Essays On Student Loans, Higher Education And Inequality

Marc Folch Cordoncillo
University of Pennsylvania

Follow this and additional works at: https://repository.upenn.edu/edissertations

Part of the Educational Administration and Supervision Commons, and the Labor Economics Commons

Recommended Citation
https://repository.upenn.edu/edissertations/3724

This paper is posted at ScholarlyCommons. https://repository.upenn.edu/edissertations/3724
For more information, please contact repository@pobox.upenn.edu.
Essays On Student Loans, Higher Education And Inequality

Abstract
This dissertation consists of two essays at the intersection of the economics of education and inequality. The first chapter analyzes how increasing levels of student debt affect career and housing choices of bachelor's degree recipients in the United States. Using within-cohort across-school variations in financial aid policies, it shows that higher student debt balances cause a front loading of earnings, lower earnings growth, lower graduate school enrollment, and earlier entry into home ownership. Then, a life-cycle model is estimated to analyze the mechanisms behind the interaction between student debt, career and housing choices. The structural model shows that post-school credit constraints generate a trade-off between career and housing choices for highly indebted graduates, playing a key role in explaining lifetime earnings inequality. Relative to the baseline 10-year fixed repayment plan, an income based repayment plan (or a more ambitious student loan forgiveness plan) increases human capital accumulation and earnings growth, while postponing entry into home ownership. The second chapter studies the evolving role of the higher education market in shaping earnings inequality in the United States. Using detailed institution-level data linked to administrative students earnings records, I document an increasing trend in post-school earnings inequality among students attending different four-year colleges and universities between 1997 and 2009. I then estimate the school quality production function as a composite of instructional expenditure per student and average (and dispersion of) students ability. The point estimates suggest that college readiness has a higher weight on explaining variation in post-school earnings relative to expenditure per student. Finally, I show a growing variation in both quality inputs, consistent with an increase in inequality in net tuition revenue and increasing relative demand for private non-profit doctoral universities.

Degree Type
Dissertation

Degree Name
Doctor of Philosophy (PhD)

Graduate Group
Economics

First Advisor
Jesus Fernandez-Villaverde

Subject Categories
Economics | Educational Administration and Supervision | Labor Economics

This dissertation is available at ScholarlyCommons: https://repository.upenn.edu/edissertations/3724
ESSAYS ON STUDENT LOANS, HIGHER EDUCATION AND INEQUALITY

Marc Folch Cordoncillo

A DISSERTATION

in

Economics

Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2021

Supervisor of Dissertation

______________________________
Jesús Fernández-Villaverde, Professor of Economics

Graduate Group Chairperson

______________________________
Jesús Fernández-Villaverde, Professor of Economics

Dissertation Committee

Dirk Krueger, Professor of Economics

Holger Sieg, Professor of Economics
Dedicated to my family
ACKNOWLEDGEMENT

I would like to thank my advisor Jesús Fernández-Villaverde and the members of my committee Dirk Krueger and Holger Sieg my deepest gratitude for their invaluable guidance and advice. I feel very lucky to have encountered such an accessible and caring group of professors.

Many others have provided very helpful feedback and advice in different parts of this thesis. Harold L. Cole, André Victor Doherty Ludvice, Alistair Macaulay, Paolo Martellini and José-Víctor Ríos-Rull deserve a special mention for various suggestions and contributions. I am also deeply thankful to my co-author Luca Mazzone.

I also wish to thank seminar participants at the Center for the Economics of Human Development, the European Economic Association Meeting, the Young Economist Symposium Conference, the Economics Graduate Student Conference at the Washington University in St. Louis, the 2nd Bank of Italy Human Capital Workshop, and the University of Pennsylvania Economics Department for insightful comments and discussions.

During these years, I was supported by the University of Pennsylvania, "la Caixa" Foundation postgraduate fellowship and the Federal Reserve Bank of Philadelphia through my Research Assistantship for Andrew Hertzberg. I want to express my gratitude for their trust.

Completing this dissertation would not have been possible without the continuous love and support of my friends and family.
ABSTRACT

ESSAYS ON STUDENT LOANS, HIGHER EDUCATION AND INEQUALITY

Marc Folch Cordoncillo

Jesús Fernández-Villaverde

This dissertation consists of two essays at the intersection of the economics of education and inequality. The first chapter analyzes how increasing levels of student debt affect career and housing choices of bachelor’s degree recipients in the United States. Using within-cohort across-school variations in financial aid policies, it shows that higher student debt balances cause a front loading of earnings, lower earnings growth, lower graduate school enrollment, and earlier entry into home ownership. Then, a life-cycle model is estimated to analyze the mechanisms behind the interaction between student debt, career and housing choices. The structural model shows that post-school credit constraints generate a trade-off between career and housing choices for highly indebted graduates, playing a key role in explaining lifetime earnings inequality. Relative to the baseline 10-year fixed repayment plan, an income based repayment plan (or a more ambitious student loan forgiveness plan) increases human capital accumulation and earnings growth, while postponing entry into home ownership. The second chapter studies the evolving role of the higher education market in shaping earnings inequality in the United States. Using detailed institution-level data linked to administrative students earnings records, I document an increasing trend in post-school earnings inequality among students attending different four-year colleges and universities between 1997 and 2009. I then estimate the school quality production function as a composite of instructional expenditure per student and average (and dispersion of) students ability. The point estimates suggest that college readiness has a higher weight on explaining variation in post-school earnings relative to expenditure per student. Finally, I show a growing variation in both quality inputs, consistent with an increase in inequality in net tuition revenue and increasing relative demand for private non-profit doctoral universities.
# TABLE OF CONTENTS

**ACKNOWLEDGEMENT** ................................................................. iv

**ABSTRACT** ............................................................................. v

**LIST OF TABLES** ................................................................. vii

**LIST OF ILLUSTRATIONS** ................................................... ix

**CHAPTER 1 : Go Big or Buy a Home: Student Debt, Career Choices and Wealth Accumulation** .................................................. 1

- 1.1 Introduction ................................................................. 1
- 1.2 Related Literature ....................................................... 6
- 1.3 Data and Empirical Analysis ........................................... 9
- 1.4 The Model ................................................................. 19
- 1.5 Calibration and Estimation ............................................. 30
- 1.6 Results ................................................................. 40
- 1.7 Policy ................................................................. 47
- 1.8 Conclusions ............................................................. 52

**CHAPTER 2 : Upskilling and the Rise in College Inequality** .............. 54

- 2.1 Introduction ............................................................... 54
- 2.2 Upskilling ............................................................... 57
- 2.3 Data: College Scorecard ............................................... 61
- 2.4 The Increase in College Inequality ................................. 67
- 2.5 Conclusions ............................................................. 76

**APPENDIX** ................................................................. 77

**BIBLIOGRAPHY** ............................................................. 111
LIST OF TABLES

TABLE 1.1 : Summary Statistics ........................................... 11
TABLE 1.2 : First Stage Regression ........................................ 16
TABLE 1.3 : Student Debt and Post-baccalaureate Choices ............... 18
TABLE 1.4 : External Parameters ........................................... 31
TABLE 1.5 : Target Moments ................................................ 36
TABLE 1.6 : Estimated Parameters ......................................... 38
TABLE 1.7 : Student Debt, Earnings, Career and Housing ............... 41
TABLE 1.8 : Sorting into Post-Bachelor Degrees ......................... 42
TABLE 1.9 : Entry into Home Ownership .................................. 45
TABLE 1.10 : Enrollment with and without Housing ...................... 46
TABLE 2.1 : Summary Statistics (4-year non-profit colleges) ............ 64
TABLE 2.2 : College Quality Production Function Estimates ............. 71
TABLE 2.3 : Carnegie Classification (2018) .............................. 75
TABLE A1.2 : Homeownership by Educational Attainment (1965-1979 cohorts) . 96
TABLE A2.1 : Gini Coefficient (by states) ................................. 108
TABLE A2.2 : College Quality Estimates (ACT and by sector) .......... 109
TABLE A2.3 : College Quality Estimates (graduation rate) ............. 110
### LIST OF ILLUSTRATIONS

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>FIGURE 1.1</td>
<td>Percentage and Cumulative Amount Borrowed for BA Degree</td>
<td>2</td>
</tr>
<tr>
<td>FIGURE 1.2</td>
<td>Earnings and Homeownership by Educational Attainment</td>
<td>10</td>
</tr>
<tr>
<td>FIGURE 1.3</td>
<td>Institutional Grants and Student Loans</td>
<td>14</td>
</tr>
<tr>
<td>FIGURE 1.4</td>
<td>Change in Grant to Aid</td>
<td>15</td>
</tr>
<tr>
<td>FIGURE 1.5</td>
<td>Indirect Utility</td>
<td>21</td>
</tr>
<tr>
<td>FIGURE 1.6</td>
<td>Interest Rate on Student Loan and Mortgage Debt</td>
<td>32</td>
</tr>
<tr>
<td>FIGURE 1.7</td>
<td>Earnings and Enrollment Life Cycle Profiles (model and data)</td>
<td>39</td>
</tr>
<tr>
<td>FIGURE 1.8</td>
<td>Homeownership Life Cycle Profiles (by career and debt levels)</td>
<td>40</td>
</tr>
<tr>
<td>FIGURE 1.9</td>
<td>Decomposing Earnings Growth: Low versus High Debt</td>
<td>43</td>
</tr>
<tr>
<td>FIGURE 1.10</td>
<td>Life-Cycle Earnings (Baseline vs. no Homeownership)</td>
<td>47</td>
</tr>
<tr>
<td>FIGURE 1.11</td>
<td>Evolution of Repayment Plans</td>
<td>48</td>
</tr>
<tr>
<td>FIGURE 1.12</td>
<td>Baseline vs. Alternative Repayment Plans</td>
<td>51</td>
</tr>
<tr>
<td>FIGURE 2.1</td>
<td>Earnings inequality (college-educated workers)</td>
<td>55</td>
</tr>
<tr>
<td>FIGURE 2.2</td>
<td>College Enrollment and Attainment</td>
<td>57</td>
</tr>
<tr>
<td>FIGURE 2.3</td>
<td>High School Graduates and Financial Aid</td>
<td>58</td>
</tr>
<tr>
<td>FIGURE 2.4</td>
<td>Distribution of Occupations (by BA requirements)</td>
<td>61</td>
</tr>
<tr>
<td>FIGURE 2.5</td>
<td>Distribution of Outcomes (4-year colleges, 2014)</td>
<td>65</td>
</tr>
<tr>
<td>FIGURE 2.6</td>
<td>Post-college Earnings Inequality (4-year colleges)</td>
<td>66</td>
</tr>
<tr>
<td>FIGURE 2.7</td>
<td>Earnings, Expenditure per Student, ACT Score (2014)</td>
<td>69</td>
</tr>
<tr>
<td>FIGURE 2.8</td>
<td>Evolution of College Ratios</td>
<td>74</td>
</tr>
<tr>
<td>FIGURE A1.1</td>
<td>Net Cost of College</td>
<td>87</td>
</tr>
<tr>
<td>FIGURE A1.2</td>
<td>Student Debt Burden</td>
<td>87</td>
</tr>
<tr>
<td>FIGURE A1.3</td>
<td>Graduate School Attainment</td>
<td>88</td>
</tr>
<tr>
<td>FIGURE A1.4</td>
<td>Model Identification</td>
<td>89</td>
</tr>
</tbody>
</table>
FIGURE A1.5: Graduate School Educated Workers and Downpayment Constraint 95
FIGURE A2.1: College Enrollment and Attainment (2-year colleges) 100
FIGURE A2.2: Earnings Inequality (10 years after enrollment) 100
FIGURE A2.3: Gini Coefficient (by region) 101
FIGURE A2.4: Earnings, Expenditure per Student, ACT Score (2014) 102
FIGURE A2.5: Evolution of College Ratios (without upper and lower percentile) 103
FIGURE A2.6: Evolution of College Ratios (doctoral vs. non-doctoral) 104
FIGURE A2.7: Evolution of College Ratios (without grouping by sector) 105
FIGURE A2.8: International Students Enrollment 106
FIGURE A2.9: Evolution of Endowment per Student (4-year colleges) 107
CHAPTER 1 : Go Big or Buy a Home: Student Debt, Career Choices and Wealth Accumulation

This chapter is co-authored with Luca Mazzone.

1.1. Introduction

Student loans have become the new normal for bachelor’s degree recipients in the United States\(^1\). Between 1993 and 2016, the percentage of students who had borrowed at any time during their undergraduate years rose from 45 percent for 1993 graduates to 68 percent for 2016 graduates (Figure 1.1). Among borrowers, the median cumulative amount borrowed rose from $13,000 to $27,000 in real terms, with 25% of graduating seniors having borrowed more than $40,000 in 2016. Student borrowing is now more likely to be a burden for a higher percentage of college graduates and a relevant factor they take into account in their economic and financial decisions\(^2\).

In presence of credit constraints, student debt affects post graduation choices by making further borrowing more difficult. We highlight one often overlooked cost of additional investment in human capital for college graduates, that is the postponing of first-time home-ownership. As enrollment in graduate school generally implies increasing student loan balances, the relative value of additional human capital investment decreases throughout the life cycle due to strong horizon effects induced by home-ownership. Thus, when housing is taken into account, the negative impact of student debt on human capital accumulation is amplified, causing lower earnings growth and lifetime earnings for bachelor’s degree recipients with higher levels of student debt.

---

\(^1\)Most of the increase in undergraduate student debt has been attributed to the substantial rising cost of college over the last decade (see Looney and Yannelis (2015)). Between 1993-2016, median real net cost of college increased 3% for 4-year college seniors (see Figure A1.1). Increasing federal loan limits as well as a relaxation of eligibility criteria helped to moderate the impact of higher cost on college enrollment.

\(^2\)Although real incomes for bachelor’s degree recipients rose as well, monthly student loan payment as a percent of monthly income (one year after graduation) also increased between 1993 and 2008 (see Figure A1.2).
We use the Baccalaureate and Beyond Longitudinal Study (B&B), a restricted access dataset compiled by the National Center for Education Statistics. The B&B surveys cover a representative sample of U.S. college graduates interviewed on successive waves.

In order to empirically examine how college borrowing affects earnings, career and housing choices in the years after graduation, we need to overcome notorious identification problems, as the amount borrowed may be determined by unobserved individual characteristics, which in turn would affect all post graduation choices. On the one hand, if low ability students are less likely to receive grants or come from a low-income background, the OLS estimator will reflect the latent negative correlation between ability and student borrowing. On the other hand, students with higher earnings expectations could be more willing to borrow, resulting in debt being positively selected.
We address the identification problem by introducing an instrument based on variations in colleges’ financial aid. Composition of aid at the college level is calculated using public access data from the Integrated Postsecondary Education Data System (IPEDS). We focus on institutional grants, which are funded from net revenues and assets and experience significant variations year-by-year for many institutions. We use these supply changes in financial aid during college enrollment to extract variation in student debt that is not correlated with post bachelor choices through unobserved students characteristics.

Our empirical results suggest that lower net wealth amplifies the trade-off between career and housing choices for college graduates. Higher levels of student debt cause a front loading of earnings, while significantly and persistently deterring additional human capital investment (measured as graduate school enrollment). This, in turn, contributes to lower earnings growth. Indebted graduates earn 0.29% more for each percentage increase in student debt in the first year. Over time, this effect is compensated by wage growth being 0.1% lower. We also find that higher levels of student debt generate an earlier entry into home ownership, consistent with lower graduate school enrollment for college graduates.

We develop a model with endogenous (risky) human capital accumulation enriched by career choices and housing decisions to rationalize the empirical evidence and understand the importance of credit constraints in shaping post graduation outcomes. After graduating from college, individuals enter the labor market and are heterogeneous in innate ability, human capital, liquid wealth and student debt. They sequentially decide on human capital investment, savings, non housing and housing consumption while they pay for student debt. At any point, they can enroll in a post bachelor program and, if they do, take on additional student debt. Workers can buy a house through a 30-year fixed-rate mortgage as long as they make a down-payment and their debt payments do not exceed a proportion of their income.

3More than one third of Bachelor’s degree graduates hold a graduate degree (Master’s, Professional school or Doctorate degree) at age 35-40 (see Figure A1.3) for recent trends in graduate school attainment.
Our theoretical framework implies that, because of introducing debt-funded human capital and post-college housing credit constraints, student debt plays a key role in explaining earnings inequality throughout the life cycle. College graduates with higher student debt balances sort into careers with lower compensation for human capital accumulation, and then experience lower earnings growth but earlier entry into home ownership. As enrollment in graduate school implies more educational debt, indebted college graduates anticipate that investing more in human capital might imply a substantial delay in first-time home-ownership, as they might later face borrowing constraints in the mortgage market.

The model is estimated by Simulated Method of Moments using a combination of data from B&B and Current Population Survey (CPS). Structural estimation highlights substantial non monetary returns to post bachelor education, that yields consumption-equivalent utility consisting of more than $50,000. On the other hand, skill premium for the post bachelor degree educated workers corresponds to 14% of the earnings differential with respect to workers with just a bachelor degree. Individuals with relatively higher ability are thus able to afford the cost of higher education given the mentioned composition of returns, while others postpone or choose the alternative career path.

The model also speaks to the effects of borrowing on home ownership. On the one hand, college graduates who choose a career with a steeper earnings path (as is the case for those who enroll in graduate programs) are at the same time more likely to postpone their investment in housing. On the other hand, workers that choose to remain in a career that implies lower earnings growth consider housing a relatively more attractive investment. Hence, home ownership is relatively higher for young graduates that started their career with more student debt, a finding consistent with our empirical results. Credit constraints in the housing market for highly leveraged households are crucial in determining these results.

These results point to an easy counterfactual exercise that helps highlighting the way in which, conversely, housing affects career sorting. In a setting without home ownership, distortions to human capital accumulation induced by student debt are smaller, enrolment in
post bachelor programs is higher, and lifetime earnings inequality decreases. The main reason for this effect is that postponing homeownership is costly: higher debt forces graduates to postpone additional education, or to invest less in human capital to accumulate higher savings. While doing so, workers realize that enrolling at a later age would mean a further postponement of homeownership, as the downpayment and/or debt-to-income constraints would likely bind for an even longer period of time, and choose to give up on additional education and choose careers with a flatter income profile.

Finally, we use the model to evaluate the impact of the recent widespread adoption of income based repayment plans (IRP) and compare it to a forgiveness plan. We find that the introduction of the IRP provides the foundation for reducing the unintended consequences of student loan debt. By lowering an individual’s monthly payments, IRP provides a consumption smoothing mechanism that reduces the need to choose a higher paying job.

The implementation of a forgiveness plan and the widespread adoption IRP yield similar outcomes, both increasing enrollment in graduate programs and earnings over the life cycle. In both cases, however, alleviating the debt burden does imply a delay in homeownership. This is particularly true in the case of IBR, due to a combination of increased post bachelor degree attendance, longer time horizon for debt repayment, and higher overall repayments for borrowers with higher balances.

The paper is organized as follows. Section 1.2 summarizes the literature, Section 1.3 describes the data and presents the empirical results, Section 1.4 provides an overview of the model and the life cycle choices of individuals, Section 1.5 calibrates and estimates the model to observed data patterns, Section 1.6 presents the main results of the model, Section 1.7 analyzes policy counterfactuals, and, Section 1.8 concludes with some policy discussion and future work.
1.2. Related Literature

The empirical evidence on how student debt affects earnings mostly points to a positive relationship, at least in the short run\(^4\). Based on a natural experiment in an elite university, Rothstein and Rouse (2011) show that student debt causes college graduates to choose jobs with an initial higher salary and reduces the probability that they choose "public" low paid jobs. Luo and Mongey (2019) find that a version of these results generalizes to the cross section of the U.S. colleges. In particular, they find that higher student debt causes college graduates to take jobs with higher wages, lower job satisfaction, and more on the job search.

Using a difference-in-difference approach, Gerald and Smythe (2019) study the impact of student debt on various labor market outcomes (income, hourly wages, and hours worked). They conclude that indebted students have initial higher earnings due to higher work hours rather than higher wage rates. Chapman (2015) finds that exogenously increasing the loan burden of a college graduate by $1,000 increases their income by $400-$800 one year after graduation. Field (2009) shows that law students who were offered loans were more likely to accept jobs in higher paying corporate law rather than public interest law.

Nonetheless, higher initial earnings may not necessarily lead to higher lifetime earnings if they are not followed by further human capital investment (Becker (1962), Ben-Porath (1967), Hause (1972) and Mincer (1974)). In this line of thought, Fos, Liberman and Yannelis (2017) investigate the effects of student debt on additional human capital investment measured as graduate school enrollment. They find that a $4,000 increase in student debt reduces the likelihood of enrollment in graduate school by 1.5 percentage points. We contribute to this empirical literature by analyzing the impact of student debt on both earnings and educational outcomes one, four and ten years after graduation for a nationally representative sample of college graduates.

\(^4\)For empirical studies that conclude a negative or neutral effect of student debt on earnings see: Weidner (2016), Akers (2012), Zhang (2013).
Another set of empirical articles have analyzed the role of student loans on first time home ownership. Controlling for multiple factors, [Houle and Berger (2015), Cooper and Wang (2014) and Gicheva and Thompson (2014)] show that student debt reduces the likelihood of homeownership for young households.[5] Using variations in tuition for public four-year colleges, [Bleemer et al. (2020)] find that the recent increase in student debt could explain between 11 and 35 percent of the decline in young’s homeownership over 2007-2015.

Using a similar approach, [Mezza et al. (2020)] estimate that a $1,000 increase in student debt decreased first time homeownership by approximately 1.5 p.p. for public 4-year college graduates who left school between 1997 and 2005. Differently from this literature, we focus on the impact of student debt on first-time home-ownership for bachelor’s degree recipients,[6] and show that, consistent with a delay in graduate school enrollment, the initial average impact of student debt on homeownership is positive (while the effect is negative within educational groups).

Our analysis also relates to the literature that study student loan program design within a quantitative framework. For example, [Ionescu (2009)] finds that repayment flexibility increases college enrollment significantly, whereas relaxation of eligibility requirements has little effect on enrollment or default rates. In a similar framework, [Ionescu and Simpson (2016)] find that tuition subsides increase aggregate welfare by increasing college investment and reducing default rates in the private market. [Johnson (2013)] also shows that tuition subsidies provide larger increases in college enrollments than increasing borrowing limits. Compared to this literature, our model provides a more detailed characterization of college graduates career and educational choices and post schooling consumption and housing decisions.

---

5 Household formation and homeownership choices are strongly linked. For example, [Sieg and Wang (2018)] show that student debt has negative effects on marriage prospects of young female lawyers.

6 In [Mezza et al. (2020)], only 11% of the sample holds a bachelor’s degree, while 50% have no college degree.
In a related paper, Di Maggio, Kalda and Yao (2019) examine the effect of student debt forgiveness on individual credit and labor market outcomes. Using hand collected lawsuits filings matched with individual credit bureau information, they find that borrowers experiencing the debt relief shock reduce their overall indebtedness by 26%. They also find that borrowers’ probability to change jobs increase after the discharge and this leads to an increase in earnings by more than $4000 over a three year period. We examine the effects of an hypothetical student debt forgiveness plan on graduate school enrollment, lifetime earnings and home-ownership choices.

Finally, our paper relates to the quantitative literature that analyzes how initial conditions affect lifetime earnings inequality. In particular, this literature focuses on the importance of initial conditions relative to shocks over the life cycle. Huggett, Ventura and Yaron (2011) study how heterogeneity in initial wealth and human capital affect lifetime inequality by modelling earnings growth through a Ben-Porath production function. They find that initial conditions, as measured at age 23, determine more than 60 percent of variation in lifetime utility, and that the majority of this variation is determined by initial human capital differences.

The role of initial conditions in shaping long term human capital accumulation has been addressed in the search and matching literature as well. Using a model with directed search and heterogeneous asset holdings, Griffy (2019) finds that initial wealth plays a crucial role in determining life cycle inequality, and heterogeneity in skills has a relatively smaller impact. This difference is caused by the inclusion of frictional labor markets, which makes wealth have a first order effect on earnings. In a similar vein, Eeckhout and Sepahsalar (2019) show that there is positive sorting between workers with net asset holdings and more productive firms. Differently from this strand of literature, we model the labor market as career paths with different (debt-funded and risky) human capital requirements and include housing as a mechanism through which career choices interact with post-college credit constraints.
1.3. Data and Empirical Analysis

1.3.1. Description of Data

Our main source of data comes from the restricted use dataset from the National Center for Education Statistics (NCES) Baccalaureate and Beyond Survey (B&B). The survey follows several cohorts of bachelor’s degree recipients over time and contains a mix of administrative and self-reported data about their income, student debt, occupation, graduate school enrollment and homeownership (among other variables).

B&B draws its cohorts from the National Postsecondary Student Aid Study (NPSAS), which collects data from large, nationally representative samples of postsecondary students and institutions to examine how students pay for postsecondary education. B&B samples are representative of graduating seniors in all majors and colleges. Our analysis focuses on the most recently available cohort (2007/08), which was followed up one and four years after graduation and was interviewed again in 2018 (forthcoming). We also use as robustness students that graduated from college in 2016 and were followed one and four years after graduation (forthcoming).

We restrict the sample to traditional college students: students who attended only one college, enrolled between 2002 and 2004, and graduated at age 21-24. In terms of colleges, we focus on four year public and private non-profit colleges, excluding private for-profit and special focus institutions. After imposing these restrictions, we also remove all colleges for which we do not have more than 3 students - this is necessary since we use an instrument that is based on college level variation and we need enough students per college for the sample to be representative.

Table 1.1 provides the main statistics for the whole sample and for the restricted sample. The table also provides CPS statistics for individuals with at least a BA degree and aged 22-25 in 2009 and 25-28 in 2012. Measures of earnings for Baccalaureate & Beyond for college
graduates are similar to the ones in Census: the average earning for a college graduate in the restricted sample was $30,282 1 year after college, while $41,380 four years after graduation. Around 40 percent of college graduates owned a house and 22 percent had a Graduate Degree by 2012.

In order to understand the role of career sorting and its relationship with earnings and entry into home ownership, we characterize career paths based on workers’ educational attainment using the Current Population Survey (CPS). As expected, college graduates with a graduate degree typically have higher earnings growth (Figure 1.2a). However, initial earnings are similar to those with a Bachelor’s degree (with a high percentage of post-BA graduates taking on graduate student debt). Indeed, educational choices are inextricably linked with housing choices: Figure 1.2b shows that workers investing more in human capital have significantly lower initial home ownership rate.7

Figure 1.2: Earnings and Homeownership by Educational Attainment

(a) Median Earnings  (b) Homeownership

Source: Current Population Survey (CPS, 2005-2020) - males with at least a Bachelor’s Degree. Workers with a graduate degree (in red) and with a bachelor’s degree (in blue). Lines represent smoothed trend.

7 Table [A1.1] shows the average marginal effect of graduate school attainment on first-time home-ownership (estimated with a probit model). Although higher mobility rates might be a contributing factor to lower home-ownership for graduate degree recipients, the estimated marginal effects are lower for older cohorts with lower amounts of student debt (Table [A1.2]).
### Table 1.1: Summary Statistics

<table>
<thead>
<tr>
<th></th>
<th>B&amp;B 08/09/12</th>
<th>IPEDS</th>
<th>CPS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Full Panel</td>
<td>Restricted Panel</td>
<td>Full Sample</td>
</tr>
<tr>
<td><strong>Institutional Characteristics</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Public, non-doc.</td>
<td>16%</td>
<td>14%</td>
<td>18%</td>
</tr>
<tr>
<td>Public, doc.</td>
<td>43%</td>
<td>49%</td>
<td>44%</td>
</tr>
<tr>
<td>Private non-profit, non-doc</td>
<td>15%</td>
<td>13%</td>
<td>15%</td>
</tr>
<tr>
<td>Private non-profit, doc</td>
<td>17%</td>
<td>20%</td>
<td>18%</td>
</tr>
<tr>
<td>Other</td>
<td>10%</td>
<td>-</td>
<td>6%</td>
</tr>
<tr>
<td><strong>Student Characteristics</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Percentage Indebted</td>
<td>66%</td>
<td>67%</td>
<td>65%</td>
</tr>
<tr>
<td>Total debt balance</td>
<td>$ 25,048</td>
<td>$ 23,017</td>
<td>$ 24,900</td>
</tr>
<tr>
<td>Dependent</td>
<td>61%</td>
<td>67%</td>
<td></td>
</tr>
<tr>
<td>Expected Family Contribution</td>
<td>$ 16,309</td>
<td>$ 18,214</td>
<td></td>
</tr>
<tr>
<td>SAT</td>
<td>1,081</td>
<td>1,091</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>43%</td>
<td>45%</td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>73%</td>
<td>80%</td>
<td></td>
</tr>
<tr>
<td><strong>Outcomes</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>1 year after graduation</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Current primary job salary</td>
<td>$ 34,554</td>
<td>$ 30,282</td>
<td>$ 30,362</td>
</tr>
<tr>
<td><strong>4 years after graduation</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Current primary job salary</td>
<td>$ 44,975</td>
<td>$ 41,380</td>
<td>$ 42,110</td>
</tr>
<tr>
<td>Home-ownership</td>
<td>38%</td>
<td>40%</td>
<td>43%</td>
</tr>
<tr>
<td>With a Graduate Degree</td>
<td>22%</td>
<td>22%</td>
<td>20%</td>
</tr>
<tr>
<td>College Obs.</td>
<td>1,440</td>
<td>750</td>
<td></td>
</tr>
<tr>
<td>Individual Obs.</td>
<td>14,250</td>
<td>7,030</td>
<td></td>
</tr>
</tbody>
</table>

Source: Baccalaureate and Beyond Longitudinal Study (B & B 2008/2012), CPS 2009-2012, and IPEDS.
1.3.2. Empirical Specification

The relationship between debt and post college outcomes can be expressed in the following reduced form Equation:

\[ y_{i,c,t+\tau} = \alpha_c + \beta d_{i,c,t} + \Gamma w_{i,c,t} + \epsilon_{i,c,t+\tau} \]  

(1.1)

where \( y_{i,c,t+\tau} \) is the individual’s outcome \( \tau \) years after graduating from college \( c \) in year \( t \), \( \alpha_c \) is a vector capturing college characteristics, \( d_{i,c,t} \) is the log of the cumulative amount of loans (federal and private) borrowed for undergraduate degree at time of graduation \( t \), and \( w_{i,c,t} \) is a vector of individual level observables.

With few students per college, college fixed effects \( (\alpha_c) \) cannot be precisely estimated. We therefore group colleges into different categories based on their sector (public or private non-profit), selectivity (minimally, moderately, very selective), and their Carnegie Classification (Doctorate granting Universities, Master’s or Baccalaureate Colleges). This groups together similar colleges, controlling for college quality. We cluster standard errors at the college group level.

In addition, we also include a rich set of individual controls \( (\Gamma) \). We use individuals characteristics that are included in the FAFSA financial aid application form (expected family contribution and dependency status), the year they started college (i.e. whether students spent 4, 5 or 6 years in college), gender and ethnicity. We also include the student college entrance test score (SAT) and the major of study in order to account for student ability. Finally, we add the ratio of the value of total institutional grants issued by the college to the sum of grants and student loans in the year the student enrolled in college \( (x_{c,t0(i)}) \) to control for financial aid availability (see Section 1.3.3).

\(^8\)See Appendix A1.1 for more details about how these variables are defined.
Nonetheless, unobserved students characteristics could still be relevant in determining access to different forms of aid and have a direct impact on students’ post baccalaureate decisions; this makes $d_{i,c,t}$ a potential endogenous variable, and thus, the OLS estimator of Equation 1.1 could be biased. The bias could go in either direction. On the one hand, if low ability students are less likely to receive grants or come from a low-income background, $\beta$ will reflect the latent negative correlation between ability and student borrowing. On the other hand, students with higher earnings expectations could be more willing to borrow, resulting in debt being positively selected.

1.3.3. Instrumental Variable: Institutional Grants

In this section, we show that supply side variations in the financial aid options faced by all students in a particular college offer a way to overcome the identification problem. In practice, students usually receive a year-by-year financial aid package that is determined by college financial aid officers, but is not known in advance at the time of application. It includes student loans, scholarships, and, grants from the government and the institution itself.

Differently from government grants and student loans, institutional grants are funded from net revenues and assets of the institution. Since institutional grants do not require repayment, they are preferred to loans and are the first to be added into a financial aid package. Loans are therefore the marginal source of funds to most students. In order to capture this substitution between institutional grants and loans, we follow Luo and Mongey (2019) and compute the ratio of the value of total institutional grants issued by the college to the sum of grants and student loans (grant to aid henceforth):

$$x_{c,t} = \left( \frac{\text{inst.grant}_{c,t}}{\text{inst.grant}_{c,t} + \text{loan}_{c,t}} \right) \times 100$$
Figure 1.3: Institutional Grants and Student Loans

(a) Public 4y Schools  (b) Private non-profit 4y Schools

Figure 1.3 shows how variations in the grant to aid ratio capture the substitution between the two measures of funding for both public and private non-profit colleges.

Nevertheless, the exclusion restriction may be violated if the grant to aid ratio is correlated with unobserved students characteristics and those characteristics have a direct impact on students’ post baccalaureate decisions. This may happen because students are not randomly assigned to a college and they choose college based on a bundle of college characteristics, which include financial aid availability.

In order to reduce this source of bias, we exploit variation in grant to aid policies during college enrollment. These changes (Figure 1.4) are likely unexpected for the student at the enrollment stage as they come from changes in the net revenues and assets of the university.
We thus take as our instrument \((z_{c,t})\) the average variation in grant to aid during college enrollment (where \(t_0(i)\) represents the year when the individual first enrolled in college \(c\)):

\[
z_{c,t} = \bar{x}_{c,t} - x_{c,t_0(i)} = \frac{\sum_{t=t_0(i)+1}^{t=T} x_{c,t}}{T - t_0(i) - 1} - x_{c,t_0(i)}
\]

The amount of college debt (first stage regression) is modeled as an outcome of individual demand for debt and these supply side college variations in institutional grants, represented by:

\[
d_{i,c,t} = \mu_c + \delta z_{c,t} + \Pi w_{i,c,t} + u_{i,c,t}
\]
We estimate the model by two-stage least squares regression (for earnings) and two-stage probit model (for graduate school attainment and homeownership). Therefore, it is important that there is significant variation of the instrument with student debt across institutions.

Table 1.2 shows that such condition is satisfied. The results imply that, on average, a 10 percentage points increase in grant to aid during college enrollment induces a corresponding 30% decrease in cumulative student debt at graduation, all else equal.

<table>
<thead>
<tr>
<th></th>
<th>log(debt\textsubscript{2008})</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td></td>
</tr>
<tr>
<td>∆ Grant to Aid</td>
<td>-0.031***</td>
</tr>
<tr>
<td></td>
<td>[0.001]</td>
</tr>
<tr>
<td>Grant to Aid (at enrollment)</td>
<td>-0.046***</td>
</tr>
<tr>
<td></td>
<td>[0.001]</td>
</tr>
<tr>
<td>Student Controls</td>
<td>Y</td>
</tr>
<tr>
<td>College Group FE</td>
<td>Y</td>
</tr>
<tr>
<td>Observations</td>
<td>7,030</td>
</tr>
</tbody>
</table>

Source: U.S. Department of Education, National Center for Education Statistics, Integrated Postsecondary Education Data System (IPEDS 2007/08) and Baccalaureate and Beyond Longitudinal Study (B & B 2008/2012). Colleges are weighted by total full-time first-time students.
1.3.4. Empirical Results

This section presents our main empirical results. We show the causal effect of student loans on earnings one, four and ten (forthcoming) years after college graduation. At the same time, we analyze the role of debt on other post baccalaureate choices intrinsically linked to earnings such as graduate school attendance and housing. We use the estimation strategy proposed in the previous section, that is, using variation in student debt coming from colleges that changed grant to aid policies while our sample of students were enrolled.

Results from the estimation of Equation (1.1) on earnings are given in the first two columns of Table 1.3. The first column shows the OLS and IV estimates for earnings one year after bachelor’s degree completion. Column (1) implies that, on average, increasing a student’s debt by 30% (roughly a $7,000 increase) would lead to an increase in annual earnings of 9%. Column (2) runs the same equation on earnings growth 4 years after graduation and shows that the effect turns significantly negative: increasing a student’s debt by 30% would lead to a reduction in earnings growth of 3%.

The front loading effect of borrowing on earnings is thus consistent with the hypothesis that highly indebted graduates need to boost their initial earnings to ease the burden of repaying their loans. The negative growth on earnings might point to indebted workers under-investing in additional education after college.

Columns (3) and (4) show the relationships between student debt and post baccalaureate education. Column (3) shows the average marginal effects of debt on post BA attainment 4 years after graduation. Ceteris paribus, a 30% increase in college debt reduces the probability of having a post baccalaureate degree four years after graduation by 1.5 percentage points. In addition, Column (4) shows the average marginal effect of debt on cumulative post BA enrollment since college graduation. As we can see, more indebted students do not catch up on graduate enrollment even four years after graduation.
Column (5) shows the average marginal effect of debt on first-time homeownership: a 30% increase in college debt increases the probability of being home owner by 1.5 p.p. (Column (5)). Such a positive relationship with housing is consistent with indebted students working more initially after graduating and under-investing in (risky debt-funded) additional human capital.

Table 1.3: Student Debt and Post-baccalaureate Choices

<table>
<thead>
<tr>
<th></th>
<th>Earnings</th>
<th>Education (t+4)</th>
<th>Housing (t+4)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>log(wage)</td>
<td>Growth</td>
<td>Graduate</td>
</tr>
<tr>
<td>t+1</td>
<td>t+4</td>
<td>Completion</td>
<td>Enrollment</td>
</tr>
<tr>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
</tr>
<tr>
<td>OLS / Probit</td>
<td>0.012</td>
<td>-0.010</td>
<td>-0.001</td>
</tr>
<tr>
<td></td>
<td>[0.013]</td>
<td>[0.030]</td>
<td>[0.001]</td>
</tr>
<tr>
<td>IV</td>
<td>0.298***</td>
<td>-0.102***</td>
<td>-0.054***</td>
</tr>
<tr>
<td></td>
<td>[0.126]</td>
<td>[0.029]</td>
<td>[0.015]</td>
</tr>
<tr>
<td>Student Controls</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>College Group FE</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Observations</td>
<td>7,030</td>
<td>7,030</td>
<td>7,030</td>
</tr>
</tbody>
</table>

Standard errors, clustered by college groups, in brackets.

1.4. The Model

Our empirical estimates show that more indebted graduates front load earnings and give up graduate school, but tend to access home ownership relatively earlier. How is this possible? The crucial factor, as we will show, is that imperfect credit markets create a trade-off between career and housing that, at certain values of parameters and debt, is consistent with our empirical findings.

1.4.1. A simple 2-period model

We assume a deterministic two period model in which workers can only choose in a binary fashion between two careers (i.e. graduate school or not, \( j \in \{B, G\} \)), and between buying a house or not (\( H \in \{0, 1\} \)) without any access to the credit market\(^9\). Homeowners (\( H=1 \)) receive a utility premium (\( v_H \)), but they have to pay for the housing price in the first period (\( p_h \)). If workers choose to go to graduate school (\( j = G \)) they find themselves in a steeper earnings profile: they receive lower earnings in the first period (\( w^- \)) but higher earnings in the second period (\( w^+ \)). For simplicity, we assume no monetary cost for graduate school and, thus, no additional student loans. Workers with student debt (\( d \)) maximize the following inter-temporal utility problem:

\[
V(d) = \max_{j, H} \log(c_1) + \log(c_2) + \log(v_H)H
\]

\[
c_1 = w_1(j) - pH - d, \quad w_1(B) = w, \quad w_1(G) = w^- < w
\]

\[
c_2 = w_2(j), \quad w_2(B) = w, \quad w_2(G) = w^+ > w
\]

This formulation yields simple analytic expressions for indirect utility and parameter ranges for each career and housing choice.

\(^9\)We are indebted to Alistair Macaulay for suggesting us a way to discuss our results under this simple two-period model.
Let’s first consider the decision of a worker to choose to go to graduate school \( (j = G) \) conditional on his housing decision (we normalize \( w \) to 1). The worker follows a threshold rule:

\[
d < d_j^* = \begin{cases} 
\frac{w^- - w^+ - 1}{w^+ - 1} - p, & \text{if } H = 1 \\
\frac{w^- - w^+ - 1}{w^+ - 1}, & \text{if } H = 0
\end{cases}
\]

This implies that given the housing choice \( (H) \), workers are less likely to choose the steep career path as student debt rises. At the same time, homeowners are less likely to go to graduate school \( (d_j^*(H = 1) < d_j^*(H = 0)) \). This happens as the decision to go to graduate school while being homeowner pushes further down consumption in the first period, increasing \( U'(c_1) \). Hence, going to graduate school is more expensive in utility terms. On the other side, conditional on career choice, a recent college graduate chooses to buy a house \( (H = 1) \) if:

\[
d < d_H^* = \begin{cases} 
- \frac{p v_H}{v_H - 1}, & \text{if } j = G \\
1 - \frac{p v_H}{v_H - 1}, & \text{if } j = B
\end{cases}
\]

In a similar vein, this implies that given the career choice \( (j) \), workers are less likely to buy a house as student debt rises and buying a house is more expensive in utility terms for those who attended graduate school. But, as these choices are taken jointly, there will be four distinct regions for the first period discrete action: \{\( H = 1, j = G \), \( H = 0, j = G \), \( H = 1, j = B \), and \( H = 0, j = B \)\}. We will denote the threshold value for each as \( d_{j,H}^* \).

In particular, our empirical findings are consistent with the following:

**Proposition 4.1:** Assume the first period costs of graduate school \( (j = G) \) are larger than the cost of buying a house, i.e. \( w - w^- > p \). Then, for middling values of student debt, workers will choose housing over graduate school \( (d_{G,0}^* < d_{B,1}^*) \).
To prove the result, it is sufficient to use the four thresholds introduced above. As we can see in Figure 1.5a, for very low values of student debt, workers choose both to go to graduate school and buy a house, while for very high levels of debt workers choose to stay as renters and the flat career profile. However, there is a region of intermediate debt values for which increasing balances induces workers to give up graduate school and buy a home. Hence, $U'(c_1)$ rises faster for lower $c_1$. However, as Figure 1.5b shows, flattening the graduate school earnings profile yields the opposite outcome: for intermediate debt values, increasing student loan balances induces workers to give up home ownership.

The existence of this result strongly depends on the assumption of imperfect financial markets. In the next section, we analyze this career-housing trade-off using a life cycle model with more dimensions of heterogeneity. In addition, we add (1) idiosyncratic human capital and unemployment risk, (2) monetary cost (and amenity value) to attend graduate school and switch to steeper earnings profile (which implies more student debt), and, (3) imperfect mortgage market.

Figure 1.5: Indirect Utility

(a) Baseline
(b) Alternative (higher $w^-$, lower $w^+$)

Indirect utility functions as a function of undergraduate student debt. Baseline assumes $w^- < v_H < w^+$. Colors represent different discrete choices: in black $\{H = 1, j = G\}$, in red $\{H = 0, j = G\}$, in blue $\{H = 1, j = B\}$, in green $\{H = 0, j = B\}$.
1.4.2. The life-cycle Model

The model described in this section builds on important contributions to the human capital literature, as the career choice model of Keane and Wolpin (1997) and the Ben-Porath (1967) model presented in Huggett, Ventura and Yaron (2011), extended to include student debt and durable consumption (Fernández-Villaverde and Krueger (2011)). The aim is to build a richer model that not only replicates the intuition of the previous section, but can also be used to perform policy counterfactual analysis.

A unit measure of college graduates enter the labor market in each period and are heterogeneous in ability (a), human capital (h), liquid wealth (k) and student debt (d). Each household lives for T periods deterministically. During working age, workers can decide to enroll in graduate school: if they do, they access a different career path. Workers also sequentially decide labor and human capital investment within their career, savings and housing and non housing consumption while they pay for student debt (if any).

1.4.3. Setting

Preferences. Each agent maximizes expected lifetime utility over non durable consumption (c) and durable consumption (s) (see Fernández-Villaverde and Krueger (2011) and Kaplan, Mitman and Violante (2019)):

\[ u(c, s) = \frac{(c^{\zeta_1} s^{1-\zeta_1})^{1-\sigma}}{1-\sigma} \]  

(1.1)

where \( c > 0 \) and \( s = 1 + \zeta_2 \), and \( \zeta_2 \) is the housing service from owned housing.

Labor Income. When individuals work, hourly earnings are priced competitively to reflect their marginal productivity. Assuming a representative firm that uses human capital from workers in both careers, earnings are given by the human capital augmented number of hours worked, multiplied by the equilibrium rental rate \( R_t \):
where $w_{j,t}(l_t, h_t) = R_t l_t h_t^{\nu_j}$.  

Workers are also exposed to unemployment risk: they can be separated from their job with probability $\rho$; while unemployed, they earn home production $b$, but cannot invest in human capital. When workers retire, they are assigned pension transfers that are proportional to their last earnings.

**Careers and human capital.** We restrict career choice to two different paths. In each career path, their compensation is equal to the marginal product of hours. Formally, normalizing rental rate $R_t = 1$, we get hourly wage $w_{j,t} = h_t^{\nu_j}$, with $j = \{B, G\}$. The two paths differ in how workers' human capital accumulation translates into productive human capital. Human capital is less productive ($\nu_B < \nu_G$) for workers without graduate school education. Therefore, assuming workers make identical human capital investments, differences in earnings would grow as workers accumulate human capital.

After the career choice is made, individuals sequentially choose how much hours to work ($l_t$) and invest in further human capital ($1 - l_t$). Human capital evolves according to the following Ben-Porath law of motion:

$$h_{t+1} = e^{zt+1}(h_t + a((1 - l_t)h_t)^\alpha), \quad z_{t+1} \sim N(\mu_z, \sigma_z^2)$$

which depends on individual’s ability ($a$) and with risk coming from human capital idiosyncratic shocks. The Ben-Porath formulation implies that switching to the ’’steeper’’ career path that follows graduate school has three contrasting effects on human capital investment decisions. Firstly, since earnings in the steeper career path loads more on human capital, investments are riskier. Secondly, higher marginal product of human capital gives weaker incentives for graduate school educated worker to invest in human capital because of a simple wealth effect. Thirdly, $\nu_G > \nu_B$ generates a strong substitution effect, in that every unit
of consumption today that is foregone in order to invest in human capital generates higher returns in the future. The third effect seems to be dominant in the data, suggesting that difference in career paths are amplified by endogenous human capital investment.

**Graduate School.** Individuals can enroll in graduate school until age 32: if they do, they attend for S periods, and then start to work in their new career. While enrolled, human capital grows in every period at rate $\gamma$, and workers pay tuition ($p_g$) and consume using a combination of their liquid savings ($k$) and graduate student loans ($d_g$). They also get non monetary utility $\xi$, which is the amenity value of being in school as opposed to working.

**Financial Markets.** Agents can save in liquid assets $k$. Workers are allowed to borrow short term, using the rate $r_-$, but they face a credit card borrowing constraint that depends on their current income ($\phi$). If $k > 0$, savings yield a constant risk free rate $r_+$.

**Student Loans.** There are several options for repaying student loans, but the traditional and still most common is the 10-year fixed payment plan. Similar to a mortgage, the borrower makes constant payments over 120 months until the balance of principal and interest is paid off. Student loan payments ($P_d$) can be obtained as:

$$P_d = \frac{d_0}{\frac{(1+r_d)^{\tau} - 1}{r_d(1+r_d)^{\tau}}}$$

(1.4)

where $d_0$ is the student debt at the time of college graduation and $r_d$ is the gross interest rate on student loans. If a worker enrolls in graduate school, payments are suspended while enrolled and graduate school debt is added to the students’ balance. After graduation, undergraduate and graduate student debt is consolidated and a new standard repayment plan is started, giving the worker 120 months to repay the full amount.

**Housing.** Workers can buy a house at any moment of their life - except when they are enrolled in graduate school - as long as their life span is long enough that they can cover the 30-year mortgage and they have enough liquid assets to use as a downpayment. Workers
are also subject to housing preference shocks, which capture shifts in life events (household formation or divorce). We model those shifts as taste shocks, i.e. additively separable choice specific random taste shocks, and assume they are i.i.d. Extreme Value type I distributed with scale parameter $\sigma_c$. If a worker chooses not to own their house, she has to rent ($P_r$). The rental price is tied to the price of the house, $P_o$, and is set to match a given price to rent ratio. Individuals can ask for a 30-year fixed mortgage ($m$) to pay the price of the house ($P_o$).

There is no possibility of default or asking for a second mortgage. Home ownership is treated as an absorbing state, so if an individual is homeowner in a given year, then it will stay as homeowner at all future dates. At the time of buying the house, individuals face two borrowing constraints: (1) they must make a downpayment $(1-\lambda)$, (2) their monthly debt payments (student and mortgage debt) cannot exceed a proportion of their income ($\psi$). We assume that both constraints must be enforced at origination only.

Home owners must always pay the mortgage payment ($P_m$) until mortgage balances are zero, following:

$$P_m = \frac{(1 - \lambda)p_h}{\frac{(1+r_d)^{30} - 1}{r_d(1+r_d)^{30}}}$$  \hspace{1cm} (1.5)

1.4.4. Recursive formulation

We will illustrate the problem for agents of different stages of life, as the recursive formulation will differ according to it. The unit of time is a semester. The choice is motivated by several facts: it corresponds to the length of the initial grace period (when student loan payments must not be made), it allows for a reasonable accounting of separation risk, and yet it reduces the time dimension enough so that we can solve and estimate the model.
We write future values in recursive expressions by adding a $t$ to them. The choice-specific value functions are denoted indicating the discrete state - for instance, $V^B_o$ indicates the value function of a home-owner without a post-bachelor degree education.

**Retired workers:**

At retirement age $t = t_R$, workers are assigned pension transfers ($p$) that are proportional to their last earnings ($w = w_{t_R-1}$). Retired workers make consumption and saving decisions using their savings from working age ($k_{t_R-1}$). If they are home owners ($o$), they have to pay the residual parts of their mortgage ($m$) in equal payments ($P_m$) until mortgage debt is fully paid off. Otherwise, if they are renters ($r$), they need to rent and pay $P_r$ every period. Retired workers cannot buy a house, as mortgage duration exceeds their life expectancy. We assume no bequests and terminal condition for liquid assets to be equal to zero. Finally, we impose a non-negativity constraint on consumption on all agents.

Recursive Problem for renters, for $t = t_R, ..., T$, is:

$$V_{r,t}(a, w, k) = \max_{k'} u(c, s) + \beta V_{r,t+1}(a, w, k')$$

$$c + k' + P_r = (1 + r) \cdot k + pw$$

$$m_T = 0, k_T = 0, k' \geq \phi(pw), c \geq 0,$$

The Problem for home owners for $t = t_R, ..., T$, with mortgage payment $P_m$ is:

$$V_{o,t}(a, w, k, m) = \max_{k'} u(c, s) + \beta V_{o,t+1}(a, w, k', m')$$

$$c + k' + P_m = (1 + r) \cdot k + pw$$

$$m' = (1 + r_d)m - P_m \geq 0$$

26
\[ k_T = 0, \; k' \geq \phi(pw), \; c \geq 0 \]

In both cases, \( r = r_+ \) if \( k \geq 0 \), and \( r = r_- \) otherwise.

**Workers:**

Agents enter working age \((t = 1, \ldots, t_{R-1})\), and face two discrete choices every period: which career to pursue, i.e. whether to enroll in graduate school \((j = \{B,G\})\), and whether to buy a house or not \((H = \{r,o\})\). In both cases, workers are subject to preference shocks - respectively, denote the preference shock for the housing choice as \( \sigma_{\varepsilon_H} \), and the preference shock for the schooling choice as \( \sigma_{\varepsilon_j} \). Both preference shocks are i.i.d. Extreme Value type I distributed with scale parameter \( \sigma_{\varepsilon} \) [McFadden (1973) and Iskhakov et al. (2017)].

Workers’ problem entails saving and choosing how much hours to work \((l)\) and invest in further human capital \((1 - l)\) in every period. Human capital investment is risky and subject to an independent and identically distributed idiosyncratic shock every period \((z)\). Earnings are given by the human capital augmented number of hours worked multiplied by the equilibrium rental rate as defined in (1.2).

There is also unemployment risk: while working \((u = 0)\), they can be exogenously separated from their job with probability \( \rho \); while unemployed \((u = 1)\), they earn home production \( b \), but cannot invest in human capital, so that \( h' = h \). In order to apply for a mortgage and thus become a homeowner, workers have to satisfy the down-payment constraint (governed by the ratio \( \lambda \)) and at the same time satisfy the debt to income constraint (determined by the value \( \psi \)). Once the mortgage is approved, the payments \((P_m)\) are fixed for the next 30 years as defined in (1.5).

Workers that enter the labor market with any positive amount of student debt \((d_0 > 0)\) are by default enrolled in a 10-year fixed rate repayment plan and don’t have the option of defaulting or deferring on student loan payments.
For notational convenience, we can collect shocks and exogenous states in \( e = \{z, u, \epsilon\} \), and all the other idiosyncratic states in \( x = \{a, h, k, d, m\} \), where \( d \) indicates residual student debt balances. The recursive problem for renters without graduate school education, while employed\(^\text{10}\) is thus:

\[
\begin{align*}
V_{r,t}^B(x, e) &= \max_{k', l} \left\{ u(c, s) + \sigma \epsilon_{B,r} + \beta \mathbb{E}[\hat{V}_{t+1}^B(x', e')] \right\} \\
&= \max_{k', l} \left\{ u(c, s) + \sigma \epsilon_{B,r} + \beta \mathbb{E}[\hat{V}_{t+1}^B(x', e')] \right\} \\
&= \max_{k', l} \left\{ u(c, s) + \sigma \epsilon_{B,r} + \beta \mathbb{E}[\hat{V}_{t+1}^B(x', e')] \right\} \\
&= \max_{k', l} \left\{ u(c, s) + \sigma \epsilon_{B,r} + \beta \mathbb{E}[\hat{V}_{t+1}^B(x', e')] \right\}
\end{align*}
\]

(1.8)

\[
c + k' + (P_r + P_d) = (1 + r) \cdot k + w_B(l, h)
\]

\[
h' = e^{\gamma}(h + a((1 - l)h)^{\alpha})
\]

\[
d' = (1 + r_d)d - P_d \geq 0
\]

\[
k' \geq \phi, \ c \geq 0
\]

where:

\[
\mathbb{E}[\hat{V}_{t+1}^B(x', e')] = \mathbb{E} \left[ \max \left\{ V_{r,t+1}^B + \sigma \epsilon_{B,r}, V_{G,t+1}^G + \sigma \epsilon_{G,r}, V_{B,o,t+1}^B + \sigma \epsilon_{B,o}' \right\} \right]
\]

and \( V_{G,t+1}^G \) and \( V_{B,o,t+1}^B \) are the value function of a worker enrolled in grad school and homeowner, respectively. Unemployed workers’ problem is analogous, with earnings replaced by \( b \) and no human capital investment decision. Unemployed workers can find a job in the same career with probability \( 1 - \rho \).

Home owners with housing payment \( P_m \) face the following problem:

\[
V_{o,t}^B(x, e) = \max_{k', l} \left\{ u(c, s) + \sigma \epsilon_{B,o} + \beta \mathbb{E}[\hat{V}_{t+1}^B(x', e')] \right\} \\
&= \max_{k', l} \left\{ u(c, s) + \sigma \epsilon_{B,o} + \beta \mathbb{E}[\hat{V}_{t+1}^B(x', e')] \right\} \\
&= \max_{k', l} \left\{ u(c, s) + \sigma \epsilon_{B,o} + \beta \mathbb{E}[\hat{V}_{t+1}^B(x', e')] \right\}
\]

(1.9)

\[
c + k' + (P_m + P_d) = (1 + r) \cdot k + w_B(l, h)
\]

\(^{10}\)See Appendix A1.2 for details about how we compute this problem.
\[ h' = e^{z'}(h + a((1 - l)h)^\alpha) \]
\[ d' = (1 + r_d)d - P_d \geq 0 \]
\[ m' = (1 + r_d)m - P_m \]

\[ P_m = \begin{cases} 
\lambda p_h, & \text{if } m = 30 \cdot P_m \\
\frac{r_d(1 + r_d)^{(1 - \lambda)p_h}}{(1 + r_d)^{(30 - 1)}} & \text{if } m < 30 \cdot P_m 
\end{cases} \]

\[ k' \geq \phi, \ c \geq 0, \frac{P_m + P_d}{w_B} \leq \psi \]

where:
\[ \mathbb{E}[\tilde{V}_{t+1}^B(x', e')] = \mathbb{E} \left[ \max \left\{ V_{o,t+1}^B + \sigma \epsilon_{B,o}^r, V_{o,t+1}^G + \sigma \epsilon_{G,o}^r \right\} \right] \]

If the worker is in the first period of home ownership, \( P_m \) equals to the downpayment required to buy the house. After that period, housing payments are determined by the mortgage equation (1.5), as before.

At this point we want to characterize the recursive problem of the individual attending graduate school. For simplicity, we will characterize only the problem of the renter. Define \( \bar{S} \) as the number of periods required to get the degree. For \( s \leq \bar{S} \):

\[ V_{r.t}^G(x, e, s) = \max_{k'} \left\{ u(c, s) + \xi + \sigma \epsilon_{G,r} + \beta \mathbb{E}[\tilde{V}_{t+1}^G(x', e', s')] \right\} \]  \hspace{1cm} (1.10)

\[ c + k' + P_r + p_g = (1 + r) \cdot k + d_g \]
\[ h' = (1 + g) \cdot h \]
\[ d' = (1 + r_d) \cdot d + d_g \cdot 1_{s=1} \]
We assume that, during graduate school, the borrowing constraint with liquid assets is tighter - since the individual is not working she has to keep her liquid assets positive. When \( s > \bar{S} \), the recursive problem is analogous to the problem of a worker with student loans, conditional on career earnings’ slope \( \nu_G \).

1.5. Calibration and Estimation

In this section, we discuss how we determine the parameters required for the analysis. We set these parameters in two ways. First, we set some parameters from elsewhere in the literature or by using data estimation (Table 1.4). The remaining parameters are estimated using indirect inference through the model.

1.5.1. External Parameters

**Timing.** Each period time in the model represents two quarters. Individuals start making decisions when they graduate from college. After finishing college, they start working and repaying their student debt. Agents retire at the age of 65 and die when they are 80.

**Preferences.** Preferences are set using standard calibration in the macroeconomics literature. The yearly discount factor is set to be 0.99. We set the constant relative risk aversion in the utility function to 2.

**Career and Human Capital.** Following Huggett, Ventura and Yaron (2011), we set the mean shock of human capital to -0.029, with 0.111 variance and the production function parameter \( \alpha \) to 0.66. We assume that, when unemployed, worker gains access to unemployment benefits that sum up to \( b \) calibrated to the Federal poverty threshold for an individual living alone in 2008 ($991 USD a month).
Table 1.4: External Parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Description</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Timing and Preferences</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T</td>
<td>126</td>
<td>Model periods (semesters)</td>
<td>-</td>
</tr>
<tr>
<td>R</td>
<td>88</td>
<td>Working Periods</td>
<td>-</td>
</tr>
<tr>
<td>β</td>
<td>0.99</td>
<td>Discount Factor</td>
<td>-</td>
</tr>
<tr>
<td>σ</td>
<td>2</td>
<td>Coeff. of Risk Aversion</td>
<td>-</td>
</tr>
<tr>
<td><strong>Careers, Human Capital</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>α</td>
<td>0.66</td>
<td>Ben-Porath Production Function</td>
<td>Huggett, Ventura and Yaron (2011)</td>
</tr>
<tr>
<td>μz</td>
<td>-0.029</td>
<td>Mean Shock of Human Capital</td>
<td>Huggett, Ventura and Yaron (2011)</td>
</tr>
<tr>
<td>σz</td>
<td>0.111</td>
<td>Riskiness of Human Capital Investment</td>
<td>Huggett, Ventura and Yaron (2011)</td>
</tr>
<tr>
<td>p₀</td>
<td>$ 50,000</td>
<td>Cost of Graduate School (one period)</td>
<td>IPEDS</td>
</tr>
<tr>
<td>S</td>
<td>4</td>
<td>Graduate School periods</td>
<td>NCES</td>
</tr>
<tr>
<td><strong>Labor Income</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ρ</td>
<td>{6%, 4.5%}</td>
<td>Separation Probability</td>
<td>Menzio, Telyukova and Visschers (2016)</td>
</tr>
<tr>
<td>b</td>
<td>$991</td>
<td>Home Production (monthly)</td>
<td>Federal poverty threshold (2008)</td>
</tr>
<tr>
<td><strong>Financial Markets</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ϕ</td>
<td>-$ 5,000</td>
<td>Credit Card Borrowing Limit</td>
<td>Survey of Consumer Finances (SCF)</td>
</tr>
<tr>
<td>r⁺</td>
<td>2%</td>
<td>Interest on liquid assets</td>
<td>Federal Reserve Board of Governors (2014)</td>
</tr>
<tr>
<td>r⁻</td>
<td>11%</td>
<td>Borrowing Rate</td>
<td>Federal Reserve Board of Governors (2014)</td>
</tr>
<tr>
<td>r₅</td>
<td>6%</td>
<td>Interest on student loans and mortgages</td>
<td>Department of Education and PMMS</td>
</tr>
<tr>
<td><strong>Housing</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>λ</td>
<td>0.15</td>
<td>Downpayment (fraction of house price)</td>
<td>Greenwald (2018)</td>
</tr>
<tr>
<td>ψ</td>
<td>0.43</td>
<td>Debt-to-Income Ratio</td>
<td>Dodd-Frank limit</td>
</tr>
<tr>
<td>pₜ</td>
<td>$ 250,000</td>
<td>House price</td>
<td>Case-Shiller house price index (2008-2012)</td>
</tr>
<tr>
<td>ν</td>
<td>0.02</td>
<td>price-to-rent</td>
<td>Case-Shiller house price index (2008-2012)</td>
</tr>
</tbody>
</table>

**Labor Income.** We set the rental rate to a yearly rate of 5% of the house price, and pension to be 45 percent of the last earned income. Finally, exogenous separation risk is set to 6 percent per year for bachelor-educated workers, and 4.5 percent for workers with a post-Bachelor degree, matching the average number of employment to unemployment transition of the two groups (see Menzio, Telyukova and Visschers (2016)).

**Financial Markets and student debt.** The annual interest rate for student loans and a 30-year fixed rate mortgage is calibrated to the 2004-2008 average rate of 6 percent (Figure 1.6). The risk free interest rate for savings is set at 0 following null real returns after 2008.
and credit card borrowing rate is fixed at an annual 10 percent. We set a credit card borrowing limit of $-5,000, targeting a median rate of credit limit to annual labor income for college graduates of 20 percent.

**Housing.** We set the price of the house at the average home price in the U.S. ($250,000) during the years 2008-2012. The rental price is set to match the price to rent ratio (0.02). The parameters that determine the LTV and DTI are chosen to match institutional features of the US mortgage market.

For the LTV parameter, we fix a downpayment constraint of $0.15 \cdot P_0$. This value is intended to reflect the distribution of the LTV in Freddie Mac data, which has two masses point around 80% and 90%, where the first mass point is typically populated by younger buyers and thus seems more appropriate for pinning down the problem of first home ownership (see [Greenwald (2018)]). In order to qualify for a Qualified Mortgage under CFPB guidelines, a borrower’s total debt to income ratio, including the mortgage payment and all other recurring debt payments, cannot exceed 43 percent.

---

**Figure 1.6: Interest Rate on Student Loan and Mortgage Debt**

![Interest Rate on Student Loan and Mortgage Debt](image)

1.5.2. Distribution of Initial Characteristics

In order to simulate the model, we have to make parameter choices regarding starting values of liquid assets, ability, human capital and student debt.

While college graduates typically do not have substantial wealth of their own, they may have access to alternative sources of wealth that are not directly measured. We assume students leave college with zero liquid assets, but receive an exogenous transfer from their parents. We use the Expected Family Contribution (EFC)\textsuperscript{11}, with an average transfer of $16,309 ($\mu_k$) and a standard deviation of $16,321 ($\sigma_k$). We also use B&B data for the student debt distribution. We use an average debt balance of $16,619. This figure is composed by a percentage of 66% of borrowers, with cumulative average balances of $25,048 ($\mu_d$) and a standard deviation of $18,223 ($\sigma_d$). We also take the correlation between parental transfers and student debt of -0.17 ($\rho_{k,d}$).

Following Huggett, Ventura and Yaron (2011), we assume that the distribution of ability and human capital is jointly log normally distributed. We calibrate the initial mean ($\mu_h$) and standard deviation ($\sigma_h$) of human capital to match the mean and standard deviation of earnings one year after graduation from B&B data - respectively at $34,554 and $20,600. We calibrate the mean level of ability ($\mu_a$) in order to match a yearly average growth rate of earnings for college graduates of 3%. We take the correlation between human capital and ability ($\sigma_{a,h}$) from Athreya et al. (2019), who estimate a life cycle model of education choice and report a correlation of 0.67.

Finally, we need to determine the correlation between ability and student debt ($\rho_{a,d}$). This parameter has an important interpretation because, if correctly identified, it informs about the bias that an econometrician would be subject to when estimating Equation (1.1) with least squares. We estimate this parameter, jointly with other structural parameters, to match the key properties of the earnings and homeownership profiles on CPS and B&B.

\textsuperscript{11}The EFC is calculated according to a formula established by law. Student’s family’s taxed and untaxed income, assets, and benefits (such as unemployment or Social Security) are considered in the formula.
1.5.3. Estimation

Simulated Method of Moments

Parameters $\Theta = \{\xi, \gamma, \nu_g, \zeta_1, \zeta_2, \rho_{a,d}\}$ are jointly estimated by Simulated Method of Moments (see Gourieroux, Monfort and Renault (1993), Smith Jr (1993) and Gallant and Tauchen (1996)). Let $x_i$ be an i.i.d. data vector, $i = 1, ..., n$, and $y_{i,s}(\Theta)$ be an i.i.d. simulated vector from simulation $s$, so that $i = 1, ..., N$, and $s = 1, ..., S$. The goal is to estimate $\Theta$ by matching a set of simulated moments, denoted as $h(y_{i,s}(\Theta))$, with the corresponding set of actual data moments, denoted as $h(x_i)$.

Define:

$$g_n(\Theta) = \frac{1}{n} \left[ \sum_{i=1}^{n} h(x_i) - \frac{1}{S} h(y_{i,s}(\Theta)) \right]$$

Building $g_n(\Theta)$ in this case faces an important challenge. In classic SMM estimation, exploration of the state space requires the model to be solved more than 10000 times. In the case of a model with a large state space like ours, this could be very computationally expensive.

To overcome the curse of dimensionality, we discretize the parameter space using sparse grids (see Bungartz and Griebel (2004)). A similar approach in structural modelling has been using in the context of maximum likelihood estimation, see for instance Heiss and Winschel (2008).

By using functions with support restricted to a neighborhood of each point to build $h(y_{i,s}(\Theta))$, our approach is suitable for approximating the parameter-moment mapping even in cases of sharp behavior, like large fluctuations of the gradient (see Stoyanov (2013)).

---

12Using a cluster with 144 CPUs, we manage to obtain a full solution of the model and simulate it in about 14 minutes.
Having $h(y_{i,s}(\Theta))$ at hand, we can construct an objective function that looks like:

$$
\hat{\Theta} = \arg \min_{\Theta} g_n'(\Theta)\hat{W}_n g_n(\Theta)
$$

where $\hat{W}_n$ is a positive definite matrix that converges in probability to a deterministic positive definite matrix $W$. There are many feasible choices for the covariance matrix, and it is common to simply rely on an identity matrix for $W$. To construct the optimal weight matrix, we use the influence function technique from [Erickson and Whited (2002)](Erickson2002) (see also [Bazdresch, Kahn and Whited (2017)](Bazdresch2017) for an application closer to our case). The derivation is explained in detail in Appendix A1.3. Finding a solution to (1.12) faces the issue of the possible presence of many local minima: to make sure our solution is robust, we restart our optimization routine using multiple sets of starting values. Each routine solves its problem using a Nelder-Mead algorithm. Having an estimate of $h(y_{i,s}(\Theta))$ also allows us to obtain standard errors of parameter estimates, as they can be calculated knowing that:

$$
a\text{Var}(\hat{\Theta}) = \left(1 + \frac{1}{S}\right) \left[ \frac{\partial g_n(\Theta)}{\partial \Theta} W \frac{\partial g_n(\Theta)}{\partial \Theta'} \right]^{-1}
$$

(1.13)

Identification and Parameter Estimates

The model generates a large number of moments that can be used for estimation. Since interactions between each choice are quite complex, global identification is not possible even if one can attempt a one to one mapping between model parameters and empirical moments. Local identification, however, simply requires that the gradient of the model implied moments with respect to the parameter, $\partial h(y_{i,s}(\Theta))/\partial \Theta$, has full rank. As noted in [Bazdresch, Kahn and Whited (2017)](Bazdresch2017), this condition suggests that for a parameter to be identified, some subset of the vector of implied moments, must change when that particular parameter moves (see Figure A1.4).
Table 1.5 shows the empirical moments we target in our estimation. We use the ratio of home ownership at age 40 by education groups (when the ratio is 1) and the constant and slope of life cycle home-ownership as target moments. Housing preferences ($\zeta_1$ and $\zeta_2$) are identified by shifts in these moments. The reason for this is that housing demand of workers with a post-bachelor degree is more sensitive to changes in the flow value: as it grows, not only less graduates enroll in post-bachelor programs, but workers with additional education try to enter into home ownership earlier, thus matching the home ownership of workers that only have a bachelor.

We also use as a moment the debt-to-income ratio at age 30. The most related parameter is the amenity value of graduate school ($\xi$): this is the value that mostly relates to the degree of sorting into additional education by ability, which in turn determines graduate school enrollment and student loan balances. We can also identify $\xi$ by looking at shifts in the relative level of incomes by educational attainment. We use the ratio of the constant and the slope of life cycle earnings calculated from individuals aged 24-40 in CPS.

Table 1.5: Target Moments

<table>
<thead>
<tr>
<th>Moments</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Data</td>
</tr>
<tr>
<td>A. Sample Means</td>
<td></td>
</tr>
<tr>
<td>Graduate to Bachelor Homeownership at age 40</td>
<td>1.0</td>
</tr>
<tr>
<td>Debt to Income at age 30</td>
<td>0.59</td>
</tr>
<tr>
<td>B. Regression Coefficients</td>
<td></td>
</tr>
<tr>
<td>Home ownership, constant</td>
<td>0.474</td>
</tr>
<tr>
<td>Home ownership, slope</td>
<td>0.019</td>
</tr>
<tr>
<td>Graduate to Bachelor earnings ratio, constant</td>
<td>1.10</td>
</tr>
<tr>
<td>Graduate to Bachelor earnings ratio, slope</td>
<td>1.71</td>
</tr>
</tbody>
</table>

Source: Current Population Survey (individuals with at least a bachelor degree) and model estimates.
We want to know about the two earnings parameters (\( \gamma \) and \( \nu_g \)); \( \gamma \) ultimately determines the value of attending a post-bachelor degree. In the model, it also affects sorting, enrollment, and overall home ownership. However, only the relationship with the constant term in the home ownership life cycle profile is monotonous; as \( \gamma \) grows, it first increases graduate school attendance. Then, it also starts to allow more indebted workers to postpone enrollment, and so the ratio decreases without affecting the ratio of the earnings profiles.

The most straightforward identifying relationship is a result of higher \( \gamma \) simply increasing returns to graduate studies, thus allowing more workers to enter into home ownership at some point in time. More intuitively, the skill premium \( \nu_g \) identifies the ratio of debt to income. The reason is that the skill premium, besides affecting earnings, is the main reason for increasing or decreasing early enrollment in the model.

Finally, we find that the correlation between debt and ability, \( \rho_{a,d} \) is clearly identified by the ratio in the slopes of life cycle profiles of earnings. This is also intuitive: as the relationship between debt balances and ability becomes stronger (i.e. more negative), sorting into post bachelor degrees will unambiguously increase, other things not varying much. Hence, our model implies that growth in earnings differentials are mostly coming from increased borrowing of graduates with lower learning ability.

Table 1.6 displays parameter estimates. The amenity value of graduate school is expressed in dollar terms, but does not correspond to \( \xi \). To obtain it, we assume individuals in grad school are renters and have zero net liquid assets. Then the value is obtained by solving for the amount of consumption increase that would yield equivalent flow utility to grad school attendance. Standard errors tells us that the estimates are precise - the only minor concern being represented by the objective function being less sensitive to changes in \( \xi \).
Table 1.6: Estimated Parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
<th>Value</th>
<th>Standard Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\xi$</td>
<td>Amenity Value of Graduate School</td>
<td>$55.171$</td>
<td>$16.795$</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>Graduate School Growth in Human Capital</td>
<td>$8.36%$</td>
<td>$0.15%$</td>
</tr>
<tr>
<td>$\nu_G$</td>
<td>Graduate School Skill Premium</td>
<td>$14.25%$</td>
<td>$3.9%$</td>
</tr>
<tr>
<td>$\zeta_1$</td>
<td>Elasticity to Housing Service</td>
<td>$0.539$</td>
<td>$0.0069$</td>
</tr>
<tr>
<td>$\zeta_2$</td>
<td>Housing Service</td>
<td>$20.484$</td>
<td>$695$</td>
</tr>
<tr>
<td>$\rho_{a,d}$</td>
<td>Correlation (ability, debt)</td>
<td>$-0.124$</td>
<td>$0.01$</td>
</tr>
</tbody>
</table>

Source: Model estimates using Simulated Method of Moments.

Model Fit

The model replicates well overall earnings dynamics (Figure 1.7). The model matches pretty well average yearly income growth (1.9% in the data and 1.9% in our model), and earnings growth naturally slows because of income effects in the Ben-Porath function.

The patterns in enrollment replicates gradual entry into post graduate studies, and the level slightly more than a third of college educated workers pursuing further education. Because of the extreme assumption that human capital accumulated while working in one career is destroyed when switching\textsuperscript{13}, workers in the model tend to enroll slightly earlier than in the data.

The slope in the life cycle pattern of home ownership in our model is higher than in the data: especially in early years, home ownership is substantially lower, and then it catches up later in the working life. This can be explained with the choice of abstracting from bequest shocks in the model, as they would allow households to bring forward home ownership by relaxing their budget constraint.

\textsuperscript{13}This choice is appropriate for some post-bachelor degrees, in particular the professional ones, where previous experience is hardly useful in the career implied by the degree. But it is clearly less appropriate to capture the role some other degrees, as MBAs and executive MBAs, play in the career of workers with some years of experience.
Another factor that limits earlier home purchases is our assumption of having just one size (and thus one price) available to workers. Extending the choice set by allowing individuals to differentiate their purchases with respect to price is the next step for improving the fit of our model to the data. Decomposing the rate of home ownership by graduate school attainment and debt levels, we can see that the delay in purchases is almost entirely attributable to workers that pursue graduate studies (see Figure 1.8).

Figure 1.7: Earnings and Enrollment Life Cycle Profiles (model and data)

Source: Model estimates (solid line) and data estimates (dashed line) computed using the Current Population Survey (CPS 2010-2020), workers with at least a Bachelor’s Degree aged 23-50.
1.6. Results

1.6.1. The Role of Student Debt

There are two main trade-offs involved in the initial career choice. First, workers that not pursue additional education start with higher disposable income but then have lower income growth compared to the more human capital intensive careers. Second, income paths of bachelor educated workers are less volatile as human capital accumulation is a risky investment. Workers whose undergraduate borrowing is above the average level start with higher earnings, because they are most likely to be working rather than being enrolled. After some years, the sorting effects of student loans start to affect earnings, and thus create a wide and persistent earnings gap.

Table 1.7 shows the estimated impact of an increase in debt on earnings, education and housing and compares it with the empirical estimates from Section 1.3.4.
In the model, indebted graduates have 0.13% higher earnings for each 1% of additional student borrowing, but 0.21% lower earnings growth in the following four years. The model also replicates well the impact of debt on graduate school attendance and home ownership.

In absence of frictions to borrowing, student loans should have no effect on career choices and human capital investment. In our model, the effects of student loans on career choices and lifetime earnings are ultimately the result of two main frictions. First, young workers face credit constraints that limit their ability to smooth durable and non-durable consumption. Second, student loans follow a predetermined fixed repayment schedule and alternative repayment schemes are limited.\textsuperscript{14}

<table>
<thead>
<tr>
<th>Table 1.7: Student Debt, Earnings, Career and Housing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Moments</td>
</tr>
<tr>
<td>---------</td>
</tr>
<tr>
<td>Earnings</td>
</tr>
<tr>
<td>$\partial y_{t+1}/\partial d_t$</td>
</tr>
<tr>
<td>$\partial \Delta y_{t+4}/\partial d_t$</td>
</tr>
<tr>
<td>Graduate School</td>
</tr>
<tr>
<td>$\partial P(G)_{t+4}/\partial d_t$</td>
</tr>
<tr>
<td>Homeownership</td>
</tr>
<tr>
<td>$\partial P(H)_{t+4}/\partial d_t$</td>
</tr>
</tbody>
</table>

$\Delta y_{t+4}$ is the 4-year growth in earnings after graduation

Source: Instrumental Variable regression results using Baccalaureate and Beyond Longitudinal Study (BB:2007/08) and model estimates.\textsuperscript{14}

\textsuperscript{14}Our empirical analysis is focused on graduates that entered the labor market in 2008: during those years, less than 7% of borrowers enrolled in plans that allowed payments to be linked to earnings. After a series of reforms, enrollment in income based plans has increased substantially in the following decade.
Table 1.8 shows how entry into graduate school is affected by borrowing conditional on ability levels. More indebted students are significantly less likely to enroll. This happens for two reasons: on the one hand, while attending school allows to postpone payments, new debt is added to the existing one. Adding the burden of additional borrowing has compounding effects which put considerable pressure on future disposable consumption, thus discouraging enrollment. On the other hand, workers still have the possibility of starting to repay, while working, and then enrolling when their debt burden has reduced.

The value of switching, however, decreases with tenure due to a horizon effect: as the worker gets older, and approaches the age where it would be optimal to start a mortgage, attending graduate school would imply a postponement of entry into home ownership because of the binding credit constraint, reducing the value of additional education. Interestingly, the largest impact of borrowing on enrolment is on individuals with higher ability. A dampened sorting into careers points on the second effect being dominant: in fact, while relatively low ability individuals enroll smoothly over post graduation years, high ability individuals who would postpone getting additional education and switching career find the option becoming increasingly costly as they get older.

<table>
<thead>
<tr>
<th>Graduate School Enrollment</th>
<th>Student Debt</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt; $22,560</td>
<td>&gt; $22,560</td>
<td></td>
</tr>
<tr>
<td>Low Ability</td>
<td>2.8%</td>
<td>2.33%</td>
</tr>
<tr>
<td>High Ability</td>
<td>46.01%</td>
<td>32.25%</td>
</tr>
</tbody>
</table>

Overall Graduate School Enrollment

| Low Ability | 13.52% | 11.20% | 12.74% |
| High Ability | 55.84% | 43.80% | 51.86% |

Source: model estimates.
1.6.2. Lifetime Earnings

In order to understand the relative importance of different channels in affecting earnings over the life cycle, we decompose earnings growth differential between workers. To do so, we group workers based on different percentiles of student debt distribution. Notice one can obtain average earnings growth from Equation (1.3). Define $s_G$ as the share of post bachelor educated workers in a given group, $\bar{a}$ as the average ability and $F(h)$ as the distribution of human capital in that group. Average earnings growth can be defined as:

$$\Delta(w) = \int \left( s_G \nu_G + (1-s_G) \nu_B \right) \bar{a} ((1-l)h)^\alpha dF(h)$$  \hspace{1cm} (1.14)

As argued above, highly indebted workers choose flatter earnings profiles. In Figure 1.9 we decompose the earnings growth differentials between the lowest and the highest tercile of workers ordered by undergraduate borrowing. Interestingly, ability plays a minor role in determining earnings growth differentials. This comes from two aspects. First, model estimates deliver small correlation between initial ability and debt. More importantly, though, there is ex post sorting that depends on the fact that human capital accumulation is risky at the individual level.

Figure 1.9: Decomposing Earnings Growth: Low versus High Debt

Source: model estimates.
High ability - high debt individuals with good human capital realizations experience both high initial wage growth and lower rate of enrollment in post bachelor degrees, as the option value of switching career decreases substantially after their human capital (and thus earnings) reach a higher level, hence they stay in the labor force while workers with lower ability and the same realizations find enrollment more valuable and thus enroll (disappearing temporarily from the workforce). As highly indebted students catch up on enrollment, the contribution of skill premium decreases, but remains positive and eventually becomes the main factor driving earnings growth differentials as human capital investment behavior reaches a plateau for most workers.

Finally, we find that endogenous human capital accumulation contributes for the lion share of earnings growth differentials. Two effects go in the same direction in determining this result. As one can see from the policy function of workers for the Ben-Porath human capital investment choice, highly indebted workers simply choose to invest less in order to have higher earnings in the current period. This is reinforced by career choices, as the same investment has higher returns for workers that enjoy a higher skill premium, consistent with earnings growth differentials being highest during the earlier years.

1.6.3. The Importance of Housing

In our model, student loans affect home ownership through two main channels. On the one hand, highly indebted students are less likely likely to pursue extra education, which has lower returns to human capital, thus lower expected growth but also lower income risk. Thus, housing is a relatively more attractive investment at the start of the working career. On the other hand, student loan borrowers might face more difficulties in satisfying both the downpayment and the debt to income requirements for a mortgage. Since student loan payments reduce workers’ disposable income, both investment in human capital and savings will be smaller. In addition, higher borrowing sorts workers into less human capital intensive careers, which negatively affects their lifetime earnings.
Table 1.9: Entry into Home Ownership

<table>
<thead>
<tr>
<th>Age of First Purchase</th>
<th>Non Borrowers</th>
<th>Borrowers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group</td>
<td></td>
<td></td>
</tr>
<tr>
<td>All Workers</td>
<td>30</td>
<td>29</td>
</tr>
<tr>
<td>Only Bachelor(^a)</td>
<td>25</td>
<td>26</td>
</tr>
</tbody>
</table>

\(^a\)= includes those who do not enroll in grad school at any point in time

Source: model estimates.

As shown in Table 1.9, all those effects play a decisive role in determining the age at which households purchase their first home. From the second row it is possible to see that, for those workers who don’t choose to enroll in graduate studies, borrowing affects home ownership mostly through the wealth effect. Hence, borrowers enter into home ownership later, with the delay growing nonlinearly in debt balances.

In the aggregate, however, the role of post-bachelor enrollment dominates. As we can observe from the first row, the larger share of enrollment of non borrowers pushes home ownership to later in life. As balances grow, the two effects compensate each other.

To understand how relevant the housing channel is in determining education and career choices, we compare model predictions in a counterfactual scenario when workers are not allowed to access home ownership and remain renters during their whole life. This way we are reducing available choices to workers compared to the baseline model, but we allow them to fully re-optimize given the new constraints they face. In this exercise, absent housing, agents can make different decisions about the timing of their investment in education (as well as about how much time to spend on human capital accumulation) than in the baseline.

Two clear trends emerge in Table 1.10: first, graduate school enrollment increases for both groups (low and high indebted graduates). Second, while highly indebted students still choose to postpone enrollment, they do eventually enroll in the following years, while the

\(^{15}\) An equivalent assumption is that we are imposing \(\zeta_2 = 0\) while leaving all other parameters unchanged from the baseline estimation
baseline model suggests strong horizon effects. Switching costs (i.e. limited transferability of human capital) and borrowing constraints still matter, and determine the difference in enrollment patterns between graduates with different debt balances. Notice, however, that even in this context enrollment should not necessarily be identical along the debt distribution, to the extent that correlation with learning ability is different from zero - as turns out to be the case according to the estimates of our baseline model.

Increased enrollment in post-bachelor programs and the missing concern of savings in order to respect the downpayment constraint and then pay the mortgage have strong earnings effects, as shown in Figure 1.10. In this case, the change takes place mostly on the human capital investment side. Yearly earning differentials between high and low indebted graduates at age 30 goes from $9,300 to $5,200, and at age 40 from $7,700 to $4,600.

<table>
<thead>
<tr>
<th>Graduate School Enrollment</th>
<th>Student Debt</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt; $22,560</td>
<td>&gt; $22,560</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Graduate School Enrollment at Age 26</th>
<th>Student Debt</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>24.4%</td>
<td>17.29%</td>
</tr>
<tr>
<td>No Housing</td>
<td>52.83%</td>
<td>29.4%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Overall Graduate School Enrollment</th>
<th>Student Debt</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>34.68%</td>
<td>27.50%</td>
</tr>
<tr>
<td>No Housing</td>
<td>65.29%</td>
<td>45.84%</td>
</tr>
</tbody>
</table>

Source: model estimates.
1.7. Policy

1.7.1. A "Debt to Equity Swap": Income Based Repayment

Income Based Repayment plans are a popular solution to broadening access to higher education, as countries like Australia and Great Britain made them their baseline program for student finance (see Chapman (2016)). During recent years, changes to old repayment plans and the creation of new ones have expanded generous income-based repayment options to a growing number of borrowers in the US (Figure 1.11). Unlike fixed payment plans, there is no set horizon of loan repayment; instead, the borrower pays a percent of discretionary income each month until the loan is paid off.

In this section, an income repayment plan in every period is introduced in the model as an alternative repayment scheme. The income repayment plan is defined to replicate the Pay As You Earn Repayment Plan (PAYE) introduced in 2012: 10 percent of discretionary income for 20 years and at the end of the repayment period, remaining balances are forgiven and the forgiven amount is considered as additional income, to be taxed at a 25% rate.
Figure 1.11: Evolution of Repayment Plans

Source: Federal Student Aid Data. Percentage of student loan borrowers enrolled in repayment plans as well as the percentage amount of student debt each repayment plan represents. Includes outstanding principal and interest balances of Direct Loan borrowers in Repayment, Deferment, and Forbearance.

We rewrite the recursive problem in (1.8), as the other problems are analogous:

\[
V_{r,t}^{B}(x,e) = \max_{k',l} \left\{ u(c,s) + \sigma \epsilon_{B,r} + \beta \mathbb{E}[V_{r,t+1}^{B}(x',e')] \right\}
\]

\[
c + k' + P_r = (1 + r) \cdot k + 0.9 \cdot w_B(l, h)
\]

\[
h' = e^{z'}(h + a((1 - l)h)^{\alpha})
\]

\[
d' = (1 + r_d)d - 0.1 \cdot w_B(l, h) \geq 0
\]

\[
k' \geq \phi
\]

\[
c \geq 0
\]

48
where:

$$\mathbb{E}[V_{t+1}^B(x', e')] = \mathbb{E}[\max\{V^B_{r,t+1} + \sigma \epsilon^r, V^G_{r,t+1} + \sigma \epsilon^G, V^B_{o,t+1} + \sigma \epsilon^B\}]$$

A quantitative exercise is necessary to assess the extent to which income based repayment plans moderate the effects of initial student loan debt. On the one hand, enrollment in income driven repayment plans reduces the ratio of student loan payments to monthly wages, increasing disposable income. On the other hand, it can extend the repayment period significantly relative to a 10-year plan, thereby potentially increasing the total interest paid by the student loan borrower over the life of the loan.

The latter effect is the main reason why enrollment under IBR rises, but due mostly to higher enrollment by high ability graduates. However, facing increasing payments during age 25-35, and a small risk of having to pay a lump sum tax in the late 30s because of residual balance forgiveness, workers under IBR delay entry into home ownership even more (see Figure 1.12). After age 45, income effects start to dominate and overall home ownership grows compared to baseline.

As shown in this section, linking repayment to income does help alleviating financial constraints. Even if the program did not achieve full participation of graduates, the growth in IBR enrollment can be credited with moderating the impact of the significant growth in undergraduate debt balances occurred between 2008 and 2016.

1.7.2. Evaluating a Radical Policy: Debt Forgiveness for All

As student debt became a prominent issue in the public debate, various political actors have called for some sort of forgiveness plan. In this chapter we introduce debt forgiveness under a balanced budget constraint, assuming the government can forgive all debt and then finance this program by spreading lump sum taxation over the life cycle of workers. This
exercise wipes out all of undergraduate debt, and replaces it with lump sum taxes levied over the life cycle to keep the reform under a balanced budget.

According to our model, a forgiveness plan would have a large impact on post bachelor enrollment. While adoption of an IBR plan would increase enrollment in post-bachelor plans to 38%, forgiveness would bring it to 44% (see Figure 1.12). It would both increase overall participation in graduate programs, and do it in particular during early years. The second effect comes from the disappearance of the delaying motive that induces indebted graduates to postpone enrollment, while the first is a result of the relaxing of borrowing constraints on the same group.

Given larger enrollment, it is not surprising that entry into home ownership is almost unchanged, as income effects move workers in the opposite direction. What happens, in fact, is that there is a small delay in access to home purchases, driven by increased enrollment. The overall impact on earnings and later age home ownership, however, is not substantially larger than under the Income Based Repayment alternative plan.

This effect on earnings comes from the differential impact the two plans have on sorting into graduate school: IBR achieves higher enrollment by a sharp increase in the enrollment of high ability individuals. On the other hand, forgiveness has negative effects on sorting, as it also increases the participation of workers with lower learning ability. This has a large impact on endogenous human capital accumulation which is the main driver of earnings growth. Lower ability workers enroll at a higher rate (and borrow for graduate studies), but their net monetary gain is small, and their endogenous human capital investment at age 30-35 is reduced compared to the baseline scenario where they had repaid their residual debt by that age.
Figure 1.12: Baseline vs. Alternative Repayment Plans

Source: model estimates.
1.8. Conclusions

What are the implications of higher levels of student debt on college graduates choices? In accordance with previous empirical studies, we find that graduating with higher levels of student debt causes higher earnings right after college. Nevertheless, we also show that higher indebtedness implies lower earnings growth in the years after graduation. We then argue that this is the result of student debt influencing human capital and career choices of college graduates. In particular, we find that individuals with higher levels of student debt are less likely to enroll in post bachelor degree programs.

We contribute to the existing literature by showing that horizon effects determined by preferences for home-ownership are an important channel for obtaining this result. While credit constraints are a necessary ingredient for initial financial conditions to affect life cycle outcomes, their interaction with a strong value attached to homeownership is able to create a wide gap between outcomes of workers that start their careers with different student debt balances.

Several policies have been advocated to help student loan borrowers. We contribute to the policy debate by showing the merits of two alternative proposals. One, that is redistributive in nature, is to operate with a widespread forgiveness plan of all undergraduate debt, financed by lump sum taxes to be repaid over an extended period of time by the same cohort whose debt was forgiven.

The other, that resembles closely the path chosen so far, aims at alleviating the burden of student debt by linking repayments to earnings. We show that the income based repayment plan (IBR) is able to achieve results that are quantitatively similar to more ambitious forgiveness programs - namely, that the IBR that already attracts a significant number of graduates is already an effective policy to reduce career and human capital accumulation distortions induced by student borrowing.
In future work, we plan to move in two directions. The first is to endogenize the college borrowing decision, by modeling undergraduate attendance, and nest our life cycle structure into a general equilibrium framework. After doing that, we will aim at comparing more comprehensive policies regarding education financing, human capital, and life cycle decisions.

Another important question to address requires extending the housing decision part of the model to allow for heterogeneity in housing size and location choice. The decline of interstate migration in the U.S. has long been associated with reduced labor market dynamism, although recent research pointed at it resulting from a reduction of the component of occupation specific human capital. In presence of location choices, housing becomes not only an investment, but can also be an obstacle to geographical and labor mobility. Changes in labor markets can thus have interactions with financial constraints, and generate interesting macroeconomic implications.
CHAPTER 2 : Upskilling and the Rise in College Inequality

2.1. Introduction

The college wage premium has increased substantially in recent decades in the United States, with a recent slowdown since 2005. Much of the overall increase in inequality has been due to an increase within education groups (Juhn, Murphy and Pierce (1993), Katz (1999) and Hoxby and Terry (1999)). Indeed, both the Gini coefficient as well as the ratio between the upper and lower earnings quartiles have increased for college-educated workers (Figure 2.1). This growth in inequality has happened at the same time of expansion in participation in Higher Education, especially in 4-year colleges (Figure 2.2).

There seem to be three possibilities to explain this growing dispersion in earnings of college-educated workers (Hoxby and Terry (1999)): (1) a growing variation in the quality of education (intensive margin), (2) a widening inequality in the abilities of students enrolling in college (extensive margin) and (3) an increasing rate of returns to human capital or a greater variance in the occupations that college-educated workers select into. In this article, I analyze the evolving role of the higher education market (intensive margin) in shaping earnings inequality between 2000 and 2018 in the United States.

Inequality among colleges has long been acknowledged (Bowen (1981) and Desrochers and Wellman (2011)). However, there is disagreement about college inequality trends in the past decades. Several papers have found that inequality has been stable over the period 2001-2010 (Halffman and Leydesdorff (2010) and Lau and Rosen (2016)). Taking a longer-term perspective, Davies and Zarifa (2012), Clotfelter (2017) and Capelle (2019) document a widening gap and increase in the Gini coefficient in college revenues and expenditures per student since 1970.

Lindley and McIntosh (2015) focuses on the second and third channel and show that both have played an important role in explaining growing college graduate wage inequality in the United Kingdom. They show that average college entrance test scores within majors declined over time, suggesting that the widening variance of test scores is due to universities accepting individuals from lower in the ability distribution.
Figure 2.1: Earnings inequality (college-educated workers)

(a) Annual Earnings (with some college)  (b) Annual Earnings (with a 4-year degree)

Source: CPS-ASEC supplement (age 25-30, employed full-time). Figure (a) includes workers with some college but no degree, Associate’s degree, Bachelor’s degree and post-Bachelor’s degree. Figure (b) includes workers with a Bachelor’s degree. Lines represent smoothed trend.

The main source of data used in this article comes from the College Scorecard Data, an online tool created by the United States Department of Education designed to provide reliable college information for students. College-level data files contain information for each university from 1996 to 2018. It contains information on college revenues, expenditures, cost of attendance, and average graduation rates (among many other variables). Data on federally aided students is linked to earnings data from administrative tax records to estimate median earnings among workers 6 to 10 years after first enrolling in college.

Following the literature on education and industrial organization (Epple, Romano and Sieg (2006), Epple et al. (2017) and Gordon and Hedlund (2019)), I assume that colleges maximize the quality they offer to students subject to a budget constraint. I assume a Cobb Douglas functional form of college quality: colleges are a technology that deliver to students a quality that depends on instructional expenditure per student and average ability of students. I also assume that dispersion in students ability affects college quality.
I model median post-college students earnings as a function of college quality. There is empirical evidence that relates post-college earnings to college quality. Black and Smith (2006) estimate a model in which college quality is a latent variable affecting post-college wages of students. Hoekstra (2009) looks at earnings of white males ten to fifteen years after high school graduation and finds a premium of 20% for students who attended the most selective state university. Andrews, Li and Lovenheim (2016) finds substantial heterogeneity in post-college earnings attributable to variation in college quality.

According to this earnings specification, the parameters of the production function of college quality can be identified using regression analysis. The point estimates suggest that college preparedness has a higher weight on explaining variation in college earnings relative to expenditure per student. On average, an increase of 10% in instructional expenditure per student (roughly a $1,100 increase, from an average of $11,000) is associated with an increase in 1% on median earnings 10 years after enrollment. On average, an increase in 10% in midpoint ACT composite score (approximately an increase of 2.3 points from an average of 23) is associated with an increase in 6% on median earnings 10 years after enrollment.

I then document three main trends in college inequality between 2000 and 2018. First, inequality in post-college students earnings has increased both within public and within private non-profit institutions. Second, inequality in both instructional expenditure per student and average ability of students have also increased within sectors (and maintains an increasing trend). Third, this growing variation in the quality of higher education is consistent with an increase in inequality in net tuition revenue and an increase in relative demand (especially for private non-profit doctoral universities).

The remaining of the paper is organized as follows. Section 2.2 reviews the major reasons behind the recent increase in college participation, section 2.3 describes the data and the construction of the variables, section 2.4 explains the methodology used and shows the main empirical results for the trends in college inequality, section 2.5 concludes.
2.2. Upskilling

Increasing participation in Higher Education, and the increasing number of college-educated workers in the labor market, is likely to change the post-college earnings distribution. College enrollment and attainment have increased substantially in the United States, especially for 4-year colleges. As we can see in Figure 2.2a, first-time enrollment at 4-year colleges has increased from 1.1 to 1.7 million students and the percentage of recent 4-year college graduates (aged 25-30) increased from 25% to 40% between 1996 and 2019. The increase in enrollment and attainment in community colleges has been much less pronounced\(^2\) with a recent decline in enrollment (Figure A2.1).

![Figure 2.2: College Enrollment and Attainment](image)

(a) 4-year College  (b) Some College

Source: IPEDS/College Scorecard and CPS-ASEC supplement. Figure (a) includes first-time undergraduate students enrolled in a predominantly Bachelor’s degree granting institution and individuals aged 25-30 with at least a Bachelor’s degree. Figure (b) includes first-time undergraduate students enrolled in some college and individuals aged 25-30 with some college but no degree, Associate’s degree, Bachelor’s degree and post-Bachelor’s degree. Lines represent smoothed trend.

\(^2\)The 2-year degree program in the US is also known as "Associate Degree". Schools offering such program are known as a community college. They are 2-year colleges which provide affordable post-secondary education as a pathway to a 4-year degree. These colleges are regulated by the state government or local districts.
Researchers from a multitude of fields have studied factors promoting college enrollment, focusing both on academic preparation and financial access to college. In 2001, the federal government passed a reform of American elementary and secondary education known as the *No Child Left Behind Act* (NCLB). In particular, NCLB required schools to close the achievement gaps in K-12 education. Since then, there’s been a significant increase in the percentage of high school graduates (Figure 2.3a).

There have also been several educational reforms promoting financial access to college, increasing both grant and loan eligibility and limits. Between 1990 and 2019, average grants per full-time equivalent (FTE) undergraduate student increased from $3,218 to $9,518 and average loans (federal and non-federal) per FTE undergraduate student increased from $1,730 to $5,366 in 2018 dollars. Overall, average aid (grants and loans) per FTE undergraduate student (in 2018 dollars) tripled over this period (Figure 2.3b).

---

3For example, the re-authorization of the Higher Education Act (1992) introduced an unsubsidized loan program and increased the borrowing limit from $17,250 to $23,000 for dependent students. In 2008, the Ensuring Cont. Access to Student Loans Act increased borrowing limit to $31,000 for dependent students.
The empirical literature on the impact of college financial aid has mostly focused on its effects on enrollment and attainment, and concludes that $1,000 increase in grants increases the likelihood of enrollment between 3 and 5 percentage points.\(^4\)\(^5\) Dynarski (2003) shows that increasing $1,000 in grants increases by 0.16 years the completed schooling of high school graduates that attend college. Bettinger (2004) provides evidence that the Pell Grants increases college persistence and completion.

However, Heckman, Lochner and Taber (1998) find that once allowing for equilibrium considerations, the impact of subsidies on college enrollment is one order of magnitude smaller than previously reported. A key characteristic of the model used to examine the equilibrium effect of educational policies is that it is set in a competitive equilibrium environment, with the impact on earnings resulting from the diminishing rate of return to labor in an aggregate production function. In this vein, Shephard and Sidibe (2019) build a general equilibrium search model and find that $1,000 tuition subsidy generates a 2 percentage point increase in college graduation.

Other studies look at the effect of student loans on college enrollment and attainment. For example, the models in Ionescu (2009) and Ionescu (2011) show that flexibility in the loan repayment increases college enrollment significantly. Dynarski (2002) finds that an increase of $1,000 in federal subsidised loans increases college attainment by 1.7 percentage points (for students from households who own housing equity).

Nevertheless, this recent expansion in financial aid has been accompanied by an increasing cost of college. Between 2000–01 and 2017–18, prices for undergraduate tuition, fees, room, and board at public 4-year colleges rose by 63 percent, and prices at private nonprofit 4-year colleges by 39 percent, after adjustment for inflation.\(^6\)

---

\(^4\)See Deming and Dynarski (2009) for a review of estimates.

\(^5\)Mattana and Joensen (2016) analyze how financial aid design (grants vs. loans) affect student behavior and achievement. A growing literature provides evidence that student loans have a negative impact on post-college outcomes. For a review of this literature see Folch and Mazzone (2021).

Gordon and Hedlund (2019) develop a model of higher education to test explanations for this increase in college tuition. Their model extends the static framework in Epple, Romano and Sieg (2006) and Epple et al. (2017) of imperfectly competitive, quality maximizing colleges, embedded in an incomplete market life-cycle environment. They analyze three driving forces: supply-side changes (Baumol’s cost disease and exogenous changes to non-tuition revenue), demand-side changes (expansion in financial aid, i.e. the Bennett Hypothesis), and macroeconomic forces (higher college earnings premium). Their model shows that demand-driven factors are the largest driving force behind the increase in college enrollment and tuition revenue.

Finally, structural shifts in the labor market can also have an impact on the educational choices of workers. The college earnings premium has increased substantially in recent decades, with a recent slowdown since 2005 (Beaudry, Green and Sand (2016), Valletta (2018) and Autor, Goldin and Katz. (2020)). Nevertheless, as shown in Figure 2.2, this recent stagnation has occurred during a period of rapid expansion in college enrollment.

Higher demand for college educated workers (employer skill upgrading) is consistent with the increasing supply of college educated workers and the recent flattening of the college wage premium. As Figure 2.4 shows, the percentage of jobs requiring a Bachelor’s degree has increased, from 25% of jobs requiring a BA degree in the typical occupation in 2003-2006 to 35% in 2014-2019. In particular, using detailed job vacancy data, Blair and Deming (2020) show that the share of job vacancies requiring a bachelor’s degree increased by more than 60 percent between 2007 and 2019, suggesting that upskilling is probably here to stay, and that four-year college enrollment will keep growing.

To sum up, three major factors have contributed to the substantial increase in college participation over the last decades: increasing proportion of high school graduates, expansion of college financial aid, and, structural shifts in the labor market.
2.3. Data: College Scorecard

2.3.1. Data Description

The main source of data used in this article comes from the College Scorecard Data. The College Scorecard is an online tool, created by the United States Department of Education, designed to provide reliable information for students to compare the cost and quality of universities in the United States. The data provide information about colleges that receive federal financial aid\(^\text{7}\) and the outcomes of the students attending those universities. The data is obtained through three main sources: federal reporting from colleges, federal financial aid, and tax-level information\(^\text{8}\).

---

\(^7\)Some colleges that do not participate in Title IV programs are also included if they meet similar criteria to Title IV participating colleges.

\(^8\)For more information about the data see the technical documentation for institutional-level data files (Department of Education [2021]).
The data files contain information for each college from 1996-97 to 2018-19 academic years. The data cover college and students characteristics, college expenditures and revenues, financial aid, cost of attendance, and graduation rates (among many other variables). In addition, I merge the College Scorecard data with college-level data from the Integrated Postsecondary Data System (IPEDS) to add additional variables that are not included in the Scorecard database.\textsuperscript{9}

To gain insight into the labor market outcomes of students, data on federally aided students (those who receive federal grants and loans) is linked to earnings data from administrative tax records. These data are used to produce aggregated estimates of college-level statistics, such as the median earnings among workers 6 to 10 years after first enrolling in an institution.\textsuperscript{10} Students who are identified as currently enrolled (determined by having a federal loan that is in in-school deferment) are excluded from the calculation.

To ensure data are stable from year to year and representative of a certain number of students, the data are pooled across two cohorts. For example, median earnings 6 years after enrollment in 2014 is reflective of all federally aided students that enrolled in academic years 2007-08 and 2008-09. Moreover, colleges with fewer than 30 federally aided students are not included to ensure data are representative.

Since 2014, earnings data are limited to one-year post-completion, and are calculated for individual fields of study (where sufficient data is available). Hence, aggregated college-level median earnings 6 and 10 years after enrollment cover the period between 2003-2014. Other college-level data (cost of attendance, expenditure per student, tuition revenue...) are available for the full panel 2000-2019.\textsuperscript{11}

\textsuperscript{9}See Appendix 2.1 for a detailed description of the data and the variables used. 
\textsuperscript{10}Earnings are defined as the sum of wages and deferred compensation from all W-2 forms received for each student, plus self-employment earnings from Schedule SE. 
\textsuperscript{11}All nominal variables (median earnings, instructional expenditure per student, net tuition revenue per student and endowment assets) are inflation adjusted to 2018 prices.
I keep non-profit colleges\textsuperscript{12} that are predominantly bachelor’s-degree granting institutions. I drop from the sample community colleges, tribal and special focus colleges (engineering schools, business, arts and music...), U.S. Service schools, non-main campuses and universities in Outlying Areas (AS, FM, GU, MH, MP, PR, PW, VI). I keep colleges with earnings data for all the years of analysis (2002-2014). The resulting panel consists of 1,266 colleges.

2.3.2. Summary statistics

Table 2.1 displays the main summary statistics for the first year that contains post-college earnings information (2003/04) and the last academic year in the sample (2018/19). Importantly, statistics differ substantially depending if they are measured at the college-level or student-level. For example, 66% of non-profit 4-year colleges are private, but only 33% of first-time undergraduate degree-seeking students attended those in 2003. For this reason, all measures presented henceforth are student weighted\textsuperscript{13}.

Over time, the distribution of students across sectors, Carnegie classifications and regions has remained stable. Instructional expenditure per student has increased both for public and private non-profit universities. This increase in expenditure per student is consistent with the increase in the net tuition revenue per student in both sectors. The average and standard deviation of college entrance test score (ACT composite score) have increased, as well as the percentage of low-income students (those with family income lower than $30,000).

As we can see in Figure 2.5a, there is substantial dispersion in the distribution of college-level median earnings 6 years after enrollment in 2014 (the last year with earnings data available). The distribution becomes more unequal ten years after enrollment, suggesting that differences amplify over time. As expected, the distribution of median earnings 6 years after enrollment is more dispersed for private non-profit universities (Figure 2.5b).

\textsuperscript{12}Only 6% (2%) of first-time students enrolled in a (4-year) for-profit college in 2018.

\textsuperscript{13}Weights are computed using analytic weights, treating each observation as if each observation is a mean computed from a sample of size n, where n is the total number of first-time undergraduate degree-seeking students in a given year.
Table 2.1: Summary Statistics (4-year non-profit colleges)

<table>
<thead>
<tr>
<th></th>
<th>2003/04</th>
<th>2018/19</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>College-level</td>
<td>Student-level</td>
</tr>
<tr>
<td><strong>Institutional characteristics</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Public</td>
<td>37%</td>
<td>67%</td>
</tr>
<tr>
<td>Private non-profit</td>
<td>63%</td>
<td>33%</td>
</tr>
<tr>
<td>Baccalaureate Colleges</td>
<td>29%</td>
<td>11%</td>
</tr>
<tr>
<td>Master’s Colleges &amp; Universities</td>
<td>43%</td>
<td>30%</td>
</tr>
<tr>
<td>Doctoral Universities</td>
<td>29%</td>
<td>59%</td>
</tr>
<tr>
<td><strong>Instructional expenditure per FTE student (2018 Dollars)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Public</td>
<td>$8,160.9</td>
<td>$9,133.6</td>
</tr>
<tr>
<td>Private non-profit</td>
<td>[3781.8]</td>
<td>[4214.6]</td>
</tr>
<tr>
<td><strong>Net tuition revenue per FTE student (2018 Dollars)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Public</td>
<td>$5,633.8</td>
<td>$6,174.1</td>
</tr>
<tr>
<td>Private non-profit</td>
<td>[2928.6]</td>
<td>[2758.3]</td>
</tr>
<tr>
<td><strong>Regions (Bureau of Economic Analysis)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>New England (CT, ME, MA, NH, RI, VT)</td>
<td>9%</td>
<td>7%</td>
</tr>
<tr>
<td>Mideast (DE, DC, MD, NJ, NY, PA)</td>
<td>19%</td>
<td>17%</td>
</tr>
<tr>
<td>Great Lakes (IL, IN, MI, OH, WI)</td>
<td>16%</td>
<td>18%</td>
</tr>
<tr>
<td>Plains (IA, KS, MN, MO, NE, ND, SD)</td>
<td>11%</td>
<td>9%</td>
</tr>
<tr>
<td>Southeast (AL, AR, FL, GA, KY, LA, MS, NC, SC, TN, VA, WV)</td>
<td>25%</td>
<td>24%</td>
</tr>
<tr>
<td>Southwest (AZ, NM, OK, TX)</td>
<td>7%</td>
<td>9%</td>
</tr>
<tr>
<td>Rocky Mountains (CO, ID, MT, UT, WY)</td>
<td>3%</td>
<td>4%</td>
</tr>
<tr>
<td>Far West (AK, CA, HI, NV, OR, WA)</td>
<td>10%</td>
<td>11%</td>
</tr>
<tr>
<td><strong>Students characteristics and outcomes</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ACT Composite score</td>
<td>22.9</td>
<td>23.5</td>
</tr>
<tr>
<td>Percentage with family income lower than $30,000</td>
<td>29.3%</td>
<td>29.7%</td>
</tr>
<tr>
<td>Graduation rate 6 years after enrollment</td>
<td>52%</td>
<td>56%</td>
</tr>
<tr>
<td>Median earnings 6 years after enrollment (2018 Dollars)*</td>
<td>$38,832.1</td>
<td>$40,183.7</td>
</tr>
<tr>
<td>Observations</td>
<td>1,266</td>
<td>1,233,607</td>
</tr>
</tbody>
</table>

Source: College Scorecard Data and IPEDS. *Median earnings observed in 2003/04 and 2014/15.
Given that the measure of earnings is not conditional on educational attainment, the observed dispersion in post-college earnings also reflects dispersion in graduation rates. As Figure 2.5c–d shows, the dispersion in graduation rates is also important, with private non-profit colleges having higher graduation rates than public colleges.
2.3.3. Trends in post-college students earnings inequality

In this section, I show recent trends in college-level earnings inequality. I focus on median earnings 6 years after enrollment as it contains more years for the panel of colleges (2003, 2005, 2007, 2009, 2011-2014) than 10 years after enrollment (2007, 2009, 2011-2014). I consider two measures of inequality, the Gini coefficient and the ratio of earnings between the 95th (90th) percentile and 5th (10th) percentile. All trends have been smoothed.

As we can see in Figure 2.6a, earnings inequality 6 years after enrollment for 4-year non-profit colleges has increased over the period 2003-2014. Most of the increase happened between the tails of the distribution: the Gini coefficient increased from 0.09 to 0.11, the ratio 95/5 from 1.7 to 1.9 and the ratio 90/10 from 1.5 to 1.6. Within sectors (Figure 2.6b), the increase in the Gini coefficient is more pronounced in private non-profit colleges, from 0.105 to 0.14.

![Figure 2.6: Post-college Earnings Inequality (4-year colleges)](image)

(a) Gini coefficient and ratio of percentiles  
(b) Gini coefficient (by sector)

Source: College Scorecard Data. College-level median earnings are weighted using total first-time undergraduate degree-seeking students.

14These results also hold using median earnings 10 years after enrollment. See Figure A2.2
The majority of undergraduate students attend a college within the same state of residency\textsuperscript{15} and hence, it is also important to understand the trend in inequality within regions. The increase in inequality is generalized within regions but presents substantial heterogeneity (\textbf{Figure A2.3}). The regions with the highest Gini coefficient (in 2014) are New England, Mideast and Southeast, with Georgia, Washington, D.C, Massachusetts, North Carolina, Pennsylvania, South Carolina and New York being the states with higher post-college students earnings inequality (\textbf{Table A2.1}).

2.4. The Increase in College Inequality

2.4.1. \textit{Modeling college quality}

Given that I focus on non-profit 4-year colleges (which account for 95\% of 4-year colleges in the US), I follow the literature on education and industrial organization and assume that colleges maximize the quality they offer to students, denoted by \((q_{c,t})\), subject to a budget constraint \((\Psi_{c,t})\). This micro-foundational of the college sector is borrowed from a literature that estimates equilibrium models of higher education (\textit{Epple, Romano and Sieg (2006), Epple et al. (2017) and Gordon and Hedlund (2019)}).

Colleges can be modelled as clubs, as the ability of students attending matter for the quality delivered to all students. There is empirical evidence that peer effects enter the college quality production function. For example, \textit{Smith and Stange (2016)} and \textit{Mehta, Stinebrickner and Stinebrickner (2018)} provide evidence of the effects of peer-effects (from roommates) on college outcomes. \textit{Zimmerman (2019)} shows that social networks built during college matter for the labor market. In addition, peer effects also arise from colleges competing for high ability students \textit{(Hoxby (2009) and Hoxby (2013))}.\textsuperscript{15}

\textsuperscript{15}72\% of first-time undergraduate students enrolled in a Bachelor’s degree program in 2011/12 enroll in the same state of residency (Beginning Postsecondary Students Longitudinal Study (BPS), NCES)
Following this literature, I assume a Cobb Douglas functional form: colleges are a technology that deliver to students a quality that depends on expenditure per student \((x_{c,t})\) and the average ability of students \((a_{c,t})\). I also assume that the more heterogeneous the pool of students in terms of ability, the more difficult it is for a college to deliver a given quality to its students (as in Capelle (2020)). I define \((\sigma_{c,t})\) as the logarithm of the within-college dispersion of students ability. Hence, the production function of quality of college \(c\) measured at time \(t\) is given by:

\[
q_{c,t}(\theta_c, x_{c,t}, f(a_{c,t})) = \theta_c \cdot x_{c,t}^{\nu} \cdot \frac{a_{c,t}}{\sigma_{c,t}}^{\alpha_2}
\]  

(2.1)

Following Azuero and Zarruk Valencia (2017), I treat median earnings for students that attended college \(c\), \(\tau\) years after enrollment as a function of college quality:

\[
y_{c,t+\tau} = y_{t+\tau} \cdot (1 + \gamma_t \cdot q_{c,t})
\]

(2.2)

where \(y_{t+\tau}\) represents the median salary for non-college workers and \(\gamma_t\) the returns to education over time. According to this specification, the parameters of the production function of college quality can be identified using regression analysis. I use median students earnings 6 (and 10) years after enrollment for a panel of 1,299 colleges in the United States through 2009-2014 (see Section 2.3). To measure the average \((a_{c,t})\) and dispersion \((\sigma_{c,t})\) of students ability, I use the midpoint of the college entrance ACT composite score and the difference between percentile 75\(^{th}\) and percentile 25\(^{th}\). For \(x_{c,t}\), I use instructional expenditures divided by the number of full-time equivalent students. As we can see in Figure 2.7, both instructional expenditure per student and ACT scores have a strong positive correlation with median earnings 6 years after first-time enrollment within public and private non-profit colleges.\(^{16}\)

\(^{16}\)See Figure A2.4 for the correlation with median earnings 10 years after enrollment.
Private colleges are privately funded and funded mostly by endowment and donations as well as tuition. They also tend to have highly selective admittance practices, offering a smaller class size and therefore higher expenditure per student. Public colleges are instead federally funded, offering lower tuition and higher admittance rates than private schools. On average, public colleges have larger class sizes, lower expenditure per student and lower college entrance scores.
For every college (c) in the panel, the following earnings equation holds:

\[
\ln\left(\frac{y_{c,t+\tau}}{y_{t+\tau}} - 1\right) = \ln(\gamma_t) 
\]

Rearranging:

\[
\ln\left(\frac{y_{c,t+\tau}}{y_{t+\tau}} - 1\right) = \ln(\gamma_t) + \ln(\theta_c) + \nu \ln(x_{c,t}) + \alpha_1 \ln(a_{c,t}) - \alpha_2 \ln(\sigma_{c,t}) + W_{c,t} + \epsilon_{c,t+\tau}
\]

Where \( y_{c,t+\tau} \) represents median earnings of undergraduate degree-seeking students attending college c, \( \tau \) years after first-time enrollment, \( \gamma_t \) is a vector with cohort fixed effects, \( \theta_c \) is a vector of college controls (Sector, Carnegie classification and state fixed effects), \( x_{c,t} \) measures average instructional expenditure per student (6-year average since enrollment) for students attending college c, \( a_{c,t} \) is the midpoint test score (ACT) of admitted students in college c, \( \sigma_{c,t} \) is the ACT quartile deviation of admitted students and \( W_{c,t} \) is a set of students controls: percentage of low-income students (aided students with family income lower than $30,000), percentage of high-income students (aided students with family income higher than $75,000) and percentage of dependent students.

Family income is an important control to be included in the earnings regression as parental background matters for achievements and access to college (Epple, Romano and Sieg (2006), Bailey and Dynarski (2011) Chetty et al. (2019), Hoxby and Turner (2019)). In addition, the ACT score theoretically provide an objective way to compare students from different high schools. However, one concern about measuring ability of students by standardized test scores is that these tests favor students with higher parents' income and wealth (Dixon-Román, Everson and McArdle (2013)).
The results of the estimation are displayed in Table 2.2. Columns (1) and (4) show the parameter estimates including sector and cohort fixed effects, columns (2) and (5) add state and Carnegie fixed effects, and, columns (3) and (6) include students controls.

<table>
<thead>
<tr>
<th></th>
<th>log(Median Earnings)</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>6 years after enrollment</td>
<td>10 years after enrollment</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( \tau = 6 )</td>
<td>( \tau = 10 )</td>
</tr>
<tr>
<td>( \nu )</td>
<td>0.08***</td>
<td>0.06***</td>
<td>0.09***</td>
</tr>
<tr>
<td></td>
<td>[0.005]</td>
<td>[0.004]</td>
<td>[0.005]</td>
</tr>
<tr>
<td>( \alpha_1 )</td>
<td>0.81***</td>
<td>0.77***</td>
<td>0.60***</td>
</tr>
<tr>
<td></td>
<td>[0.021]</td>
<td>[0.019]</td>
<td>[0.023]</td>
</tr>
<tr>
<td>( \alpha_2 )</td>
<td>-0.06***</td>
<td>-0.05***</td>
<td>-0.04***</td>
</tr>
<tr>
<td></td>
<td>[0.0.009]</td>
<td>[0.008]</td>
<td>[0.008]</td>
</tr>
<tr>
<td>Sector and Cohort FE</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>State and Carnegie FE</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Student Controls</td>
<td>N</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td>Adj. R-squared</td>
<td>0.56</td>
<td>0.71</td>
<td>0.76</td>
</tr>
<tr>
<td>Observations</td>
<td>4,619</td>
<td>4,619</td>
<td>4,462</td>
</tr>
</tbody>
</table>

Standard errors in brackets.

Source: College Scorecard Data. Regressions are weighted using total first-time undergraduate degree-seeking students.
The estimates show that the elasticities $\nu$ and $\alpha_1$ are both statistically significant in all the specifications. As expected, the parameter $\alpha_2$ is negative and statistically significant. This means that, on average, the dispersion in students’ ability is associated with lower earnings after college enrollment.

The point estimates suggest that college preparedness has a higher weight on explaining variation in college earnings compared to expenditure per student. On average, an increase of 10% in instructional expenditure per student (roughly a $1,100 increase, from an average expenditure of $11,000) is associated with an increase in 1% on median earnings 10 years after enrollment. On average, an increase in 10% in the median ACT composite score (approximately an increase of 2.3 points from an average score of 23) is associated with an increase in 6% on median earnings 10 years after enrollment. Given that the measure of earnings is not conditional on educational attainment, the observed variation in post-college earnings also reflects variation in graduation rates. Median earnings $\tau$ years after enrollment can be expressed as follows:

$$y_{c,t+\tau} = (1 - p_{c,t+\tau}) \cdot y_{c,ng,t+\tau} + p_{c,t+\tau} \cdot y_{c,g,t+\tau}$$  \hspace{1cm} (2.5)

Where $p_{c,t+\tau}$ reflects the percentage of students that graduated within $\tau$ years after first-time enrollment, $y_{c,ng,t+\tau}$ median earnings of drop-outs and $y_{c,g,t+\tau}$ median earnings of graduates. Table A2.3 shows the parameter estimates when using as dependent variable graduation rates 6 and 8 years after first-time enrollment.

---

17 One potential bias in the measure of ability using ACT test scores is that the percentage of first-time undergraduate students that took the ACT might be low in some colleges. In Table A2.2 I restrict the sample to colleges with at least 50% of first-time degree/certificate-seeking students submitting ACT scores.

18 Table A2.2 shows similar coefficient results within public and private non-profit universities.
2.4.2. Evolution of college inequality

Understanding the production function of college quality helps to explain the contribution of the intensive margin channel on increasing post-college earnings inequality. In this section, I group 4-year non-profit colleges (within sector) between those above (below) the upper (lower) earnings decile in 2014. I use this criteria to show the evolution in the ratio of students post-college earnings (6 and 10 years after enrollment) over the period 2002-2014 as well as the ratios of college quality (expenditure and ACT scores) and tuition revenue over the period 2002-2018.

Following Equation 2.4 these differences between groups (within sector, s) can be modeled as follows:

$$
\Delta y_{s,t} = \ln(\bar{y}_{s,90,t}) - \ln(\bar{y}_{s,10,t}) = \nu(\ln(\bar{x}_{s,90,t}) - \ln(\bar{x}_{s,10,t})) + \alpha_1(\ln(\bar{a}_{s,90,t}) - \ln(\bar{a}_{s,10,t})) + u_{s,t}
$$

I show three main results. First, as shown in section 2.3, the average difference in post-college earnings between the upper and lower deciles has increased substantially both within public and private non-profit universities (Figures 2.8a-b). Second, the ratios in average expenditure per student and ACT test scores between these two groups have also increased and maintain an increasing trend (Figures 2.8c-d). This increasing trend points to a continued increase in college quality and post-college earnings inequality for 2014-2018. Third, according with colleges maximizing the quality they offer subject to a budget constraint, (Figures 2.8e-f) show that this increase in college inequality is consistent with an increase in net tuition revenue per student and an increase in relative demand (measured as a decrease in the difference of the admitted-to-applicants ratio)⁴⁹

⁴⁹In Appendix A2.2 I remove universities in the upper and lower percentile (Figure A2.5). Alternatively, I group colleges by their Carnegie classification instead of earnings deciles (Figure A2.6). The trends remain in both cases. Figure A2.7 shows the evolution of ratios without grouping by sectors.
Figure 2.8: Evolution of College Ratios

(a) $\Delta^V_{s,t}$ (6 years after enrollment)  
(b) $\Delta^P_{s,t}$ (10 years after enrollment)  
(c) $\Delta^X_{s,t}$  
(d) $\Delta^Q_{s,t}$  
(e) $\Delta^T_{s,t}$  
(f) $\Delta^S_{s,t}$
When comparing the Carnegie classification between deciles (Table 2.3), colleges at the upper earnings decile in 2014 consist mostly of research/doctoral universities: doctoral universities in the lower decile represent 25.72% (4.71%) of public (private non-profit) colleges, while 89.68% (91.22%) in the upper decile.

One additional factor contributing to the increasing demand and net tuition revenue for doctoral universities is the increase in enrollment from international students. International students usually pay full tuition price, increasing net tuition revenues for colleges. The number of international students attending 4-year colleges has increased substantially in recent years (Figure A2.8), from 1.7% (4.1%) to 3.2% (7.5%) in public (private non-profit) doctoral universities.

In addition to tuition revenue, there are two additional revenue sources for U.S. universities: government funding and endowment assets (and private gifts). Capelle (2019) shows that both the average and the progressivity of government transfers to 4-year public colleges have declined over 1980-2016. In addition, Lerner, Schoar and Wang (2008) document growing differences in endowments for private colleges. As Figure A2.9 shows, endowment per student for doctoral private non-profit universities has increased substantially.

Table 2.3: Carnegie Classification (2018)

<table>
<thead>
<tr>
<th></th>
<th>Public</th>
<th>Private non-profit</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Lower</td>
<td>Upper</td>
</tr>
<tr>
<td>Baccalaureate Colleges</td>
<td>12.20%</td>
<td>1.45%</td>
</tr>
<tr>
<td>Master’s Colleges &amp; Universities</td>
<td>62.08%</td>
<td>8.87%</td>
</tr>
<tr>
<td>Doctoral Universities</td>
<td>25.72%</td>
<td>89.68%</td>
</tr>
<tr>
<td>First-time enrolled students</td>
<td>85,578</td>
<td>114,174</td>
</tr>
</tbody>
</table>

Source: College Scorecard Data and IPEDS.

Desrochers and Wellman (2011) report trends in college revenues over the period 1999–2009. Private colleges rely mostly on tuition revenue and endowments, while public colleges rely mainly on government funding and tuition revenue.
2.5. Conclusions

The college wage premium has increased in recent decades, with a recent slowdown since 2005, despite substantial increases in 4-year college enrollment. Simple focus on wage differentials between educational groups can miss some of the overall recent change in earnings inequality. Indeed, wage inequality within college-educated workers has increased substantially in the United States. In this paper, I show how a growing variation in the quality of 4-year colleges (the intensive margin channel) has contributed to this increase in wage inequality for college-educated workers.

I define colleges as a technology that deliver to students a quality that depends on instructional expenditure per student and average (and dispersion of) students ability. Using college-level data linked to administrative students earnings records, I model post-college students earnings as a function of college quality. According to this specification, the parameters of the production function of college quality can be identified using regression analysis. The point estimates suggest that college preparedness has a higher weight on explaining variation in college earnings relative to expenditure per student.

I then document three main trends in college inequality between 2002 and 2018. First, inequality in post-college students earnings has increased both within public and within private non-profit institutions. Second, inequality in both instructional expenditure per student and average ability of students have also increased within sectors (and maintain an increasing trend). Third, according with colleges maximizing the quality they offer subject to a budget constraint, this growing variation in the quality of higher education is consistent with an increase in inequality in net tuition revenue and an increase in the relative demand (especially for private non-profit doctoral universities).

This increasing self-segregation of students across colleges emphasizes how attending college per se does not compensate for a lower initial ability level and provide additional perspective on why human capital investment before college is critical in shaping earnings inequality.
A1.1. Data

This section describes in more detail the data sources for the variables covered in Section 1.3.

A1.1.1. B&B data

We use the restricted-use data and keep observations that have a positive value in the weight variable wte000, which represents the students who received a bachelor’s degree in the 2007-08 academic year and responded to all interviews (2007-08, 2009, and 2012). We use the following variables:

Debt:

**Cumulative loan amount borrowed for undergraduate through 2007-08 (b1borat):** Indicates the cumulative amount borrowed from all sources for the respondent’s undergraduate education through June 30, 2008. Does not include Parent PLUS loans. We log-transform this variable and deal with zero values by adding $1.

Post College outcomes:

**2012 Current Primary Job Salary (b2cjsal):** Indicates the respondent’s annualized salary from their current or most recent primary job. Primary job is defined as the respondent’s current or most recent job that lasted more than 3 months. We replace with a zero value the earnings of those who were not working at the time of the interview but reported the most recent earnings. We log-transform this variable and deal with zero values by adding $1.
2009 Current Primary Job Salary (b1erninc): Indicates the respondent’s income from their current job as of the B&B:09 interview. For respondents with multiple jobs, salary is only for the primary job, the job at which the respondent worked the most hours. We log-transform this variable and deal with zero values by adding $1.

2012 Current Value of Primary Residence (b2fhomval): Indicates the approximate current value of the respondent’s home(s), as reported by the respondent in the B&B:12 interview. We classify as home owners those observations with a value higher than zero. For the value of the house, we consider houses with a value higher than $100,000 and log-transform the variable.

Highest degree attained since bachelor’s as of 2012 (b2hideg): Identifies the highest postsecondary degree or certificate the respondent had obtained after completing the 2007-08 bachelor’s degree, as of the BB:12 interview. Variable categories are: Did not earn degree, Undergraduate certificate or diploma, Associate’s degree, Additional bachelor’s degree, Post-baccalaureate certificate...and Doctoral degree - other.

College Group Fixed Effects:

Institution Sector in 2007/08 (sector4): Indicates the sector of the 2007-08 bachelor’s degree-granting institution, using five categories. Variable categories are: Public 4-year, Private nonprofit 4-year, Public 2-year, For-profit, and Others or attended more than one institution. WE keep Public 4-year and Private nonprofit 4-year colleges.

Carnegie code (2005 basic, collapsed) for 2007-08 institution (cc2005c): Indicates the Carnegie basic institution classification code, using collapsed categories, of the 2007-08 bachelor’s degree-granting institution. Variable categories are: Associate’s, Research and doctoral, Master’s, Baccalaureate, and Special focus and other. We drop Associate’s and Special focus and other institutions.
Selectivity of 2007-08 bachelor’s degree institution (selectv2): Indicates the level of selectivity of the 2007-08 bachelor’s degree-granting institution. Only applies to public and private nonprofit 4-year institutions; other institutions were set to zero. The selectivity measure was developed for IPEDS, for public or private nonprofit 4-year institutions using the following criteria: whether the institution was open admission (no minimal requirements), the number of applicants, the number of students admitted, the 25th and 75th percentiles of ACT and/or SAT scores, and whether or not test scores were required. Variable categories are: Very selective, Moderately selective, Minimally selective, and Open admission.

Individual Controls:

Date of first postsecondary enrollment (pse_date): Identifies the year and month, in YYYYMM format, when the respondent first enrolled in postsecondary education. We keep those students that enroll between 2002 and 2004.

Expected Family Contribution in 2007-08 (efc): Indicates the composite estimate of the federal expected family contribution (EFC) used in 2007-08 need analysis.

Dependency status in 2007/2008 (depend): Indicates the respondent’s dependency status during the 2007-08 academic year.

SAT I score (tesatder): Indicates the respondent’s SAT I combined score, derived as either the sum of SAT I verbal and math scores or the ACT composite score converted to an estimated SAT I combined score using a concordance table from the following source: Dorans, N.J. (1999). Correspondences Between ACT and SAT I Scores (College Board Report No. 99-1).

Field of Study (majors4y): Indicates the respondent’s major or field of study, using 10 categories, for the 2007-08 bachelor’s degree. Variable categories are: Computer and information sciences; Engineering and engineering technology; Bio and phys science; sci
tech, math, agriculture; General studies and other; Social Sciences, Humanities, Health-
care, Business, Education and Other Applied. We classify them in three categories: STEM 
and health-care, Social Sciences and Business, Other.

**Race/Ethnicity (race):** Indicates the respondent’s race/ethnicity with Hispanic or Latino 
origin as a separate category. Variable categories are: White, Black or African American, 
Hispanic or Latino, Asian, American Indian or Alaska Native, Native Hawaiian, other and 
More than one race. We classify them into four categories: White, Black, Latino, Asian, 
Other.

**Gender (gender):** Indicates the respondent’s sex.

*1.1.2. IPEDS data*

Using harmonized college identifiers, we merge the B&B individual level data with insti-
tution level from the Institutional Post-Secondary Database (IPEDS). We use the IPEDS 
data in order to get information about the cost of attendance as well as the amount of 
grants and loans at the institutional level. We use the following variables for 2004-2007 
from the IPEDS data center:

**Student Debt:**

**Average amount of student loans awarded to full-time first-time undergraduates**
*(loan):* Any monies that must be repaid to the lending institution for which the student 
is the designated borrower. Includes all Title IV subsidized and unsubsidized loans and 
all institutionally- and privately-sponsored loans. Does not include PLUS and other loans 
made directly to parents.

**Percent of full-time first-time undergraduates awarded student loans** *(ploan):* 
Percentage of full-time, first-time degree/certificate-seeking undergraduate students who 
were awarded student loans.
Institutional Grants:

**Average amount of institutional grant aid awarded to full-time first-time undergraduates (grant):** Scholarships and fellowships granted and funded by the institution and/or individual departments within the institution, (i.e., instruction, research, public service) that may contribute indirectly to the enhancement of these programs.

**Percent of full-time first-time undergraduates awarded institutional grant aid (pgrant):** Percentage of full-time, first-time degree/certificate-seeking undergraduate students who were awarded institutional grants (scholarships/fellowships).

Grant-to-Aid:

Some of the institutions have a missing value in grants or loans and at the same time the percentage of students who were awarded grants or loans is zero. We substitute these observations with a zero value in grants or loans. We then drop all colleges with a grant-to-aid of 0 or 100 in any of the six years (2002-2007). Given that the average sum (and percent) of institutional grant and loan amounts are not available for 2002-2007, we construct the total institutional grant-to-aid ratio in the following way:

\[
\text{aid}_{j,t} = \text{ploan}_{j,t}\text{loan}_{j,t} + \text{pgrant}_{j,t}\text{grant}_{j,t} = (\frac{\text{TotalDebt}_{j,t}}{\text{Indebted}_{j,t}})(\frac{\text{Indebted}_{j,t}}{\text{Students}_{j,t}}) + (\frac{\text{Grant}_{j,t}}{\text{Recipient}_{j,t}})(\frac{\text{Recipient}_{j,t}}{\text{Students}_{j,t}})
\]

\[
x_{j,t} = \frac{(\frac{\text{Grant}_{j,t}}{\text{Recipient}_{j,t}})(\frac{\text{Recipient}_{j,t}}{\text{Students}_{j,t}})}{\text{aid}_{j,t}}
\]
A1.2. Solution Method

A1.2.1. Discrete-Continuous Choices

We illustrate how we take into account discrete choices with the problem of an employed renter with student loans, as in the Bellman Equation (1.8). For illustrative purposes only, we assume no borrowing constraints. If the worker had no discrete choices to make, the Bellman equation for the optimal consumption of a worker would satisfy the following first order condition known as the Euler equation:

\[ 0 = u'_c(c, s) - \beta (1 + r) \mathbb{E} \left( u'_c(c', s') \right) \]  
(A.1)

However, since at any period the renter worker can choose two discrete choices (to become a homeowner or switch career), the problem at the state vector point \{a, h, j, d, e, t\} involves solving for all the possible combinations of available discrete choices.

Following Iskhakov et al. (2017), we assume instead that the discrete choices are affected by choice-specific taste shocks, \( \sigma_{\varepsilon} \varepsilon_t \), i.i.d. Extreme Value type I distributed with scale parameter \( \sigma_{\varepsilon} \) as in McFadden (1973).

Taking again the value function in (1.8). Abstracting from career and repayment choice, and focusing only on the home-ownership decision, the expected value of the future value function becomes:

\[
\mathbb{E}[V'] = \max \mathbb{E}[V_r(k', h', j', m', d', e', t + 1)], \mathbb{E}[V_{o,\lambda}(k', h', j', m', d', e', t + 1)] = \\
= \max \mathbb{E}[V_r(\cdot, t + 1) + \sigma_{\varepsilon} \varepsilon(\cdot)], \mathbb{E}[V_{o,\lambda}(\cdot, t + 1) + \sigma_{\varepsilon} \varepsilon(\cdot)] = \\
= \sigma_{\varepsilon} \log \left( \exp \{V_r(\cdot, t + 1)/\sigma_{\varepsilon} \} + \exp \{V_{o,\lambda}(\cdot, t + 1)/\sigma_{\varepsilon} \} \right) \\
\]  
(A.2)
Thus, the Euler equation for a renter can then be written as:

\[ 0 = u'_c(c, s) - \beta(1 + r)E[u'_c(c', s' > 1) \cdot P(s' > 1|k', h', j', m', d', e') + u'_c(c', s' = 1) \cdot P(s' = 1|k', h', j', m', d', e')] \]  

(A.3)

where \( P(s' > 1) \) and \( P(s' = 1) \) are conditional choice probabilities given by the binomial logit formula:

\[
P(s' > 1|k', h', j', m', d', e') = \frac{\exp\{V_o,\lambda(\cdot, t + 1)/\sigma_\varepsilon\}}{\exp\{V_o,\lambda_H(\cdot, t + 1)/\sigma_\varepsilon\} + \exp\{V_r(\cdot, t + 1)/\sigma_\varepsilon\}}
\]

\[
P(s' = 1|k', h', j', m', d', e') = \frac{\exp\{V_r(\cdot, t + 1)/\sigma_\varepsilon\}}{\exp\{V_o,\lambda(\cdot, t + 1)/\sigma_\varepsilon\} + \exp\{V_r(\cdot, t + 1)/\sigma_\varepsilon\}}
\]

(A.4)

A1.2.2. Borrowing constraints

Solving (1.8) requires taking care of an additional issue. Formally, given the state \( S \) and indicating the Euler equation as \( \phi : S \times \mathbb{R}^m \to \mathbb{R} \), and the policy function as \( k' : S \times \mathbb{R}^m \to \mathbb{R} \), one needs to find policy and multiplier \( (k', \mu) \in \mathbb{R} \times \mathbb{R} \) s.t.

\[
\phi(S, k', \mu) = 0, \quad k' \geq \phi, \quad \mu \geq 0
\]

(A.5)

Following [Garcia and Zangwill (1981)], this problem can be transformed into a system of two equations, and can then be solved using standard solution algorithms for root finding.
Define a variable $\alpha$ such that:

$$
\alpha \equiv \begin{cases} 
\mu, & \text{if } \mu \geq 0, \ k' = \phi \\
-k', & \text{if } \mu = 0, \ k' \geq \phi
\end{cases}
$$

(A.6)

and

$$
\alpha^+ = (\max(0,\alpha))^k \\
\alpha^- = (\max(0,-\alpha))^k
$$

(A.7)

where $k \in \mathbb{N}^+$. The variable acts like a "penalty" when the constraint is violated, forcing the algorithm to search in the feasible set. The problem can be rewritten as finding policies and $\alpha$ such that:

$$
\phi(S, k', \alpha^+) = 0 \ , \ k' - \alpha^- = 0
$$

(A.8)
A1.3. Optimal Weight Matrix for GMM

We follow Erickson and Whited (2002) in computing the optimal weight matrix $\hat{\Omega}^{-1}$ from the following formula for clustered covariance:

$$
\hat{\Omega} = \frac{1}{nT} \sum_{i=1}^{n} \left( \sum_{t=1}^{T} \psi_{h(x_{i,t})} \right) \left( \sum_{t=1}^{T} \psi_{h(x_{i,t})} \right)^t
$$

(A.9)

in which $\psi_{h(x_{i,t})}$ is the vector of influence functions for the empirical moments $h(x_{i,t})$.

Deriving the influence functions for choice of moments is relatively straightforward. Take any subset of $h(x_{i,t})$ and denote it as $\theta$. For those moments that are obtained from simple averages, i.e. $\hat{\theta} = E(x_i)$, the influence function can be computed simply as:

$$
\psi_{\hat{\theta}}(x) = x - E(X)
$$

(A.10)

In the case of linear regression coefficients, we need to get influence function for the slope and the constant. The slope is $\hat{\theta}(\beta) = \frac{\text{Cov}(X,Y)}{\text{Var}(X)}$. Then:

$$
\psi_{\hat{\theta}(\beta)}(x, y) = \frac{(x - E(X))(y - E(Y)) - \text{Cov}(X,Y)}{\text{Var}(X)} - \frac{\left((x - E(X))^2 - \text{Var}(X)\right) \text{Cov}(X,Y)}{(\text{Var}(X))^2} = 
$$

$$
= \frac{(x - E(X))(y - E(Y)) - \beta (x - E(X))}{\text{Var}(X)} = \frac{(x - E(X))}{\text{Var}(X)} [(y - E(Y)) - \beta (x - E(X))]
$$

The constant is instead $\hat{\theta}(\alpha) = E(y) - \frac{\text{Cov}(X,Y)}{\text{Var}(X)} E(x) = E(y) - \frac{E(XY)E(X) - (E(X))^2E(Y)}{\text{Var}(X)}$. Then:

$$
\psi_{\hat{\theta}(\alpha)}(x, y) = - \frac{(xy - E(XY)) E(X) + (x - E(X)) E(XY) - (y - E(Y)) (E(X))^2 - 2(x - E(X)) E(X) E(Y)}{\text{Var}(X)} +
$$

$$+ \left(\frac{(x - E(X))^2 - \text{Var}(X)}{(\text{Var}(X))^2} \right) \frac{(E(XY)E(X) - (E(X))^2 E(Y))}{\text{Var}(X)} = y - E(Y) -
$$

$$- \frac{(xy - y E(X)) E(X) + (x - E(X)) \left(\text{Cov}(X,Y) - E(X) E(Y)\right)}{\text{Var}(X)} - \frac{\left((x - E(X))^2 - \text{Var}(X)\right)}{(\text{Var}(X))^2}$$
Finally, we use the ratio of regression coefficients. Take the ratio of slopes $\hat{\theta}(\beta_g/\beta_b)$. Then by the chain rule:

$$
\psi_{\hat{\theta}(\beta_g/\beta_b)}(x_g, y_g, x_b, y_b) = \frac{\psi_{\hat{\theta}(\alpha)}(x_g, y_g)\beta_b - \psi_{\hat{\theta}(\alpha)}(x_b, y_b)\beta_g}{\beta_b^2} \tag{A.11}
$$

And similarly, one can obtain the influence function for the ratio of constants.
A1.4. Additional Figures

Figure A1.1: Net Cost of College

![Figure A1.1: Net Cost of College](image)

Figure A1.2: Student Debt Burden

![Figure A1.2: Student Debt Burden](image)

Source: U.S. Department of Education, National Center for Education Statistics, 1993/94, 2000/01, 2007/08 and 2015/16 Baccalaureate and Beyond Longitudinal Study (B&B:93/94, B&B:2000/01, B&B:2007/08, B&B:2015/16). Figure A1.1 shows student budget minus all grants for graduating seniors and Figure A1.2 shows graduate’s monthly student loan payment as a percent of their monthly income.
Figure A1.3: Graduate School Attainment

Source: CPS-ASEC supplement (age 35-40), percentage of Bachelor’s Degree graduates with a post-Bachelor’s Degree. Line represents smoothed trend.
Figure A1.4: Model Identification
Distribution of yearly disposable income, i.e. labor wages plus net liquid asset holdings, minus debt payments and housing expenditures for workers with graduate school education. The red line represents the downpayment constraint.
### Table A1.1: Homeownership by Educational Attainment (1980-1994 cohorts)

<table>
<thead>
<tr>
<th></th>
<th>Homeownership for post-BA graduates</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>age 25</td>
</tr>
<tr>
<td></td>
<td>(1)</td>
</tr>
<tr>
<td>Graduate School</td>
<td>-0.150**</td>
</tr>
<tr>
<td></td>
<td>[0.028]</td>
</tr>
<tr>
<td>Cohort and State FE</td>
<td>Y</td>
</tr>
<tr>
<td>Individual Controls</td>
<td>Y</td>
</tr>
<tr>
<td>Observations</td>
<td>3,264</td>
</tr>
</tbody>
</table>

Source: CPS 2005-2020 (males, with a BA degree or higher).

### Table A1.2: Homeownership by Educational Attainment (1965-1979 cohorts)

<table>
<thead>
<tr>
<th></th>
<th>Homeownership for post-BA graduates</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>age 25</td>
</tr>
<tr>
<td></td>
<td>(1)</td>
</tr>
<tr>
<td>Graduate School</td>
<td>-0.06*</td>
</tr>
<tr>
<td></td>
<td>[0.036]</td>
</tr>
<tr>
<td>Cohort and State FE</td>
<td>Y</td>
</tr>
<tr>
<td>Individual Controls</td>
<td>Y</td>
</tr>
<tr>
<td>Observations</td>
<td>2,796</td>
</tr>
</tbody>
</table>

Source: CPS 1992-2014 (males, with a BA degree or higher).
A2.1. Data

This section describes in more detail the variables used in Section Section [2.3].

Cohort Definition and Weights:

**Earnings and Graduation Rates cohorts:** earnings measures are based on the universe of Title IV students in a given entry cohort at each college. Graduation rates are based on the set of full-time first-time students in a given entry cohort at each institution. To reduce variability, cohorts are pooled across two years.

**College weight:** total number of undergraduate degree-seeking students attending for the first time at the institution. Includes students enrolled in academic or occupational programs. Also includes students enrolled in the fall term who attended college for the first time in the prior summer term, and students who entered with advanced standing (college credits earned before graduation from high school). Weights are pooled across two years.

**Variables:**

**Median Earnings (in 2018 Dollars):** median earnings among workers (i.e., among those with positive yearly earnings) 6 to 10 years after first enrolling in an institution. Individuals who are identified as currently enrolled (determined by having a federal loan that is in in-school deferment) are excluded from the calculation. To reduce variability, data are pooled across two years. Earnings 6 years after enrollment are observed in 2003, 2005, 2007, 2009, 2011-2014. Earnings 10 years after enrollment are observed in 2007, 2009, 2011-2014.

**Graduation rates:** institutions report (via the IPEDS Graduation Rates component) on the completion rates for full-time, first-time students who complete within 100 or 150 percent of the expected time to completion. To reduce variability from year to year, graduation rates are pooled across two years on a rolling basis. Graduation rates are observed for the full panel.
Carnegie Classification: the Carnegie Foundation classifies institutions in several ways. I use the basic classification and classify colleges in three categories: Doctoral Universities (Very High Research Activity, High Research Activity, Doctoral/Professional Universities), Master’s Colleges Universities (Larger Programs, Medium Programs and Small Programs) and Baccalaureate Colleges (Arts Sciences Focus, Diverse Fields and Mixed Baccalaureate/Associate’s). I drop Associate’s and Special focus and other colleges.

ACT (American College Testing Program) score: the ACT assessment program measures educational development and readiness to pursue college-level coursework in English, mathematics, natural science, and social studies. The data include the 25th and 75th percentiles of ACT composite score. I use the midpoint of the 25th and 75th percentile. I drop colleges with a midpoint lower than 15.

Net tuition revenue per FTE student (in 2018 Dollars): Tuition and fees are revenues from all tuition and fees assessed against students (net of refunds and discounts allowances) for educational purposes. If tuition or fees are remitted to the state as an offset to the state appropriation, the total of such tuition or fees should be deducted from the total state appropriation and added to the total for tuition and fees. If an all-inclusive charge is made for tuition, board, room, and other services, a reasonable distribution is made between revenues for tuition and fees and revenues for auxiliary enterprises. Tuition and fees excludes charges for room, board, and other services rendered by auxiliary enterprises. FTE enrollment covers the 12-month period ending June 30 prior to the IPEDS collection year.

Instructional Expenditure per FTE student (in 2018 Dollars): total expenses is the sum of all operating expenses associated with the colleges, schools, departments, and other instructional divisions of the institution and for departmental research and public service that are not separately budgeted. This would include compensation for academic instruction, occupational and vocational instruction, community education, preparatory and adult basic education, and remedial and tutorial instruction conducted by the teaching faculty for the institution’s students. FTE enrollment covers the 12-month period ending
June 30 prior to the IPEDS collection year. I drop colleges with expenditure per FTE student lower than $100. For the regression analysis, I use a 6 year average of expenditure per student.

**Selectivity**: ratio of applicants to admitted students. Applicant: An individual who has fulfilled the institution’s requirements to be considered for admission (including payment or waiving of the application fee, if any) and who has been notified of one of the following actions: admission, nonadmission, placement on waiting list, or application withdrawn (by applicant or institution). Admitted: applicants that have been granted an official offer to enroll in a college or university.

**Endowment assets (in 2018 Dollars)**: consists of gross investments of endowment funds, term endowment funds, and funds functioning as endowment for the institution and any of its foundations and other affiliated organizations.

**Percent of low-income and high-income students**: percentage of aided students whose family income is between $0-$30,000 (low-income), percentage of aided students with family incomes higher than $75,001 in nominal dollars (high-income).
A2.2. Additional Figures

Figure A2.1: College Enrollment and Attainment (2-year colleges)

Source: IPEDS/College Scorecard and CPS-ASEC supplement. Figure includes first-time undergraduate students enrolled in a predominantly Associate's Degree granting institution and individuals aged 25-30 with an Associate’s degree. Line represents smoothed trend.

Figure A2.2: Earnings Inequality (10 years after enrollment)

(a) Gini and ratio of percentiles

(b) Gini (by sector)

Source: College Scorecard Data. College-level median earnings are weighted using total first-time undergraduate degree-seeking students.
Figure A2.3: Gini Coefficient (by region)

Source: College Scorecard Data. College-level median earnings are weighted using total first-time undergraduate degree-seeking students.
Figure A2.4: Earnings, Expenditure per Student, ACT Score (2014)

(a) Expenditure (public)  
(b) Expenditure (private non-profit)  
(c) ACT score (public)  
(d) ACT score (private non-profit)  

Source: College Scorecard Data. Data is weighted using total first-time undergraduate degree-seeking enrollment (10 years before).
Figure A2.5: Evolution of College Ratios (without upper and lower percentile)

(a) $\Delta_{s,t}^y$ (6 years after enrollment)  
(b) $\Delta_{s,t}^y$ (10 years after enrollment)

(c) $\Delta_{s,t}^x$  
(d) $\Delta_{s,t}^o$

(e) $\Delta_{s,t}^t$  
(f) $\Delta_{s,t}^S$

103
Figure A2.6: Evolution of College Ratios (doctoral vs. non-doctoral)

(a) $\Delta y_{s,t}^y$ (6 years after enrollment) 
(b) $\Delta y_{s,t}^n$ (10 years after enrollment)

(c) $\Delta x_{s,t}$  
(d) $\Delta a_{s,t}$

(e) $\Delta t_{s,t}$  
(f) $\Delta S_{s,t}$
Figure A2.7: Evolution of College Ratios (without grouping by sector)

(a) $\Delta_y^i$ (6 years after enrollment)  
(b) $\Delta_y^p$ (10 years after enrollment)

(c) $\Delta_y^i$  
(d) $\Delta_y^s$

(e) $\Delta_x^i$  
(f) $\Delta_y^s$
Figure A2.8: International Students Enrollment

Source: College Scorecard and IPEDS.
Figure A2.9: Evolution of Endowment per Student (4-year colleges)

Source: College Scorecard and IPEDS.
### Table A2.1: Gini Coefficient (by states)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>GA</td>
<td>0.101</td>
<td>0.144</td>
<td>2.19%</td>
<td>OR</td>
<td>0.06</td>
<td>0.074</td>
<td>1.14%</td>
</tr>
<tr>
<td>DC</td>
<td>0.098</td>
<td>0.141</td>
<td>0.63%</td>
<td>OH</td>
<td>0.063</td>
<td>0.0744</td>
<td>4.43%</td>
</tr>
<tr>
<td>MA</td>
<td>0.097</td>
<td>0.126</td>
<td>3.37%</td>
<td>WA</td>
<td>0.045</td>
<td>0.069</td>
<td>1.69%</td>
</tr>
<tr>
<td>NC</td>
<td>0.074</td>
<td>0.125</td>
<td>3.21%</td>
<td>FL</td>
<td>0.046</td>
<td>0.067</td>
<td>3.65%</td>
</tr>
<tr>
<td>PA</td>
<td>0.101</td>
<td>0.122</td>
<td>5.86%</td>
<td>WV</td>
<td>0.076</td>
<td>0.059</td>
<td>0.80%</td>
</tr>
<tr>
<td>SC</td>
<td>0.08</td>
<td>0.121</td>
<td>1.73%</td>
<td>NM</td>
<td>0.037</td>
<td>0.054</td>
<td>0.46%</td>
</tr>
<tr>
<td>MO</td>
<td>0.087</td>
<td>0.116</td>
<td>2.26%</td>
<td>UT</td>
<td>0.033</td>
<td>0.052</td>
<td>0.90%</td>
</tr>
<tr>
<td>NY</td>
<td>0.090</td>
<td>0.116</td>
<td>6.97%</td>
<td>MN</td>
<td>0.048</td>
<td>0.049</td>
<td>1.79%</td>
</tr>
<tr>
<td>ME</td>
<td>0.099</td>
<td>0.116</td>
<td>0.47%</td>
<td>MT</td>
<td>0.039</td>
<td>0.048</td>
<td>0.50%</td>
</tr>
<tr>
<td>MD</td>
<td>0.066</td>
<td>0.110</td>
<td>1.34%</td>
<td>NE</td>
<td>0.049</td>
<td>0.048</td>
<td>0.84%</td>
</tr>
<tr>
<td>NH</td>
<td>0.060</td>
<td>0.110</td>
<td>0.80%</td>
<td>HI</td>
<td>0.051</td>
<td>0.047</td>
<td>0.27%</td>
</tr>
<tr>
<td>VA</td>
<td>0.089</td>
<td>0.103</td>
<td>3.11%</td>
<td>IA</td>
<td>0.034</td>
<td>0.045</td>
<td>1.44%</td>
</tr>
<tr>
<td>IN</td>
<td>0.084</td>
<td>0.101</td>
<td>3.16%</td>
<td>SD</td>
<td>0.045</td>
<td>0.039</td>
<td>0.39%</td>
</tr>
<tr>
<td>MI</td>
<td>0.073</td>
<td>0.099</td>
<td>3.38%</td>
<td>KS</td>
<td>0.035</td>
<td>0.039</td>
<td>1.09%</td>
</tr>
<tr>
<td>IL</td>
<td>0.074</td>
<td>0.098</td>
<td>3.44%</td>
<td>ND</td>
<td>0.053</td>
<td>0.026</td>
<td>0.41%</td>
</tr>
<tr>
<td>CA</td>
<td>0.061</td>
<td>0.095</td>
<td>9.07%</td>
<td>AZ</td>
<td>0.026</td>
<td>0.025</td>
<td>1.47%</td>
</tr>
<tr>
<td>NJ</td>
<td>0.046</td>
<td>0.094</td>
<td>1.98%</td>
<td>ID</td>
<td>0.034</td>
<td>0.007</td>
<td>0.47%</td>
</tr>
<tr>
<td>AL</td>
<td>0.072</td>
<td>0.093</td>
<td>1.91%</td>
<td>AK</td>
<td>-</td>
<td>-</td>
<td>0.20%</td>
</tr>
<tr>
<td>RI</td>
<td>0.053</td>
<td>0.091</td>
<td>0.66%</td>
<td>DE</td>
<td>-</td>
<td>-</td>
<td>0.38%</td>
</tr>
<tr>
<td>CO</td>
<td>0.067</td>
<td>0.090</td>
<td>1.78%</td>
<td>WY</td>
<td>-</td>
<td>-</td>
<td>0.11%</td>
</tr>
<tr>
<td>CT</td>
<td>0.055</td>
<td>0.088</td>
<td>1.29%</td>
<td>NV</td>
<td>-</td>
<td>-</td>
<td>0.51%</td>
</tr>
<tr>
<td>AR</td>
<td>0.063</td>
<td>0.086</td>
<td>0.96%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TX</td>
<td>0.070</td>
<td>0.085</td>
<td>6.82%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LA</td>
<td>0.081</td>
<td>0.079</td>
<td>1.68%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TN</td>
<td>0.062</td>
<td>0.078</td>
<td>2.15%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VT</td>
<td>0.061</td>
<td>0.077</td>
<td>0.36%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>KY</td>
<td>0.060</td>
<td>0.076</td>
<td>1.75%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OK</td>
<td>0.055</td>
<td>0.076</td>
<td>1.24%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MS</td>
<td>0.08</td>
<td>0.075</td>
<td>0.89%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WI</td>
<td>0.05</td>
<td>0.074</td>
<td>2.35%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Source: College Scorecard Data and IPEDS.
Table A2.2: College Quality Estimates (ACT and by sector)

<table>
<thead>
<tr>
<th></th>
<th>6 Years after enrollment</th>
<th></th>
<th>10 Years after enrollment</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$P_{ACT} &gt;50%$ Public</td>
<td>$P_{ACT} &gt;50%$ Private</td>
<td>$P_{ACT} &gt;50%$ Public</td>
<td>$P_{ACT} &gt;50%$ Private</td>
</tr>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
</tr>
<tr>
<td>$\nu$</td>
<td>0.10***</td>
<td>0.10***</td>
<td>0.10***</td>
<td>0.10***</td>
</tr>
<tr>
<td></td>
<td>[0.006]</td>
<td>[0.008]</td>
<td>[0.007]</td>
<td>[0.007]</td>
</tr>
<tr>
<td>$\alpha_1$</td>
<td>0.57***</td>
<td>0.58***</td>
<td>0.62***</td>
<td>0.67***</td>
</tr>
<tr>
<td></td>
<td>[0.032]</td>
<td>[0.036]</td>
<td>[0.035]</td>
<td>[0.041]</td>
</tr>
<tr>
<td>$\alpha_2$</td>
<td>-0.04***</td>
<td>-0.06***</td>
<td>-0.09***</td>
<td>-0.04***</td>
</tr>
<tr>
<td></td>
<td>[0.011]</td>
<td>[0.011]</td>
<td>[0.012]</td>
<td>[0.014]</td>
</tr>
<tr>
<td>Sector and Cohort FE</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>State and Carnegie FE</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Students Controls</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Adj. R-squared</td>
<td>0.77</td>
<td>0.77</td>
<td>0.76</td>
<td>0.80</td>
</tr>
<tr>
<td>Observations</td>
<td>2,310</td>
<td>1,688</td>
<td>2,774</td>
<td>1,263</td>
</tr>
</tbody>
</table>

Standard errors in brackets.

Source: College Scorecard Data. Regressions are weighted using total first-time undergraduate degree-seeking enrollment.
### Table A2.3: College Quality Estimates (graduation rate)

<table>
<thead>
<tr>
<th></th>
<th>6 years after enrollment</th>
<th>8 years after enrollment</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( \tau = 6 )</td>
<td>( \tau = 8 )</td>
</tr>
<tr>
<td>( \nu )</td>
<td>(1) 0.06***</td>
<td>(2) 0.03***</td>
</tr>
<tr>
<td></td>
<td>(3) 0.04***</td>
<td>(4) 0.06***</td>
</tr>
<tr>
<td></td>
<td>(5) 0.035***</td>
<td>(6) 0.04***</td>
</tr>
<tr>
<td></td>
<td>[0.002]</td>
<td>[0.002]</td>
</tr>
<tr>
<td></td>
<td>[0.002]</td>
<td>[0.002]</td>
</tr>
<tr>
<td>( \alpha_1 )</td>
<td>(1) 1.00***</td>
<td>(2) 1.07***</td>
</tr>
<tr>
<td></td>
<td>(3) 0.84***</td>
<td>(4) 0.95***</td>
</tr>
<tr>
<td></td>
<td>(5) 1.02***</td>
<td>(6) 0.83***</td>
</tr>
<tr>
<td></td>
<td>[0.009]</td>
<td>[0.008]</td>
</tr>
<tr>
<td></td>
<td>[0.009]</td>
<td>[0.001]</td>
</tr>
<tr>
<td>( \alpha_2 )</td>
<td>(1) -0.041***</td>
<td>(2) -0.017***</td>
</tr>
<tr>
<td></td>
<td>(3) -0.011**</td>
<td>(4) -0.041***</td>
</tr>
<tr>
<td></td>
<td>(5) -0.024***</td>
<td>(6) -0.015***</td>
</tr>
<tr>
<td></td>
<td>[0.004]</td>
<td>[0.003]</td>
</tr>
<tr>
<td></td>
<td>[0.003]</td>
<td>[0.004]</td>
</tr>
</tbody>
</table>

#### Additional Information

- **Sector and Cohort FE:** Y Y Y Y Y Y
- **State and Carnegie FE:** N Y Y N Y Y
- **Students Controls:** N N Y N N Y
- **Adj. R-squared:** 0.78 0.85 0.88 0.77 0.85 0.87
- **Observations:** 10,187 10,187 9,811 8,083 8,083 7,796

Standard errors in brackets.

Source: College Scorecard Data. Regressions are weighted using total first-time undergraduate degree-seeking enrollment.
BIBLIOGRAPHY


Fernández-Villaverde, J., and D. Krueger. 2011. “Consumption and saving over the life cycle: how important are consumer durables?” *Cambridge University Press*.


Garcia, CB, and WI Zangwill. 1981. “Pathways to solutions, fixed points, and equilibria.”


