6 are combined into a single category. Specialty schools (type 7) are omitted.

^cSelected sample is presented in table 4.1 and is derived from NLSY 1997 respondents (Bureau of Labor Statistics, 2013). Individuals included are those from the cross-sectional sample who completed high school, participated in the college choice section of the survey, were admitted to at least one four-year institution, and are not missing critical data.

(b) Predicted post-college income, by school attended. Type 1 indicates most selective schools. $\text{avgIncome} = \alpha_0 + \alpha_1 \text{attend} + \epsilon$.^{a, b, c}

Fixed effects regression model			
	Adjusted R^2 :		0.1341
	Estimate	St. Error	p-value
(Intercept)	25,256	1,375	
Type 1	24,387	3,901	7.99×10^{-10}
Type 2	22,800	2,987	9.81×10^{-14}
Type 3	9,078	2,000	6.91×10^{-6}
Type 4	7,015	1,786	9.60×10^{-5}
Types 5/6	1,710	2,347	0.4670

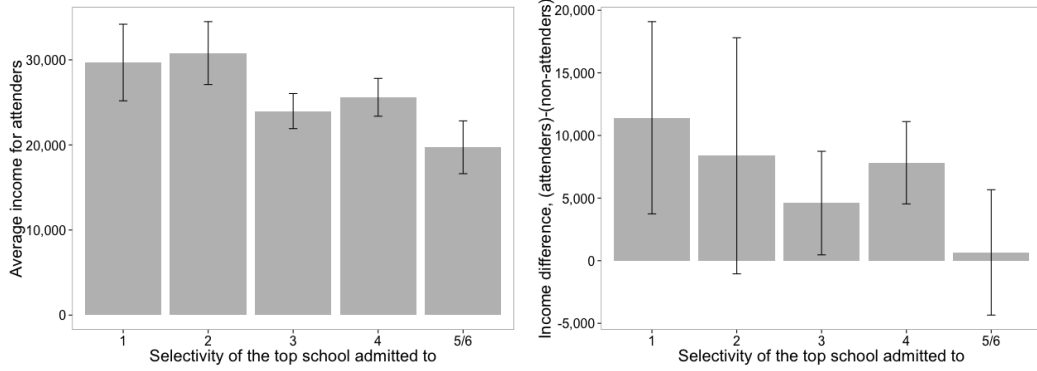
^aBase case is non-attenders.

^bSelectivity scores are based on Barron's selectivity index, which measures the acceptance rate of students who have applied for study (Barron's, 2009). Types 5 and 6 are combined into a single category. Specialty schools (type 7) are omitted.

^cSelected sample is presented in table 4.1 and is derived from NLSY 1997 respondents (Bureau of Labor Statistics, 2013). Individuals included are those from the cross-sectional sample who completed high school, participated in the college choice section of the survey, were admitted to at least one four-year institution, and are not missing critical data.

admitted groups (p-value = 0.7422). However, consistent with above results, salary of attenders differs across top admitted groups (p-value = 8.66×10^{-5}). Figure 4.2 shows these differences; income for attenders is higher for those whose top admitted school is more selective. The second panel in figure 4.2 examines whether there is a significant difference in the benefit of attending school at all across top admitted groups. We plot the difference between earnings of attenders and non-attenders for each top admitted group with 95% confidence intervals, and see that these differences are in fact indistinguishable across groups. This implies that the choice of whether or not to attend school has a similar effect on future earnings across ability groups.

Overall, we are able to show that differences exist between those who attend school



(a) Average post-college income for attenders at any college, by top admitted category. Error bars show 95% confidence intervals on the mean. Type 1 indicates most selective schools.

(b) Differences in post-college income for attenders versus non-attenders, by top admitted category. Error bars show 95% confidence intervals on the difference. Type 1 indicates most selective schools.

Figure 4.2: Average post-college earnings by selectivity of top school admitted to. No significant differences exist in average earnings for non-attenders across top admitted categories. ^{a,b,c}

^aSelectivity scores are based on Barron's selectivity index, which measures the acceptance rate of students who have applied for study (Barron's, 2009). Types 5 and 6 are combined into a single category. Specialty schools (type 7) are omitted.

^bSelected sample is presented in table 4.1 and is derived from NLSY 1997 respondents (Bureau of Labor Statistics, 2013). Individuals included are those from the cross-sectional sample who completed high school, participated in the college choice section of the survey, were admitted to at least one four-year institution, and are not missing critical data.

^cSignificant differences in lifetime earnings for attenders exist between the following top admitted groups (p-values): Type 1 and type 3 (0.0092), type 1 and type 4 (0.0601), type 1 and type 5/6 (0.0005), type 2 and type 3 (0.0007), type 2 and type 4 (0.0091), type 2 and type 5/6 (5.1×10^{-5}), type 3 and type 5/6 (0.0906), and type 4 and type 5/6 (0.0189).

at all versus those who do not pursue higher education, and that incomes are higher for those students who are admitted to an elite school and choose to attend it. Figure 4.3 summarizes these results by providing the best estimate of average post-college income for each group. Post-college income is the same for all non-attenders. Income is the same across different attendance decisions within each top admitted category except for students admitted to top institutions. For these students, attendance at the elite institution increases income.

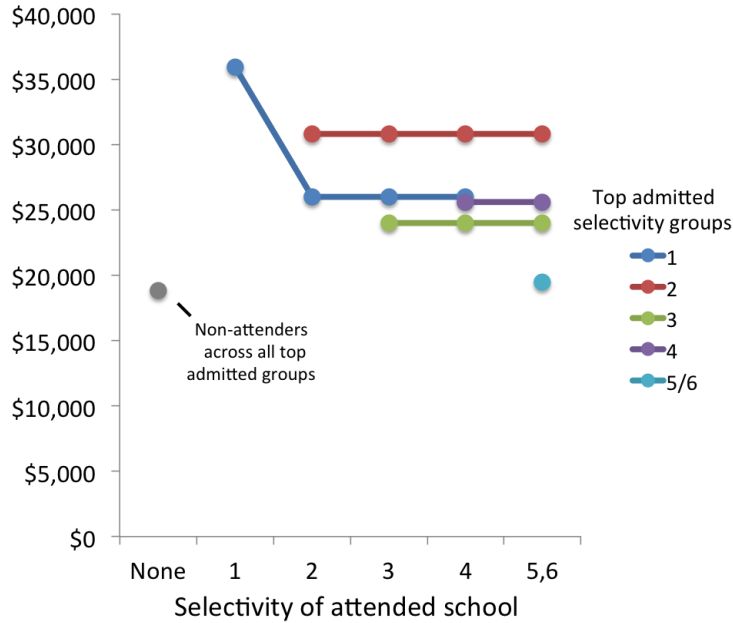


Figure 4.3: Best estimates of post-college income by selectivity of the best school admitted to and attendance decision. Type 1 indicates most selective schools.^{a, b, c, d}

^aExpected college graduation is considered to be four years after the beginning of a college education for those attending college, and four years after the start of the academic year for which students applied to college for those not attending.

^bSelectivity scores are based on Barron's selectivity index, which measures the acceptance rate of students who have applied for study (Barron's, 2009). Types 5 and 6 are combined into a single category. Specialty schools (type 7) are omitted.

^cSelected sample is presented in table 4.1 and is derived from NLSY 1997 respondents (Bureau of Labor Statistics, 2013). Individuals included are those from the cross-sectional sample who completed high school, participated in the college choice section of the survey, were admitted to at least one four-year institution, and are not missing critical data.

^dOnly significant differences shown.

4.5.2 Effect of college attendance and other factors

Selectivity of the best school admitted to is not the only important predictor of post-college income. We use the individual student data available in the NLSY 1997 to examine which other factors may affect income. As shown in table 4.3, we find that mathematical ability (measured by math SAT score), household income, and major choice have an effect on income. The top school admitted to is not a significant predictor of income after adjusting for these student and family characteristics. Students choosing majors in business and, especially, engineering exhibit consistently

higher post-college salaries across admission groups. The positive income benefit from majoring in engineering or business fields remains significant for students whose top admitted school is of type 1 (p-value = 9.43×10^{-5}) and for the subset of these individuals who attend a type 1 school (p-value = 0.0077). We also investigate the impact of gender, race, and location of residence, but find no significant effect from these. We do not include the effect of income variance across groups in this model since we do not have sufficient data to fully assess this variance. Income variance is highest among students whose top admitted school is of type 1, but this may be driven by the fact that these students have a wider range of attendance options. After adjusting for top school admitted to, there is no significant difference in income variance based on the school type attended. Income volatility should be analyzed further as more income data is collected on NLSY 1997 respondents. Even with the inclusion of student characteristics, this model does not explain a large portion of overall income variance, suggesting that many other factors matter in determining income levels.

As stated in section 4.4, the samples used in this analysis include any student that initially enrolled in a school of a given quality. Adjusting this sample to only students that successfully graduate from institutions in each category reduces the sample size, but suggests similar results. We also investigate the effect of attending the best school one was admitted to versus a lower quality school, and find no difference in expected earnings between these groups except for those admitted to type 1 schools.

4.5.3 Lifetime income

We expand on the analysis of raw income data by developing a model of the expected value of the future lifetime income stream that results from attending a school with a particular selectivity type. The goal of this analysis is to provide an illustrative example of future work that may be performed as more income data is collected from the NLSY 97. We project the shape of each individual's most likely

Table 4.3: Effect of student characteristics on post-college income. $\text{avgIncome} = \alpha_0 + \alpha_1\text{mathSAT} + \alpha_2\text{householdIncome} + \alpha_3\text{engineeringIndicator} + \alpha_4\text{humanitiesIndicator} + \epsilon$. ^{a, b}

Regression model			
	Adjusted R^2 :	0.1192	
	Estimate	St. Error	p-value
(Intercept)	10,790	3,897	
Math SAT score	2,876	879	0.0011
Household income	0.0320	0.0130	0.0142
Engineering/business indicator	10,300	2,039	6.33×10^{-7}
Humanities/arts indicator	-4,003	2,258	0.0769

^aSelected sample is presented in table 4.1 and is derived from NLSY 1997 respondents (Bureau of Labor Statistics, 2013). Individuals included are those from the cross-sectional sample who completed high school, participated in the college choice section of the survey, were admitted to at least one four-year institution, and are not missing critical data. The sample in this table excludes individuals for whom regression data is missing. Math SAT scores are missing for about one third of the students in our sample, reducing the sample size for this model. A model using top school admitted to instead of top math SAT score gives similar results with worse fit.

^bMajor choice is based on the initial program decision made upon entering college. It does not necessarily reflect the final major in which a degree was attained, since some students change majors or do not complete their studies. Engineering/business majors include engineering, business, and computer-related fields. Social sciences/arts includes social sciences such as psychology, archeology, education, political science, sociology, and theology as well as fine arts studies.

labor income curve over the lifetime. Income estimates are based on employment status, school enrollment status, and employment income reported in the NLSY 97. All reported employment income is used, so long as students claim full-time working status and report at least 12,000 hour of employment and at least \$10,000 of labor income in a given year. The salary lower bound helps reduce noise in the data caused by unusual employment situations such as unpaid internships, volunteer work, or career training. Income from years of unemployment or severe under-employment is not used in the estimation of the shape of the lifetime income curve.⁷

To extrapolate from the early career income data provided in NLSY 97 to the expected lifetime income curve for the students in our sample, we supplement NLSY 97 data with current synthetic work-life earnings projections from the US Census Bureau (Census Bureau, 2012). Synthetic earnings estimates are adjusted by level of education received (Dya and Newburger, 2002). The method described in this section allows us to gain maximal insight into future earnings by

- using all income data points available for a particular individual,
- taking advantage of information about the shape of lifetime income curves from US Census data,
- accounting for differences in the number of years worked between those who attend college and those who do not, and
- accounting for earnings over the whole lifecycle instead of taking the net present value of only the available data points.

⁷Income projection was also performed without the \$10,000 upper bound, without restrictions on full-time employment status, and allowing use of income from sources such as unemployment insurance and investments. The approach used here was chosen because it maximizes the number of individuals for whom income projection is possible. The result that income parameters do not vary across school type attended after adjusting for the top attended school is consistent across all projection approaches.

This supplements the analysis of raw data explained in the previous section, and establishes a model for future analysis to be performed as more income data is collected. Analysis of our lifetime income projections yields results similar to those found in the raw data, suggesting that student ability rather than schooling decisions drive differences in income. Unfortunately, the data requirements for the projection technique used here require us to drop a significant portion of the original sample. Analysis performed using lifetime income estimates include a sample of only 525 individuals who provide sufficient income data for projection. This sample is presented in table 4.4.

Table 4.4: Number of students in sample, by attendance and admission groups. Type 1 indicates most selective schools.^{a,b}

			Selectivity of school attended					Non-attenders	Total
			More selective		Less selective				
			1	2	3	4	5,6		
Best school admitted	More	1	15	8	11	9	0	7	50
	selective	2		15	24	16	3	8	66
		3			77	35	11	24	147
	Less	4				123	13	57	193
selective	5,6					40	29	69	
Total			15	23	112	183	67	125	525

^aSample is derived from NLSY 1997 respondents (Bureau of Labor Statistics, 2013). Individuals included are those from the cross-sectional sample who completed high school, participated in the college choice section of the survey, were admitted to at least one four-year institution, and are not missing critical data, including income data required for lifecycle income projection.

^bSelectivity scores are based on Barron's selectivity index (Barron's, 2009). Types 5 and 6 are combined into a single category. Specialty schools (type 7) are omitted.

As we reduce our sample from the complete set of NLSY 97 respondents, we ensure that the original demographics of our data are preserved. Students with very few data points are removed from the sample, and we must ensure that this does not introduce severe bias. Students with fewer post-college income data points may include high caliber students who attend graduate school or pursue unpaid learning opportunities as well as lower ability students who take longer to complete college or have very low paying jobs. As shown in table 4.5, we find that the rate of data deletion is similar

across attendance and admission groups.

Table 4.5: Percentage of selected NLSY 1997 sample used in analysis, by attendance and admission groups. Individuals removed from sample do not report sufficient income data points. Type 1 indicates most selective schools.^{a,b}

			Selectivity of school attended					Non-attenders	Total
			More selective		Less selective				
			1	2	3	4	5,6		
Best school admitted	More selective	1	35%	36%	41%	50%	NA	44%	40%
		2		25%	47%	53%	43%	38%	39%
		3			48%	52%	58%	38%	48%
	Less selective	4				51%	50%	50%	51%
		5,6					51%	55%	53%
Total			35%	28%	47%	52%	52%	46%	46%

^aSelected sample is presented in table 4.1 and is derived from NLSY 1997 respondents (Bureau of Labor Statistics, 2013). Individuals included are those from the cross-sectional sample who completed high school, participated in the college choice section of the survey, were admitted to at least one four-year institution, and are not missing critical data.

^bSelectivity scores are based on Barron's selectivity index (Barron's, 2009). Types 5 and 6 are combined into a single category. Specialty schools (type 7) are omitted.

Lifetime income estimates are created by fitting a restricted Nelson-Siegel equation to synthetic lifetime earnings curves created by the US Census Bureau.⁸ We then calibrate the model, estimating a single shape parameter for each individual in our sample that captures the likely shape of his or her lifetime income curve. We study differences in income across student ability and school quality groups by comparing these shape parameters and the lifetime income they imply. Details of this analysis are provided in appendix C.3, and results confirm the hypothesis that the selectivity of the school attended does not affect income after adjusting for selectivity of the top school admitted to. As more data is collected from the NLSY 97 survey, more sophisticated income projection methods may allow for further study of the effects of lifetime income levels and variance.

⁸See Nelson and Siegel (1987). The Nelson-Siegel equation is commonly used to model yield curves.

4.6 Drivers of college choice

We fail to show that there is a difference in post-college income based on selectivity of the school attended after adjusting for the selectivity level of the top school a student is admitted to. This result from the previous section implies that we cannot test for the effect of future income on the schooling decision in general, since the schooling decision has no sizable impact on future income. We can test for the effect of elite institution status, since students attending elite institutions exhibit higher salaries. In this section, we test which other factors affect the way students make decisions about whether and where to attend college. We see that concerns about school quality and selectivity as well as monetary interests are key drivers of the decision, even though school quality does not appear to predict post-college earnings in most cases. School quality is important to students outside of elite school status, and elite status alone does not appear to affect schooling decisions.

Candidate drivers consist of individual, school, and interaction characteristics. School characteristics include list tuition, financial aid including in-state residency discounts, instructional spending per student, public versus private status, financial aid received by the student body, school size, and the student body gender ratio. Individual characteristics include student gender, ethnicity, parents' educational attainment, household size and income, and high school type. Interaction terms compare the most preferred school's selectivity to that of the student's top admitted school, examine urban versus rural status of the student's home and school, and capture distance between home and the college campus.

The college decision is broken down into two components: First, we use a binary logistic model to capture the decision of whether or not to attend college at all. Second, we use a conditional logistic model to determine factors important in choosing between colleges for those students that in fact do attend school.⁹ Additional

⁹We use R to construct all models. First stage binary logit models are estimated using the glm

background on binary and conditional logistic choice models is provided in appendix C.6.

4.6.1 Choice of attendance: The first stage binary logit model

The first decision in the college choice process is the decision of whether or not to attend college at all. The first stage input dataset contains one entry per student which stores individual characteristics, information about the most preferred school admitted to, and interaction terms quantifying the fit of the student at this school. For students attending college, the most preferred school admitted to is the school attended. For students that do not attend school, we use the predicted utility of each candidate school from the second stage model to determine which school is the most preferred (see table 4.7 below for the second stage model). The second stage model is estimated first, allowing us to derive an equation for the utility of attendance at a particular school for a particular person. Once this equation is known, we can determine which of the possible schooling options facing each non-attender would have provided them with the highest utility. This school is used as the most preferred school for non-attenders in the first stage model. This method assumes that non-attenders have similar preferences over school characteristics as attenders, even though not attending is the highest utility option for non-attenders.

In the first stage of the model, we are able to include both individual and school characteristics, since each choice situation is defined by a particular student and his or her most preferred school. Individual student data include student ability measured by the selectivity level of the top school admitted to, gender, ethnicity, parents' education, household size, and wealth. We convert the categorical data on the selectivity of the top admitted school to a continuous variable by replacing each Barron's category with the acceptance rate with which it is associated. For example,

package, while second stage conditional choice models are constructed using the `mlogit` package and `clgit` function from the `survival` package (R Core Team, 2013; Therneau, 2013; Elff, 2013).

a type 1 school is measured as having an acceptance rate of 33%. This reduces the impact of small sample size by reducing the number of parameters that must be estimated. We call this continuous version of admitted school selectivity “top admitted school acceptance rate.” We do not include the expected earnings premium associated with college attendance in our model, since we have shown in the previous section that this is the same across students. Previous work suggests that differences in the expected college wage premiums would have an effect on enrollment probability if they were significant, as in e.g. Dillon (2012) who shows that changes in the college premium over time have an effect on enrollment probability. We also test for the significance of higher order terms and likely interactions between variables.

The probability of attending college at all given that the most preferred school is j^* is given by the binary logistic model below:

$$\begin{aligned}
 P_{i,j^*=1} &= \frac{e^{\delta'x_{i,j^*}}}{e^{\delta'x_{i,j^*}} + 1} & (4.1) \\
 &= \frac{1}{1 + e^{-\delta'x_{i,j^*}}} \\
 &= \frac{1}{1 + e^{-\delta'(\text{topAdmit}_{j^*} + \text{graduationRate}_{j^*})}}
 \end{aligned}$$

Model coefficients are provided in table 4.6.

We find that student ability (the top school admitted to) is a strong predictor of college attendance, with students admitted to schools with low acceptance rates (high selectivity) being more likely to attend college. After conditioning on student ability, only a few variables provide additional explanatory power. Graduation rate is highly correlated with other metrics of school quality, including selectivity, list tuition, average faculty salary, and private versus public control status. While it is not clear that students making the college decision focus explicitly on the graduation rate of the school they are considering, the significance of this variable captures students’ assessment of the overall quality of the institution. Graduation rate is

Table 4.6: First stage model: Binary logit model of whether to attend college or not. Equation (4.1).^a

Logit regression model			
	McFadden's R^2 :	0.0521 ^b	
	AIC:	1,049.4 ^c	
	Estimate	St. Error	p-value
(Intercept)	-1.2100	1.0900	
Top admitted school acceptance rate ^d	-0.0180	0.0060	0.0028
Graduation rate ^e	2.0543	0.5835	0.0004

^aSelected sample is presented in table 4.1 and is derived from NLSY 1997 respondents (Bureau of Labor Statistics, 2013). Individuals included are those from the cross-sectional sample who completed high school, participated in the college choice section of the survey, were admitted to at least one four-year institution, and are not missing critical data.

^bMcFadden's $R^2 = 1 - \frac{\ln(\hat{L}_{full})}{\ln(\hat{L}_{intercept})}$ where \hat{L}_{full} gives the estimated likelihood of the model with predictors and $\hat{L}_{intercept}$ gives the estimated likelihood of the intercept-only model.

^cAkaike Information Criterion, see e.g. Akaike (1969).

^dTo limit the effects of small sample size, we change the categorical data on attendance and admission to interval data based on the admission criteria of schools within each category. Specifically, using documentation on Barron's selectivity index, we categorize selectivity groups based on the acceptance rate of applicants. Type 1 schools have acceptance rates of at most 33% while type 5 and 6 schools admit 100% of their applicants.

^eGraduation rate as reported in IPEDS. Graduation rate is highly correlated with other measures of school quality, such as faculty salary, expenditures, and selectivity.

treated here as an exogenous characteristic of the school in question, even though quality of student body is clearly determined by institutional policies. A study of the supply-side decisions that affect graduation rate is outside of the scope of this work, and we treat it as a set characteristic of each institution. The high level of significance of this variable suggests that the likelihood of attendance is affected by student ability and the quality of the top school they were admitted to even within the top admitted groups defined using Barron's selectivity index. As shown in the next section, students tend to prefer schools that perform highly on quality metrics relative to other schools in their choice set. In this model, we see that having such a high quality school available as a possible choice makes it more likely for a student to attend school. Gender was also a significant predictor of attendance in preliminary models, with females being slightly more likely to attend college than males.

Overall, the model in table 4.6 does not explain very much of the variance in the college attendance decision, and does not successfully predict whether or not a student will attend school. Student and school quality appear to be the primary drivers of attendance, and other variables fail to provide additional information on attendance probability after adjusting for these factors.

This model closely parallels the work of Long (2004b) and allows us to update Long's models using a more recent cohort of the NLSY. Comparison to Long's work provides some context for our study and lets us verify consistency with the literature. Long studies changes in the factors affecting college attendance and the choice between schools over time, studying data from the high school graduating classes of 1972, 1982, and 1992. Our work adds a new set of models for the high school graduating class of circa 2002. Comparing our first stage model to the attendance decision model for 1992 graduates in Long (2004b), we find that our model provides more significant predictors of college attendance. Additionally, the variables which predict attendance in our model are not candidate variables in Long's work. Attempting to recreate Long's

model with our data yields models with largely insignificant parameters. Where significant, signs are generally consistent with Long’s original model. These results are presented in appendix C.7. Both Long’s original models and the model presented in this section fail to capture the majority of variance of the binary college attendance decision. Long finds that tuition price became less important in the attendance decision between 1972 and 1992. Our analysis confirms these findings, as tuition is not a significant predictor of attendance in the first stage model in table 4.6.

4.6.2 Choice between colleges: The second stage conditional logit model

The second stage of our model captures the choice between colleges *given that an individual has decided to attend school at all*. In this model, alternatives are defined entirely by school characteristics, allowing us to include specific details about each of the colleges represented in our sample. School characteristics include price information, financial aid availability, student body characteristics, school wealth, and school quality metrics including an indicator for elite (type 1) school status. The data include one sample point for each student/school combination. Data are stratified by student, with each student’s stratum including one set of data per alternative school considered. The likelihood of selecting a particular school is calculated relative to the other schools in a specific student’s choice group, meaning that we cannot account for individual characteristics in this portion of the model. This means that we cannot adjust explicitly for student ability (top school admitted to), since this parameter is constant across a choice situation (for a given student). However, we can include student ability information through the use of interaction effects. Selectivity difference captures the difference between the selectivity level of the top school to which a student was admitted and the selectivity of the current candidate school, providing a metric of whether a school is a “stretch” or “safety” choice for the particular stu-

dent.¹⁰ We also test for the significance of an elite school indicator, since we show that attendance at elite (type 1) schools provides a unique income benefit. Other interaction terms we consider include distance between a student’s home and the college and the proportion of tuition that is covered by financial aid. It is this portion of the model that makes significant use of the Independence of Irrelevant Alternatives (IIA) property, which states that the relative utility of schools in the choice set does not change mathematically if some alternative schools fail to be included in the sample. Additional details on the IIA property and logistic choice models are provided in appendix C.6.

Since our data on school characteristics are very rich, we begin by examining collinearity between school characteristics that may measure similar traits. As mentioned in the previous section, graduation rate is highly correlated with other metrics of school quality. Our sample also includes several variables that measure expenditures, tuition and net cost, and federal aid access. Given the interrelatedness of some variables, we develop a robust method of model selection for the second stage model. We investigate candidate models that include all combinations of variables predicting tuition and net cost, selectivity, expenditures and wealth, and school quality, selecting the combination of variables that yields the best fit. Appendix C.8 provides more details on pre-estimation analysis of the relationships between predictors and the model selection process. Models using alternate configurations of related predictor variables (e.g. list tuition and aid amounts instead of net tuition) generate similar results. We also investigate the use of higher order and interaction terms.

We find that net tuition and school quality as measured by graduation rate predict the likelihood of choosing a particular institution. Net tuition is calculated by subtracting financial aid from list out-of-state tuition. Financial aid includes the in-state discount, if it applies, financial aid awarded by the school, and financial aid awarded

¹⁰Selectivity difference = selectivity of this schooling option - selectivity of the top school admitted to.

by other sources for use at this school or any school attended. Students prefer less expensive schooling. Alternative models that include both list tuition and financial aid amounts also support this hypothesis. We do not find any evidence that higher list tuition with more aid is preferred to lower list tuition. Graduation rate is a metric of school quality, and is highly correlated with related variables such as faculty salary, instructional spending, and selectivity. The significance of this set of variables implies that students prefer the highest quality schools in their choice portfolio. An indicator for elite (type 1) status of an institution is not significant after including graduation rate, suggesting that the preference for higher quality institutions is not driven by the income benefit associated with attendance at elite institutions. The preference for higher quality institutions exists across top admitted groups, even for students who should not expect an income benefit from attendance at a higher quality institution. The model equation is given below:

$$\begin{aligned}
 P_{i,j} &= \frac{e^{\delta'x_{i,j}}}{\sum_{k \in J} e^{\delta'x_{i,j}}} & (4.2) \\
 &= \frac{e^{\delta_1 \text{netTuition}_j + \delta_2 \text{netTuition}_j^2 + \delta_3 \text{graduationRate}_j}}{\sum_{k \in J} e^{\delta_1 \text{netTuition}_j + \delta_2 \text{netTuition}_j^2 + \delta_3 \text{graduationRate}_j}}
 \end{aligned}$$

Model coefficients are provided in table 4.7. As described in the previous section, this model is also used to predict which of their candidate schools each non-attender would most prefer. This information is used as an input in the first stage model.

The second stage model has more explanatory power than the first stage model, but still fails to capture all of the variance in the choice between schools. As with the first stage model, we see that school quality plays a key role, with students picking higher quality institutions within their choice sets. This is countered by a desire to pay a lower tuition price, and decisions are likely a trade-off between these two main drivers. The first stage model shows that higher ability students are more likely to

Table 4.7: Second stage model: Conditional logit model of college choice given college attendance. Equation (4.2).^a

Conditional logit regression			
	McFadden's R^2 :		0.0710 ^b
	Estimate	St. Error	p-value
Net tuition ^c	-6.59x10 ⁻⁵	1.07x10 ⁻⁵	6.11x10 ⁻¹⁰
(Net tuition) ²	-9.11x10 ⁻¹⁰	3.17x10 ⁻¹⁰	4.06x10 ⁻³
Graduation rate ^d	2.52	0.566	8.40x10 ⁻⁶

^aSelected sample is presented in table 4.1 and is derived from NLSY 1997 respondents (Bureau of Labor Statistics, 2013). Individuals included are those from the cross-sectional sample who completed high school, participated in the college choice section of the survey, were admitted to at least one four-year institution, and are not missing critical data.

^bMcFadden's $R^2 = 1 - \frac{\ln(\hat{L}_{full})}{\ln(\hat{L}_{intercept})}$ where \hat{L}_{full} gives the estimated likelihood of the model with predictors and $\hat{L}_{intercept}$ gives the estimated likelihood of the intercept-only model.

^cNet tuition is calculated as list out-of-state tuition price minus financial aid. Financial aid includes financial aid awarded for attending the particular school in question, financial aid for use at all schools, and the in-state tuition discount.

^dGraduation rate as reported in IPEDS. Graduation rate is highly correlated with other measures of school quality, such as faculty salary, expenditures, and selectivity.

attend school. In the second stage model, we see that students tend to select the higher quality institutions within their choice set across student ability categories.

As with the first stage model, we compare our model to that presented for 1992 graduates in Long (2004b). The model in table 4.7 does not fit the data as well as Long's model, likely due to the smaller size of our sample. However, we do find that related variables have signs consistent with Long's model. Results from an attempt to recreate Long's original model with our data are presented in appendix C.7.

4.7 Conclusions

The models presented in this work provide evidence for the hypothesis that post-college labor income is driven primarily by individual characteristics and major rather than institutional choice. The unique data available through the college choice interview section of the NLSY 1997 allow us to study income differences using a new debiasing method. The method of debiasing introduced in this chapter accounts for

differences in ability levels across groups of students attending schools of varying selectivity by conditioning on the selectivity of the best school an individual was admitted to. With the exception of the decision to attend college at all, attendance at an elite (type 1) institution, and major choice, we fail to show significant differences in income based on schooling decisions. Earnings expectations are improved by the decision to attend college at all. Earnings for those who do not attend any college are the same across top admitted categories, while earnings for attenders are highest for those admitted to top schools. The difference in salary between attenders versus non-attenders is the same across top admitted categories. Engineering, business, and computer-related majors offer a wage premium while humanities and arts majors exhibit lower salaries even after adjusting for ability. This suggests that wage increases are not simply due to students of different abilities self-selecting into these majors.

Even though school quality does not appear to have a strong effect on post-college income for most groups, it is a primary driver of school choice. The decision to attend school at all is driven by student quality (measured by the top school admitted to) and the quality of the most-preferred (top choice) school to which a student was admitted (measured by graduation rate). Increased cost of schooling decreases the likelihood of selecting a particular school, while school quality (measured by graduation rate) increases the likelihood of attending that school. School quality is therefore important in both portions of the college decision process (attendance and choice between schools), even though there is no obvious income benefit related to quality outside of students choosing to attend an elite (type 1) institution. The schooling decision is a tradeoff between desires for attendance at a more selective and a cheaper institution, yet the benefit of paying for increased selectivity is unclear.

4.8 Future work

The primary challenges in the development of the models presented here are small sample size and the lack of a long time series of income information for each student. As the NLSY 1997 continues to be administered and more interview data are collected, our model can be updated to reflect additional income information. As NLSY 1997 respondents continue to report income data, educational outcomes, and attained degrees, projection methods similar to the one presented in section 4.5.3 could be used to create a model that focuses on income across the whole lifetime, modeling difference in income curve shape as well as income variance between admission and attendance groups. Additionally, income differences based on institutional choices may exist for particular subsets of the population (such as a particular major). These differences could be included in the choice models, modeling schooling decisions as selection between future income streams.

Recently, several groups have suggested the use of sophisticated risk-adjusted lifecycle choice models for evaluating returns to higher education. In a lifecycle choice model, employment and savings decisions are made over time, and the expected value and variance of future income is used to determine the worth of income streams resulting from a particular educational choice. The lifecycle choice approach can be applied to the school choice problem by viewing school choice as selection between different risky income streams and estimating empirically the income dynamics for each possible choice. As more data are collected as part of NLSY 1997, it will become possible to estimate the necessary dynamics of realized and hypothetical future income streams for each individual and each educational choice they face.

The lifecycle choice model was introduced in two seminal papers by Robert C. Merton (Merton, 1969, 1971). Individuals face a risky future income stream and must make savings and consumption decisions so as to maximize their overall lifetime consumption. More complex version of Merton's models include multiple hedging

assets, incomplete markets, multiple sources of income uncertainty, and a variety of other adjustments that must in many cases be solved numerically or iteratively (see e.g. Gourinchas and Parker (1999), Henderson (2005), and Carroll (2004)). Early papers applying these lifecycle choice models to questions in higher education include work by Brown et al. (2012) and Bhuller et al. (2011). Brown et al. (2012) show that proper adjustment for income risk over the lifetime reduces estimates of the benefit of college education. Bhuller et al. (2011) model differences in the lifetime earnings curve based on different levels of education, pointing out that lifecycle bias arises if post-graduation income is used as a proxy for all future earnings. Cocco et al. (1998) study the effect of human capital and education on investment and savings decisions throughout the lifecycle, examining the impact of credit constraints and labor income uncertainty. As new data become available, the current model will benefit from increased sample size and more information about earnings differences between groups.

A second natural extension of this work is a more in-depth study of non-income drivers of the college decision. Since we find that income is not significantly affected by school choice except for students choosing to attend elite colleges, it is somewhat surprising to see that school quality metrics such as graduation rate and faculty salary are the main drivers of institutional choice even for those not facing the choice of attendance at an elite institution. Future research should investigate whether these are indeed the true drivers of students' decisions, or whether a perceived boost in income from attendance at a better institution is behind these choices.

CHAPTER V

Conclusion

This dissertation tackles three important questions related to higher education markets and the role of monetary factors and financial aid in these markets. Chapter II provides context for a more thorough examination of US higher education by studying the role of public and private funding in twelve countries in North America and Europe. Chapter III looks at the effect of financial factors on the supply and demand balance in the US higher education market, studying drivers of enrollment on both sides. Finally, chapter IV provides a glimpse into the individual-level college choice process, helping us understand whether and how financial considerations play into student decisions.

The US higher education market function quite differently from most Western European systems. In the United States, most institutions of higher education expect students to pay tuition and fees with private funds, though aid is often available. In many European countries, government subsidies have eliminated or severely reduced tuition and fees for tertiary studies, with students and families responsible for only small enrollment fees, living expenses, and specialty program charges. Throughout the Western world, governments are interested in providing high quality education to as many qualified students as possible. In addition to developing a well-educated workforce, high quality universities attract foreign talent, funding, and businesses.

The study presented in chapter II examines drivers of higher education enrollment, and finds that overall wealth (measured by GDP) is the primary driver of enrollment changes. Especially in countries like the United States, where private funds form the basis for educational spending, GDP increases are translated quite directly into increased enrollment at institutions of higher education. We do not see significant differences in the effect of GDP on enrollments as a result of recent European reforms, most of which lead to increased privatization and autonomy in the higher education market.

The United States government has reacted to the problem of ever increasing tuition prices by vastly expanding its federal financial aid programs. Private aid has also grown, and individual institutions are stepping up their support through scholarships, tuition waivers, work study, and discounts. Chapter III presents a macro-level analysis of supply and demand in the changing higher education market from 1976 through 2007. We find that financial factors are important on both the supply and demand sides. Revenue and cost concerns appear to drive supply-side (institutional) decisions, while students focus on post-college benefit of higher education and debt levels. We find support for the hypothesis that schools are incentivised to continually increase spending, though increased debt aversion amongst consumers may lead to increased price sensitivity.

Concerns about college affordability are becoming more pervasive in individual decisions about higher education. In chapter IV, we study the factors driving the college attendance and school choice decisions, with a focus on the effect of future income. We find that affordability is a primary decision driver, with college quality providing a significant counter-force to the desire to attend an affordable institution. However, we do not see that attendance at a higher quality institution translates directly into higher post-college earnings except for students who attend top schools. Student ability and major choice appear to drive post-college income even after ad-

justing for ability differences, suggesting that the choice of major has a bigger effect on income expectations than the choice of school for most students. As more income data are collected on the individuals studied in this chapter, increasingly sophisticated models of income expectations over the lifetime will allow us to develop further insights into the effect of post-college payoffs on the attendance and choice decisions.

APPENDICES

APPENDIX A

Appendices for Chapter II: Higher education funding in the Western world

A.1 Starting point for alternate models

This appendix provides two alternate baseline models considered during analysis. These models use population (table A.1) and spending (table A.2) instead of GDP to construct the baseline relationship between country size and enrollment. Models developed from these starting points produce results very similar to those presented in the main analysis of chapter II. Note that the portion of both spending and population that is correlated with GDP provides most of the predictive power.

Table A.1: Effect of population correlated and uncorrelated with GDP on new enrollment. $\text{newEnrollment}_{i,t} = \beta_0 \text{uncorrelatedPopulation}_{i,t} + \beta_1 \text{correlatedPopulation}_{i,t}$.^{a,b}

Ordinary least squares regression			
Adjusted R ²	0.3352		
Variable ^c	Coefficient value	St. error	p-value
(Intercept)	2905.1	2506.7	
Population uncorr. w. GDP	-916.8	664.3	0.17
Population corr. w. GDP	18087.5	2262.3	6.38x10 ⁻¹³

^aPopulation is broken out into components that are correlated and uncorrelated with GDP. The portion of population correlated with GDP measures increases in wealth associated with population growth.

^bData include yearly GDP, population, and enrollment figures for countries listed in table 2.1 from 1997 to 2008.

^cFor detailed variable descriptions see table 2.2.

Table A.2: Effect of spending on tertiary education correlated and uncorrelated with GDP on new enrollment. $\text{newEnrollment}_{i,t} = \beta_0 \text{uncorrelatedSpending}_{i,t} + \beta_1 \text{correlatedSpending}_{i,t}$.^{a,b}

Ordinary least squares regression			
Adjusted R ²	0.3032		
Variable ^c	Coefficient value	St. error	p-value
(Intercept)	1058.2	2680.9	
Spending uncorr. w. GDP	-2013.2	809.6	0.01
Spending corr. w. GDP	7643.4	997.1	3.77x10 ⁻¹²

^aSpending is broken out into components that are correlated and uncorrelated with GDP. The portion of spending correlated with GDP measures increases in tertiary education spending that are related to population growth.

^bData include yearly GDP, population, and enrollment figures for countries listed in table 2.1 from 1997 to 2008.

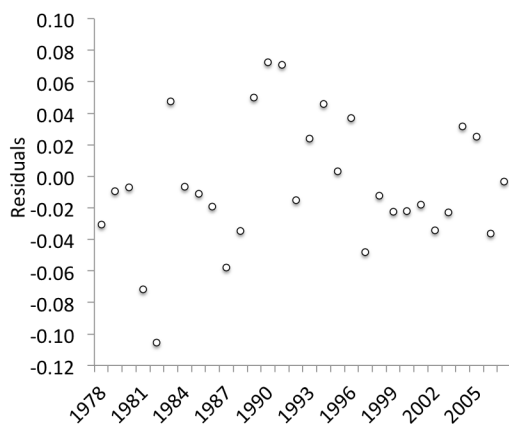
^cFor detailed variable descriptions see table 2.2.

APPENDIX B

Appendices for Chapter III: The effect of credit growth and financial aid on college tuition and fees

B.1 Final model residual plots

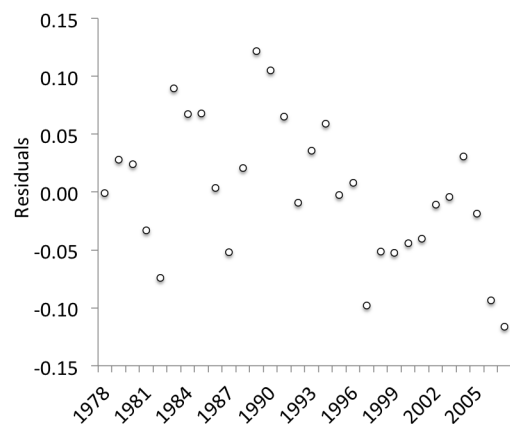
This appendix provides plots of residuals over time for the final models in tables 3.4 and 3.5 that make both the two-stage least squares adjustment for endogeneity of net tuition price and the AR(1) autoregressive error term adjustment for heteroskedasticity. These plots do not show the second stage model residuals Y and W but rather the white noise terms from the autoregressive equation for estimation of residuals, e_D and e_S , as seen in equations (3.2) and (3.3). Heteroskedasticity is dramatically reduced, with errors appearing to be white noise.



(a) Supply model.^{a,b}

^aModel is provided in table 3.4.

^bThis plots does not show the second stage supply model residuals Y but rather the white noise term from the autoregressive equation for estimation of residuals, e_S , as seen in equation (3.2).



(b) Demand model.^{a,b}

^aModel is provided in table 3.5.

^bThis plot does not show the second stage demand model residuals W but rather the white noise term from the autoregressive equation for estimation of residuals, e_D , as seen in equation (3.3).

Figure B.1: Residuals over time, final models in tables 3.4 and 3.5.

B.2 Demand model with offerings change tuition component

In the model below, we examine the effect of the component of net tuition that is uncorrelated with the cost of operations (HEPI index). We refer to this as offerings change, since it measures tuition fluctuations caused by changes in offerings to students, meaning changes in service level or quality of the educational experience. We see that new offerings increase demand. The component of tuition perfectly correlated with the cost of operations appears to have no effect.

Table B.1: Demand model, offerings change only (enrollment per high school graduate).^{a, b}

Two-stage least squares regression			
Correlation (model estimates vs. actual)	0.56		
Exclusion restrictions	Test statistic	p-value	
Sargan	4.8347	0.0892	
Basmann	4.6194	0.0993	
Variable	Coefficient value	St. error	p-value
(Intercept)	0.0031	0.0173	
Benefit of college	0.0230	0.0201	0.2624
Credit effects per student	-0.1478	0.0489	0.0055
Offerings change per student ^c	0.0022	0.0012	0.0640

^aModel equation is given in equation (3.2), with net tuition per student replaced by offerings change per student.

^bFor detailed variable descriptions see table 3.1.

^cOfferings change refers to the portion of net tuition that is uncorrelated with the cost of operations of schools, as measured by the HEPI index. This measures changes in tuition caused by changes in offerings to students instead of price increases for services already being offered.

B.3 Main dataset

This appendix provides the input dataset for all models described in chapter III. Note that not all variables are included in the final models. The data provided here are extrapolated from publicly available data as indicated in footnotes. These data points have not been detrended according to equation (3.1).

Table B.2: Main dataset

Category		Quantity	Cost of operations	Non-tuition revenue			
Variable	Year	Enrollment at all institutions (all years)/high school graduate ^a	HEPI index value ^b	Current fund expenditures per student ^{c,d}	Yale/Harvard endowment market value per student ^{e,f}	Yale endowment spending per student ^{g,h}	Donations (voluntary support) per student ⁱ
Unit			Index value in 1983=100	\$	Thousand \$	Thousand \$	Thousand \$
	1976	2.645	59.65	8,388.06	109.18	5.850	0.548
	1977	2.680	63.60	8,410.72	110.34	5.362	0.580
	1978	2.681	68.10	8,433.38	107.04	4.942	0.610
	1979	2.763	74.00	8,456.04	101.41	4.486	0.598
	1980	2.909	81.65	8,478.70	98.13	3.979	0.557
	1981	2.997	89.85	8,655.86	98.97	3.663	0.555
	1982	3.091	96.95	8,987.53	95.56	3.573	0.591
	1983	3.242	102.40	9,319.20	109.52	3.631	0.637
	1984	3.288	107.80	9,650.86	118.98	3.551	0.693
	1985	3.362	113.55	9,982.53	122.93	3.630	0.736
	1986	3.397	118.60	10,385.20	151.39	3.970	0.773
	1987	3.377	123.55	10,699.16	177.90	4.300	0.786
	1988	3.431	129.50	10,855.76	183.87	4.827	0.786
	1989	3.668	136.80	10,957.11	182.90	5.312	0.772
	1990	3.929	144.50	11,021.73	186.48	5.596	0.760
	1991	4.168	150.85	11,054.75	183.60	6.017	0.745
	1992	4.210	155.70	11,146.74	184.41	6.483	0.762
	1993	4.187	160.60	11,422.95	197.98	7.016	0.788
	1994	4.153	165.70	11,750.52	213.88	7.711	0.811
	1995	4.103	170.55	11,991.40	231.88	8.298	0.857
	1996	4.086	175.70	12,217.89	269.77	8.984	0.920
	1997	3.993	181.55	12,581.53	324.67	9.984	1.010
	1998	3.917	186.90	13,097.61	381.26	11.325	1.112
	1999	3.914	193.00	13,664.01	421.41	12.834	1.196
	2000	3.967	201.70	14,336.99	496.26	14.033	1.222
	2001	4.089	210.75	15,030.70	541.76	15.725	1.154
	2002	4.164	218.10	15,707.21	514.57	18.665	1.071
	2003	4.181	226.35	16,499.41	512.24	21.227	1.028
	2004	4.221	235.65	17,353.85	558.92	22.627	1.018
	2005	4.235	245.80	18,270.53	625.43	23.956	1.040
	2006	4.195	256.05	19,249.45	695.85	25.623	1.069
	2007	4.171	266.75	20,290.60	800.92	27.417	1.073

^aNational Center for Education Statistics (2008), Table 219.

^bNational Center for Education Statistics (2008), Table 31, Commonfund Institute (2009).

^cNational Center for Education Statistics (2008), Table 360.

^dWe assume linear growth between 1976-1979 and 1981-1984. 1997-2001 values are based on regression estimation from public expenditures. We assume quadratic growth 2002-2007 based on best fit from historical data.

^eHarvard University Office of the Provost (2009), Yale University Office of Institutional Research (2009).

^fHarvard endowment values 1975-1977 assume linear growth at the same rate as 1978-82, and Harvard enrollment 1975-87 assume linear growth at 1988-91 rate. Harvard and Yale are used because their investment models are used by other schools and for data availability reasons.

^gYale University Office of Institutional Research (2009).

^hWe assume Yale endowment 1976-1987 values grow linearly at same rate as 1988-1992. Yale and Harvard values are used because their investment models are used by other schools and for data availability reasons.

ⁱNational Center for Education Statistics (2008), Table 358.

Table B.3: Main dataset (continued)

Category		Financial aid to schools	Benefit of college			Household credit
Variable	Year	Total government support per student ^{a, b}	Earnings difference, (bachelor's or higher minus high school) ^c	Fraction of population with bachelor's degree ^{d, e}	Unemployment rate ^f	Household debt per student ^g
Unit		\$	\$	% persons 25 and older	Proportion of 16+ civilian labor force	Million \$
	1976	4,444	6,091	0.147	0.077	173.14
	1977	4,520	6,253	0.153	0.071	185.64
	1978	4,543	6,013	0.159	0.061	203.08
	1979	4,528	6,039	0.166	0.058	207.10
	1980	4,417	6,000	0.171	0.071	192.10
	1981	4,303	5,868	0.176	0.076	183.93
	1982	4,245	6,053	0.181	0.097	179.68
	1983	4,350	6,440	0.186	0.096	189.71
	1984	4,561	6,407	0.190	0.075	208.94
	1985	4,843	7,288	0.194	0.072	236.70
	1986	5,114	7,729	0.195	0.070	255.41
	1987	5,258	7,545	0.200	0.062	262.77
	1988	5,419	7,727	0.205	0.055	271.84
	1989	5,556	8,265	0.212	0.053	275.13
	1990	5,564	7,917	0.213	0.056	275.73
	1991	5,575	6,975	0.214	0.068	268.27
	1992	5,629	7,096	0.215	0.075	272.11
	1993	5,648	7,797	0.219	0.069	282.43
	1994	5,753	7,887	0.224	0.061	296.49
	1995	5,835	7,524	0.231	0.056	308.35
	1996	5,867	7,976	0.236	0.054	315.77
	1997	6,000	8,401	0.240	0.049	322.49
	1998	6,247	9,034	0.246	0.045	339.49
	1999	6,555	9,297	0.253	0.042	351.91
	2000	7,011	9,917	0.257	0.040	361.34
	2001	7,511	9,779	0.262	0.047	368.71
	2002	7,956	9,573	0.269	0.058	383.78
	2003	8,307	9,380	0.273	0.060	406.94
	2004	8,492	9,283	0.277	0.055	430.64
	2005	8,470	9,632	0.277	0.051	455.65
	2006	8,479	9,542	0.282	0.046	478.02
	2007	8,505	9,283	0.289	0.046	481.68

^aNational Center for Education Statistics (2008), Tables 348, 349, 353.

^bValues in 2001-2003 are based on the sum of support for public and private schools reported in Digest of Education Statistics Tables 349 and 350. We assume linear growth rate in the percentage of financial aid going to public schools from 2004 to 2007. We assume linear growth in 2007 based on observations from 2004 to 2006.

^cUS Census Bureau and Bureau of Labor Statistics (2009), Table A-3. Mean Earnings of Workers 18 Years and Over, by Educational Attainment, Race, Hispanic Origin, and Sex (1975 to 2007).

^dNational Center for Education Statistics (2008), Table 8, US Census Bureau and Bureau of Labor Statistics (2009).

^eWe assume linear growth during 1976-1979 and 1981-1984.

^fUS Census Bureau and Bureau of Labor Statistics (2009).

^gUS Federal Reserve (2009), Table D.3.

Table B.4: Main dataset (continued)

Category		Household wealth		Financial aid to students		
Variable	Year	Per capita disposable income ^a	Standard & Poor's Case-Shiller Home price index (non-seasonally adjusted, real) ^{b,c}	Federal grants per student ^d	Federal student loans per student ^e	Federal work study award per student ^{f,g}
Unit		\$	Index value in 1983=100	\$	\$	\$
	1976	10,496	1.46	3,160	373.12	72.42
	1977	10,749	1.41	3,648	404.48	83.82
	1978	11,074	1.35	4,281	474.85	83.38
	1979	10,959	1.25	5,773	592.27	84.93
	1980	10,672	1.13	7,643	760.59	84.60
	1981	10,700	1.05	7,817	857.48	75.53
	1982	10,767	1.01	6,986	814.11	66.40
	1983	11,140	1.00	7,763	804.56	66.01
	1984	11,773	0.99	8,379	901.61	67.03
	1985	11,999	0.97	8,350	935.24	63.20
	1986	12,354	0.98	8,443	926.34	60.07
	1987	12,452	0.96	9,934	995.53	56.48
	1988	12,854	1.00	10,078	1,057.23	52.80
	1989	13,011	1.02	9,761	1,014.19	49.93
	1990	13,010	0.98	9,633	971.34	50.31
	1991	12,872	0.93	10,104	952.48	49.57
	1992	13,140	0.91	10,423	983.36	49.08
	1993	13,086	0.89	13,686	1,164.52	48.43
	1994	13,278	0.89	16,071	1,437.81	46.45
	1995	13,432	0.88	17,554	1,626.74	45.02
	1996	13,611	0.88	18,732	1,730.86	48.50
	1997	13,866	0.89	19,255	1,789.28	50.97
	1998	14,438	0.93	19,627	1,817.12	53.68
	1999	14,619	0.98	20,136	1,816.73	51.81
	2000	15,066	1.03	20,363	1,797.24	49.59
	2001	15,136	1.09	21,614	1,783.82	48.84
	2002	15,452	1.17	24,379	1,864.86	49.33
	2003	15,655	1.26	27,574	2,047.41	48.76
	2004	16,033	1.40	29,574	2,197.85	46.09
	2005	16,036	1.56	30,316	2,268.44	42.91
	2006	16,447	1.59	30,806	2,280.17	40.01
	2007	16,613	1.47	33,877	2,346.54	37.90

^aBureau of Economic Analysis (2009), Table 2.1.

^bStandard & Poor's (2009).

^cWe assume linear growth during 1976-1987, based on growth rate from 1987 to 1996.

^dThe College Board (2009), Table 1.

^eThe College Board (2009), Table 1.

^fThe College Board (2009), Table 1.

^gValues prior to 1996 are discounted by 11% to reflect that these are commitments instead of disbursements, which are about 11% higher.

Table B.5: Main dataset (continued)

Category		Institutional aid	Tuition	Interest rate
Variable	Year	Institutional grants per student ^a	Tuition and fees (all institutions) per student ^b	Real interest rate (Treasury interest rate - change in CPI) ^c
Unit		\$	\$	One year real treasury rate
	1976	261.77	1,544	0.0012
	1977	250.02	1,575	-0.0042
	1978	242.84	1,578	0.0075
	1979	237.50	1,540	-0.0070
	1980	229.10	1,488	-0.0150
	1981	219.23	1,510	0.0448
	1982	220.18	1,597	0.0611
	1983	239.16	1,711	0.0637
	1984	270.58	1,813	0.0659
	1985	297.32	1,936	0.0486
	1986	328.20	2,050	0.0459
	1987	355.53	2,099	0.0312
	1988	361.93	2,162	0.0351
	1989	383.22	2,216	0.0371
	1990	444.06	2,240	0.0249
	1991	488.42	2,313	0.0165
	1992	531.14	2,424	0.0088
	1993	581.06	2,541	0.0044
	1994	624.11	2,656	0.0276
	1995	660.41	2,750	0.0311
	1996	689.03	2,837	0.0257
	1997	728.09	2,903	0.0334
	1998	779.91	2,996	0.0349
	1999	825.27	3,077	0.0287
	2000	842.39	3,082	0.0275
	2001	822.27	3,112	0.0064
	2002	801.52	3,238	0.0042
	2003	828.43	3,427	-0.0104
	2004	872.23	3,634	-0.0077
	2005	914.44	3,769	0.0023
	2006	955.91	3,892	0.0171
	2007	982.15	4,002	0.0168

^aThe College Board (2009), Table 1.^bNational Center for Education Statistics (2008), Table 331.^cUS Federal Reserve (2011).

B.4 Pre-estimation test

The pre-estimation test provides an indication of the value of a particular proxy variable in predicting enrollment. The quantity metric, enrollment per high school graduate, is regressed on each proxy variable individually using ordinary least squares regression. Although coefficients and standard errors are not necessarily correct because of the clear presence of omitted variable bias, these regressions provide some additional consistency checks for the final models provided in tables 3.4 and 3.5. We compare pre-estimation test signs with the signs found in the final models to ensure consistency.

Signs are generally consistent between the pre-estimation t-tests and the final models. One exception is financial aid to schools (per student) in the supply model, which has an insignificant pre-estimation test result (p-value = 0.39) but becomes significant once we include other relevant variables to eliminate omitted variable bias. Net tuition (per student) cannot be adequately pre-tested due to the presence of endogeneity, and because there is omitted variable bias in the pre-estimation model regressing enrollment on tuition price alone. This pre-estimation model cannot account for the fact that tuition price should have a different sign in the supply and the demand model. We also see some sign inconsistencies between the first principal component of college benefit in the demand model and the pre-tests of the underlying variables of this principal component. However, since the variables in the principal component have different signs, and because we do not have a strong theoretical basis for the “correct” sign of these metrics of the benefit of college, we cannot accurately compare the principal component sign to the sign of underlying variables.¹ A pre-estimation test of the principal component itself yields the same positive sign we see

¹For example, an increase in the percentage of the adult population holding a bachelor’s degree may make a degree less valuable, as described by Fortin (2006), thus presumably decreasing demand. On the other hand, prevalence of bachelor’s degrees in a society may make a college education seem more relevant and increase the perceived importance of education.

in the final demand model.

Table B.6: Pre-estimation test results (p-values are p-values of coefficients).^{a, b}

Variable name	Coefficient sign	p-value
Aid to schools per student	-	0.3861
Bachelor's percentage	-	0.1833
Disposable income per student	-	0.0189
Donations per student	-	0.0023
Earnings difference	-	0.1674
Endowment distribution per student	-	0.0067
Endowment value per student	-	0.0001
Grants per student	-	0.0166
HEPI	+	0.6928
Household debt per student	-	0.0016
Student loans per student	-	0.0073
Unemployment	+	0.0057

^aFor detailed variable descriptions, see table 3.1.

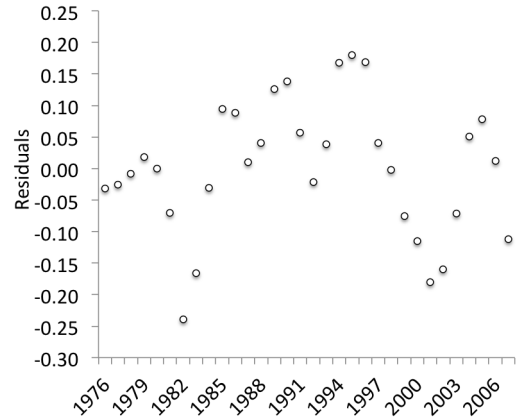
^bSigns and p-values are the result of regressing enrollment per high school graduate on each candidate variable individually. These values are used to validate coefficient signs in the final models in tables 3.4 and 3.5.

B.5 Preliminary model residual plots

This appendix provides plots of residuals over time for the preliminary models that make the two-stage least squares adjustment for endogeneity of net tuition price but do not make the AR(1) autoregressive error term adjustment for heteroskedasticity. Note the cyclical (non-random) nature of these plots, indicating that heteroskedasticity is present. This is due to the autocorrelated nature of the underlying dataset.



(a) Supply model.^a



(b) Demand model.^a

^aThis plot shows the second stage supply model residuals before the AR(1) adjustment, Y , as seen in equation (3.2).

^aThis plot shows the second stage demand model residuals before the AR(1) adjustment, W , as seen in equation (3.3).

Figure B.2: Residuals over time, preliminary models.

B.6 Comparison of parameter estimates in data subsets

In order to examine robustness of the models in tables 3.4 and 3.5, we construct models of the same structure using subsets of the data. We estimate separate models for the period 1976 to 1990 and the period 1991 to 2007. We estimate the autocorrelation component separately for each period, and then estimate coefficients for each variable. We find that effects of some variables change over time, as indicated in tables B.7 and B.8. These tables also show the result of a set of t-tests examining whether coefficients change significantly between the two periods. We find that some significant changes occur, but conclude that the second half of the data (1991 to 2007) yields results very close to those attained from the full dataset.

Table B.7: Parameter estimates in data subsets, supply model.^a

Two-stage least squares regression			
Period 1976 to 1990			
b ₀ (residual autocorrelation)	0.290		
Correlation (model estimates vs. actual)	0.602		
Variable ^b	Coefficient value	St. error	p-value
(Intercept)	-0.1231	0.0899	
Cost of operations (HEPI)	-0.0064	0.0504	0.9018
Non-tuition revenue per student	-0.7053	0.3258	0.0556
Financial aid to schools per student	-0.0004	0.0007	0.5171
Net tuition per student	0.0015	0.0021	0.4946
Period 1991 to 2007			
b ₀ (residual autocorrelation)	-0.314		
Correlation (model estimates vs. actual)	0.877		
Variable	Coefficient value	St. error	p-value
(Intercept)	0.0042	0.0124	
Cost of operations (HEPI)	0.0462	0.0113	0.0018
Non-tuition revenue per student	-0.2897	0.0534	0.0002
Financial aid to schools per student	0.0001	0.0000	0.0114
Net tuition per student	-0.0008	0.0003	0.0279
t-test for coefficient difference, H ₀ : Coefficients for first and second period are the same			
	p-value	Result	
Cost of operations (HEPI)	0.0738	Fail to reject H ₀ at $\alpha = .05$	
Non-tuition revenue per student	0.1177	Fail to reject H ₀ at $\alpha = .05$	
Financial aid to schools per student	0.0399	Reject H ₀ at $\alpha = .05$	
Net tuition per student	0.7716	Reject H ₀ at $\alpha = .05$	

^aModel equation is given in equation (3.2), with data split into two time periods.

^bFor detailed variable descriptions see tables B.2-B.5.

Table B.8: Parameter estimates in data subsets, demand model.^a

Ordinary least squares regression

Period 1976 to 1990

d ₀ (residual autocorrelation)	0.460		
Correlation (model estimates vs. actual)	0.879		
Variable ^b	Coefficient value	St. error	p-value
(Intercept)	-0.0142	0.0258	
Benefit of college	-0.0786	0.0504	0.1475
Credit effects per student	-0.1840	0.0433	0.0014
Net tuition per student	-0.0026	0.0018	0.1876

Period 1991 to 2007

d ₀ (residual autocorrelation)	-0.146		
Correlation (model estimates vs. actual)	0.553		
Variable	Coefficient value	St. error	p-value
(Intercept)	-0.0446	0.0160	
Benefit of college	0.0243	0.0338	0.4866
Credit effects per student	-0.0656	0.0511	0.2233
Net tuition per student	0.0007	0.0008	0.4121

t-test for coefficient difference, H₀: Coefficients for first and second period are the same

	p-value	Result
Benefit of college	0.0000	Reject H ₀ at $\alpha = .05$
Credit effects per student	0.0031	Reject H ₀ $\alpha = .05$
Net tuition per student	0.1659	Fail to reject H ₀ at $\alpha = .05$

^aModel equation is given in equation (3.3), with data split into two time periods.

^bFor detailed variable descriptions see tables B.2-B.5.

B.7 Ljung-Box test of residual autocorrelation

Table B.9 provides results of the Ljung-Box test of residual autocorrelation for supply and demand models with and without the first-order autoregressive error terms. As we see here, the AR(1) adjustment eliminates autocorrelation from the supply model. Demand model residuals still exhibit some autocorrelation, but it is reduced by the AR(1) adjustment.

Table B.9: Ljung-Box test for autocorrelation of residuals, p-values with and without AR(1) adjustment.^a

H ₀ : Residuals uncorrelated at lag 12	
Supply model	p-values
With AR(1) adjustment	0.301 (fail to reject H ₀ at $\alpha = .05$)
Without AR(1) adjustment	0.166 (reject H ₀ at $\alpha = .05$)
Demand model	p-values
With AR(1) adjustment	0.0001 (reject H ₀ at $\alpha = .05$)
Without AR(1) adjustment	1.87x10 ⁻⁶ (reject H ₀ at $\alpha = .05$)

^aResiduals for models with the AR(1) adjustment are presented in appendix B.1. Residuals for models without this adjustment are presented in appendix B.5.

APPENDIX C

Appendices for Chapter IV: An examination of educational choices, student characteristics, and ability-adjusted income

C.1 Effect of selectivity of school attended on post-college income, unconditional and conditional on selectivity of best school admitted to

Figure C.1 shows average income over the first four years with respect to the selectivity of the school an individual attended. Income across the whole sample (bar graph) declines as the selectivity of the school attended declines. This trend becomes somewhat less obvious after conditioning on the selectivity level of the best school to which an individual was admitted (lines). The overall downward salary trend in the unconditional bar graph of figure 4.1a appears stronger than the trend in C.1. This difference is likely an artifact of the sample used here, but is not surprising given the fact that the top school attended carries more information than the top school admitted to. For example, all students who attend a type 1 school must have

been admitted to at least a type 1 school, meaning we capture a great deal of the admissions information from data on the school attended.

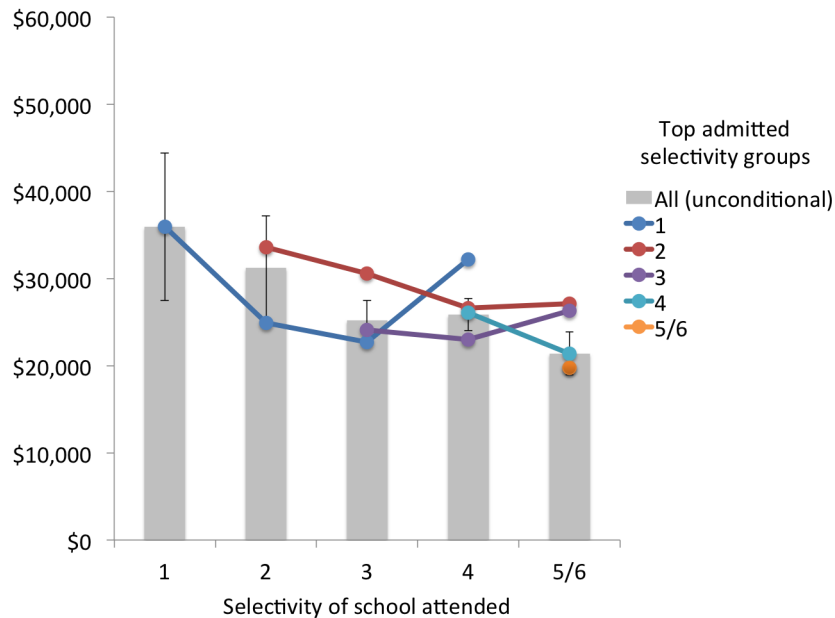


Figure C.1: Average income over the first four years after expected college graduation by selectivity of the school attended (bars), with 95% confidence intervals. Lines indicate income by selectivity of school attended after conditioning on selectivity of the top school admitted to. Type 1 indicates most selective schools.^{a, b, c}

^aExpected college graduation is considered to be four years after the beginning of a college education for those attending college, and four years after the start of the academic year for which students applied to college for those not attending.

^bSelectivity scores are based on Barron's selectivity index, which measures the acceptance rate of students who have applied for study (Barron's, 2009). Types 5 and 6 are combined into a single category. Specialty schools (type 7) are omitted.

^cSelected sample is presented in table 4.1 and is derived from NLSY 1997 respondents (Bureau of Labor Statistics, 2013). Individuals included are those from the cross-sectional sample who completed high school, participated in the college choice section of the survey, were admitted to at least one four-year institution, and are not missing critical data.

C.2 Variables necessary for building discrete choice model

Table C.1 provides an overview of candidate variables considered in constructing the models outlined in this chapter. Source information, descriptions of missing data, and details about each variable are provided.

Table C.1: Candidate variables used in model construction. Candidate variables are used for models in tables 4.6 and 4.7.

Variable Name	Description	Notes	Source
Student characteristics			
Top school admitted to	Barron's selectivity index for top admitted school	1 indicates most selective schools, 5/6 indicates least selective. Note that Barron's original categories 5 and 6 are combined into type 5/6. Specialty schools (7) are excluded.	NLSY 97, Barron's
School attended	Barron's selectivity index of attended school	1 indicates most selective schools, 5/6 indicates least selective. Note that Barron's original categories 5 and 6 are combined into category 5/6. Specialty schools (7) are excluded. School attended is blank for those who do not attend school.	NLSY 97, Barron's
Graduated	Indicates whether individual has graduated from college by the current survey date	Latest survey data available are from 2010. Students attending college right after high school and taking 4 years to complete a degree would have expected graduation of 2007.	NLSY 97
β_0	Nelson Siegel income curve parameter	Parameter is a measure of expected lifetime income. Larger parameters indicate a higher lifetime earnings curve. Derivation of this parameter is explained in section 4.5.	NLSY 97, Census Bureau
Gender	Student's gender		NLSY 97
Race/Ethnicity	Student's race and ethnicity	Possible values are White, Black, American Indian/Eskimo/Aleut, Asian/Pacific Islander, Other.	NLSY 97

Grades	High school grades	Categorical high school grades, with 1 indicating mostly A's received and 8 indicating mostly failing grades received.	NLSY 97
SAT Math	Math SAT score	Categorical Math SAT score. 1: 200-300, 2: 301-400, 3: 401-500, 4: 501-600, 5: 601-700, 6: 701-800, 0: no score.	NLSY 97
SAT Verbal	Verbal SAT score	Categorical Verbal SAT score. 1: 200-300, 2: 301-400, 3: 401-500, 4: 501-600, 5: 601-700, 6: 701-800, 0: no score.	NLSY 97
Major	Major categories for major at attended school	Possible values are sciences, social sciences, business, engineering, and computers, health, and humanities.	NLSY 97
College year	Year first attended college	Used to ensure student attends college within two years of high school graduation.	NLSY 97
Choice year	Year of the college choice (generally senior year of high school)	Used to ensure IPEDS data from the student's choice year are used.	NLSY 97
Father education	Number of years of father's formal education		NLSY 97
Mother education	Number of years of mother's formal education		NLSY 97
Household size	Number of individuals (including student) living in household	Data point closest in time to the choice year is used.	NLSY 97
Household income	Gross annual household income	Data point closest in time to the choice year is used.	NLSY 97
Urban-rural	Residence in rural or urban area	Data point closest in time to the choice year is used.	NLSY 97
High school type	High school type	Public, private, religious, alternative, etc.	NLSY 97

School characteristics

List tuition	List out-of-state tuition	The advertised out-of-state tuition price, often called the sticker price.	IPEDS
Loan percentage	Percentage of the student body receiving student loans.	Indicates whether or not this school participates in loan programs.	IPEDS

Federal grant percentage	Percentage of students receiving federal grants.	Serves as an indicator of the wealth of the student body, since many grants are specifically for lower-income students (Pell grants).	IPEDS
Gender ratio	Gender ratio	Enrolled males/enrolled females.	IPEDS
Total students	Number of enrolled full-time students.		IPEDS
Control	Publicly or privately controlled institution		IPEDS
Carnegie class	Adjusted Carnegie Class of institution.	Level 1: Research/Ph.D. granting, 2: Other graduate, 3: Bachelor's/liberal arts schools, 4: Specialty and associates-type degrees.	IPEDS
Selectivity	Selectivity from Barron's selectivity index	Specialty schools (Barron's index = 7) are excluded.	Barron's
Expenditures	Total expenditures per full-time student		IPEDS
Instructional expenditures	Instructional expenditures per full-time student	IPEDS definition of instructional spending is used. Total expenses associated with instructional divisions of the institution, including compensation of instructors, community and adult education, tutoring, and other academic-related activities.	IPEDS
Faculty per student	Faculty per student per full-time student		IPEDS
Average salary	Average salary of full-time faculty	Proxy for faculty quality and the percentage of faculty holding a Ph.D.	IPEDS
Division	Athletic division	Indicator if whether school is in top athletic division (division 1 NCAA).	IPEDS
Graduation rate		Graduation rate within 150% of normal completion time.	IPEDS

Student/school interactions

Financial aid	Total financial aid received for the first year of schooling, including aid offered specifically for a particular school, aid offered no matter which school is attended, and in-state tuition discounts	Students that claim that they do not know financial aid amounts are treated as receiving zero financial aid, since it does not factor into the decision. If aid information is missing in the school choice section, aid from the school attended section is supplemented. This aid amount is adjusted based on the relationship between reported numbers where they are available in both sections. Aid from the schools attended section $\ast 0.445 + 963.86$ yields an estimate of aid reported in the schools admitted to section (adjusted $R^2 = 0.2522$).	NLSY 97
Distance	Distance between home and school zip codes	Students that do not provide a zip code are located based only on state data. Zip code data points are translated into latitude and longitude. For schools that fail to report zip information to IPEDS, latitude and longitude and longitude data from Google Maps are used.	NLSY 97, IPEDS
In-state status	Student has in-state residency	If no residence state data point is provided, we assume student does not have in-state status.	NLSY 97, IPEDS
Urban rural match	Match of student residence and school's urban/rural characteristics	For schools that do not report urban/rural information to IPEDS, schools were researched on their individual websites and Google Maps.	NLSY 97, IPEDS

Selectivity difference	Selectivity of current candidate school minus selectivity of top admitted school of student	Selectivity of school is sometimes missing in the Barron's selectivity index of the correct year, in which case it can be predicted by 25 th and 75 th percentile SAT scores, percentage of applicants admitted, Carnegie Class, and faculty salaries. R^2 of this regression is 0.4893.	NLSY 97, IPEDS, Barron's
------------------------	---	--	--------------------------

C.3 Projecting lifetime income

C.3.1 Functional form of the income curve

Synthetic work-life earnings curves suggest the shape of the income curve over the lifetime. We experiment with several functional forms that allow us to approximate this shape. The traditional Mincer model of lifetime earnings calls for a quadratic age-income profile model for log income:

$$\log(y(t)) = \gamma_0 + \gamma_1 t + \gamma_2 t^2 + v \quad (\text{C.1})$$

where t is age and v gives the (potential) number of years of experience, to allow for an adjustment for length of schooling. Past literature has shown that this functional form provides bad fit, requiring at least a fourth order model (Polachek, 2007). As a replacement for Mincer's model in this application, we suggest a restricted Nelson-Siegel curve. The Nelson-Siegel equation allows us to estimate each individual's lifetime expected earnings profile with only a few data points and provides close fit to Census synthetic work-life earnings curves for all levels of educational attainment.¹

The Nelson-Siegel equation has four parameters, β_0 , β_1 , β_2 , and τ , and the function is given by:

$$y(t) = -\beta_0 + \beta_1 \left(\frac{1 - e^{-t/\tau}}{t/\tau} \right) + \beta_2 \left(\frac{1 - e^{-t/\tau}}{t/\tau} - e^{-t/\tau} \right) \quad (\text{C.2})$$

where

- β_0 (level parameter) gives the long-term level of the function.
- β_1 (slope parameter) is found in the second term of the Nelson-Siegel function.

The second term is an exponential decay function. A positive parameter yields

¹Plots of these fits by educational attainment are provided in figures C.3 and C.4. More information on the fit of these models is provided in section C.3.2.

a downward slope, a negative parameter an upward slope.

- β_2 (curvature parameter) is found in the third term of the Nelson-Siegel function. The third term is a Laguerre function (product of an exponential and a polynomial function) that produces the hump (positive parameter) or valley (negative parameter) in the function. A larger absolute value means the hump or valley is more pronounced. This parameter gives the rate at which the slope and curvature parameters decay to zero.
- τ (shape parameter) determines the steepness of the slope factor and the location of the maximum or minimum of the Laguerre function (third/curvature term).

Estimation of the model parameters is done numerically by using nonlinear optimization, grid search, Kalman filters and/or state space modeling. However, since income curves have a fairly common shape across the lifetime of all individuals, we are able to fix several of the parameters of the Nelson-Siegel curve. This not only ensures that lifetime income curves have the correct shape, but also reduces data requirements by limiting the number of parameters we must estimate.

C.3.2 Parameter estimation

We begin by fixing τ so that maximum income is achieved at the correct point during an individual's career. In the synthetic work-life income data for people ranging from high school graduates to master's recipients, this maximum yearly income occurs at around age 50 (Census Bureau, 2012). We want to determine the value of τ that forces the maximum value of the third term in equation (C.2) to occur when t equals the desired value. The third term is the Laguerre function curvature term, $f(t, \tau) = \frac{1 - e^{-t/\tau}}{t/\tau} - e^{-t/\tau}$. Let t^* be the time at which we would like the maximum to

occur and τ^* be the τ value that achieves this maximum, as shown below:

$$f(t, \tau) = \frac{1 - e^{-t/\tau}}{t/\tau} - e^{-t/\tau} \quad (\text{C.3})$$

$$t^* = \arg \max_t [f(t, \tau^*)]$$

Setting $\tau = \tau^*$, equation (C.2) reduces to:

$$y(\tau) = -\beta_0 + \beta_1 A(t, \tau^*) + \beta_2 B(t, \tau^*) \quad (\text{C.4})$$

where $A(t, \tau^*) = \frac{1 - e^{-t/\tau^*}}{t/\tau^*}$ and $B(t, \tau^*) = \frac{1 - e^{-t/\tau^*}}{t/\tau^*} - e^{-t/\tau^*}$ and A and B can be computed directly from the age input t and the fixed curvature parameter τ^* . Parameters β_0 , β_1 , and β_2 are estimated using ordinary least squares.

Figure C.2 shows the surface of the value of the third (curvature) term, $\frac{1 - e^{-t/\tau}}{t/\tau} - e^{-t/\tau}$, over time and tau. We determine numerically that a maximum of 50 occurs when $\tau = 27.88$. With τ fixed, we can achieve close shape fits for the Census synthetic work-life earnings data, as shown in figure C.3.

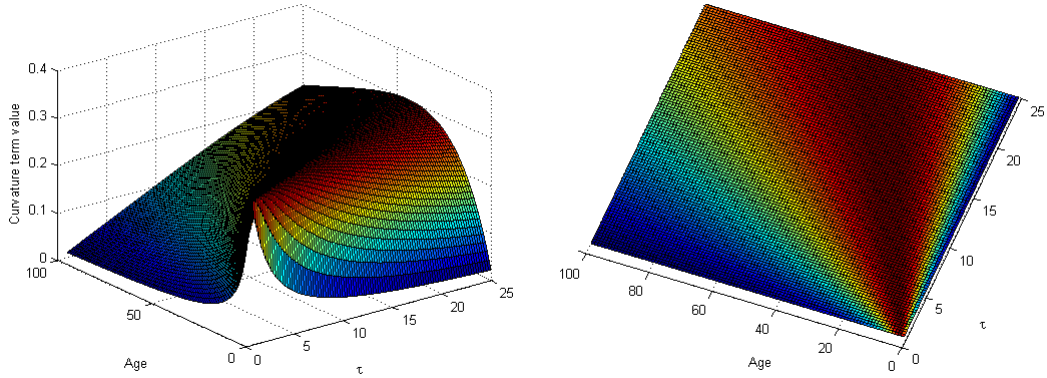


Figure C.2: Third (curvature) term in equation (C.2), $\frac{1 - e^{-t/\tau}}{t/\tau} - e^{-t/\tau}$, over τ and age (t). Value of τ determines where maximum of equation (C.2) occurs.

Since we wish to further reduce the number of parameters to be estimated, we take advantage of the consistent relationship between the remaining Nelson-Siegel

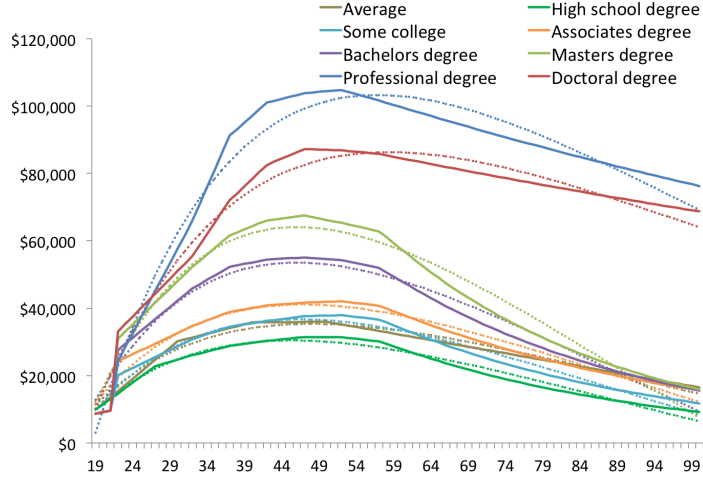


Figure C.3: Census synthetic lifetime income earnings curve fits based on equation (C.2) with fixed $\tau = \tau^*$ only. Dotted lines indicate Nelson-Siegel estimates and solid lines indicate Census data.^{a, b}

^aCensus Bureau (2012).

^b $\tau^* = 27.88$ ensures that lifetime income maximum occurs at the desired age of 50.

parameters across education groups. Specifically, we notice that the ratios $\frac{\beta_0}{\beta_2}$ and $\frac{\beta_0}{\beta_1}$ are fairly consistent across education groups. We investigate three possible approaches of parameter reduction and test their ability to fit Census synthetic work-life income curves. The regression equation fit for each method is provided below.

- (a) Setting β_1 to equal its average value across high school, some college, associates, and bachelors degrees.

$$\beta_1 = 29,332, y(\tau) - \beta_1 A = -\beta_0 + \beta_2 B$$

- (b) Fixing the relationship between β_0 and β_2 based on the strong relationship seen in Census data.

$$\beta_2 = -m\beta_0 + b \text{ where } m = -3.8149 \text{ and } b = 36,241, R^2 = 0.9967, y(\tau) - bB = -(1 + mB)\beta_0 + \beta_1 A$$

- (c) Fixing the relationship between β_0 and β_2 as above and the relationship between β_0 and β_1 based on the strong relationship seen in Census data.

$$\beta_1 = -n\beta_0 + a \text{ where } n = -0.2445 \text{ and } a = 2,234.3, R^2 = 0.9605, y(\tau) - bB -$$

$$aA = -(1 + mB + nA)\beta_0$$

The resulting model fits are shown in figure C.4. We determine that method (c) is able to fit synthetic work-life income curves very closely using only a single parameter, β_0 .² We project β_0 for each individual in our dataset, and thereby attain an estimate of the future earnings expectations for each student.

C.3.3 Income by educational attainment

Based on the income parameter β_0 for each individual, we analyze the differences in expected lifetime incomes between different attendance and admission selectivity groups. We find very similar trends to those seen in the raw post-college income data presented in section 4.5. It is sufficient to examine the value of β_0 across groups since this parameter fully defines future income expectations under the assumptions of the previous section. Note that we also examine differences in the expected lifetime earnings implied by β_0 , with similar results.³ Unemployment is again not explicitly modeled in the projected income for future years since we do not have future employment status information.

As in the analysis of raw post-college income data, we find that both the quality of the best school admitted to and the school attended to predict the income parameter β_0 . However, the predictive ability of the selectivity of school attended is lost if we control by best school admitted. Table C.2 shows that attendance information that

²Mean squared errors between estimated income and actual Census data for those with high school degrees, some college, associate's degrees, and bachelor's degrees increases only slightly when parameters are fixed. Mean squared error with only τ fixed is 3,933,487 across these groups. Mean square errors for methods a) through c) in figure C.4 are 4,600,744, 4,845,014, and 4,828,948, respectively. Income estimates using methods a) and b) were also used to conduct the analysis described in section C.3.3. Results are similar to those presented here but income projection is possible for a smaller number of individuals.

³We calculate discounted expected lifetime earnings based on β_0 and equation (C.4), assuming that non-attenders follow the estimated income curve from ages 22 to 100, while attenders follow it from 18 to 100. Earnings are discounted by a yearly factor of 0.96. The value of the discount factor is based on estimates found in past literature, for example in Cocco et al. (1998) and Gourinchas and Parker (1999).

is uncorrelated with the best school admitted to (“attendance error”) provides no additional information on income expectations after adjusting for best school admitted to. In contrast, “admission error” remains a significant predictor of income even after adjusting for the school attended; students that are admitted to top schools and choose to attend lower caliber institutions exhibit significantly higher earnings than their classmates. Overall, we do not have sufficient income data to develop a complete understanding of lifetime income expectations across groups. Therefore, we are not able to include lifetime income expectations in the choice models presented in section 4.6. As more data is collected, income projection can be improved and income expectations included in choice models.

Table C.2: Separating effects of top school admitted to and school attended selectivity on Nelson-Siegel income parameter using interval selectivity data.

(a) Predicted β_0 , by selectivity of top school admitted to and attendance error. $\beta_0 = \alpha_0 + \alpha_1 \text{topAdmit} + \alpha_2 \text{attendError} + \epsilon$.^{a,b}

Ordinary least squares regression			
	Adjusted R^2 :		0.0508
	Estimate	St. Error	p-value
(Intercept)	345,627	24,074	
Top admitted school acceptance rate	-1,518	315	2.00×10^{-16}
Attendance error	-250	584	0.6690

^aParameter β_0 in the restricted Nelson-Siegel function derived from equation (C.2) fully determines the lifetime income curve under the assumptions made in section C.3.2.

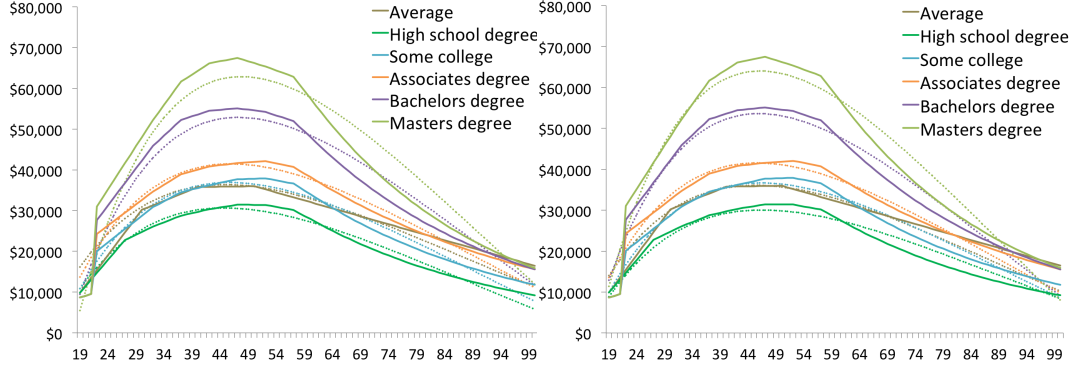
^bSelectivity scores are based on Barron’s selectivity index (Barron’s, 2009). Types 5 and 6 are combined into a single category. Specialty schools (type 7) are omitted. Interval scale is created by replacing category label with the percentage of applicants admitted in each selectivity category. Type 1 schools have admissions rates of at most 33% while type 5 and 6 schools admit 100% of their applicants.

(b) Predicted β_0 , by selectivity of school attended and admission error. $\beta_0 = \alpha_0 + \alpha_1 \text{attend} + \alpha_2 \text{admitError} + \epsilon$.^{a,b}

Ordinary least squares regression			
	Adjusted R^2 :		0.0508
	Estimate	St. Error	p-value
(Intercept)	313,268	22,440	
Attended school acceptance rate	-21,939	5,935	0.0002
Admission error	-1,352	451	0.0029

^aParameter β_0 in the restricted Nelson-Siegel function derived from equation (C.2) fully determines the lifetime income curve under the assumptions made in section C.3.2.

^bSelectivity scores are based on Barron’s selectivity index (Barron’s, 2009). Types 5 and 6 are combined into a single category. Specialty schools (type 7) are omitted. Interval scale is created by replacing category label with the percentage of applicants admitted in each selectivity category. Type 1 schools have admissions rates of at most 33% while type 5 and 6 schools admit 100% of their applicants.



(a) β_1 set to its average value across high school, some college, associates, and bachelors degrees. $\beta_1 = 29,332$, $y(\tau) - \beta_1 A = -\beta_0 + \beta_2 B$. Dotted lines indicate Nelson-Siegel estimates and solid lines indicate Census data.^{a,b}

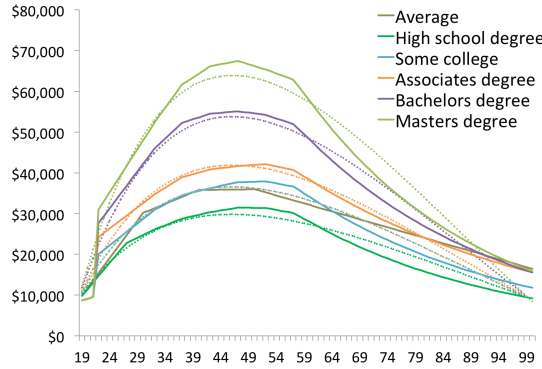
^aCensus Bureau (2012).

^b $\tau^* = 27.88$ ensures that lifetime income maximum occurs at the desired age of 50.

(b) Fixed relationship between β_0 and β_2 based on trends seen in Census data. $\beta_2 = -m\beta_0 + b$ where $m = -3.8149$ and $b = 36,241$, $R^2 = 0.9967$, $y(\tau) - bB = -(1 + mB)\beta_0 + \beta_1 A$. Dotted lines indicate Nelson-Siegel estimates and solid lines indicate Census data.^{a,b}

^aCensus Bureau (2012).

^b $\tau^* = 27.88$ ensures that lifetime income maximum occurs at the desired age of 50.



(c) Fixed relationship between β_0 and β_2 as in (b) and relationship between β_0 and β_1 based on trends seen in Census data. $\beta_1 = -n\beta_0 + a$ where $n = -0.2445$ and $a = 2,234.3$, $R^2 = 0.9605$, $y(\tau) - bB - aA = -(1 + mB + nA)\beta_0$. Dotted lines indicate Nelson-Siegel estimates and solid lines indicate Census data.^{a,b}

^aCensus Bureau (2012).

^b $\tau^* = 27.88$ ensures that lifetime income maximum occurs at the desired age of 50.

Figure C.4: Census synthetic lifetime income earnings curve fits based on equation (C.2) for different estimation procedures.

C.4 Conditional post-college income effects using categorical selectivity data

This appendix presents fixed effects models using categorical admission and attendance categories that investigate effects on income after conditioning. School categories 1, 2, 3, 4, and 5/6 are treated as factor levels instead of converting them to interval data based on acceptance rates defining these categories. This results in the need to estimate more parameters, making it difficult to see significant differences between groups. In table C.3a, we compare post-college income for students in each attendance category based on their top admitted school. In table C.3b, we compare post-college income for students in each top admitted category based on the school they attended. In general, we cannot show that a difference exists between incomes after conditioning.

Table C.3: Separating effects of top school admitted to and school attended selectivity on post-college income, categorical selectivity data.

(a) Predicted post-college income by selectivity of top school admitted to, adjusted by attended category. $\beta_{0_k} = \alpha_{0_k} + \alpha_{1_k} \text{topAdmit} + \epsilon_k$ for each attendance category k .^a

Fixed effects regression								
Category			Coefficients, by top school admitted ^b					
Type at- tended	#	Types admitted to	(Intercept)	2	3	4	5/6	
none	257	1,2,3,4,5	18,273	4,150	1,106	-4,958	792	
1	39	NA	NA	NA	NA	NA	NA	NA
2	77	1,2	24,915 (*) ^c	8,691	NA	NA	NA	NA
3	226	1,2,3	22,698 (*)	7,869	1,431	NA	NA	NA
4	333	1,2,3,4	32,192 (*)	-5,576 (.) ^d	-9,187 (*)	-6,115	NA	NA
5/6	121	2,3,4,5	27,152 (*)	-859	-5,795	-7,426	NA	NA

^aSelectivity scores are based on Barron's selectivity index, which measures the acceptance rate of students who have applied for study (Barron's, 2009). Types 5 and 6 are combined into a single category. Specialty schools (type 7) are omitted.

^bCoefficients are relative to baseline. Baseline is top admitted school of type 1 except for attendance category 5/6; in attendance category 5/6/, base case is top admitted school of type 2.

^c* indicates significance at the 0.05 level.

^d. indicates significance at the 0.10 level.

(b) Predicted post-college income by selectivity of school attended, adjusted by top admitted category. $\beta_{0_k} = \alpha_{0_k} + \alpha_{1_k} \text{attend} + \epsilon_k$ for each top admitted category k .^a

Fixed effects regression								
Category			Coefficients, by school type attended ^b					
Top admitted	#	Types attended	(Inter- cept)	1	2	3	4	5/6
1	121	none, 1, 2, 3, 4	18,273 (*) ^c	17,670 (*)	6,642	4,424	13,918 (.) ^d	NA
2	155	none, 2, 3, 4, 5	22,423	NA	11,183	8,143	4,192	4,729
3	297	none, 3, 4, 5	19,379 (*)	NA	NA	4,750 (*)	3,626	6,915
4	355	none, 4, 5	17,778 (*)	NA	NA	NA	8,298 (*)	3,579
5,6	125	none, 5	19,065 (*)	NA	NA	NA	NA	661

^aSelectivity scores are based on Barron's selectivity index, which measures the acceptance rate of students who have applied for study (Barron's, 2009). Types 5 and 6 are combined into a single category. Specialty schools (type 7) are omitted.

^bCoefficients are relative to baseline. Baseline is attending no school at all.

^c* indicates significance at the 0.05 level.

^d. indicates significance at the 0.10 level.

C.5 Effect of attending college at all

Table C.4 examines the effect of the decision to attend college at all on post-college income for each top admitted category. Results indicate that attending college at all may have a positive effect on post-college income for all top admission categories, especially for those admitted to higher caliber schools. Further analysis reveals that there is no difference in the expected salary coefficient for students who choose not to attend school at all between different top admitted categories. The average post-college annual salary for non-attenders is \$18,798. For college attenders, salary differences do exist between different top admitted groups. These differences are explained further in figure 4.2.

Table C.4: Effect of binary decision to attend college on post-college income.^{a,b}

Summary of fixed effects models by top admitted category					
Top admitted	# in category	Coefficients		p-values	
		Non-attenders ^c	Attenders	Non-attenders	Attenders
1	121	18,273	29,688	0.0015	0.0607
2	155	22,423	30,800	1.31x10 ⁻⁰⁵	0.1170
3	297	19,379	23,980	3.99x10 ⁻¹⁹	0.0433
4	355	17,778	25,599	3.38x10 ⁻²⁵	4.77x10 ⁻⁵
5,6	125	19,065	19,726	1.66x10 ⁻¹⁷	0.7937

^aSelectivity scores are based on Barron's selectivity index, which measures the acceptance rate of students who have applied for study (Barron's, 2009). Types 5 and 6 are combined into a single category. Specialty schools (type 7) are omitted.

^bSelected sample is presented in table 4.1 and is derived from NLSY 1997 respondents (Bureau of Labor Statistics, 2013). Individuals included are those from the cross-sectional sample who completed high school, participated in the college choice section of the survey, were admitted to at least one four-year institution, and are not missing critical data.

^cNon-attenders are the base case in this regression.

C.6 Logistic choice models

In a discrete choice framework, we model a decision-maker that makes a choice or set of choices over a set of alternatives. For a decision-maker i , we would like to understand what causes him or her to select a particular alternative, j . The choice is based on a set of observable factors, x , as well as a set of unobservable factors, ϵ . These inputs fully determine the agent's choice by defining the utility of each alternative to person i , written as $U_{i,j}$. To simplify this model, we assume that the non-random component of utility is linear in parameters, allowing us to write down the utility function as follows:

$$U_{i,j} = \delta' x_{i,j} + \epsilon_{i,j} \quad (\text{C.5})$$

where δ is the vector of parameters to be estimated and x is the vector of input characteristics of the decision maker i , the choice j , and interaction terms between i and j . To estimate parameters of the model, we analyze the probability of a given choice j being made when a certain set of observable factors x is present. Let us define an indicator function that indicates the outcomes of ϵ for which j is the optimal choice. Let $h(x, \epsilon)$ be the behavioral process that fully determines the agent's choice given x and the realization of ϵ . The indicator function can be written as:

$$I[h(x, \epsilon) = j] = \begin{cases} 1 & \text{if } h(x, \epsilon) = j, \\ 0 & \text{otherwise} \end{cases} \quad (\text{C.6})$$

We then examine all possible realizations of ϵ for which j is the choice outcome: $P(j|x) = P(\epsilon \text{ s.t. } h(x, \epsilon) = j)$. We can formulate $P(j|x)$ as an integral over the

possible realizations of ϵ :

$$\begin{aligned} P(j|x) &= P(I[h(x, \epsilon) = j] = 1) \\ &= \int I[h(x, \epsilon) = j]f(\epsilon)d\epsilon \end{aligned} \tag{C.7}$$

The integral in equation (C.7) can be expressed in closed form only in a few cases. One such case is when the components of ϵ are distributed according to i.i.d. type 1 extreme value distributions, making the choice model a logistic choice model (see e.g. McFadden (1974)). When we have this type of distribution for ϵ in equation (C.5), we can write the probability of selecting a particular choice j only in terms of the non-random component of utility, as shown below:

$$P_{i,j} = \frac{e^{\delta'x_{i,j}}}{\sum_{k \in J} e^{\delta'x_{i,k}}} \tag{C.8}$$

Logistic choice models have several useful characteristics for this application. Firstly, the logistic models exhibits what is called the Independence of Irrelevant Alternatives (IIA) property. This means that preferences between alternatives do not change as the choice set changes. For any two alternatives a and b , the ratio of their logit probabilities depends only on these two alternatives:

$$\frac{P_{i,a}}{P_{i,b}} = \frac{e^{\delta'x_{i,a}} / \sum_K e^{\delta'x_{i,k}}}{e^{\delta'x_{i,b}} / \sum_K e^{\delta'x_{i,k}}} = e^{\delta'x_{i,a} - \delta'x_{i,b}} \tag{C.9}$$

The IIA property is useful for this application since the choice set available to students depends on a previous decision, namely where to apply for school. The schools a student did not apply to may shift the utility of the schools to which they did apply. However, due to the IIA property the relative probability of attending one school over another remains constant. This allows us to model choice between colleges even when we are unsure that every possible school considered by the student is provided in the

NLSY 97 dataset.

A second critical advantage of the logistic choice model is that we can create a choice model focused entirely on the characteristics of the alternatives, called a conditional choice model. A conditional choice model is used in the second stage of our model. Since every student's choice set is unique and may include schools that do not appear anywhere else in the sample, we define schools solely based on their individual characteristics, such as tuition price, school size, and metrics of school quality. The alternative-based conditional logit model, described further in section 4.6.2, allows us to do exactly this.

C.7 Comparison to models in Long (2004b)

The models presented in this chapter closely follow the structure developed in Long (2004b). Our models are generally consistent with Long's previous findings, but results differ due to the use of different datasets. The dataset used in this chapter is much smaller than Long's, but contains richer data on individual qualities, school characteristics, and income over time. Additionally, the current work focuses exclusively on four-year schools, while Long includes two-year institutions. Tables C.5 and C.6 presents Long's original models as well as recreations of these models using our data. Signs are consistent, with the exception of distance from the student's home to the school. We find a very small positive effect from distance when replicating Long's model, but distance is not significant in other model configurations. Our selectivity difference metric implies that students choose to attend more selective institutions relative to their own performance. This is the same directional preference indicated by Long's SAT percentile variables, which indicate that students prefer colleges whose SAT percentiles are higher than their own. Model fit is slightly better in the current model than in Long's for the first stage, while Long's fit is better in the second stage model. We use average faculty salary as a proxy for the percentage of faculty with a PhD, and selectivity difference between the school in question and top admitted school as a proxy for SAT performance relative to the student body. The comparison models from Long's work used here are those for the 1992 cohort. Note that coefficients in Long's original work are presented as odds ratios (e^β) but have been adjusted here for consistency with the rest of the chapter.

Table C.5: Comparison model for first stage logit regression model of decision to attend college at all.

(a) First stage model from Long (2004b), 1992 cohort.		(b) First stage model to replicate Long (2004b) using the NLSY 97 data. ^a	
Logit regression		Logit regression	
	Estimate		Estimate
Tuition price (aid adjusted)	0.0009	(Intercept)	-0.007
Tuition price ²	0	Tuition price (aid adjusted) ^b	-5.37x10 ⁻⁶
Distance	-0.1122 (.) ^a	(Tuition price) ²	1.13x10 ⁻⁹ (.) ^c
Distance ²	0.0175	Distance	1.50x10 ⁻⁵ (*) ^d
Instructional expenditures	0.0040	(Distance) ²	-2.32x10 ⁻¹¹ (*)
Instructional expenditures ²	0.0000	Instructional expenditures	-2.77x10 ⁻⁶
County unemployment rate ^b	0.0173	(Instructional expenditures) ²	-2.65x10 ⁻¹²

^a. indicates significance at the 0.10 level.

^bCounty unemployment rate is not readily available for NLSY 97 data and is omitted from the comparison model.

^aSelected sample is presented in table 4.1 and is derived from NLSY 1997 respondents (Bureau of Labor Statistics, 2013). Individuals included are those from the cross-sectional sample who completed high school, participated in the college choice section of the survey, were admitted to at least one four-year institution, and are not missing critical data.

^bThis is list tuition-financial aid, where financial aid includes in-state discounts, aid awarded by the individual school, and aid awarded for use at all schools.

^c. indicates significance at the 0.10 level.

^d* indicates significance at the 0.05 level.

Table C.6: Comparison model for second stage conditional logit regression model of choice between candidate schools.

(a) Second stage model from Long (2004b), 1992 cohort.		(b) Second stage model to replicate Long (2004b). ^a	
Conditional logit regression		Conditional logit regression	
Pseudo R^2 :	0.4719	Pseudo R^2 :	0.035
	Estimate		Estimate
Tuition price (aid adjusted) ^a	-0.4234 (*)	Tuition price (aid adjusted)	-5.91x10 ⁻⁵ (*) ^b
(Tuition price) ²	0.0146 (*) ^b	(Tuition price) ²	-7.59x10 ⁻¹⁰ (*)
Distance	-1.3213 (*)	Distance	3.06x10 ⁻⁶
(Distance) ²	8.15x10 ⁻² (*)	(Distance) ²	-3.18x10 ⁻¹²
Instructional expenditures	9.85x10 ⁻² (*)	Instructional spending	-2.72x10 ⁻⁷
(Instructional expenditures) ²	-2.10x10 ⁻³ (*)	(Instructional spending) ²	-2.57x10 ⁻¹⁴
Student/faculty ratio	1.26x10 ⁻²	Average faculty salary ^c	9.35x10 ⁻⁶
% faculty with Ph.D.s	5.98x10 ⁻³ (*)	Selectivity difference ^d	-1.17x10 ⁻² (.) ^e
Student SAT %ile larger than school's	-0.3384 (*)		
Student SAT %ile smaller than school's	0.1663 (*)		
2-year college dummy	1.580 (*)		

^aThis is list tuition-financial aid, where financial aid includes in-state discounts, aid awarded by the individual school, and aid awarded for use at all schools.

^b* indicates significance at the 0.05 level.

^aSelected sample is presented in table 4.1 and is derived from NLSY 1997 respondents (Bureau of Labor Statistics, 2013). Individuals included are those from the cross-sectional sample who completed high school, participated in the college choice section of the survey, were admitted to at least one four-year institution, and are not missing critical data.

^b* indicates significance at the 0.05 level.

^cAverage faculty salary is used as a proxy for faculty quality, similar to Long (2004b)'s use of % faculty with Ph.D.s.

^dSelectivity difference is the difference in Barron's selectivity index between the top school admitted to for the current student and the selectivity of the school in question (selectivity of this school minus selectivity of top school admitted to). This is a proxy for ability fit of the student, similar to the SAT percentiles used in Long (2004b).

^e. indicates significance at the 0.10 level.

C.8 Collinearity and model selection

As part of the model selection process, we determine which variables are highly correlated in order to eliminate collinearity from our models. We use correlation and condition numbers to identify problem areas, and create a robust model selection process that tests all permutations of variables in highly collinear categories. Table C.7 provides an overview of the approach used.

Table C.7: Overview of model selection process.^a

	Category	Description of variables
		Net tuition ^{*b}
A	Tuition/aid	List out of state tuition and financial aid amount List out of state tuition, financial aid amount without the instate discount, and instate discount amount
B	Selectivity	Interval-measured selectivity of school in question Interval-measured difference between top admitted and school in question selectivity*
C	Expenditures/wealth	Instructional expenditures per student* Expenditures per student Faculty per student
D	School quality	Total students Graduation rate* Principal component of quality variables: total students, graduation rate, percentage receiving federal grants, average faculty salary, and selectivity ^c

^aCategories of highly related variables tested in model selection for second stage model (choice between schools).

^{b*} indicates configuration used in final model.

^cWhile the principal component approach yields slightly better model fits, graduation rate is used in final models for the sake of easy interpretation.

Note that we also test for linear separation of each predictor variable to ensure reasonable model results. Outliers are found in some cases. For example, an all male college adds an extreme data point to the gender ratio variable. In this and similar cases, we carefully investigate data for accuracy and ensure that inclusion/exclusion of the data point does not have large effects on the final model parameters.

BIBLIOGRAPHY

BIBLIOGRAPHY

- Akaike, H. (1969). A Method of Statistical Identification of Discrete Time Parameter Linear Systems. *Annals of the Institute of Statistical Mathematics*, 21(1):225–242.
- Angrist, J. and Krueger, A. (2010). Instrumental Variables and the Search for Identification: From Supply and Demand to Natural Experiments. *Journal of Economic Perspectives*, 15(4):69–85.
- Arcidiacono, P. (2004). Ability Sorting and the Returns to College Major. *Journal of Econometrics*, 121(1-2):343–375.
- Avery, C. and Hoxby, C. M. (2003). Do and Should Financial Aid Packages Affect Students' College Choices? Working Paper 9482, National Bureau of Economic Research.
- Barron's (2009). Admissions Competitiveness Index. <http://www.barronspac.com/index.cfm?anonymous&page=cs>.
- Becker, G. and Chiswick, B. (1966). Education and the Distribution of Earnings. *American Economic Review*, 56(2), 358-369.
- Becker, G. S. (1994). *Human Capital: A Theoretical and Empirical Analysis with Special Reference to Education (3rd Edition)*. Number beck94-1 in NBER Books. National Bureau of Economic Research, Inc.
- Becker, G. S. and Murphy, K. M. (2007). The Upside of Income Inequality. *The American*, May/June 2007.
- Behrman, J. and Rosenzweig, M. (1999). Ability Bias in Schooling Returns and Twins: A Test and New Estimates. *Economics of Education Review*, 18(2):159-167.
- Bennett, W. and Wilezol, D. (2013). *Is College Worth It?: A Former United States Secretary of Education and a Liberal Arts Graduate Expose the Broken Promise of Higher Education*. Thomas Nelson, Inc.
- Bennett, Jr., W. J. (Feb. 18, 1987). Our Greedy Colleges. *New York Times*.
- Berger, B. (2013). Why A College Degree May Not Be Worth It. *U.S. News and World Report*.

- Best, K. and Keppo, J. (2012). The Credits that Count: How Credit Growth and Financial Aid Affect College Tuition and Fees. *Education Economics*.
- Bhuller, M., Mogstad, M., and Salvanes, K. G. (2011). Life-Cycle Bias and the Returns to Schooling in Current and Lifetime Earnings. IZA Discussion Papers 5788, Institute for the Study of Labor (IZA).
- Brewer, D. J., Eide, E. R., and Ehrenberg, R. G. (1999). Does It Pay to Attend an Elite Private College? Cross-Cohort Evidence on the Effects of College Type on Earnings. *The Journal of Human Resources*, 34(1):pp. 104–123.
- Brown, J., Fang, C., and Gomes, F. (2012). Risk and Returns to Education. <http://faculty.london.edu/fgomes/bfg.pdf>.
- Bureau of Economic Analysis (2009). National Economic Accounts.
- Bureau of Labor Statistics (2009). Consumer Price Index.
- Bureau of Labor Statistics (2013). National Longitudinal Survey of Youth. <http://www.bls.gov/nls/nlsy79.htm>.
- Campbell, R. and Siegel, B. N. (1967). The Demand for Higher Education in the United States, 1919-1964. *The American Economic Review*, 57(3):pp. 482–494.
- Carneiro, P., Heckman, J. J., and Vytlacil, E. J. (2010). Estimating Marginal Returns to Education. Working Paper 16474, National Bureau of Economic Research.
- Carnevale, A., Strohl, J., and Melton, M. (2013). What’s It Worth: The Economic Value of College Majors. *Georgetown University Center on Education and the Workforce*.
- Carnevale, A. P., Rose, S. J., and Cheah, B. (2009). The College Payoff. *Georgetown University Center on Education and the Workforce*.
- Carroll, C. (2004). Theoretical Foundations of Buffer Stock Saving. NBER Working Papers 10867, National Bureau of Economic Research, Inc.
- Census Bureau (2012). American Community Survey Data on Educational Attainment. <https://www.census.gov/hhes/socdemo/education/data/acs/index.html>.
- Central Intelligence Agency (2011). CIA World Factbook. <https://www.cia.gov/library/publications/the-world-factbook/>.
- Centro de Estudios en Gestión de la Educación Superior (2007). Rates of Return and Funding Models in Europe: Final Report to the Directorate-General for Education and Culture of the European Commission. *Education and Culture DG*.
- Chaker, A. M. (Sep. 4, 2009). Students Borrow More than Ever for College. *The Wall Street Journal*.

- Chatfield, C. (2009). *The Analysis of Time Series*. Chapman & Hall.
- CHELPS, IOE London, Technopolis Group (2010). Funding Reform. *Progress in Higher Education Reform Across Europe*, Vols. 1 & 2.
- Cocco, J., Gomes, F., and Maenhout, P. (1998). Consumption and Portfolio Choice over the Life-Cycle. Center for Economic Studies - Discussion papers ces9805, Katholieke Universiteit Leuven, Centrum voor Economische Studien.
- College Board Advocacy and Policy Center (2010). Trends in College Pricing 2010. *Trends in Higher Education Series*.
- College Measures (2013). Improving Higher Education Outcomes in the United States. <http://collegemeasures.org>.
- Commonfund Institute (2009). 2009 HEPI: Higher Education Price Index.
- Cunha, J. M. and Miller, T. (2012). Measuring Value-Added in Higher Education: Possibilities and Limitations in the Use of Administrative Data. *Working Paper*.
- Curs, B. R., Singell, L. D., and Waddell, G. R. (2005). Money for Nothing? The Institutional Impact of Changes in Federal Financial Aid Policy. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=616784.
- Dale, S. and Krueger, A. B. (2011). Estimating the Return to College Selectivity over the Career Using Administrative Earnings Data. Working Paper 17159, National Bureau of Economic Research.
- Dale, S. B. and Krueger, A. B. (1999). Estimating the Payoff to Attending a More Selective College: An Application of Selection on Observables and Unobservables. Working Paper 7322, National Bureau of Economic Research, Inc.
- Davis, S. J. and von Wachter, T. M. (2011). Recessions and the Cost of Job Loss. Working Paper 17638, National Bureau of Economic Research.
- Deming, D. and Dynarski, S. (2009). Into College, Out of Poverty? Policies to Increase the Postsecondary Attainment of the Poor. *Targeting Investments in Children: Fighting Poverty When Resources are Limited*.
- Department of Education (2013). Cohort Default Rates. <http://www2.ed.gov/offices/OSFAP/defaultmanagement/index.html>.
- Department of Education: College Affordability and Transparency Center (2013). College Scorecard. <http://www.whitehouse.gov/issues/education/higher-education/college-score-card>.
- Dillon, E. W. (2012). The College Earnings Premium and Changes in College Enrollment. *Working Paper, University of Michigan*.

- Dillon, E. W. and Smith, J. A. (2013). The Determinants of Mismatch Between Students and Colleges. Working Paper 19286, National Bureau of Economic Research.
- Drewes, T. and Michael, C. (2006). How Do Students Choose a University?: An Analysis of Applications to Universities in Ontario, Canada. *Research in Higher Education*, 47(7):781–800.
- Dya, J. C. and Newburger, E. C. (2002). The Big Payoff: Educational Attainment and Synthetic Estimates of Work-Life Earnings. *US Census Bureau Current Population Reports, Special Studies*.
- Dynarski, S. M. (2004). Who Benefits from the Education Saving Incentives? Income, Educational Expectations, and the Value of the 529 and Coverdell. Working Paper 10470, National Bureau of Economic Research, Inc.
- Dhring, B. and Mordonu, A. (2007). What Drives Inflation Perceptions? A Dynamic Panel Analysis. Technical report, European Commission: Directorate-General for Economic and Financial Affairs.
- Economist, T. (2010). Foreign University Students: Will They Still Come? <http://www.economist.com/node/16743639>.
- EDALO Education Promotional Services S.L. (2011). Compare Tuition Fee Schemes in Europe. <http://www.studyineurope.eu/tuition-fees>.
- Elff, M. (2013). mclogit: Mixed Conditional Logit. <http://cran.r-project.org/web/packages/mclogit/index.html>.
- European Commission (2011). Eurostat. <http://epp.eurostat.ec.europa.eu/portal/page/portal/eurostat/home/>.
- Flabbi, L., Paternostra, S., and Tiongson, E. R. (2011). Returns to Education in the Economic Transition: A Systematic Assessment Using Comparable Data. *Economics of Education Review*, 27: 724-240.
- Fortin, N. M. (2006). Higher-Education Policies and the College Wage Premium: Cross-State Evidence from the 1990s. *American Economic Review*, 96(4):959–987.
- Fu, C. (2012). Equilibrium Tuition, Applications, Admissions and Enrollment in the College Market. Working Papers 2012-002, Human Capital and Economic Opportunity Working Group.
- Gourinchas, P.-O. and Parker, J. A. (1999). Consumption Over the Life Cycle. NBER Working Papers 7271, National Bureau of Economic Research, Inc.
- Greene, J. P. (2010). Administrative Bloat at American Universities: The Real Reason for High Costs in Higher Education. *Goldwater Institute Policy Report*.
- Grinblatt, M., Keloharju, M., and Linnainmaa, J. (2009). IQ and Stock Market Participation. *Working Paper*.

- Hacker, A. and Dreifus, C. (2010). *Higher Education?: How Colleges Are Wasting Our Money and Failing Our Kids—and What We Can Do About It*. Times Books, Henry Holt and Company.
- Hämäläinen, U. and Uusitalo, R. (2008). Signalling or Human Capital: Evidence from the Finnish Polytechnic School Reform. *The Scandinavian Journal of Economics*, 10:755–775.
- Harris, W. C. and Beschloss, S. C. (2011). Don't We Want the Best and Brightest to Come and Stay? *Strategic News Service LLC*.
- Harvard University Office of the Provost (2009). Harvard University Factbook.
- Hayashi, F. (2000). *Econometrics*. Princeton University Press.
- Heckman, J. J., Lochner, L. J., and Todd, P. E. (2008). Earnings Functions and Rates of Return. Working Paper 13780, National Bureau of Economic Research.
- Heinrich, G. and Hildebrand, V. (2005). Returns to Education in the European Union: A Reassessment from Comparative Data. *European Journal of Education*, Vol. 40, No. 1.
- Henderson, V. (2005). Explicit Solutions to an Optimal Portfolio Choice Problem with Stochastic Income. *Journal of Economic Dynamics and Control*, 29(7):1237–1266.
- Hight, J. E. (1975). The Demand for Higher Education in the US 1927-72; The Public and Private Institutions. *The Journal of Human Resources*, 10(4):pp. 512–520.
- Hipp, J. R., Marquart-Pyatt, S., and Paxton, P. M. (2011). *Nonrecursive Models: Endogeneity, Reciprocal Relationships, and Feedback Loops*. SAGE.
- Hoekstra, M. (2009). The Effect of Attending the Flagship State University on Earnings: A Discontinuity-Based Approach. *The Review of Economics and Statistics*, 91(4):717–724.
- Hsing, Y. and Chang, H. S. (1996). Testing Increasing Sensitivity of Enrollment at Private Institutions to Tuition and Other Costs. *American Economist*.
- Hurwitz, M. (2012). The Impact of Institutional Grant Aid on College Choice. *Educational Evaluation and Policy Analysis*, 34(3):344–363.
- ICHEFAP (2011). International Comparative Higher Education Finance and Accessibility Project. <http://gse.buffalo.edu/org/IntHigherEdFinance/>.
- John, E. P. S. (1992). Changes in Pricing Behavior During the 1980s: An Analysis of Selected Case Studies. *The Journal of Higher Education*, 63(2):pp. 165–187.
- Johnstone, D. B. (2005). Cost-Sharing and the Cost-Effectiveness of Grants and Loan Subsidies to Higher Education. *Higher Education Dynamics Douro Series*.

- Kim, E. H., Morse, A., and Zingales, L. (2009). Are Elite Universities Losing their Competitive Edge? *Journal of Financial Economics*.
- Klein, M. C. (2013). Still a Long Way to Go. <http://www.economist.com/blogs/freexchange/2012/11/americas-deleveraging>.
- Knutsen, C. H. (2012). Democracy, Technological Change and Economic Growth. Technical report, Department of Political Science, University of Oslo.
- Long, B. T. (2003). The Impact of Federal Tax Credits for Higher Education Expenses. Working Paper 9553, National Bureau of Economic Research, Inc.
- Long, B. T. (2004a). How Do Financial Aid Policies Affect Colleges? The Institutional Impact of the Georgia HOPE Scholarship. *The Journal of Human Resources*, 39(4):pp. 1045–1066.
- Long, B. T. (2006). College Tuition Pricing and Federal Financial Aid: Is There a Connection? *Testimony before the US Senate Committee on Finance, Report Card on Tax Exemptions and Incentives for Higher Education: Pass, Fail, or Need Improvement?*
- Long, T. B. (2004b). How Have College Decisions Changed Over Time? An Application of the Conditional Logistic Choice Model. *Journal of Econometrics*, 121(1-2):271–296.
- Manoli, D. (2008). Education Signaling and the Returns to Schooling. *Working Paper, University of California, Los Angeles & RAND*.
- McCluskey, N. (Sep. 14, 2010). Oh, to Be Politically Favored! *CATO @ Liberty*.
- McFadden, D. (1974). Conditional Logit Analysis of Qualitative Choice Behavior. *Frontiers in Econometrics*.
- Merton, R. C. (1969). Lifetime Portfolio Selection Under Uncertainty: The Continuous-Time Case. *The Review of Economics and Statistics*, 51.
- Merton, R. C. (1971). Optimum Consumption and Portfolio Rules in a Continuous-Time Model. *Journal of Economic Theory*, 3:373–413.
- Mincer, J. (1974). Schooling, Experience and Earnings. *National Bureau of Economic Research*.
- Mortenson, T. (2001). Trends in College Participation by Family Income. *Mortenson Research Seminar on Public Policy*.
- National Center for Education Statistics (2008). Digest of Education Statistics.
- National Center for Education Statistics (2009). Integrated Postsecondary Education Data System.

- Nelson, C. R. and Siegel, A. F. (1987). Parsimonious Modeling of Yield Curves. *The Journal of Business*, 60(4):473–89.
- OECD (1998). Paying for Tertiary Education: The Learner Perspective. *Education Policy Analysis*.
- OECD (2011). OECD.StatExtracts. <http://stats.oecd.org/Index.aspx>.
- Pandit, S. and Wu, S.-M. (1990). *Time Series and System Analysis with Applications*. Robert E. Krieger Publishing Company.
- Park, H. H. (2009). Joint Modeling of Dynamic and Cross-Sectional Heterogeneity: Introducing Hidden Markov Panel Models. Technical report, Department of Political Science, University of Chicago.
- PayScale.com (2013). PayScale Database. <http://www.payscale.com/>.
- Peña, P. A. (2006). Tuition and Wealth at American Colleges. http://economics.missouri.edu/seminars/files/2007/pena_02_06_07.pdf.
- Perna, L. (2006). Studing College Access and Choice: A Proposed Conceptual Model. In Smart, J., editor, *Higer Education*, volume 21 of *Higher Education: Handbook of Theory and Research*, pages 99–157. Springer Netherlands.
- Pilon, M. (Aug. 9, 2010). Student-Loan Debt Surpasses Credit Cards. *The Wall Street Journal*.
- Polachek, S. W. (2007). Earnings Over the Lifecycle: The Mincer Earnings Function and Its Applications. *IZA Discussion Paper Series*, (3181).
- Psacharopoulos, G. (1972). The Economic Returns to Higher Education in Twenty-Five Countries. *Higher Education*, Vol. 1, No. 2, pp 141-158.
- Psacharopoulos, G. and Mattson, R. (1998). Estimating the Returns to Education: A Sensitivity Analysis of Concepts and Methods. *Journal of Educational Planning and Administration*, 12, 3: 271-287.
- R Core Team (2013). glm: stats. <http://stat.ethz.ch/R-manual/R-patched/library/stats/html/glm.html>.
- Rizzo, M. and Ehrenberg, R. G. (2004). *Resident and Nonresident Tuition and Enrollment at Flagship State Universities*. National Bureau of Economic Research, Inc.
- Schumpeter, J. (2010). Declining by degree: Will Americas Universities Go the Way of its Car Companies? *The Economist*.
- Selingo, J. J. (2013). The Diploma’s Vanishing Value. <http://online.wsj.com/article/SB10001424127887324874204578440901216478088.html>.

- Singell, L. and Stone, J. A. (2007). For Whom The Pell Tolls: The Response of University Tuition to Federal Grants-in-Aid. *Economics of Education Review*, 26(3):285–295.
- Standard & Poor’s (2009). Standard & Poor’s/Case-Shiller Home Price Indices.
- The Chronicle of Higher Education (2013). College Completion. <http://collegecompletion.chronicle.com/>.
- The College Board (2009). 2009 Trends in Student Aid. http://trends.collegeboard.org/student_aid.
- The College Board (2012a). Education Pays. <http://trends.collegeboard.org/sites/default/files/education-pays-2010-full-report.pdf>.
- The College Board (2012b). Trends in College Pricing. <http://trends.collegeboard.org/sites/default/files/college-pricing-2012-full-report-121203.pdf>.
- The College Board (2012c). Trends in Student Aid. <http://trends.collegeboard.org/sites/default/files/student-aid-2012-full-report-130201.pdf>.
- Therneau, T. (2013). survival: Survival Analysis. <http://cran.r-project.org/web/packages/survival/index.html>.
- UNESCO (2011). Data centre. <http://stats.uis.unesco.org/unesco/tableviewer/document.aspx?ReportId=143>.
- UNESCO (2013). Institute for Statistics. <http://www.uis.unesco.org/Education/Pages/international-standard-classification-of-education.aspx>.
- US Census Bureau and Bureau of Labor Statistics (2009). Current Population Survey.
- US Federal Reserve (2009). Flow of Funds Accounts of the United States.
- US Federal Reserve (2011). Federal Reserve Interest Rates.
- U.S. News and World Report (2012). How U.S. News Calculates the College Rankings. <http://www.usnews.com/education/best-colleges/articles/2012/09/11/how-us-news-calculates-its-best-colleges-rankings>.
- Vedder, R. (2010). Why Did 17 Million Students Go to College? *The Chronicle of Higher Education*.
- World Bank (2011). Historical Gross Domestic Product (GDP) and Growth Rates of GDP. *World Development Indicators*.
- Yale University Office of Institutional Research (2009). Yale Book of Numbers.