

# Dynamic Models of Human Capital Investment

by

Jared Ashworth

Department of Economics  
Duke University

Date: \_\_\_\_\_

Approved:

---

V. Joseph Hotz, Supervisor

---

Peter Arcidiacono

---

Hugh Macartney

---

Arnaud Maurel

Dissertation submitted in partial fulfillment of the requirements for the degree of  
Doctor of Philosophy in the Department of Economics  
in the Graduate School of Duke University  
2015

ABSTRACT

Dynamic Models of Human Capital Investment

by

Jared Ashworth

Department of Economics  
Duke University

Date: \_\_\_\_\_

Approved:

---

V. Joseph Hotz, Supervisor

---

Peter Arcidiacono

---

Hugh Macartney

---

Arnaud Maurel

An abstract of a dissertation submitted in partial fulfillment of the requirements for  
the degree of Doctor of Philosophy in the Department of Economics  
in the Graduate School of Duke University  
2015

Copyright © 2015 by Jared Ashworth  
All rights reserved except the rights granted by the  
Creative Commons Attribution-Noncommercial Licence

# Abstract

My dissertation examines human capital investments and their role in individual's labor market outcomes. Chapter 2 analyzes how public school teachers decide to make human capital investments and the effects that these decisions have on their future labor market outcomes. In particular, I look at the decisions of employed teachers to obtain an advanced degree. Teachers' education and career decisions are modeled via a dynamic framework in the presence of teacher-specific unobserved heterogeneity. I find that teachers' decisions to obtain master's degrees are motivated by more than just an increase in salary. In particular, I observe teachers with master's degrees receiving a better draw on job characteristics, as measured by school quality, and that teachers are willing to pay between \$1,500 and \$20,000 to to move up one quartile in school quality. I also find that teachers value having broad access to online degree programs more than they dislike tuition costs. Counterfactual simulations by unobserved ability are consistent with a story that high-type teachers value both the salary increase and a better draw in career prospects, whereas low-type teachers are mostly interested in the salary increase.

Chapter 3 investigates the evolution over the last two decades in the wage returns to schooling and early work experience. Using data from the 1979 and 1997 panels of the National Longitudinal Survey of Youth, we isolate changes in skill prices from changes in composition by estimating a dynamic model of schooling and work decisions. Importantly, this allows us to account for the endogenous nature of the

changes in educational and accumulated work experience over this time period. We find an increase over this period in the returns to working in high school, but a decrease in the returns to working while in college. We also find an increase in the incidence of working in college, but that any detrimental impact of in-college work experience is offset by changes in other observable characteristics. Overall, our decomposition of the evolution in skill premia suggests that both price and composition effects play an important role. The role of unobserved ability is also important.

To Kristie, my love, my rock, my everything

# Contents

<b>Abstract</b>	<b>iv</b>
<b>List of Tables</b>	<b>x</b>
<b>List of Figures</b>	<b>xii</b>
<b>Acknowledgements</b>	<b>xiii</b>
<b>1 Introduction</b>	<b>1</b>
<b>2 Teacher Education Decisions and Career Outcomes</b>	<b>6</b>
2.1 Introduction . . . . .	6
2.2 Data and Descriptives . . . . .	10
2.2.1 Data and Background . . . . .	10
2.2.2 Job Characteristics . . . . .	12
2.2.3 College Characteristics . . . . .	15
2.2.4 Teacher Characteristics . . . . .	17
2.3 Model . . . . .	19
2.3.1 Choices . . . . .	19
2.3.2 Payoffs . . . . .	20
2.3.3 Value Functions . . . . .	25
2.4 Identification and Estimation . . . . .	27
2.4.1 Assumptions . . . . .	27
2.4.2 Identification and Estimation . . . . .	28

2.5	Results . . . . .	29
2.6	Counterfactuals . . . . .	33
2.6.1	Model Fit . . . . .	33
2.6.2	Simulations . . . . .	34
2.7	Conclusion . . . . .	35
	Figures and Tables . . . . .	37
<b>3</b>	<b>Changes across Cohorts in Wage Returns to Schooling and Early Work Experiences: Distinguishing Price and Composition Effects</b>	<b>51</b>
3.1	Introduction . . . . .	51
3.2	Previous Literature . . . . .	54
3.3	Trends in Wages, Skills and Skill Returns across Cohorts . . . . .	57
3.3.1	Data . . . . .	58
3.3.2	Trends . . . . .	63
3.4	The Model . . . . .	69
3.4.1	Activities and Risk Sets . . . . .	69
3.4.2	School and Work Experience . . . . .	70
3.4.3	Wages . . . . .	71
3.4.4	Activity Choice Value Functions . . . . .	72
3.4.5	Cognitive and Non-cognitive Ability . . . . .	73
3.5	Estimation . . . . .	74
3.5.1	Error Structure . . . . .	74
3.5.2	Likelihood Function . . . . .	75
3.5.3	Identification . . . . .	77
3.6	Results . . . . .	78
3.6.1	Specification of the model . . . . .	79
3.6.2	Sheepskin effects . . . . .	81



3.6.3	Returns to experience . . . . .	82
3.6.4	Factor loadings . . . . .	85
3.7	Decompositions . . . . .	85
3.7.1	Setup . . . . .	86
3.7.2	Results . . . . .	92
3.8	Conclusion . . . . .	95
	Figures and Tables . . . . .	97
<b>A</b>	<b>Data and Estimation Notes for Chapter 2</b>	<b>113</b>
A.1	Sample Selection . . . . .	113
A.2	Tuition . . . . .	114
A.3	Identification . . . . .	115
A.4	Estimation . . . . .	117
<b>B</b>	<b>Data and Estimation Notes for Chapter 3</b>	<b>120</b>
B.1	Sample Selection . . . . .	120
B.2	Formal definitions of decomposition components . . . . .	122
	<b>Bibliography</b>	<b>126</b>
	<b>Biography</b>	<b>131</b>

# List of Tables

2.1	Sample Teacher Salary Schedule, 2005-2006 . . . . .	37
2.2	Teacher Transition Rates by Bachelor's and Any Master's . . . . .	40
2.3	Teacher Transition Rates by Bachelor's and For-Profit Master's . . . . .	41
2.4	Impact on Exit Rates . . . . .	41
2.5	Demographic Statistics . . . . .	44
2.6	Structural Estimates - Ability Equation . . . . .	48
2.7	Structural Estimates - Utility Equation . . . . .	49
2.8	Model Fit and Simulations . . . . .	50
3.1	Risk sets and activities . . . . .	97
3.2	Graduation probabilities by age . . . . .	100
3.3	Average growth in full-time wages due to various experiences by final educational attainment . . . . .	101
3.4	Average growth in full-time wages due to various experiences at age 29 by final educational attainment . . . . .	102
3.5	College and HS wage premium and dispersion at age 29 for full-time workers . . . . .	103
3.6	Median AFQT score and dispersion by final educational attainment . . . . .	104
3.7	Family background characteristics by final educational attainment . . . . .	105
3.8	Local labor market conditions at various ages . . . . .	106
3.9	Measures of returns to schooling across specifications . . . . .	107
3.10	Select wage equation marginal effects (age 29) . . . . .	108

3.11	Full-time wage factor loading estimates . . . . .	109
3.12	Full-time log wage decompositions at age 29 for an additional unit of schooling experience . . . . .	110
3.13	Full-time log wage decompositions at age 29 for an additional unit of work experience . . . . .	111
3.14	Full-time log wage decompositions at age 29 for educational attainment	112
A.1	Sample Selection . . . . .	114
B.1	Choice Sample Selection . . . . .	121
B.2	Wage Sample Selection . . . . .	121

# List of Figures

2.1	Market Share of Master’s Degrees Awarded in Education by Year and College Type . . . . .	37
2.2	Annual Percent Change of Salary by Degree Decision . . . . .	38
2.3	Annual Average Salary if Teacher with BA vs Average College Graduate in County . . . . .	38
2.4	School Quality by Degree Level . . . . .	39
2.5	School Quality by Master’s Degree Type . . . . .	39
2.6	Stock of Teachers and Teachers with Master’s Degrees by Year in North Carolina . . . . .	42
2.7	Access and Tuition Metrics by College Type . . . . .	43
2.8	Odds Ratios of Praxis I Deciles on Degree Decision, Baseline is BA . . . . .	45
2.9	Decision Tree . . . . .	46
2.10	Payoff Tree . . . . .	47
3.1	Distributions of average experience by age . . . . .	98
3.2	Distributions of average end-of-panel experience by final educational attainment . . . . .	99

# Acknowledgements

First and foremost I would like to thank my dissertation advisor, Joe Hotz. His patience was tried and tested, but it held.

I also received outstanding support and advice from the my other committee members, Peter Arcidiacono, Hugh Macartney, and Arnaud Maurel.

There is a vibrant community of applied microeconomists at Duke across a wide array of fields. I thank them collectively, especially those attending the Labor Lunch, as well as Tyler Ransom, who in addition to being one of my coauthors, is also a walking Matlab repository.

I am grateful for the Southern Economic Association for allowing me to present at their meetings and for partially funding my travels.

I thank the local congregation of my church, the Church of Jesus Christ of Latter-Day Saints. This was a great place to find friends and support, especially other Duke graduate parents striving to balance school, family and church.

I am grateful to my Lord for the strength He sent me.

From a young age my father, the Professor, instilled in me the love of learning and the joy of work. And as I have grown, he has been a constant source of encouragement, comfort, and humor. Le agradezco mucho.

My dear mother, the Teacher, passed away after my first year of graduate school. She was a constant guide and example to never stop learning, epitomizing this with receiving a Master's degree in her 50's. I thank, miss and love her, but I also know that she is looking down at me, smiling, dancing, and tickling those celestial ivories.

I had the joy and privilege to have two little children while at graduate school, Alia and Daxton. Their smiles and cheers when I returned home each day filled my heart and restored my energy.

Finally, I am grateful to my wife Kristie. The support, love and butt-kicking that I received from her were instrumental in being able to complete this long and lovely journey. My life has been immeasurably blessed since the fateful day when I called up her apartment and her roommates were gone. I am grateful for her every day of my life.

# 1

## Introduction

Over the last few decades, there have been significant changes to the U.S. education and labor markets. There have been increases in education across the board, from high-school to college to graduate school. Alongside this, the wage returns to education have generally increased, as have various non-pecuniary benefits associated with the labor market. In all of these education and employment decisions, workers are not just trying to get the most out of their current period, but also maximize their future (expected) payoffs.

This dissertation examines the dynamic nature of many of the education and career decisions individuals face. I first explore the decision to get a master's degree by analyzing how public school teachers choose to attain more education and how these decisions affect their labor market outcomes. Examining teachers provides many unique opportunities, including a homogenous work force, a lack of competition on wages, and a known return to the wages of receiving a master's degree.

I then look at how the drastic changes in the returns to schooling and experience interact with the changes in individuals schooling and experience decisions. The questions I ask are: What is the relative importance of changes in skill price versus

skill composition over the past 20 years? What are the trends in the wage returns to in-school work experience? How much of the evolution in the college wage premium actually reflects an increase of in-school, and, more generally, early work experience?

In Chapter 2, I look at teachers using data from North Carolina. Teachers in North Carolina are compensated via a state-wide salary schedule. This schedule is a fixed function of teaching experience and highest degree completed. As such, there is little to none wage competition. This leads to more of an emphasis on compensating differentials that jobs may offer, such as measures of school quality. Human capital investments are well-known ways to improve both salary and job quality, and should have the same important role here. I am able to observe the costs and timing of these investments, and as such am able to characterize the entire decision and outcome processes that teachers face.

I first show that human capital investments by teachers are associated with job quality and that teachers place a high value on different aspects of job quality that they face. Teachers who receive master's degrees from traditional institutions transition to higher-quality schools at higher rates than teachers without.<sup>1</sup> My results indicate that teachers have a willingness to pay between \$1,700 and \$20,000 to be able to move up a quartile in school quality, which is another motivation for receiving a master's degree. It is important to note, however, that teachers who receive master's degrees from for-profit colleges generally transition to higher-quality schools at lower rates, implying that where you get your degree from may be a good indicator of one's motivations.

I also show that teachers' choices are driven in large part by the ease of access that many colleges provide. I find large, significant, and positive effects of having many distance-learning options available. This is in contrast to much smaller effects

---

<sup>1</sup> Quality measures are defined relative to average test scores and average socio-economic status of students.



for having the same number of local brick-and-mortar universities accessible, though they are still significant and positive. Additionally, the positive effect of having online education dominates the negative effect of tuition.

I conceptualize my model using a dynamic framework because teachers make these decisions with their entire career in mind, not just the current period. Future choice sets are allowed to vary by the current period decision, allowing teachers to choose among multiple career and education paths. In addition, state variables evolve over time based on previous choices and variables. This allows me to explicitly include in the model the probability of moving to a better school conditional on current period choices. Even more, this model allows me to consider counterfactual scenarios, such as how would teachers make schooling decisions if the premium to a master's degree was removed.

I estimate this model using a rich and detailed dataset from North Carolina on teachers, schools and school districts, dating back to 1995. This dataset contains precise information on every education and career decision that teachers make, which allows me to explicitly model the teachers' forward-looking behavior. Additionally, a unique feature about North Carolina is that the wage schedule that teachers face is very centralized, meaning all teachers face the same state salary schedule. As mentioned, this greatly reduces the heterogeneity in this labor market. This further simplifies the analysis and allows me to focus on teachers' other career outcomes.

Finally, I am able to use my model and results to run counterfactual simulations for various policy effects. I find that teachers with high-unobserved ability are more likely to get master's degrees, but are also more likely to leave teaching unless those degrees are from a private, non-profit college. I also find that equating the costs to attending college across all universities and options leads to an increase in the incidence of black teachers obtaining master's degrees from traditional universities, which leads to more black teachers in the higher quality schools.

Chapter 3 looks at the changes over time in the composition and return to schooling and experience. Joint with V. Joseph Hotz, Arnaud Maurel and Tyler Ransom, my coauthors and I find that the relative importance of the price and composition effects varies dramatically across skills. Regarding in-school work experience, we find that the direct returns to working while in college have decreased over time, with the earlier decrease attributable mostly to price effects, and the more recent decrease attributable mostly to composition effects. Further, composition effects explain little in the evolution of the college wage premium. Related to this, we find that there is both a significant increase in the incidence of in-college work over time and a decrease in the wage return to in-college work. These combine to produce a negative impact on the composition effect, which is offset by a positive net impact of the remaining skill correlates. Finally, and consistent with other studies (e.g. Taber, 2001), we find that almost all the increase in the college wage premium in the 1980s is due to a change in the returns to and composition of unobserved skills. These unobserved effects have diminished greatly in recent years and have contributed to the declining growth in the college wage premium.

Our analysis makes use of two longitudinal data sets, the 1979 and 1997 panels of the National Longitudinal Surveys of Youth (NLSY). We divide our analysis among three cohorts of individuals: (i) NLSY79 respondents born in years 1959 and 1960; (ii) NLSY79 respondents born in years 1961 through 1964; and (iii) NLSY97 respondents, all of whom were born in years 1980 through 1984. As will be shown, these three cohorts differ markedly in their human capital investment decisions and the market conditions they faced while making such decisions.

While ours is not the first study to examine labor market trends over this time period, our use of longitudinal rather than repeated cross-sectional data allows us to more accurately measure early-career work experience and account for its endogeneity. From each of the NLSY surveys, we construct comparable measures of schooling,

work, and military histories from ages 16-29, along with comparable measures of earnings, educational attainment, ability, local labor market conditions, and personal and family background characteristics. From these histories, we are able to construct refined measures of human capital including whether or not work experience occurred simultaneously with schooling. Following many studies in the literature, we restrict our analysis to males.

In order to obtain wage estimates that reflect selection-free, causal effects of human capital accumulation, we specify and estimate a dynamic model of schooling and work decisions that controls for person-specific unobserved heterogeneity. We linearly approximate the value functions (see Eckstein and Wolpin, 1989), but allow the idiosyncratic shocks to be correlated across choice alternatives. This correlation is induced by our factor-analytic approach inspired by Cameron and Heckman (1998, 2001) and Heckman et al. (2006b). We use comparable cognitive test scores and the panel structure of the data to identify the heterogeneity factors. We then use the model estimates to conduct counterfactual simulations (decompositions) which allows us to assess the role of price and composition effects in unobserved skills as well as observed skills.

## Teacher Education Decisions and Career Outcomes

### 2.1 Introduction

Over the last two decades, there has been a significant increase in the incidence of teachers obtaining master's degrees in education. New master's degrees in education have grown at a rate almost 5 times that of the rate of total teachers.<sup>1</sup> The percentage of teachers with a master's degree increased from around 45% to almost 60%.<sup>2</sup> There are many factors that would motivate a teacher to get a master's degree. The extra salary that teachers with master's degree earn is potentially a main factor, even though the premium has been fairly constant nationwide over this time period at around 15-17%.<sup>3</sup> Additionally, over the last decade the availability of distance education and online courses has increased dramatically. There are now almost 40% more colleges awarding master's degrees in education than there were in 1995. And the types of institutions awarding degrees have evolved as well, especially with the

---

<sup>1</sup> Integrated Postsecondary Education Data System (IPEDS) and U.S. Department of Education (2013)

<sup>2</sup> U.S. Department of Education (1996, 2003, 2009, 2013, 2014)

<sup>3</sup> U.S. Department of Education (2014), Table 211.20.

emergence of for-profit universities.<sup>4</sup>

This paper explores the decision to get a master's degree by analyzing how public school teachers choose to attain more education and how these decisions affect their labor market outcomes. Most labor markets are very heterogeneous, and a worker may make human capital investments for a myriad of reasons. These include a better draw from the wage distribution, improved job characteristics, better access to health services, access to a different marriage market, improved job security, and the list goes on. Understanding how a human capital investment affects these benefits is quite difficult. However, by focusing on a very unique situation, I able to remove a lot of the noise from the analysis.

Teachers in North Carolina are compensated via a state-wide salary schedule. This schedule is a fixed function of teaching experience and highest degree completed. As such, there is little to none wage competition. This leads to more of an emphasis on compensating differentials that jobs may offer, such as measures of school quality. Human capital investments are well-known ways to improve both salary and job quality, and should have the same important role here. I am able to observe the costs and timing of these investments, and as such am able to characterize the entire decision and outcome processes that teachers face.

I first show that human capital investments by teachers are associated with job quality and that teachers place a high value on different aspects of job quality that they face. Teachers who receive master's degrees from traditional institutions transition to higher-quality schools at higher rates than teachers without.<sup>5</sup> My results indicate that teachers have a willingness to pay between \$1,700 and \$20,000 to be able to move up a quartile in school quality, which is another motivation for receiving

---

<sup>4</sup> See Figure 2.1.

<sup>5</sup> Quality measures are defined relative to average test scores and average socio-economic status of students.

a master's degree. It is important to note, however, that teachers who receive master's degrees from for-profit colleges generally transition to higher-quality schools at lower rates, implying that where you get your degree from may be a good indicator of one's motivations.

I also show that teachers' choices are driven in large part by the ease of access that many colleges provide. I find large, significant, and positive effects of having many distance-learning options available. This is in contrast to much smaller effects for having the same number of local brick-and-mortar universities accessible, though they are still significant and positive. Additionally, the positive effect of having online education dominates the negative effect of tuition.

I conceptualize my model using a dynamic framework because teachers make these decisions with their entire career in mind, not just the current period. Future choice sets are allowed to vary by the current period decision, allowing teachers to choose among multiple career and education paths. In addition, state variables evolve over time based on previous choices and variables. This allows me to explicitly include in the model the probability of moving to a better school conditional on current period choices. Even more, this model allows me to consider counterfactual scenarios, such as how would teachers make schooling decisions if the premium to a master's degree was removed.

I estimate this model using a rich and detailed dataset from North Carolina on teachers, schools and school districts, dating back to 1995. This dataset contains precise information on every education and career decision that teachers make, which allows me to explicitly model the teachers' forward-looking behavior. Additionally, a unique feature about North Carolina is that the wage schedule that teachers face is very centralized, meaning all teachers face the same state salary schedule. As mentioned, this greatly reduces the heterogeneity in this labor market. This further simplifies the analysis and allows me to focus on teachers' other career outcomes.

Finally, I am able to use my model and results to run counterfactual simulations for various policy effects. I find that teachers with high-unobserved ability are more likely to get master's degrees, but are also more likely to leave teaching unless those degrees are from a private, non-profit college. I also find that equating the costs to attending college across all universities and options leads to an increase in the incidence of black teachers obtaining master's degrees from traditional universities, which leads to more black teachers in the higher quality schools.

My paper contributes to a rich body of research on occupational and educational choice. In addition to the many, seminal works on occupational choice (see Miller (1984), Siow (1984), Heckman and Sedlacek (1985), Dolton et al. (1989), and Sullivan (2010)), there is an array of literature focusing on teacher's career decisions (see Manski (1987), Stinebrickner (2001a), Stinebrickner (2001b) and Wiswall (2007)). The main contribution of this paper to this vein of literature is to analyze an education decision of a working professional, rather than the choice of a job. Rather than have the individual choose a bundle of job characteristics, I allow these characteristics to arrive probabilistically, based on the teacher's education choice. This is more closely related to educational choice research (see Card (1999) and Arcidiacono et al. (2008)). But as mentioned, because of the unique situation where there is little to no wage competition, I am able to focus instead on the other potential returns to a degree.

This research is also among the first to disentangle outcomes for individuals by type of master's degree, namely public, for-profit and non-profit. Previous work mainly focused on the the baccalaureate level and below (see Deming et al. (2012) and Cellini and Chaudhary (2011)), finding in general that students who attended for-profit performed the same or slightly worse than other students. Finally, there is a large amount of literature looking at classroom returns to a master's degree, which finds that graduate degrees do not play a large role in student outcomes (see Harris

and Sass (2009), Goldhaber and Brewer (2000), Clotfelter et al. (2007), Clotfelter et al. (2010)). This is another reason to focus on the labor market outcomes of the teacher instead.

The next section describes the data that will be used, and presents descriptive statistics of teacher characteristics and choices. Section 2.3 introduces a theoretical model of how teachers make their education decisions. Section 2.4 presents the strategy for estimation and identification and Section 2.5 reports the results. I present counterfactual simulations in Section 2.6 and Section 2.7 and concludes.

## 2.2 Data and Descriptives

This section describes the data to be used and presents descriptive results. The data sets are exceptional in that they are able to provide precise detail about the timing and location of the teachers' career and educational decisions. Additionally, the descriptive evidence points to rich variation across teachers in their characteristics and behavior. This unique data allows for a very flexible model.

### *2.2.1 Data and Background*

Information on the teachers in this analysis comes primarily from data from The North Carolina Education Research Data Center (NCERDC). This data set contains a wealth of information on teachers' choices and experiences. It is a longitudinal database on the entire North Carolina public education system, dating back to the mid-1990s. The data set contains each college degree each teacher receives, including the date received and information about the granting institution. The precise detail on each college enables me to create a rich educational history for every teacher.

In addition to college background, the NCERDC reports other teacher information such as licensure test scores, compensation, experience, school, teaching subject, previous employment, sex and race/ethnicity. There is also rich data available at the



student, school and school district level. For this analysis I use information on free lunch rates and test scores at the school level.<sup>6</sup>

One important feature about the North Carolina data is the changing landscape of education policies related to compensation and certification. The main feature regarding compensation occurred in 2000, when the Excellent Schools Act of 1997 went into effect, increasing the salary premium to a master's degree from 6% to 10%.<sup>7</sup> This salary shift provides rich variation and will help in identifying the role of salary in the decision process.

In addition, a relevant master's degree can provide additional state-level certifications. However, it cannot be used for the primary license area, rather only an additional one. This is similar for No Child Left Behind (NCLB). While the NCLB requirement for teachers to become "highly qualified" could be accomplished with a relevant master's degree, this was only viable for a secondary license area because of the state requirements.<sup>8</sup> Because the sample of potentially effected teachers is so small, I do not include controls for the effects of a degree on licensing.

Information about the relevant characteristics of colleges and universities comes primarily from the Integrated Postsecondary Education Data System (IPEDS). This contains information from 1984-2011 on institution characteristics such as college control type, degree levels offered, degrees conferred, students enrolled, tuition, and other items. As a stipulation to receive funds, all Title-IV eligible institutions of higher education are required to report this information annually to the Department

---

<sup>6</sup> Other information available includes, but is not limited to, financial data, expulsion rates, criminal activity, and student demographics.

<sup>7</sup> North Carolina, Senate Bill 272, 1997.

<sup>8</sup> See <http://sbepolicy.dpi.state.nc.us/policies/TCP-A-001.asp?pri=02&cat=A&pol=001&acr=TCP>, accessed May 17, 2013; <http://www.dpi.state.nc.us/nclb/highly/faqs/>, accessed April 23, 2012, no longer available since as of 2013, NCLB is irrelevant in North Carolina, though archived page is available.

of Education.<sup>9</sup> The match rate between the North Carolina education data and IPEDS is very high, on the order of 95%.

In order to get more precise detail on tuition, I supplement the IPEDS tuition data with information from various universities websites. Most colleges report the graduate level tuition as the tuition charged to traditional graduate students, excluding professional and MBA students. However, some report an average tuition charged to all students working on an advanced degree. Because of this, I adjust those tuitions to be more representative of the tuition charged for a master's in education. Further information is available in the appendix.

Out of the 157,533 teachers who started teaching after 1995, 16,869 received their master's degree before they started teaching. As such, I drop them since I am not able to model their education decisions. I also drop individuals with missing and/or invalid observations, mainly due to right-censoring, missing demographic information, and invalid employment history. My final sample contains 101,066 teachers and 502,216 observations. More information on the sampling process is available in the appendix.

### *2.2.2 Job Characteristics*

As mentioned, wages in this situation are very unique. The wage that teachers receive from the state government is the same for all teachers of a given level of experience and degree. Table 2.1 lists some monthly wages in the 2005-2006 school year. This shows that the salary for a teacher with an MA is 10% higher, as mentioned previously. However, the increase in wages due to an additional year of experience are not fixed. For example, note that the difference between 0 and 1 years of experience is only \$420, while the difference between 2 and 3 years of experience is \$1,560. Since

---

<sup>9</sup> United States, Higher Education Act of 1965 as amended, Title IV, 20 USC 1094, Section 487(a)(17) and 34 CFR 668.14(b)(19).

teachers can't transition within a given year's salary schedule, Figure 2.2 shows the annual percent change of wages from year to year for individuals who begin the year with only a bachelor's degree. It is broken out by whether bachelor's-only teachers choose to continue with a bachelor's or receive a master's. This shows both the 6% and 10% wage gap across degrees, but also a large amount of variation across and within year for a given degree decision, leading to a high amount of variation by experience.<sup>10</sup>

Variation in the wages also comes from the wage of the outside option, which is based on the wage in the county that the teacher is in. Figure 2.3 shows the average county wage for college graduates and the average wage for bachelor's-level teachers with 10 years of experience over the sample. The first thing to notice is that college graduates as a whole are paid much more than teachers. This is true both at the mean as well as the entire interquartile range. But in addition to the outside wage being higher, it is also a lot noisier. This is due to the variation across counties in the average wage for college graduates.

There is also rich variation in job characteristics, beginning with average socioeconomic status (SES) of students at the school. Figure 2.5a shows the distribution of the percent of the students at each teacher's school who qualify for free or reduced price lunch. The first item to note is that there is large variation in free-lunch percentages for all teachers. Additionally, teachers whose highest degree is a bachelor's are more often in low SES schools (high free lunch schools) than teachers with a master's degree. The respective average percentage of students with free lunch is 35.5% for bachelor's and 33.7% for master's. This difference and the difference in distribution are both statistically significant.

I also see variation in the other metric of job quality I examine, average school test

---

<sup>10</sup> Teachers have a very good idea of their wage the next period given that these wage schedules are generally laid out months or years in advance.

score. This is the average of the Z-score of the combined reading and math scores for all tested students in a teacher's school.<sup>11</sup> The distribution of these is in Figure 2.5b and, similar to SES, teachers with master's degrees generally find themselves with students of higher academic ability. The school mean of student's Z-scored test scores are -0.01 and 0.02 standard deviations for bachelor's and master's respectively. Note that these are standard deviations of the underlying students distribution and most of the school districts are within 0.5 standard deviations of the mean. Thus, having the average student score 0.03 standard deviations higher is fairly important. The difference of these means and the difference in distribution are also both statistically significant.

Finally, Figure 2.5 does the same comparison by type of master's degree. Doing so shows a similar story comparing for-profit master's to traditional master's, with teachers with for-profit master's being at schools with higher free lunch incidence and lower test scores. However, the differences here are much larger than the differences between bachelor's and master's. This is even more evidence of a large endogeneity effect in the degree decision.

In addition to job quality metrics generally being higher for teachers with a master's degree, having a master's is also positively correlated with the rate at which teachers are able to move to a higher quality job. Table 2.2 shows the rates at which teachers transition to different quartiles of job quality between the current period and the next period. This table focuses on those who begin the current period with just a bachelor's degree, which represents the first choice set of the model. The left side of the table looks at teachers who continue with only a bachelor's degree. The right side of the panel looks at teacher who choose to obtain a master's degree. Teachers with master's degrees are able to move upwards at a much higher rate than

<sup>11</sup> For grades 3-8, these are end-of-year tests in mathematics and reading. For 9-12, these are end-of-year tests in Algebra I and English I. For schools without and test scores (i.e., K-2 schools, the district average is used.

those without. This is especially true for moving into the top quartile of schools.

However, Table 2.3 shows that not all master's degrees provide the same benefits. Specifically, teachers obtaining master's degrees from for-profit institutions have transition rates at about the same level as those without any master's degree. Thus, while there is an observed pecuniary return associated with a for-profit degree (as with any master's degree), there does not appear to be a comparable non-pecuniary return.

There are also differentials in the transition rates into the outside option. Table 2.4 lists the odds ratios for the different type of degrees on leaving the North Carolina public teaching system. Teachers who have obtained master's degrees from public institutions exit at similar rates to those without any master's degree. However, those who choose to obtain a private master's degree are less likely to leave. Compared to choosing a public master's degree, those choosing a non-profit master's degree are about 80% as likely to leave, and those who choose a for-profit master's degree are half as likely to leave. This is consistent with a world in which a public master's degree provides more value in the non-teaching professions than others. This is also consistent with a story of selection, in that teachers who choose for-profit master's degrees are much less likely to try to find a job outside of teaching, and thus want to create a permanent increase in their salary.

### *2.2.3 College Characteristics*

As mentioned, over the past twenty years there have been drastic changes in the national market for master's degrees in education. This is also true in North Carolina. Figure 2.6 plots the year-by-year stock of total teachers and teachers with master's degrees in North Carolina. At the beginning of the studied period, slightly more than 1 in 5 North Carolina teachers held a master's degree. This jumped to almost 1 in 4 in 2007 and 1 in 3 by 2011.

In addition to an overall increase in degree attainment, there has been a shift in the type of colleges providing master's degrees. As Figure 2.1 has already shown, there has been a large increase in the incidence of individuals receiving master's degrees from for-profit universities. At the turn of the century, they had virtually no presence in this market. However, after just one decade, 10% of degrees were awarded by these institutions, with most of this increase in share coming from public universities.

There has also been a change in the cost of a master's degree, both in terms of access costs (i.e. how hard it is to receive instruction) and tuition costs. The first component of access cost relates to the physical proximity of each college type. Figure 2.8a plots the average number of institutions of a given college type within a 25-mile radius of each teacher. The average teacher had access to at least one public and one non-profit university in most every year, with access increasing over the latter part of the time period. Only in 2003 do we see for-profits begin to offer master's degrees in education at local campuses, and this is primarily driven by one institution.<sup>12</sup>

The other component of access cost is tied into distance education. Figure 2.8b reports the number of distance-learning options of each college type. Note that while there is variation over teachers in the physical index within a given year, there is no variation in this distance learning index, since all teachers face the same choice set. Public universities have offered distance learning for a long time in the form of first correspondence courses, followed by online programs. In 2003-2005, many for-profit universities began awarding degrees online in North Carolina, and now have almost as many options as public institutions.

Finally, the most straightforward measure of cost is the tuition that colleges

---

<sup>12</sup> The institution is Strayer University. While University of Phoenix has a large, physical presence in North Carolina, none of their local campuses offer education classes or degrees.

charge, and there is a strong disparity in the tuition charged at these institutions. To measure this, I create a tuition index for each teacher which is the average tuition of the colleges that have are either in physical proximity or have distance learning options.<sup>13</sup> In Figure 2.8c this average tuition index varies across all college types. However, in general, the tuition at public universities is the lowest, followed by non-profit universities, with for-profit universities charging the most. Note that a lot of non-profit institutions offer education programs at lower rates than their other programs. The high value of both the 2003 tuition and the slope for the for-profits is part of the puzzle surrounding why they are so popular of late.

#### *2.2.4 Teacher Characteristics*

Teachers vary greatly by race and gender, both overall and by degree level. Table 2.5 shows demographic statistics by degree level, experience and ability. The majority of individuals who ever teach are white and female, and they also have the longest average tenure as well. This indicates some measure of positive selection into the profession for these teachers, knowing that they are more likely to last longer. The distribution of master's degrees also varies across these demographics, with female and black teachers accounting for larger shares of master's degrees than other teachers.

There are also differences in ability across teacher's education decisions, as measured by the types of tests that teachers themselves take. The first of these are the Praxis I exams, which are basic tests in math, reading and writing. All teachers must pass these to obtain their initial license. The second of these are the Praxis II exams, which field specific tests. Notwithstanding a few exceptions, all teachers must pass these to obtain their full license by the end of their third year. In order to get a consistent measure of teacher test score for both Praxis I and Praxis II, I

---

<sup>13</sup> See Section 2.3.2 for more details.

follow Clotfelter et al. (2010) and normalize scores so that all tests of each type in each year have a mean of zero and a standard deviation of one. Doing this I create composite Praxis I and Praxis II test scores, which are simple averages of all Praxis I and Praxis II exams a teacher takes, respectively.<sup>14</sup> For my analysis I focus on Praxis I exams.

There is a large racial distribution in Praxis I scores. On average, white teachers in the sample score 0.13 standard deviations above the population mean, whereas black teachers in the sample score -0.24 standard deviations below the population mean. Further, men score higher than women, with sampled men and women averaging 0.17 and 0.05 standard deviations above the population mean, respectively.<sup>15</sup> This indicates a potential negative selection issue, since both blacks and women were the most likely to receive a master's degree, but had the lowest licensure test scores.

In order to understand more the relationship between Praxis I scores and the decision process, Figure 2.8 shows the relationship between Praxis I decile and deciding to obtain a master's degree and what type. Panel (a) only looks at Praxis I decile and the decisions to obtain a master's degree without any other controls. In this panel there is virtually no relationship between this observed measure of ability and the degree decision. However, controlling for demographic characteristics in panel (b) reveals a significant positive relationship, indicating that on average those receiving their master's do have higher ability or academic preparation. Further, panel (c) breaks out the decision to get a master's into the three different types of degree and we see an even more interesting scenario, with those in the upper deciles of the Praxis I distribution being very likely to receive a public master's and very unlikely to obtain a for-profit master's. Thus a large amount of the sorting is indeed done on

---

<sup>14</sup> See the data appendix for more information on the Praxis exams.

<sup>15</sup> The normalization of test scores happened on the entire population, before dropping invalid teachers.



ability. In the next section I will describe the model that I create in order to relate the patterns I see in the raw data with the endogenous decision making process that teachers face.

## 2.3 Model

In this section, I introduce a dynamic model where teachers make education and career decisions to maximize their utility in the presence of an unobserved type. Teachers make choices in every period, taking into account the payoffs in the current period and all future periods. The inclusion of an unobserved type controls for factors unseen by the econometrician that may be influencing the decision of how and where to get a degree. I augment each observation with an observed measure of teacher ability. This framework sets up my identification strategy to estimate the decision process these teachers face.

### 2.3.1 Choices

Choices are made by each teacher  $i \in \mathcal{N}$ , where  $\mathcal{N}$  represents all teachers who did not receive a master's degree before teaching. In each period  $t$ , teacher  $i$  chooses activity  $j$  to maximize her current and future utility. Teachers initially choose activity  $j \in \mathcal{J}^0 = \{TEACH, PUB, FOR, NON, O\}$ , where *TEACH* is only teaching, *PUB*, *FOR*, and *NON* are teaching while going back to school to get a master's degree from a public, for-profit, and non-profit institution, respectively, and *O* represents taking the outside option, which is leaving the NC education system.

Once a teacher receives a master's degree, her choice set is simplified to either continue teaching or to leave and take the outside option. However, there is a different choice set for each type of master's degree. Thus teacher  $i$  chooses  $j \in \mathcal{J}^{r_{it}} = \{TEACH, O\}$  for  $r_{it} \in \{1, 2, 3\}$ , where choice sets 1, 2, and 3 are for teachers with master's degrees from public, for-profit, and non-profit colleges, respectively. For

notational convenience, I drop the subscripts on  $r_{it}$ , and simply refer to a choice set as  $\mathcal{J}^r$ . Breaking these choice sets out by degree type is crucial as it allows the effects of the chosen degree to last beyond the period in which the degree was received. Using this notation, the initial choice set can be referred to as choice set 0. In summary, there are a total of 11 alternatives across four choice sets, with the master's-degree choice sets being mutually exclusive.

The timing of the choice goes as follows: In the beginning of period  $t$ , teacher  $i$  decides whether or not to get a master's degree. At the end of period  $t$ , teacher  $i$  receives said degree, if a degree was pursued. If no schooling was chosen, then the teacher's highest degree is still a bachelor's degree, and she still faces choice set 0 in period  $t + 1$ . If schooling was chosen, her highest degree changes to a master's degree, and the teacher decides whether to keep teaching or not within her college-type choice set. As long as the outside option is not chosen, the teacher continues making dynamic decisions. This can be seen in Figure 2.9.

### *2.3.2 Payoffs*

Teachers receive payoffs every period for choices that they make. These payoffs are dependent on various characteristics, depending on the choice. While all choices are influenced by teacher characteristics, teaching choices are influenced by job characteristics of the school whereas choosing the outside option is influenced by the outside wage. Finally, college costs only enter the payoffs for master's degree choices. In this section I breakout more fully what each of these influencing factors is and how they are unique to this environment.

#### *Teacher Characteristics and Ability*

As mentioned, payoffs for all teachers depend on the teachers' characteristics. The demographic variables I use are teacher's race and gender, where race is either white,

black or other. I include calendar year dummies to control for non-stationary trends in the data. This is especially important given how the education landscape has evolved over time. I also include a quadratic in experience. As I have alluded to, the financial returns to experience are included in the salary term (see below), thus this term is capturing any other role of experience may have on individuals tastes for a given choice. As this is measure of teaching experience, it does not enter the utility for choosing the outside option. These teacher variables are depicted in the model as  $x_{it}^d$ .

The model also accounts for measures of ability, both observed and unobserved. I establish the relationship between observed ability ( $y_i$ ) and unobserved ability ( $s_i$ ) as

$$y_i = \phi race_i + \kappa^y s_i + \nu_i, \quad (2.1)$$

where  $\nu_i$  is a well-behaved error term. The ability metric that I use here is the teacher's Praxis I test score. As mentioned, this is the test they take to get in to teaching in the first place, and as such acts as a good proxy of pre-teaching ability. I include  $race_i$  to account for any potential racial bias in the test. One issue that I mention in the data appendix is that this score is scarcely reported by NCERDC after 2006. However, under certain assumptions, it can still be used to help identify the unobserved type. This is discussed further in Section 2.4.

### *Job Characteristics*

Many models allow the choice set to include a bundle of job characteristics, as well as the education decision (see Roy (1951) and Heckman and Sedlacek (1985)). This would allow the teacher to make simultaneous choices. Similar to other works (see Arcidiacono et al. (2012)), my model does not let the agent explicitly choose certain characteristics, but rather lets these evolve dynamically based on the individuals choices. Thus teachers choosing certain degrees or degree types will have different

probabilities of receiving different job characteristics. Non-pecuniary job benefits for teaching choices take the form of quartiles of average SES of students in the school and average district test score, represented as  $x_{it}^s$ . As mentioned, these job quality characteristics evolve dynamically in the model. Let  $x_{it}$  represent all variables in the model, including  $x_{it}^d$  and  $x_{it}^s$ . If choice  $j$  is selected in period  $t$ , then the probability of  $x_{i,t+1}$  occurring is determined by the transition probability

$$f_j(x_{i,t+1} | x_{it}). \quad (2.2)$$

Depending on whether a teacher chooses a master's degree or not and which type of master's degree he receives, he will be more or less likely to see certain realizations of  $x_{i,t+1}$ . This also extends into the other choice sets ( $r > 0$ ). Allowing a different choice set for each master's degree type causes these probabilities to be different not just the period immediately following the degree, but in all subsequent periods. Thus the probability of getting into a higher quality school or school district changes with degree type, both upon receipt as well as years down the road. Note that in addition to these non-pecuniary benefits, teacher experience also evolves dynamically via this transition probability. All other variables are fixed (i.e., teacher race/gender) or deterministic (i.e., calendar year).

Like wages, non-pecuniary job benefits for the outside option are also unobserved. However, since the individual chooses to leave teaching, no school characteristics enter their utility. Note again that the outside option taken can be many things, including home production, working in North Carolina outside of the public education system, leaving the state, or unemployment.

Unlike non-pecuniary benefits, pecuniary benefits are deterministic. As mentioned earlier, how much a teacher is paid in North Carolina is calculated from a fixed wage schedule. This schedule is a function of the teacher's experience and highest degree, though it varies year to year. Thus the wage function for teachers

can be expressed as  $wage_{ijrt} = wage(year_{it}, exp_{it}, r, j)$ , where  $year_{it}$  and  $exp_{it}$  are calendar year and years of experience for teacher  $i$  in period  $t$ , and again  $r$  is the teacher's relevant choice set, which determines the level of pay. Teaching choices in the first (BA) choice set receive  $wage_{ijrt} = wage_{it}^{BA}$ , whereas teaching choices in the later (MA) choice sets receive  $wage_{ijrt} = wage_{it}^{MA}$ . Again, note that this is not a function of any other teacher characteristic, observed or unobserved. Thus we may expect teachers with high ability to have some form of compensating differential to make up for this.

The financial payoffs for the choice to leave the profession is the outside wage,  $wage_{ijrt} = wage_{it}^O$ , which is unobserved. I set it to be the average county wage of college-educated people in teacher  $i$ 's county in year  $year_{it}$ , and thus is not a function of experience or choice set. For all individuals, wage can be expressed as

$$wage_{ijrt} = \begin{cases} wage_{it}^{BA} & \text{if } r = 0 \text{ \& } j \notin \{O\} \\ wage_{it}^{MA} & \text{if } r \neq 0 \text{ \& } j \notin \{O\} . \\ wage_{it}^O & \text{if } j \in \{O\} \end{cases} \quad (2.3)$$

### *College Costs*

The final component of the payoff structure enters the master's degree choices. This cost for teacher  $i$  to choose master's degree type  $j$  in period  $t$ ,  $c_{ijt}$ , is broken out into tuition  $\overline{tu}_{ijt}$  and access costs  $m_{ijt}$ :  $c_{ijt} = c(\overline{tu}_{ijt}, m_{ijt})$ .

As mentioned in Section 2.2.3, the access term  $m_{ijt}$  contains indices for both physical proximity and distance learning. Specifically, the physical proximity index is the number of institutions of each college type within a 25-mile radius. The distance learning component is an index for the number of institutions of each college type that are offering correspondence courses or online degrees in education in North Carolina. Both of these indices are included in the model as bins ranging from zero to four or more.

Tuition is the average tuition of all colleges of type  $j$  that teacher  $i$  can attend in period  $t$ , which set I denote by  $M_{ijt}$ . This includes all institutions that are in the ease-of-access index, so those within a 25-mile radius as well as all programs with a significant online presence. I allow teachers to take advantage of the Lifelong Earning Tax Credit, which gives individuals returning to college a tax credit of 20% of eligible education expenses, up to \$2000.<sup>16</sup> The average tuition is thus calculated as

$$\overline{tui}_{ijt} = \frac{1}{M_{ijt}} \sum_{m=1}^{M_{ijt}} \left\{ (1 - \tau) tui_{mt} \right\} \quad (2.4)$$

where the Lifetime Learning Tax Credit  $\tau = 0.2$ .

#### *Utility Functions*

I combine all of these elements into the teacher's utility function, which is assumed to have the following functional form:

$$U_{ijt} = \begin{cases} \alpha_j x_{it}^d + \kappa_j s_i + \lambda x_{it}^s + \psi wage_{it}^{BA} + \varepsilon_{ijt} & \text{if } r = 0 \text{ \& } j \in \{TEACH\} \\ \alpha_j x_{it}^d + \kappa_j s_i + \lambda x_{it}^s + \psi wage_{it}^{BA} + \delta c_{ijt} + \varepsilon_{ijt} & \text{if } r = 0 \text{ \&} \\ & j \in \{PUB, FOR, NON\} \\ \alpha_j x_{it}^d + \kappa_j s_i + \lambda x_{it}^s + \psi wage_{it}^{MA} + \varepsilon_{ijt} & \text{if } r \neq 0 \text{ \& } j \in \{TEACH\} \\ \alpha_j x_{it}^d + \kappa_j s_i + \psi wage_{it}^O + \varepsilon_{ijt} & \text{if } j \in \{O\}, \end{cases} \quad (2.5)$$

where  $\varepsilon_{ijt}$  is a well-behaved error term, and all other terms were defined previously. The first case refers to the choice in the initial risk set ( $r = 0$ ) that corresponds to teaching but not returning to get a degree. It depends on the job characteristics of the school and the bachelor's level wage,  $wage_{it}^{BA}$ . The second case in Equation (2.5) are the choices in the initial risk that correspond to teaching and receiving a

---

<sup>16</sup> Note that this is not a deduction, so every teacher can get this, regardless of whether she itemizes or not (<http://www.irs.gov/publications/p970/ch03.html>, accessed April 4, 2013).

degree. These choices incur the additional costs of returning to college, but still face the same job characteristics as the first case. This is because the extra salary from having a master's degree isn't applied until the next school year. This leads into the third case, beginning the period with a master's degree of any type ( $r \in \{1, 2, 3\}$ ). Here the relevant wage is the master's-level wage,  $wage_{it}^{MA}$ . Thus there is a one-period difference between when the teacher pays the college costs and receives the wage bump. Note that I allow  $\psi$  to be different from the element in  $\delta$  that is the coefficient on tuition. These utility of money parameters are estimated separately to allow for possibilities of the tuition being subsidized by an unobserved scholarship.

The last case in equation (2.5) is the utility from choosing the outside option in any choice set ( $r \in \{0, 1, 2, 3\}$ ). It does not depend on the school characteristics, nor a teaching wage, but rather the wage for the outside option,  $wage_{it}^O$ . In addition, experience does enter this function, which is implemented by putting restrictions on the coefficients on experience. Because this is a choice model, there needs to be a base choice. For each choice set, the outside option  $j \in \{O\}$  is set to be the base choice. This has implications for the interpretation of the coefficients, which I discuss in Section 2.5. Figure 2.10 adds payoffs to the nodes from Figure 2.9 to help visualize the different utilities that each choice gives.

### 2.3.3 Value Functions

The payoffs from future periods are determined via the value functions. These are formed similar to Hotz and Miller (1993) and Arcidiacono and Miller (2011). The teacher is a forward-looking agent, and as such chooses  $d_{it} = j \in \mathcal{J}^r$  to maximize an expected discounted sum of utilities

$$\mathbb{E}_t \left[ \sum_{\tau=t}^T \sum_j^{\mathcal{J}^r} \beta^{\tau-t} \mathbf{1} \{d_{i\tau} = j\} [u_{ij\tau}(x_{i\tau}) + \varepsilon_{ij\tau}] \right], \quad (2.6)$$

where  $\beta$  is the discount factor. Flow utility  $u_{ijt}(x_{it})$  is defined from the current period utility function equation (2.5) as the non error terms:  $U_{ijt}(x_{it}) = u_{ijt}(x_{it}) + \varepsilon_{ijt}$ . The probability  $p_{jt}$  of choosing option  $j$  is the expected value of the indicator function,  $E[\mathbf{1}\{d_{it} = j\}]$ . This probability is the conditional choice probability of the agent choosing action  $j$  in period  $t$ .

Now define the value function as the expected discounted sum of utilities in equation (2.6), conditional on behaving optimally:

$$V_{it}(x_{it}) = E_t \left[ \sum_{\tau=t}^T \sum_j^{\mathcal{J}^r} \beta^{\tau-t} \mathbf{1}\{d_{i\tau}^0 = j\} [u_{ij\tau}(x_{i\tau}) + \varepsilon_{ij\tau}] \right],$$

where  $d_{i\tau}^0 = j$  is the optimal decision in each period.

I define the conditional value function for choosing option  $j$ ,  $v_{ijt}(x_{it})$ , as the flow payoff of choosing option  $j$ ,  $u_{ijt}(x_{it})$ , plus the payoffs from behaving optimally in the future. Then for a given choice, the conditional value function can be expressed as:

$$v_{ijt}(x_{it}) = u_{ijt}(x_{it}) + \beta \sum_{x_{i,t+1}} V_{i,t+1}(x_{i,t+1}) f_j(x_{i,t+1}|x_{it})$$

where the future payoffs are dependent on choice  $j$  via the probability  $f_j(\cdot)$  of being in state  $x_{i,t+1}$  given that the agent is currently in state  $x_{it}$ .

I use the *conditional* value function to express the future *unconditional* value function as the maximum of the future *conditional* value functions, and thus

$$v_{ijt}(x_{it}) = u_{ijt}(x_{it}) + \beta \sum_{x_{i,t+1}} \left[ E_t \max_{k \in \mathcal{J}^r} \{v_{ik,t+1}(x_{i,t+1}) + \varepsilon_{ik,t+1} | d_{it} = j\} \right] f_j(x_{i,t+1}|x_{it}) \quad (2.7)$$

where the expectation  $E_t$  is over the draws of  $\varepsilon_{ij,t+1}$ .

The utility maximization process can now be represented as choosing  $d_{it} = k$  such that

$$k = \operatorname{argmax}_{j \in \mathcal{J}^r} \{v_{ijt}(x_{it}) + \varepsilon_{ijt}\}. \quad (2.8)$$



## 2.4 Identification and Estimation

This section lays out the identifying assumptions and estimation procedures. I present standard assumptions to enable estimation, including assumptions on the distribution of the unobserved terms. Additional assumptions are based on the various educational policies that are in effect. Estimation of the likelihood is performed using established conditional choice procedures. The unobserved type is estimated using the Estimation-Maximization algorithm. These strategies allow for easily interpretable results with minimal assumptions.

### 2.4.1 Assumptions

Estimation begins by making an assumption on the distribution of the choice function error terms. Specifically, I assume that  $\varepsilon_{ijt}$  has a Type-I extreme value distribution, which results in the joint distribution of the degree-decision to be a multinomial logit. Assuming this functional form for the error term implicitly assumes that the independence of irrelevant alternatives (IIA) holds.

I assume that the error term from the ability equation,  $\nu_i$ , is distributed  $\mathcal{N}(0, \sigma)$ . Further, I assume that after controlling for the unobserved type, the two error terms are independent of each other.

I assume that the unobserved type  $s_i$  is assumed to have a discrete distribution. I currently estimate the model assuming  $S = 2$  types. To address the issue of missing test scores, I assume that the distribution of the unobserved type is the same across the teachers who report their Praxis I score and those who do not. This is most important across time, since fewer and fewer teachers report their Praxis I scores as time goes on.

Because wages are not observed in the outside option, I set  $wage_{it}^O$  equal to the average wage of college-educated individuals in the teacher's county. In addition, the

outside option is assumed to be a terminal choice.<sup>17</sup>

Finally, I assume that  $f(\cdot)$  follows a Markov process. As such, I calculate it as a single, one-period transition matrix. The variables that are allowed to transition dynamically are experience and the metrics of school quality. As mentioned previously, these metrics are socio-economic quartiles, proxied for by percent free-lunch at the school, and average test score quartiles.

#### 2.4.2 Identification and Estimation

Identification of the unobserved type comes from observed variation in the data. An individual who has a high Praxis I score, chooses to attend a for-profit university and then leaves the data, most likely has a low draw on unobserved ability. Similarly, a black, female with a low Praxis I score, who is consistently in high quality schools, is likely to have a high draw on unobserved ability. Situations like these and the panel nature of the data are what allow for identification of the unobserved type.

The estimation procedure is laid out in detail in the appendix, but I will provide a high-level summary of the process here. The assumption on the error term allows the  $E_{max}$  term in equation 2.7 to have a clean, closed form solution:

$$v_{ijt}(x_{it}) = u_{ijt}(x_{it}) + \beta \sum_{x_{i,t+1}} [v_{iOr,t+1}(x_{i,t+1}) - \ln(p_{Or,t+1}(x_{i,t+1})) + \gamma] f_j(x_{i,t+1} | x_{it}). \quad (2.9)$$

The future utility can then be expressed in terms of the one-period ahead conditional value function of the outside option,  $v_{iOr,t+1}(x_{i,t+1})$ , and one-period ahead choice probability for the outside option,  $p_{Or,t+1}(x_{i,t+1})$ . Both of these are functions of state variables  $x_{i,t+1}$ , that change probabilistically depending on current period choices.

By making the outside option a terminal state, I do not worry about any choices that teachers make after that point, and thus the value of the state variables after

---

<sup>17</sup> I do not observe what teachers do after they leave education, though it is most likely very broad, from teaching jobs in other states, non-education jobs or home production, to name a few.

the outside option is chosen is immaterial. Further, since I assumed the wage for the outside option was the average wage of college graduates in the teacher's county, future payouts will be the same no matter what the current period choice is. This just leaves the one-period ahead conditional choice probability of choosing the outside option, which I calculate relying on the transition rates of the state variables and using a flexible logit. Estimation takes place via maximum likelihood and the Expectation-Maximization algorithm (see Arcidiacono and Miller (2011)). See the appendix for full details on the estimation routines.

## 2.5 Results

In this section I present the results from estimating the dynamic model. I find that teachers are willing to pay significant amounts to be able to move up to a higher quality school. I also find that they appear to value the access that comes with online programs more than the tuition costs.

I first present the structural estimates from the observed ability, equation (2.1), which are found in Table 2.6. Recall that these are the effects of regressing the observed teacher test score (Praxis I) on race and unobserved ability. It is important to note that the coefficient on unobserved ability is restricted to unity. This has two implications. First, restricting it to be positive implies that being Type 1 instead of Type 0 indicates that the teacher is a high-ability teacher. Second, since the distribution of Praxis I scores was normalized to have zero mean and unit variance, the interpretation of the coefficients on type dummy in the utility equation will be that of one standard deviation of unobserved ability.

The coefficients on race in the ability equation are negative and significant for non-white teachers. This is consistent with a theory of racial bias in the exam. In particular, coupled with the intercept and the restriction on the type dummy, a black, Type-1 teacher and a white, Type-1 teacher would have expected Praxis I scores 0.17

standard deviations below the mean and 0.56 standard deviations above the mean, respectively. Thus even the high ability black teachers suffer enough of a racial bias to score below the average. This reemphasizes the importance of capturing the effects of unobserved academic preparation on the decision process.

The structural estimates from the utility equation are found in Table 2.7. The first panel lists the teacher characteristics. These columns are grouped by choice set. The first four columns of estimates are all in the initial choice set ( $r = 0$ ), where the first three indicate a teacher deciding to teach and go back to college, the next one is to teach without going back to college, and omitted is the reference choice, which again refers to the outside option of leaving the North Carolina public school system. The other three columns are the estimates for the other three choice sets ( $r \in \{1, 2, 3\}$ ), listing the estimate for choosing these to teach with the master's degree, and omitting again the reference choice of the outside option.

The coefficients on race and gender are very interesting. The coefficient on black choosing to receive a master's from a for-profit is quite large and positive (1.42). While this may be explained by black teachers having a much higher preference for for-profit universities, it is also that this is picking up other items. There are fewer application requirements for for-profit universities, thus this may be capturing a preference for a program that is easier to get into. Alternatively, if it is indeed the case that for-profit degrees require less work, as anecdotal evidence supports, this could be a preference for the ease of degree completion. But note that in addition, black teachers are much more likely to choose to receive any type of master's degree, relative to the outside option, than any other race. Thus it is likely that black teachers simply have an unobserved preference for obtaining a degree. The direction of the coefficients on female are the same as on black, implying that similar stories may also hold for female teachers.

The other teacher characteristic to examine is the type dummy. Recall from above

that the dummy is turned on for high-ability, Type 1 teachers, who by construction earn exactly one standard deviation higher than low-ability, Type 0 teachers. The biggest role this plays in the decision to get a master's degree from a public university, which is much higher than the other options. Further, all coefficients are positive for teachers who begin a period with a master's degree, indicating they are more likely to stay teaching than low types. This is also true for teacher with master's degrees from non-profit universities. However, among teachers from public and for-profit universities, high ability types are more likely to leave than low ability types. This is another very interesting story. High-ability teachers who really value teaching and want to stick around are those who choose the non-profit option.

The next panel of Table 2.7 looks at job characteristics. Note that these coefficients are restricted to be the same across all choices. Here is where we see the relative value of salary to school quality. First looking at salary, the coefficient is fairly small at 0.011 for \$10,000. The increase in salary comes the period after choosing to get a degree. The other reason is related to the non-teaching choice in the initial choice set. The salary on the outside option is set equal to the average wage of college graduates that live in the teacher's county, which for every county in North Carolina is higher than the average teacher salary. Thus, by construction, it is difficult for salary to enter too highly into the teachers utility. But it is still positive, thus an increase in teacher salary will cause teachers to continue teaching, with or without a master's degree.

The metrics of school quality are the average socio-economic status (SES) and the average test score of the students at the teacher's school. Note that the preference for these goes in opposite directions. Thus, conditional on test score, teachers prefer to be with the lowest SES students. This speaks to a general preference of wanting to help those who need the most help and are able to receive it. This implies that the most preferred school would be the lowest quartile in SES and the highest in test

score. This extreme happens rarely, with 90% of schools either being in the same quartile for both scores or being off by one quartile. This is helpful to consider the relative value of moving between quartiles to the next, which is roughly \$10,000 for each difference of 0.011. Thus the value of being at schools in both highest quartiles is  $0.102 - 0.042 = 0.59$  and that in the second highest is  $0.093 - 0.036 = 0.057$ . This difference of 0.002 translates into roughly \$1,700. Note further that the difference between the third highest and the second highest is 0.024, which is roughly \$20,000. Thus we see that teachers do indeed place a high value on job characteristics beyond simply salary.

The final panel of Table 2.7 contains the estimates of college cost. The most important characteristics are the availability of distance education. Teachers with at least two online options of a given type of college are very likely to return and get a degree. The coefficient on two options is 2.61 compared to 2.74 and 3.11 for three and four-plus options, respectively. Thus while the effect continues to increase with more options, the key factor in deciding on a master's degree type is having a couple of different options of that type. In contrast, the effect of a similar number of physical locations is quite muted, though still positive and statistically significant. The estimates are all within the range of 0.13-0.46, indicating that the number of colleges of a given type is not important, just that there is at least one. Thus the value of having increased online access is much, much more important than having colleges close by.

Finally, the coefficient on tuition is -0.06 for \$10,000. The negative value does adversely effect the decision go back to school (as it should), but even with the average tuition in 2011 ranging from \$10,000 to \$20,000, the effect of the tuition cost pales in comparison to the falling access costs. Also, recall that I estimated the coefficient on tuition and salary separately to allow for potentially unobserved tuition costs or subsidies, and the coefficient on salary is 0.01. While the tuition coefficient

being higher is consistent with higher costs than I observe (perhaps teachers do not take the tax credit), it is still possible that the sticker shock of the degree also carries some negative utility to it. I continue in the next section to examine how well the estimates fit the model and predict future simulations.

## 2.6 Counterfactuals

This section presents simulations of the data. I first show that the model fits the data quite well. I then explore how various policies and situations would change teacher outcomes.

### 2.6.1 *Model Fit*

The fit of the model can be found in Table 2.8. The first column refers to the observed choices and state variables in the data. The second column refers to the results of a forward simulation, starting at each person's initial characteristics and choice sets, and allowing the choices and state variables to evolve according to the estimates and transition rates.

The first panel of the table looks at the choice probabilities. These match up very well, especially for the choice sets once a master's degree is received ( $r \in \{1, 2, 3\}$ ). The initial choice set also matches up very well, although the forward simulation predicts slightly less teachers continuing to teach at the bachelor's degree level and slightly more leaving the profession. This is partly due to some teachers choosing the outside option in the first period in the simulation, which is not observed in the data.

The second panel compares the state variables across the two samples. First, this shows the distribution of observations at each quartile of the school quality measures, SES and ability. The forward simulation does a very good job at matching the data, with (absolute) differences ranging from 0.1 to 0.6 percentage points. In addition

to school quartile metrics, I report mean tenure. There is a difference here, with the forward simulation predicted about 1/4 of a year less than the data. This is mainly driven by those in the initial choice set choosing the outside option a bit more frequently.

### *2.6.2 Simulations*

I use the estimation results to run various counterfactual simulations. The connecting theme of these simulations is related to policies that equalize the costs of attending various college types. These policies need to address both financial costs (tuition) and accessibility costs (proximity, online). The most straight-forward policy would address the financial cost by subsidizing college types with high tuitions. The other policies to address the access costs would be more complicated, but potentially still feasible, perhaps including compensating rural teachers to spend their summer in a more urban area or compensate institutions to provide more online programs. While these policies may take different shapes, the unifying feature is the same: all teachers face the same costs across college types. In order to quantify the effects of policies like this, I compare the decisions teachers made in the data with the decisions they make in the simulation. I then compare the job characteristics that people have before and after the decision. I also break these results out by demographic groups to compare how different races and genders react to the policies.

Another simulation I consider is a simple comparison of different types of teachers in the data. Specifically, I compare how high and low draws from the ability distribution differentially affect teachers choices and outcomes (state variables). The summary of the result is that teachers with high ability draws make choices that lead them probabilistically into higher quality schools, though the final differences in school quality are small.

The current results of the simulations can be found in Table 2.8. The biggest



result of equating the college costs is seen in the non-profit choice. Before, less than 1% of teachers in the initial choice set chose to receive a degree from a non-profit. However, in the world where college costs are equated, this jumps up to 10%. This comes in partially because the intercept in the non-profit choice is less negative than the other choices. Thus, when the playing field is leveled, the natural preferences for this choice push teachers into this option. Results from other simulations are forthcoming.

## 2.7 Conclusion

There has been significant changes in the incidence and form of master's degrees in education over the past 20 years. I have presented a dynamic model which allows teachers to internalize the differing financial and labor market costs and benefits of this decision. My main results have shown that teachers are very heavily influenced by the increased availability of distance learning. I have also show that teachers face different labor market outcomes based on where they get their master's degree from. Specifically, teachers receiving their degrees from for-profit institutions start in schools with lower socio-economic status and lower test scores, and it is harder for these teachers to transition upwards to a higher quality school. The estimates of their decision making shows that teachers who attend for-profit universities are not as interested in moving up to higher academically-achieving schools as teachers who attend more traditional colleges.

I am able to utilize this model to run various counterfactual scenarios. One of the more interesting ones is to see how would teachers act if they no longer had as many options. Specifically, I can simulate a world in which for-profit universities don't exist or where online options are more wide-spread. Another interesting scenario analyzes how teachers' education decisions change in the absence of a premium for a master's degree. This is especially relevant in North Carolina since recent legislation

revoked this premium for all new teachers, which went into effect in July 2014.<sup>18</sup>

---

<sup>18</sup> Senate Bill 402, <http://www.ncleg.net/sessions/2013/budget/2013/S402-CSMDxf-9.pdf>, accessed Sept 13, 2014.

## Figures and Tables

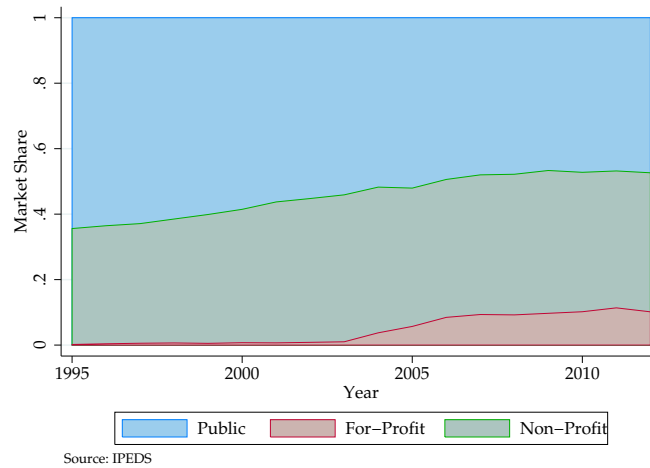


FIGURE 2.1: Market Share of Master's Degrees Awarded in Education by Year and College Type

Table 2.1: Sample Teacher Salary Schedule, 2005-2006

Experience	BA	MA
0	\$25,510	\$28,060
1	\$25,930	\$28,520
2	\$26,370	\$29,010
3	\$27,930	\$30,720
4	\$29,330	\$32,260
5	\$30,670	\$33,740
6	\$31,960	\$35,160
7	\$33,000	\$36,300
8	\$33,480	\$36,830
9	\$33,970	\$37,370
...	...	...

Source: NCERDC.

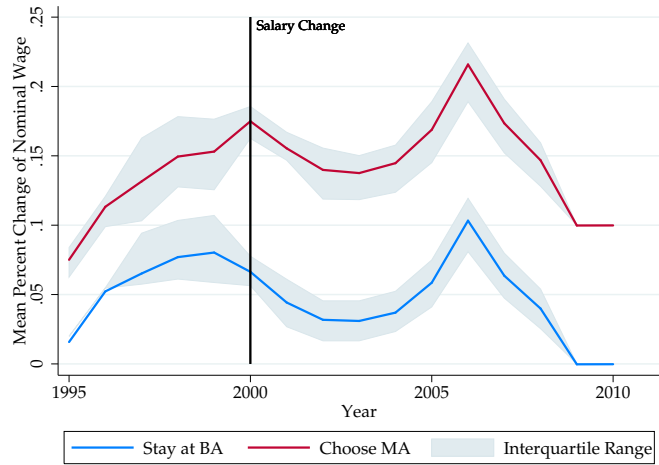


FIGURE 2.2: Annual Percent Change of Salary by Degree Decision

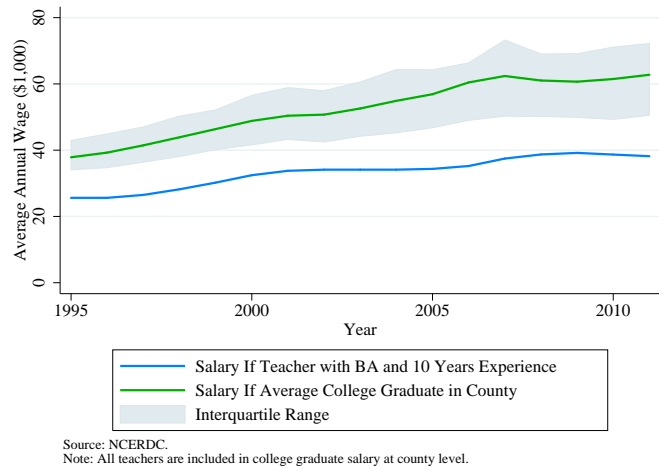
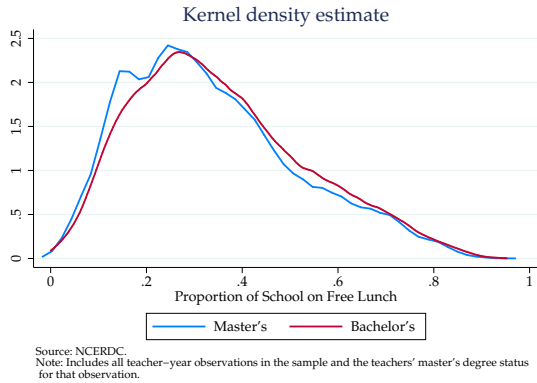
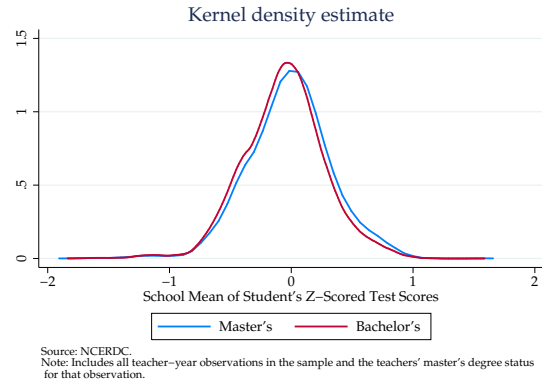


FIGURE 2.3: Annual Average Salary if Teacher with BA vs Average College Graduate in County

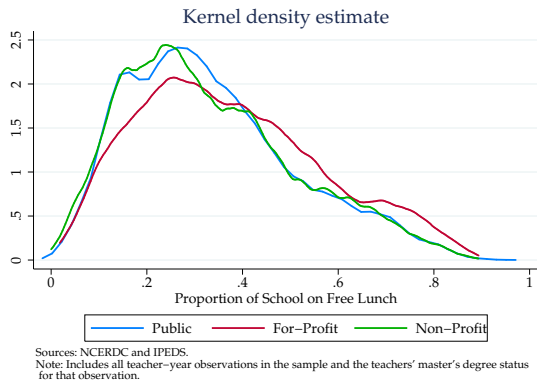


(a) School Free Lunch

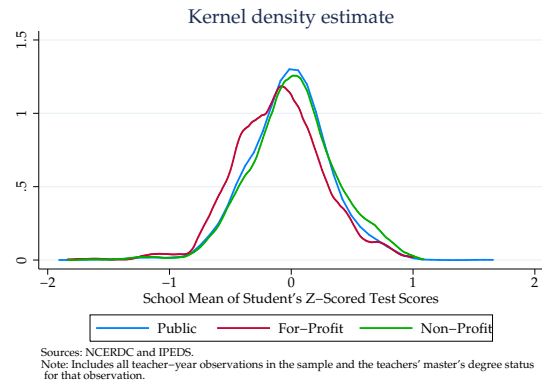


(b) School Test Score

FIGURE 2.4: School Quality by Degree Level



(a) School Free Lunch



(b) School Test Score

FIGURE 2.5: School Quality by Master's Degree Type

Table 2.2: Teacher Transition Rates by Bachelor's and Any Master's

(a) State-wide Quartile of the Average Socio-Economic Status (SES) of Students at a Teacher's School

		SES Quartile, $t + 1$							
		Bachelor's				Any Master's			
		Bottom	2nd	3rd	Top	Bottom	2nd	3rd	Top
SES Quartile, $t$	Bottom	34.3	27.4	21.7	16.5	29.1	27.2	22.1	21.6
	2nd	19.8	29.0	30.2	21.1	18.0	27.9	30.3	23.8
	3rd	12.1	23.0	32.7	32.2	11.3	20.7	30.8	37.3
	Top	8.6	14.3	27.4	49.7	6.6	13.1	27.0	53.2

(b) State-wide Quartile of the Average Test Scores of Students at a Teacher's School

		Test Score Quartile, $t + 1$							
		Bachelor's				Any Master's			
		Bottom	2nd	3rd	Top	Bottom	2nd	3rd	Top
Test Score Quartile, $t$	Bottom	29.4	31.7	22.2	16.7	29.3	27.9	21.1	21.7
	2nd	17.0	33.5	27.4	22.0	15.6	33.1	26.3	25.0
	3rd	10.3	25.2	35.0	29.5	7.0	21.8	34.8	36.4
	Top	8.2	17.3	28.5	46.1	7.8	15.8	25.4	51.0

Sources: NCERDC and IPEDS.

Notes: Rates are for white, female teachers who switch schools.

Table 2.3: Teacher Transition Rates by Bachelor's and For-Profit Master's

(a) State-wide Quartile of the Average Socio-Economic Status (SES) of Students at a Teacher's School

		SES Quartile, $t + 1$							
		Bachelor's				For-Profit Master's			
		Bottom	2nd	3rd	Top	Bottom	2nd	3rd	Top
SES Quartile, $t$	Bottom	34.3	27.4	21.7	16.5	28.6	28.7	21.9	20.8
	2nd	19.8	29.0	30.2	21.1	18.7	29.0	29.3	22.9
	3rd	12.1	23.0	32.7	32.2	10.3	20.6	32.9	36.2
	Top	8.6	14.3	27.4	49.7	6.3	12.8	26.1	54.8

(b) State-wide Quartile of the Average Test Scores of Students at a Teacher's School

		Test Score Quartile, $t + 1$							
		Bachelor's				For-Profit Master's			
		Bottom	2nd	3rd	Top	Bottom	2nd	3rd	Top
Test Score Quartile, $t$	Bottom	29.4	31.7	22.2	16.7	45.5	22.7	22.7	9.1
	2nd	17.0	33.5	27.4	22.0	22.5	30.0	32.5	15.0
	3rd	10.3	25.2	35.0	29.5	0.0	28.6	42.9	28.6
	Top	8.2	17.3	28.5	46.1	4.2	8.3	33.3	54.2

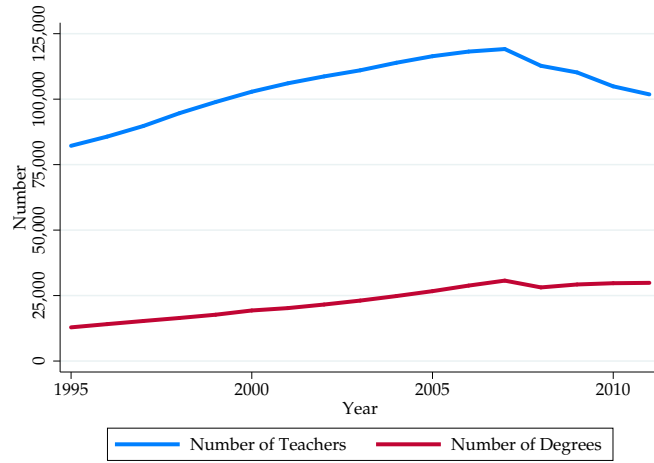
Sources: NCERDC and IPEDS.

Notes: Rates are for white, female teachers who switch schools.

Table 2.4: Impact on Exit Rates

Final Degree	Odds Ratio
BA	1.000 (base)
Public	0.996
For-Profit	0.501***
Non-Profit	0.785***

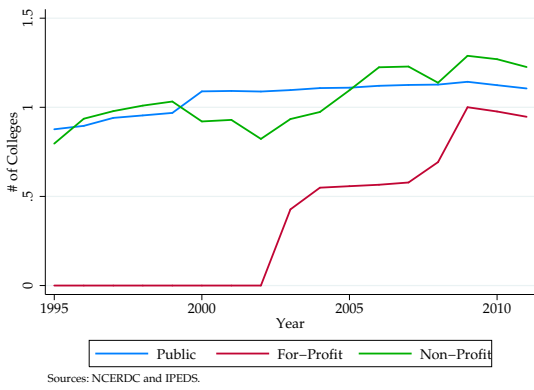
Sources: NCERDC and IPEDS.  
 Estimates are from logistic regression of choosing the outside option on degree type, race, gender and Praxis I.



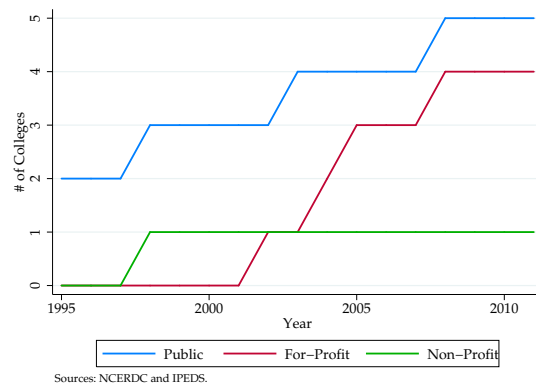
Source: NCERDC

FIGURE 2.6: Stock of Teachers and Teachers with Master’s Degrees by Year in North Carolina

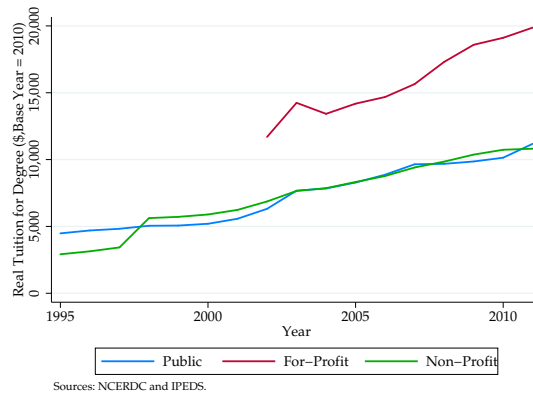




(a) # of Colleges within 25 Miles of Teacher



(b) # of Colleges with Online Education



(c) Average Tuition

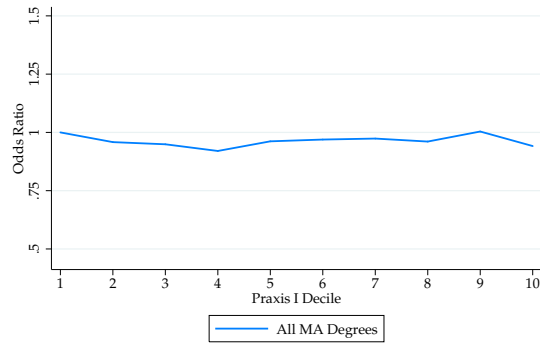
FIGURE 2.7: Access and Tuition Metrics by College Type

Table 2.5: Demographic Statistics

	White	Black	Other Race	Male	Female	Total	N
Overall %	82.1	14.9	2.9	22.0	78.0	100.0	101,066
BA %	82.3	14.6	3.0	22.9	77.1	100.0	83,476
MA %	81.0	16.4	2.6	18.0	82.0	100.0	17,590
Mean Tenure	4.80	4.25	3.98	4.13	4.80	4.70	101,066
Mean Praxis I	0.13	-0.24	-0.09	0.17	0.05	0.08	48,354

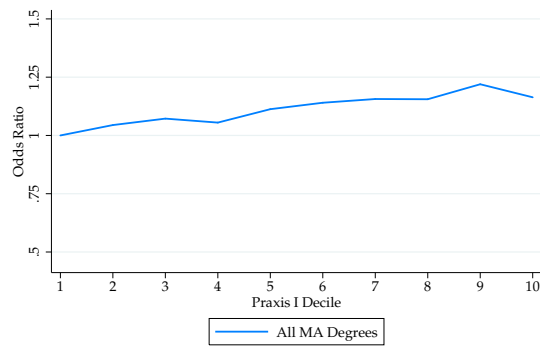
Sources: NCERDC and IPEDS

Note: Includes all teachers in the sample and their eventual master's degree status.



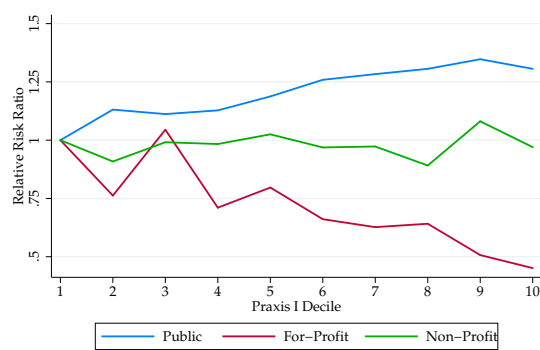
Source: NCERDC.  
 Note: Includes all teachers in the sample and their eventual master's degree status.

(a) No Controls



Source: NCERDC.  
 Note: Includes all teachers in the sample and their eventual master's degree status. Includes controls for race and gender.

(b) Demographic Controls



Sources: NCERDC and IPEDS.  
 Note: Includes all teachers in the sample and their eventual master's degree status. Includes controls for race and gender.

(c) Demographic Controls and Degree Type

FIGURE 2.8: Odds Ratios of Praxis I Deciles on Degree Decision, Baseline is BA

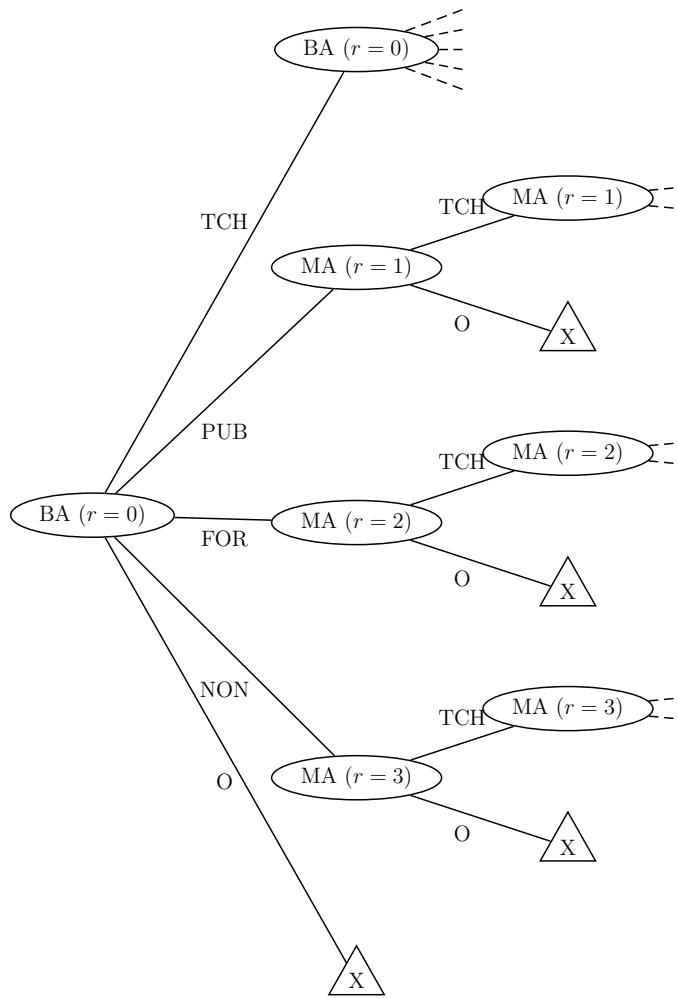


FIGURE 2.9: Decision Tree

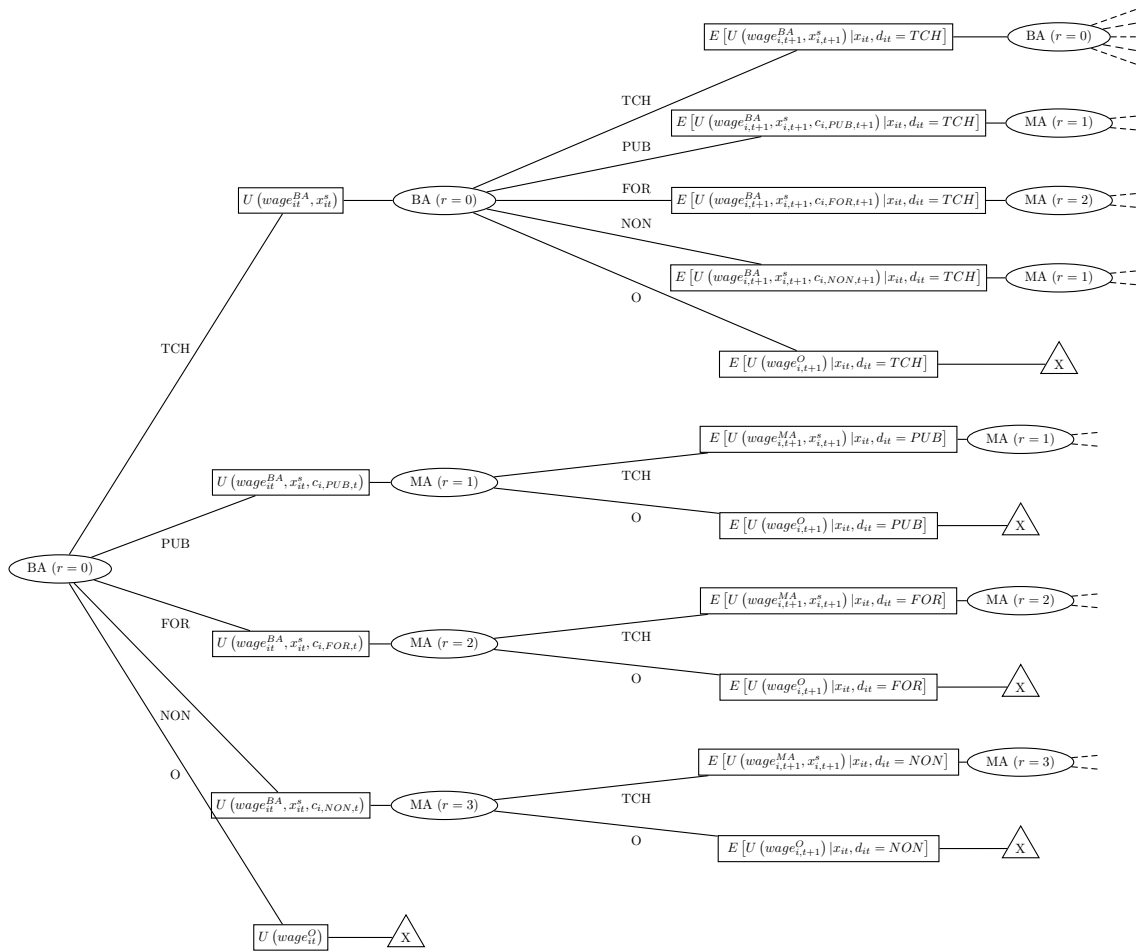


FIGURE 2.10: Payoff Tree

Table 2.6: Structural Estimates - Ability Equation

	DV = Praxis I Z-Scored
constant	-0.437***
black	-0.731***
other race	-0.358***
type dummy ( $\kappa$ )	1.00 <sup>a</sup>
standard deviation ( $\sigma$ )	0.610***

<sup>a</sup> Coefficient restricted to be 1.  
 \*\*\* p<0.01, \*\* p<0.05, \* p<0.1

Table 2.7: Structural Estimates - Utility Equation

(a) Teacher Characteristics

	Teach & Receive MA			Teach	Teach Already w/ MA		
	Public	For-Profit	Non-Profit	BA Only	Public	For-Profit	Non-Profit
constant	-7.71***	-10.07***	-5.62***	0.15***	0.08	0.44***	0.10
black	0.15***	1.41***	0.10*	-0.02	-0.05 <sup>a</sup>	-0.05 <sup>a</sup>	-0.05 <sup>a</sup>
other race	-0.02	0.15	-1.07***	-0.04	-0.03 <sup>a</sup>	-0.03 <sup>a</sup>	-0.03 <sup>a</sup>
female	0.19***	0.50***	0.27***	0.01	0.06 <sup>*a</sup>	0.06 <sup>*a</sup>	0.06 <sup>*a</sup>
type dummy ( $\kappa$ )	0.92***	0.37***	0.24***	0.16***	-0.04	-0.18	0.12**

(b) Job Characteristics

salary (\$10K)	0.011***
student SES 25th-50th <sup>b</sup>	-0.030**
student SES 50th-75th <sup>b</sup>	-0.036**
student SES 75th-100th <sup>b</sup>	-0.042**
student test score 25th-50th <sup>b</sup>	0.064***
student test score 50th-75th <sup>b</sup>	0.094***
student test score 75th-100th <sup>b</sup>	0.102***

(c) College Costs

tuition (\$10K) <sup>c</sup>	-0.067**
1 coll w/in 25 mile <sup>c</sup>	0.230***
2 coll w/in 25 mile <sup>c</sup>	0.181***
3 coll w/in 25 mile <sup>c</sup>	0.133***
4+ coll w/in 25 mile <sup>c</sup>	0.457***
1 coll w/ dist. ed <sup>c</sup>	-0.182
2 coll w/ dist. ed <sup>c</sup>	2.616***
3 coll w/ dist. ed <sup>c</sup>	2.753***
4+ coll w/ dist. ed <sup>c</sup>	3.111***

Notes: Model includes time trends and experience.

<sup>a</sup> These values are restricted to be the same across all options for those who already have an MA.

<sup>b</sup> School quality quartiles do not enter the utility for the outside options.

<sup>c</sup> College costs only enter the "Teach and Receive MA" options in the first choice set.

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

Table 2.8: Model Fit and Simulations

		Data	Model Fit	Simulation (Equate College Costs)
<i>Choice Probabilities</i>				
$r = 0$ (Initial Choice Set)	Public	2.0	2.0	0.8
	For-Profit	0.2	0.2	0.2
	Non-Profit	0.6	0.6	10.1
	BA Only	88.3	85.8	80.6
	Outside	9.0	11.5	8.4
$r = 1$ (Public)	Teach	90.2	90.6	89.3
	Outside	9.8	9.4	10.7
$r = 2$ (For-Profit)	Teach	91.9	92.4	91.1
	Outside	8.1	7.6	8.9
$r = 3$ (Non-Profit)	Teach	89.7	89.7	90.4
	Outside	10.3	10.3	9.6
<i>State Variables</i>				
SES Quartile	Bottom	21.0	21.4	21.7
	2nd	23.4	23.6	24.0
	3rd	26.8	26.7	26.1
	Top	28.9	28.3	28.1
Ability Quartile	Bottom	13.9	14.3	14.4
	2nd	24.9	25.1	25.0
	3rd	36.0	35.8	36.8
	Top	25.3	24.8	23.8
Mean Tenure		4.97	4.69	5.41

\*\*\* p&lt;0.01, \*\* p&lt;0.05, \* p&lt;0.1



# Changes across Cohorts in Wage Returns to Schooling and Early Work Experiences: Distinguishing Price and Composition Effects

## 3.1 Introduction

Since the 1970s there have been dramatic changes in the structure of the U.S. labor market. Foremost among these is a steep increase in the college wage premium during the 1980s, followed by a slower increase thereafter.<sup>1</sup> The characteristics and skill accumulation of American youth have also changed over this same time period.<sup>2</sup> College attendance has increased, college graduation has been delayed, and the amount of in-college work experience has gone up.<sup>3</sup> These underlying changes to the composition of youth are of immense importance in understanding how the overall premium for skill investment has evolved.

Our paper looks at three related research questions: What is the relative impor-

---

<sup>1</sup> See, for example, Katz and Murphy (1992); Card and Lemieux (2001); Carneiro and Lee (2011)

<sup>2</sup> For example, Altonji et al. (2012) note an increase in skills over time, but an overall widening of the skill distribution driven by trends in parental education.

<sup>3</sup> See, for example, Bacolod and Hotz (2006); Scott-Clayton (2012); Bound et al. (2012)

tance of changes in skill price versus skill composition over the past 20 years? What are the trends in the wage returns to in-school work experience? How much of the evolution in the college wage premium actually reflects an increase of in-school, and, more generally, early work experience? Failure to account for evolution in the incidence of early work experience may lead to an overestimation of the increase in the returns to schooling. In answering these questions, we control for both endogeneity and selection that plague estimates of the returns to skills. We do this by specifying and estimating, for a series of cohorts, a dynamic model of schooling and work decisions. We then decompose the evolution in the data into price and composition effects (see Oaxaca, 1973).

We find that the relative importance of the price and composition effects varies dramatically across skills. Regarding in-school work experience, we find that the direct returns to working while in college have decreased over time, with the earlier decrease attributable mostly to price effects, and the more recent decrease attributable mostly to composition effects. Further, composition effects explain little in the evolution of the college wage premium. Related to this, we find that there is both a significant increase in the incidence of in-college work over time and a decrease in the wage return to in-college work. These combine to produce a negative impact on the composition effect, which is offset by a positive net impact of the remaining skill correlates. Finally, and consistent with other studies (e.g. Taber, 2001), we find that almost all the increase in the college wage premium in the 1980s is due to a change in the returns to and composition of unobserved skills. These unobserved effects have diminished greatly in recent years and have contributed to the declining growth in the college wage premium.

Our analysis makes use of two longitudinal data sets, the 1979 and 1997 panels of the National Longitudinal Surveys of Youth (NLSY). We divide our analysis among three cohorts of individuals: (i) NLSY79 respondents born in years 1959 and

1960; (ii) NLSY79 respondents born in years 1961 through 1964; and (iii) NLSY97 respondents, all of whom were born in years 1980 through 1984. As will be shown, these three cohorts differ markedly in their human capital investment decisions and the market conditions they faced while making such decisions.

While ours is not the first study to examine labor market trends over this time period, our use of longitudinal rather than repeated cross-sectional data allows us to more accurately measure early-career work experience and account for its endogeneity. From each of the NLSY surveys, we construct comparable measures of schooling, work, and military histories from ages 16-29, along with comparable measures of earnings, educational attainment, ability, local labor market conditions, and personal and family background characteristics. From these histories, we are able to construct refined measures of human capital including whether or not work experience occurred simultaneously with schooling. Following many studies in the literature, we restrict our analysis to males.

In order to obtain wage estimates that reflect selection-free, causal effects of human capital accumulation, we specify and estimate a dynamic model of schooling and work decisions that controls for person-specific unobserved heterogeneity. We linearly approximate the value functions (see Eckstein and Wolpin, 1989), but allow the idiosyncratic shocks to be correlated across choice alternatives. This correlation is induced by our factor-analytic approach inspired by Cameron and Heckman (1998, 2001) and Heckman et al. (2006b). We use comparable cognitive test scores and the panel structure of the data to identify the heterogeneity factors. We then use the model estimates to conduct counterfactual simulations (decompositions) which allows us to assess the role of price and composition effects in unobserved skills as well as observed skills.

The remainder of the paper is organized as follows: Section 3.2 reviews the relevant literature; Section 3.3 details the construction of the data and the descriptive

trends over this time period; Sections 3.4 and 3.5 discuss the specification and estimation of our econometric model; Section 3.6 discusses results of the model estimates; Section 3.7 formulates counterfactual comparisons upon which we base our decompositions. Section 3.8 concludes.

## 3.2 Previous Literature

Our study brings together literature on a number of topics in labor economics. Such topics include estimating the returns to schooling in the presence of endogeneity and selection, analyzing and explaining the evolution of skill premia over time, understanding the effect of in-school work on future educational and labor market outcomes, and using decomposition methods to understand the extent to which differences in outcomes across groups are attributable to observable or unobservable characteristics of the groups.

The literature estimating the returns to schooling originated with Mincer (1974), who popularized the Mincer model which interprets the coefficient on schooling in a log wage equation as a rate of return. He uses 1960 Census data and finds that his model specification fits the data fairly well. More recently, Heckman et al. (2006a) survey 50 years of the Mincer model and find that the basic Mincerian earnings model holds in 1960 Census data, but not in more recent data. They conclude that using flexible methods yields more accurate estimates of the returns to schooling. They find that, between 1940-1990, the returns to college have stayed roughly constant for white men (at 16 log points), but have doubled for black men (from 12 log points to 25 log points, with most of the jump occurring during the 1980s).<sup>4</sup>

Our study follows Heckman et al. by making use of flexible polynomials of schooling and work experience in the wage equation, as well as “sheepskin effects”. We also

---

<sup>4</sup> Heckman et al. do not report estimates on the evolution of the returns to work experience because.

emphasize that we improve on other studies by using *actual* work experience rather than *potential* work experience in our calculation of the returns to schooling. This allows us to separate human capital accumulation from age effects.

Our work is also related to a number of studies who have used the NLSY cohorts to analyze changes in demographics and market structure over long periods of time. The overall theme of these papers is that skill levels have increased over time, but skill returns have decreased. Altonji et al. (2012) compare the NLSY97 with the NLSY79 and find, as we do, that the current generation has become more skilled than the previous generation, but that the skill distribution has widened. They do not look at how the prices to skill have changed over the two generations, nor do they discuss evolution in educational attainment or work experience. We emphasize that some of their documented change in the composition of skills could be due to changes in the price of skills. Bacolod and Hotz (2006) compare the NLSY79 with the earliest NLS cohorts (NLS-YW and NLS-YM) and find, similar to us, that skills have increased across generations while real wages over early working ages have decreased. However, they don't separate skill prices from skill composition. Boehm (2013) uses the NLSY79 and NLSY97 to test whether job polarization has caused the middle class squeeze. He finds that there is a strong association between job polarization and changes in returns to occupation-specific skills, but that this only explains the top part of the wage distribution. Castex and Dechter (2014) compare the NLSY97 with the NLSY79 and find that returns to cognitive skills have diminished over time but that returns to education have grown over time. They show that the decline in returns to ability can be explained by differences in the technological growth rate over the two generations.

We also add to the literature on understanding the effect of in-school work on future educational and labor market outcomes. Scott-Clayton (2012) describes the upward trend of in-college work over the past 40 years and examines a set of explana-

tions. She finds that while there is no one explanation for the rise of in-college work, borrowing constraints are a significant factor. Bacolod and Hotz (2006) document similar trends in working while in school using the NLSY79, NLS-YW and NLS-YM. Neither look at the effect of in-college work on future earnings. Hotz et al. (2002) use NLSY79 data to investigate whether in-school work is productive later in life. They find that, once controls for dynamic selection are implemented, the estimated returns diminish greatly in magnitude. Our analysis closely follows Hotz et al. (2002) but compares how wage returns have evolved over *multiple* NLSY cohorts.

Another segment of pertinent literature analyzes the long-term evolution in the college wage premium. These papers are primarily concerned with providing explanations for changes in skill prices and composition, which is different from our objective. Lee and Wolpin (2010) analyze the wage evolution from 1968-2000 and estimate a model that nests the following stories: wage inequality, increasing female labor force participation, the college wage premium, and shifts in employment from the goods-producing sector to the service-producing sector. Their model is able to explain wage dynamics arising from skill-biased technical change, capital-skill complementarities, changes in relative prices of product markets, and changing demographics. They find that no single explanation dominates. Carneiro and Lee (2011) use 1% Decennial Census samples from 1960-2000 and find that increases in college enrollment led to decreases in the average quality of college graduates, resulting in a 6% decline in the college wage premium. Cunha et al. (2011) use data from the PSID, NLS66 and NLSY79 to understand the determinants of the college wage premium. They find that shifts in the relative supply and demand of college vs. high school labor are the main reason for the increasing college wage premium, but that supply and demand shifts in skills and composition effects also play a role. Fang (2006) uses 1990 Census 5% PUMS data and finds that college education enhances students' productivity by 40%, and that this productivity enhancement accounts for up to

67% of the college wage premium. Fortin (2006) uses CPS data to understand the relationship between higher education policies, college labor supply, and the college wage premium across U.S. states. She finds that the relationship among these three variables is much weaker in states with high private enrollment rates, high migration rates, and high levels of interstate trade. Taber (2001) uses NLSY79 data and finds suggestive evidence that the increase in the college wage premium in the 1980s was due to an increase in the demand for unobserved ability.

We emphasize two important aspects in estimating the returns to schooling: *(i)* properly accounting for the fact that such decisions are endogenous; and *(ii)* accounting for accumulated work experience as opposed to potential work experience. In order to properly control for selection, we use a two-dimensional factor model to obtain selection-corrected wage estimates (see Taber, 2001; Hotz et al., 2002; Cunha et al., 2011). Within the factor model is a linear approximation of a dynamic model of early-career choices (see Eckstein and Wolpin, 1989; Keane and Wolpin, 1997). It is also important to account for work experience accumulated before graduation because such work experience may be rewarded upon post-schooling labor market entry. Failure to account for this pre-graduation work experience would bias estimates of the returns to schooling by incorrectly attributing the portion of the initial wage that corresponds to work experience.

Finally, in order to assess the reasons for why wage returns have evolved the way they have, we make use of decomposition methods in labor economics (Oaxaca, 1973; Fortin et al., 2011; Kline, 2011). Using our model, we decompose the observed evolution in skill premia into observed and unobserved price and composition effects.

### 3.3 Trends in Wages, Skills and Skill Returns across Cohorts

In this section, we discuss the data used to describe trends and estimate structural

models. We describe trends in wages and skill composition over the past 20 years. These trends form the basis of our decompositions in analyzing how much of the trends in earnings premiums are due to changes in composition versus changes in skill prices.

### *3.3.1 Data*

The data we use come from Rounds 1-15 of two panels of the National Longitudinal Survey of Youth (NLSY): the NLSY79 (calendar years 1979-1993) and the NLSY97 (calendar years 1997-2011). These surveys interview American youth beginning in their adolescent years and following them through adulthood. Topics of the survey include education, employment, marriage and fertility, health, and many others. The NLSY surveys are well-suited for our analysis for two reasons: (i) they contain rich information on human capital investment decisions early in the life cycle; and (ii) the data and surveying methodology are comparable enough to make credible inferences about how early-career human capital accumulation has evolved across the two surveys.

We divide our analysis among three cohorts of individuals: (i) NLSY79 respondents born in years 1959 and 1960 (henceforth referred to as “NLSY79 old” or “79o”); (ii) NLSY79 respondents born in years 1961 through 1964 (henceforth referred to as “NLSY79 young” or “79y”); and (iii) NLSY97 respondents (henceforth referred to as “NLSY97” or “97”). We follow other papers in the literature (e.g. Taber, 2001) that have divided the NLSY79 into multiple cohorts. This division is made primarily because swiftly changing market forces during the 1980s had a strong effect on human capital investment decisions.

### *Interview structure*

In both of the NLSY surveys, individuals are interviewed annually for the first 15 survey rounds and biannually thereafter. At each interview, respondents provide a



history of what has transpired in their life since the previous interview.<sup>5</sup> For example, the survey collects information on all jobs held between the current and previous interview, the wage and hours worked at each of those jobs, and the industry and occupation code of each job. Data related to educational attainment and schooling enrollment are similarly rich. Linking the survey reports together, it is possible to get a monthly employment (or schooling enrollment) history for each month since age 14. This data richness allows us to measure human capital at a refined level and is crucial to our analysis in two ways: (i) we can distinguish between work experience that occurred during school as opposed to over the summer between semesters; and (ii) we can differentiate work experience that occurred before and after schooling graduation.<sup>6</sup>

#### *Variable construction*

In order to answer the research questions posed at the beginning of this article, we use data on the following topics: personal and family background characteristics; local labor market conditions; earnings (if employed); and schooling and work histories, including military participation. For schooling and work histories, we observe for each calendar month the individual's schooling level and enrollment status along with his employment status and intensity (i.e. part-time or full-time). If an individual is employed, we observe his corresponding hourly wage. We discuss the exact classification of each schooling and work activity at the end of this subsection.

*Personal and family characteristics and innate ability* Personal characteristics observed in the data include the individual's Armed Services Vocational Aptitude Battery

---

<sup>5</sup> At the first interview, the survey asks extensive questions related to working and schooling history before the survey. For respondents who miss an interview, interviewers attempt to contact the individual during the next cycle.

<sup>6</sup> Measuring at the monthly level is particularly important because month of graduation from college is not as standardized as month of graduation from high school.

(ASVAB) subject test scores, race, nativity, and birth year. Family background characteristics in the data are not time-varying and are measured at the first interview. They include the education level of each of the individual’s biological parents, family income at the start of the survey, maternal co-residence status and whether or not the household had a female head when the respondent was age 14.

*Local labor market conditions* We observe local labor market conditions at the county level. These include the percentage of all residents who are employed in the individual’s county of residence along with the income per worker in the county. To create these local labor market variables, we make use of the restricted-access Geocode supplement of each of the NLSY surveys.

*Wages and educational degrees* The wage in our analysis is defined as the average hourly wage across all jobs worked in the month, weighted by the hours worked at each job. Wages are deflated using the CPI-U with a base year of 1982-84. We only include wages observed during employment spells (i.e. we discard wages reported when the individual was in the military or did not report working). We trim outliers by dropping wages outside of the range \$2-\$50 in 1982-84 dollars.

Educational attainment has three values, based on whether or not an individual holds a high school diploma or bachelor’s degree. Individuals with neither are classified as high school dropouts. Those who hold a GED or a high school diploma are considered high school graduates. Those who hold a bachelor’s degree are considered college graduates.

*School and work activity variables* In the analysis we make use of a monthly activity variable, which takes on six possible values in each of three different educational attainment sets (discussed previously, and hereafter referred to as risk sets). The

activity set contains the following choice alternatives: not working while in school; working while in school; working part-time (not in school); working full-time (not in school); military service; and all other activities (a residual category that includes home production and unemployment). The activity variable thus takes on 18 possible primary values. For example, work in school in the first risk set would be work during high school. Similarly, work in school in the second risk set would be work during college. In addition to these activities, the individual can transition to another risk set by graduating either high school or college. This results in two transition values that the activity variable can take on, one for each of the first two risk sets. The full set of possibilities is displayed in Table 3.1.

The primary monthly activity variable within each risk set is constructed as follows:

- Military if the person spent at least as many weeks in the military as working, and was not enrolled in school.
- Full-time working if the person was not in school, reported working all weeks of the months, and worked 35 or more hours per week.
- Part-time working if the person was not in school, and either reported positive weeks worked or more than 42 total hours worked in the month.
- Working while in school if the person was in school and worked at least one week in the month or at least 8 hours in the month.
- School only if the person was in school but did not report any weeks worked and reported less than 8 total hours worked in the month.
- “Other activities” if the person did not fall into any of the above categories.

### *Comparability across surveys and cohorts*

As discussed previously, the two NLSY surveys are quite comparable in their methodology and the types of information they collect. However, there are some key differences between them, which we discuss here.

Foremost among the differences is the age of respondents at the first interview. In the first wave of the NLSY79, respondents are aged 14-21 (aged 14-17 for the NLSY79 young and aged 18-21 for the NLSY79 old), in contrast to the NLSY97 where respondents are aged 12-16 at the first interview. This difference in starting ages makes it more difficult to create comparable pre-interview work and schooling histories, and ASVAB test scores.<sup>7</sup> As much as possible, we attempt to construct comparable measures of each variable of interest. As a compromise, we start measuring work history at age 16 and discard the oldest group of individuals in the NLSY79 old (i.e. those who were 20 or older at the time of the first interview).

The second difference between the two surveys has to do with attrition rates. In the NLSY97, attrition rates are much higher than in the NLSY79. For example, after 12 interviews in the NLSY79, the non-response rate was 10%, compared with about 17% for the NLSY97. While the higher attrition rate in the recent panel might be cause for concern, Aughinbaugh and Gardecki (2008) show that the additional attrition in the NLSY97 does not affect estimates of labor market outcomes. Furthermore, as discussed in Atrostic et al. (2001), attrition rates increased in six different U.S. government surveys during the 1990s. We take these conclusions as evidence that differing attrition rates between the two NSLY surveys is not a major problem for our analysis.

---

<sup>7</sup> We follow the procedure outlined in Altonji et al. (2012) to equate the ASVAB scores for both test-taking age and medium. This procedure is outlined at length in Altonji et al. (2009)

### *Sample inclusion and selection*

We construct individual monthly schooling and work histories from ages 16-29. In an attempt to minimize recall error of older respondents at the first interview, we exclude the oldest two birth cohorts of the 1979 survey (birth years 1957 and 1958), which is why the NLSY79 old has the smallest number of observations. We include all birth cohorts of the 1997 survey because these respondents were much younger. To minimize recall error within the survey time period, we drop any individuals once they miss three or more interviews. Because our model focuses on early career transitions, we follow each birth cohort until the age of 29.<sup>8</sup> Details on our sample selection are listed in Tables B.1 and B.2.

Our estimation subsample comprises 1,196 males in the NLSY79 old totaling 178,326 individual-month observations, 2,656 males in the NSLY79 young totaling 396,258 individual-month observations, and 4,443 males in the NLSY97 totaling 587,050 individual-month observations.<sup>9</sup> Our wage analysis comprises 100,293 observations in the NLSY79 old, 228,180 observations in the NLSY79 young, and 292,529 observations in the NLSY97.

### *3.3.2 Trends*

This section explains trends in the data regarding work experience, educational attainment, wages, ability, and family background characteristics. We outline the general trends in these data in order to describe how these trends have influenced earnings and human capital investment decisions over the time period under consideration.

---

<sup>8</sup> For the 1997 data, we have data on 29-year-olds for all but the youngest birth cohort (those aged 13 in 1997).

<sup>9</sup> We focus on males for two reasons: *(i)* because we would otherwise need to formally model marriage and fertility decisions, which would be too cumbersome for the present analysis; and *(ii)* because the literature that has studied human capital formation (e.g. Keane and Wolpin, 1997) has focused on males.

### *Work experience and educational attainment*

We first discuss trends across the three cohorts in schooling and work experience as well as educational attainment. Figure 3.1 displays the average age-experience profile for each of the three cohorts for a variety of work and school experiences. By age 29, individuals in the NLSY97 have significantly more months of in-college work experience than the other cohorts: eight months more than either of the NLSY79 cohorts. Pure schooling experience is actually greatest in the NLSY79 old, followed by the NLSY97 cohort, though the differences are smaller, with about two months between the cohorts. With regards to employment, total work experience at any give age is slightly smaller in the oldest NLSY79 cohort, but virtually the same in the rest. Since so much of the work experience for the NLSY97 cohort is due to in-school work, there is a very large gap in the amount of full-time work experience between the cohorts, with those in the NLSY97 cohort having between 6 months to a year less of full-time work experience.

To better understand how the trends in Figure 3.1 have evolved by education levels, we construct Figure 3.2, which presents the average education-experience profile at the end of each cohort's panel. Comparing panels (b) and (c) of this figure shows that the increase in in-school work in the NLSY97 was primarily in college rather than in high school. Panel (c) shows that this increase occurred for all college attendees, not just those who graduated.

The trends in non-school work experience show even more striking differences among the cohorts, as seen in panels (d)-(f) of Figure 3.2. We first note significant differences in the overall work experience of high school dropouts and graduates. Specifically, those in the NLSY97 cohort have much lower levels of total work experience by age 29. This implies a significant increase in either military or "other activities," most likely unemployment, given the similar levels of part-time work ex-

perience shown in panel (d). For those who began college, there is also a very large difference in full-time work experience in panel (e) between the NLSY97 and the NLSY79 cohorts. However, about half of this difference is explained by an increase in working in college. All of these trends suggest that there are large differences in the skill composition across cohorts.

Table 3.2 lists various graduation probabilities among the cohorts. High school graduation rates (or GED completion rates) improved by about two percentage points in recent years but were level between the NLSY79 cohorts. Further, the probability of beginning college has steadily increased by about 5 percentage points per cohort. This significantly outpaced the increase in high school graduation rates, implying that most of the increase in college attendance came from who previously would have graduated high school and not enrolled in college.

Examination of the college graduation rates shows another story. By age 29, we see a steady 3 percentage point increase in the graduation rate across all three cohorts. However, if we look at graduation rates by age 26, there is no increase between the NLSY79 young and the NLSY97 cohort. Together with the evidence in Figure 3.1, this shows that time to a bachelor's degree has increased over this period, a finding consistent with Bound et al. (2012).<sup>10</sup>

### *Wages*

We now turn our discussion from trends in the accumulation of experience to trends in the wage profiles associated with experience accumulation. The next set of tables and figures examine the evolution of wages profiles by experience and educational attainment.

In order to see if human capital investment has been rewarded, Table 3.3 examines the growth in full-time wages over the panel, broken out by different experiences and

---

<sup>10</sup> These trends also hold for graduation conditional on starting college.

educational attainment. [MORE HERE]

We also assess how wage premia associated with educational attainment have evolved over this time period. Table 3.5 shows the wage premia and dispersion associated with high school graduation, completion of some college, and college graduation across the three cohorts. The high school wage premium exhibits a U shape across the three cohorts, while the college wage premium exhibits a hump shape. The premium for completing some college has also fallen over time, most steeply in recent years. Our finding of a decreasing college wage premium between the NLSY79 young and the NLSY97 is unique (see Boehm, 2013; Castex and Dechter, 2014). However, it is robust to a number of different specifications, and we conclude that it is a feature of the data, and at the very least, is consistent with a slowdown in the growth of the college wage premium. Finally, we note that the evolution in the dispersion of wages varies by education group. Specifically, there has been increased wage dispersion for high school and college graduates, while there has been a tightening of the wages for high school dropouts. These findings are generally consistent with Juhn et al. (1993) and Goldin and Katz (2007), who conclude that wage dispersion has increased over time, especially for those in the upper parts of the distribution.

Each of the trends in this section has ignored the fact that selection is pervasive in the reduced-form wage profiles we have considered here. In the next section, we introduce the model that we use to account for the selection and endogeneity that are embedded in human capital investment decisions. In our final results, we present selection-corrected wage premia that tell a different story.

### *Demographics*

Before turning our attention to the econometric model, we conclude our discussion on descriptive trends in the data by focusing on how personal and family background characteristics have evolved across each of the NLSY cohorts.



We begin by discussing evolution in Armed Forces Qualifying Test (AFQT) scores.<sup>11</sup> Table 3.6 lists the change in the median AFQT score and its dispersion. The overall median AFQT score has shown a slight U shape over time, falling by 0.07 standard deviations and then rising by 0.08 standard deviations. However, this masks substantial heterogeneity by skill accumulation. The median AFQT score for college graduates has monotonically fallen, while AFQT for high school dropouts has monotonically risen. The distribution for those completing some college is largely the same across cohorts. However, for all other education groups, the variance in AFQT has increased over time. These results are consistent with the findings of Altonji et al. (2012), who find that the skill distribution has widened.

We continue our discussion on the evolution of demographics by examining the relationship between family background characteristics and educational attainment. This comparison is made in Table 3.7. Between the two NLSY79 cohorts there has been very little evolution in mother's education, whereas it has increased by about one grade level uniformly across educational groups between the NLSY79 young and NLSY97 cohorts. Increases in father's education have been highest among those with the lowest educational attainment. Of note is a sharp increase among the NLSY79 cohorts for high school dropouts, unmatched by any other educational group. Family income drops sharply between the two NLSY79 cohorts for all but high-school dropouts. However, across the NLSY79 young and NLSY97 there was a sharp increase in family income for some college and college graduates. Further, the difference in family income between high school dropouts and college graduates has also increased between the NLSY79 young and NLSY97 cohorts, by about 20%.

---

<sup>11</sup> The AFQT is a subset of the ASVAB. Specifically, AFQT scores are a weighted average of four ASVAB sub-tests: Arithmetic Reasoning (AR), Mathematics Knowledge (MK), Paragraph Comprehension (PC), and Word Knowledge (WK). In our model, we make use of six ASVAB sub-tests, the four in the AFQT as well as Coding Speed (CS) and Numerical Operations (NO). However, to maintain comparability with previous literature, we report evolution in the AFQT in this section.

Finally, female-headed households have become much more common over time, but almost exclusively for non-college graduates, who are 10 percentage points more likely to be in such a household. We emphasize that these stark differences among all these cohorts further motivate the fact that they should be treated separately.

Finally, we analyze the role of local labor market conditions in the human capital accumulation process (see Cameron and Heckman, 1998). Table 3.8 gives information about how our two county-level local labor market variables, “employment” rate and income per worker evolve over the life cycle.<sup>12</sup> There are striking differences between the NLSY97 and two NLSY79 cohorts at age 16. Over the 20 years separating these groups, the local labor market measures have grown remarkably. Over the life cycle, these measures have also increased for both of the NLSY79 cohorts, but decrease between ages 26 and 29 for the NLSY97. This result is likely due to the Great Recession, which occurred during the calendar years corresponding to this age range. Surprisingly, by age 29, the employment rate is identical across all three cohorts and the gap in income per worker has also narrowed. While we do not model endogenous migration patterns that affect these results, we do note that accounting for the general evolution of labor market conditions is important when modeling wages and human capital investment decisions over the life cycle.

The stark differences we see in schooling and work experiences (endogenous choices) as well as demographic, family, and local labor market characteristics (exogenous characteristics) are the prime motivation for our structural model in which we attempt to assess the extent to which endogenous decisions have influenced the evolution of wage returns to skills. Our model incorporates all the trends we have discussed.

---

<sup>12</sup> “Employment” rate is the number of employees reported by employers divided by population. Because individuals can hold more than one job, the numbers are much higher than the corresponding national employment-population ratio, which has ranged between 57% and 64% over the time period we consider.

### 3.4 The Model

Here we describe the underlying model of an agent’s choice of work and schooling activities over their life cycle. We use this model to account for the endogenous change across cohorts in levels of schooling and early work experience and to obtain selection-free estimates of wage returns to these experiences.

#### 3.4.1 Activities and Risk Sets

We assume that at each age  $a$  – which is measured in *months* in our case – individual  $i$ , who is a member of birth cohort  $c$ , chooses *activity*  $j$  from a *risk set* of activities, where the risk set at any point in time may vary with age and/or the occurrence(s) of one or more previous events. For simplicity, we suppress notation indexing the individual’s cohort. In practice, we estimate the model separately for each cohort  $c$ , so all the parameters should be understood as cohort-specific. Let  $R_{ia}$  denote the risk-set for individual  $i$  at age  $a$ , where we assume that there are  $K$  possible risk sets, i.e.,  $R_{ia} = r \in 1, \dots, K$ . Then, conditional on facing risk set  $R_{ia} = r$ , individual  $i$  chooses from among  $J^r$  activities at age  $a$ , where

$$d_{iaj}^r = \begin{cases} 1 & \text{if } i \text{ is in activity } j \text{ from risk set } r \text{ at age } a \\ 0 & \text{otherwise} \end{cases} \quad (3.1)$$

and  $\sum_{j=1}^{J^r} d_{iaj}^r = 1$ , for all  $i, a$  and  $r$ .

After the initial risk set ( $R_{ia} = 1$ ), we allow for *attainment-contingent* risk sets, i.e., some “attainment” activity has to occur in order to change the risk set. More formally:

$$R_{ia} = r \text{ iff } d_{i\tilde{a}j}^{R_{i\tilde{a}}} = 1 \text{ at some age } \tilde{a}, \tilde{a} < a, \quad (3.2)$$

for  $r > 1$ . In our case, the relevant activities are graduation from high school, which changes the risk set to  $R_{ia} = 2$  and graduation from college, which changes the risk

set to  $R_{ia} = 3$ . The three risk sets and the activities associated with each are given in Table 3.1.

### 3.4.2 School and Work Experience

We are interested in the effects of accumulated “experiences” on various outcomes in this model. In particular, we are interested in accumulated *years* of school attendance, years of work experience, etc., as well as educational attainments, such as high school and college graduation. The vector of experiences is given by:

$$\mathbf{x}_{ia}^r = \left( x_{1ia} \quad x_{2ia}^r \quad x_{3ia} \quad x_{4ia} \quad x_{5ia} \quad x_{6ia} \quad I_{ia}(R_{ia} < 3) \quad I_{ia}(R_{ia} = 3) \right)' \quad (3.3)$$

where the experience variables are:  $x_{1ia}$ , the number of years of schooling attendance as of age  $a$ ;  $x_{2ia}^r$ , the number of years of work and school experience in the relevant risk set  $r$ ;  $x_{3ia}$ , the total number of years of part-time (non-school) work as of age  $a$ ;  $x_{4ia}$ , the total number of years of full-time (non-school) work as of age  $a$ ;  $x_{5ia}$ , the number of years in the military as of age  $a$ ;  $x_{6ia}$ , the number spent in other activities as of age  $a$ ;  $I_{ia}(R_{ia} < 3)$ , an indicator equal to 1 if individual  $i$  has received a high school degree as of age  $a$ ; and  $I_{ia}(R_{ia} = 3)$ , an indicator equal to 1 if individual  $i$  has received a bachelor’s degree as of age  $a$ . For  $j = 1, 3, \dots, 6$ , the experience variables are accumulated since a starting age,  $a_0$ , and we use  $a_0 = 192$  (16 years old):

$$x_{jia} = \frac{1}{12} \sum_{\ell=a_0}^{a-1} d_{i\ell j}. \quad (3.4)$$

For  $j = 2$ , the experience term is either the number of years spent working in high school since  $a_0$  if in the first risk set,  $R_{ia} = 1$ , or it is the number of years spent working while in college or graduate school,  $R_{ia} > 1$ :

$$x_{jia}^r = \begin{cases} \frac{1}{12} \sum_{\ell=a_0}^{a-1} d_{i\ell j} & \text{if } R_{ia} = 1 \\ \frac{1}{12} \sum_{\ell=a_{HS_i}}^{a-1} d_{i\ell j} & \text{if } R_{ia} > 1 \end{cases} \quad (3.5)$$

where  $a_{HS_i}$  is the age of graduation from high school.

### 3.4.3 Wages

Let  $W_{iaj'}$  denote the potential hourly wage rate that  $i$  would realize at age  $a$  if he chose activity  $j'$ ,  $j' = 2, 3, 4$ . We assume that  $W_{iaj'}$  is determined by the individual's *human capital*, or *skills*,  $H_{ia}$  that he has as of the beginning of age  $a$ , measured in efficiency units; the occupation-specific skill price  $P_{aj'}$  per efficiency unit that varies across time and/or ages,  $a$ , across the local labor market in which  $i$  resides at age  $a$ ;<sup>13</sup> and idiosyncratic shocks, denoted by  $e^{\varepsilon_{iaj'}}$ , that are unanticipated by the individual:

$$W_{iaj'} = P_{aj'} H_{ia} e^{\varepsilon_{iaj'}}, \quad (3.6)$$

so that the log of wages, denoted by  $w_{iaj'}$  is the following linear function:

$$\begin{aligned} w_{iaj'} &= p_{aj'} + h_{ia} + \varepsilon_{iaj'} \\ &= w_{iaj'}^e + \varepsilon_{iaj'}, \end{aligned} \quad (3.7)$$

where  $p_{aj'} \equiv \ln P_{aj'}$ ,  $h_{ia} \equiv \ln H_{ia}$ , and  $w_{iaj'}^e \equiv p_{aj'} + h_{ia}$  is  $i$ 's *expected log wage* at age  $a$ , i.e., the wage that  $i$  expects to get if he chooses activity  $j'$ . We assume that  $p_{aj'}$  is the following function of age/time and the conditions of the local labor market in which  $i$  resides at age  $a$ ,  $\mathbf{m}_{ia}$ :

$$p_{aj'} = \beta_{0j'}^r + \beta_{\mathbf{m}} \mathbf{m}_{ia}. \quad (3.8)$$

And we assume that the (log of the) individual's stock of human capital,  $h_{ia}$ , is determined by some observed personal characteristics, e.g., one's birth year, race, etc., denoted by the vector  $\mathbf{z}_i$ , the individual's accumulated schooling and work experiences,  $\mathbf{x}_{ia}^r$ , and the individual's unobserved personal characteristics,  $\boldsymbol{\xi}_i$ , which is broken out into elements pertaining to the individual's cognitive ( $\xi_{i1}$ ) and non-cognitive ( $\xi_{i2}$ ) abilities:

$$h_{ia} = \beta_{\mathbf{z}} \mathbf{z}_i + \beta_{\mathbf{x}g} (\mathbf{x}_{ia}^r) + \beta_{\xi j'1} \xi_{i1} + \beta_{\xi j'2} \xi_{i2}. \quad (3.9)$$

<sup>13</sup> See Moretti (2011) for a survey of models of local labor markets.

It follows that

$$\begin{aligned}
w_{iaj'} &= w_{iaj'}^e + \varepsilon_{iaj'}, \\
&= \beta_{0j'}^r + \beta_{\mathbf{m}} \mathbf{m}_{ia} + \beta_{\mathbf{z}} \mathbf{z}_i + \beta_{\mathbf{x}} g(\mathbf{x}_{ia}^r) + \beta_{\xi j'1} \xi_{i1} + \beta_{\xi j'2} \xi_{i2} + \varepsilon_{iaj'} \quad (3.10)
\end{aligned}$$

where  $g(\cdot)$  contains: (i) a cubic polynomial in all types of accumulated experience, (ii) pairwise interactions between school experience and each of the work experience variables (work in school, part-time work and full-time work), and (iii) indicators for having graduated high school and for having graduated college (Heckman et al., 2006a). Note that schooling experience in  $g(\cdot)$  is the sum of school-only and work-in-school experience so as to be comparable to the literature originating with Mincer (1974).

One of our primary interests is in obtaining consistent estimates of the parameters in (3.10). This will in turn allow us to isolate the role played by skill prices in the change across cohorts in returns to schooling and early work experiences. As we make clear below, the central obstacle is that the elements of  $\mathbf{x}_{ia}^r$  are endogenous unless one conditions on the unobserved factors,  $\boldsymbol{\xi}_i$ . We now develop the nature of linkage through the sequences of activity choices individual  $i$  makes over his life cycle.

#### 3.4.4 Activity Choice Value Functions

Let the value function to individual  $i$  who is age  $a$  who engages in activity  $d_j^r$  be denoted by  $V_{iaj}^r$ . These value functions depend on the elements of the individual's information set at age  $a$ : personal characteristics,  $\mathbf{z}_{ia}$ , family background characteristics,  $\mathbf{f}_{ia}$ , local labor market characteristics,  $\mathbf{m}_{ia}$ , accumulated school and work experiences  $\mathbf{x}_{ia}^r$ , and the individual's unobserved personal characteristics,  $\boldsymbol{\xi}_i$ . For computational simplicity, we approximate the  $V_{iaj}^r$ 's for  $j = 1, \dots, J^r$  as a linear func-

tion of these characteristics:

$$\begin{aligned} V_{iaj}^r(\boldsymbol{\xi}_i) &= v_{iaj}^r(\boldsymbol{\xi}_i) + \omega_{iaj} \\ &= \boldsymbol{\alpha}_{\mathbf{f}_i}^r \mathbf{f}_i + \boldsymbol{\alpha}_{\mathbf{z}_i}^r \mathbf{z}_i + \boldsymbol{\alpha}_{\mathbf{m}_{ia}}^r \mathbf{m}_{ia} + \boldsymbol{\alpha}_{\mathbf{x}_{ia}, \mathbf{z}_i}^r b(\mathbf{x}_{ia}^r, \mathbf{z}_i) + \alpha_{\xi_{j1}}^r \xi_{i1} + \alpha_{\xi_{j2}}^r \xi_{i2} + \omega_{iaj}, \end{aligned} \quad (3.11)$$

where  $b(\cdot)$  contains: (i) a quadratic polynomial in all types of accumulated experience, (ii) linear interactions between race/ethnicity and each type of accumulated experience, and (iii) no indicators for educational attainment, since these are already embedded in the choice sets. Finally,  $\omega_{iaj}$  captures the idiosyncratic factors that affect the individual's value from choosing activity  $j$  at age  $a$ .

It follows that at each age  $a$ , individual  $i$  chooses activity  $j$  from among the activities in the current risk set,  $R_{ia} = r$  so as to maximize his utility:

$$j_{ia}^{r*} = \underset{k}{\operatorname{argmax}} V_{iak}^r, \forall r. \quad (3.12)$$

#### 3.4.5 Cognitive and Non-cognitive Ability

Our model incorporates two unobserved random factors representing unobserved cognitive and non-cognitive ability. To measure unobserved cognitive ability ( $\xi_{i1}$ ), we use six subject tests from the ASVAB, each of which has been normalized to correct for different test taking ages and test media similar to Altonji et al. (2009).<sup>14</sup> For each subject test  $s$ , the z-scored test score  $y$  for individual  $i$  is defined as a function of personal characteristics,  $\mathbf{z}_{ia}$ , family background characteristics,  $\mathbf{f}_{ia}$ , and the cognitive ability  $\xi_{i1}$

$$y_{is} = \gamma_{0s} + \boldsymbol{\gamma}_{\mathbf{f}_i} \mathbf{f}_i + \boldsymbol{\gamma}_{\mathbf{z}_i} \mathbf{z}_i + \gamma_{\xi_{s1}} \xi_{i1} + \eta_{is}, \quad (3.13)$$

where  $\eta_{is}$  captures idiosyncratic variation in test scores not related to the cognitive ability or test score determinants.<sup>15</sup> We include the observable characteristics  $\mathbf{z}_{ia}$

<sup>14</sup> The six subject tests we use are: Arithmetic Reasoning, Coding Speed, Mathematics Knowledge, Numerical Operations, Paragraph Comprehension, and Word Knowledge.

<sup>15</sup> The mean and standard deviation used to compute the z-scores are taken across all cohorts.

and  $\mathbf{f}_{ia}$  in this equation in order to capture, for example, bias in testing related to racial and family background differences.

There is little overlap in the measures of non-cognitive traits across the two NLSY surveys.<sup>16</sup> Due to this data limitation, we are unable to measure non-cognitive ability consistently across NLSY cohorts. Instead, we use the panel nature of the data to identify the non-cognitive ability factor  $\xi_{i2}$ . Thus this second factor is defined as all unobserved person-specific factors influencing the agent’s wage and decision process that are not in the clearly-defined cognitive factor.

### 3.5 Estimation

In this section we further characterize our econometric model and the strategy for estimating its parameters. In particular, we summarize the specification of the error structure of our model and the estimation procedures we employ. For now, we continue to ignore the three different cohorts—the NLSY79 old, the NLSY79 young and the NLSY97—although we allow for all of the parameters of our model to be cohort group-specific and explicitly examine the across-cohort differences in the marginal returns to schooling and work experience in wages and in counterfactual analyses of cross-cohort differences in wages below.

#### 3.5.1 Error Structure

We assume that  $\boldsymbol{\xi}_i$  is a person-specific vector of factors that is stochastically independent of the distributions of the observables,  $\mathbf{z}_i$ ,  $\mathbf{f}_i$ ,  $\mathbf{m}_{ia}$ , and of the unobservables,

---

<sup>16</sup> The NLSY79 contains the Rotter locus of control score and Rosenberg self-esteem scale for all individuals. These have been used in other studies as non-cognitive measures (Heckman et al., 2006b; Cunha et al., 2011). The NLSY97 does not collect information on any of these tests, but instead collects information on risky behavior such as school suspensions, sexual promiscuity and substance abuse. Aucejo (2014) uses school suspensions, fights, “precocious sex,” grade retention, and 8th grade GPA as non-cognitive measures.



$\boldsymbol{\omega}_{ia}$ ,  $\boldsymbol{\varepsilon}_{ia}$ , and  $\boldsymbol{\eta}_i$ , for all  $a$  and  $i$ .<sup>17</sup> At the same time, because the choice of past activities determine the accumulated experiences in  $\mathbf{x}_{ia}^r$  it is *not* the case that the elements of this vector are independent of  $\boldsymbol{\xi}_i$ , i.e.,

$$F(\mathbf{x}_i^r, \boldsymbol{\nu}_i) \neq f(\mathbf{x}_i^r)f(\boldsymbol{\nu}_i), \quad (3.14)$$

but

$$F(\mathbf{x}_i^r, \boldsymbol{\nu}_i | \boldsymbol{\xi}_i) = f(\mathbf{x}_i^r | \boldsymbol{\xi}_i)f(\boldsymbol{\nu}_i | \boldsymbol{\xi}_i), \quad (3.15)$$

where  $\boldsymbol{\nu}_i \equiv (\boldsymbol{\omega}_i + \boldsymbol{\xi}_i, \boldsymbol{\varepsilon}_i + \boldsymbol{\xi}_i, \boldsymbol{\eta}_i + \boldsymbol{\xi}_i)$ ,  $F(\cdot, \cdot)$  is the joint distribution function, and  $f(\cdot)$  is the marginal distribution function. We further assume that  $\boldsymbol{\xi}_i$  is mean zero and has identity covariance matrix. With respect to  $\boldsymbol{\omega}_{ia}$ ,  $\boldsymbol{\varepsilon}_{ia}$ , and  $\boldsymbol{\eta}_i$ , respectively, we assume that they are independently distributed both across and at each age,  $a$ , and have mean zero and constant variances. That the vector of activity shocks,  $\boldsymbol{\omega}_{ia}$ , are uncorrelated with  $\boldsymbol{\varepsilon}_{ia}$  is the result of assuming that decisions about activities are made before the actual realizations of wages are known by  $i$ .

### 3.5.2 Likelihood Function

We assume that the idiosyncratic errors in the activity payoff functions,  $\omega_{iaj}$ , have a Type I extreme value distribution so that the choice probability for this activity, conditional on  $\boldsymbol{\xi}_i$ , has the logistic form:

$$P_{iaj}^r(\boldsymbol{\xi}_i) = \frac{\exp(v_{iaj}^r(\boldsymbol{\xi}_i))}{\sum_{k=1, \dots, J^r} \exp(v_{iak}^r(\boldsymbol{\xi}_i))} \quad (3.16)$$

where  $v_{iaj}^r(\boldsymbol{\xi}_i)$  is the deterministic component of the value function, as defined in the first line of (3.11).

We assume that the idiosyncratic errors entering the wage function in (3.10) are normally distributed with zero mean and variance  $\sigma_{wj'}^2$  and its probability density

---

<sup>17</sup> Recall that  $\boldsymbol{\eta}_i$  is the vector of test scores for individual  $i$ . This vector is allowed to be correlated through the factor structure introduced in equation (3.13).

function is given by:

$$f_w(\boldsymbol{\xi}_i) = \frac{1}{\sigma_{wj'}} \phi \left( \frac{w_{iaj'} - \beta_{0j'}^r - \boldsymbol{\beta}_m \mathbf{m}_{ia} - \boldsymbol{\beta}_z \mathbf{z}_i - \boldsymbol{\beta}_x g(\mathbf{x}_{ia}^r) - \beta_{\xi j'1} \xi_{i1} - \beta_{\xi j'2} \xi_{i2}}{\sigma_{wj'}} \right), \quad (3.17)$$

$$j' = 2, 3, 4,$$

where  $\phi(\cdot)$  is the standard normal pdf.

We also assume that the idiosyncratic errors entering the ASVAB test score function in (3.13) are normally distributed with zero mean and variance  $\sigma_s^2$  and each probability density function is given by:

$$f_s(\xi_{i1}) = \frac{1}{\sigma_s} \phi \left( \frac{y_{is} - \gamma_{0s} - \boldsymbol{\gamma}_{fs} \mathbf{f}_i - \boldsymbol{\gamma}_{zs} \mathbf{z}_i - \gamma_{\xi s1} \xi_{i1}}{\sigma_s} \right). \quad (3.18)$$

It follows that the log likelihood function is given by:

$$\log \mathcal{L}(\boldsymbol{\theta}) = \sum_i \log \int \mathcal{L}_i(\boldsymbol{\theta} | \boldsymbol{\xi}_i) f_\xi(\boldsymbol{\zeta}) d\boldsymbol{\zeta} \quad (3.19)$$

where

$$\mathcal{L}_i(\boldsymbol{\theta} | \boldsymbol{\xi}_i) = \prod_s f_s(\xi_{i1}) \prod_a \prod_r \left[ \prod_{j^r=1,5,6,7} P_{iaj^r}^r(\boldsymbol{\xi}_i)^{d_{iaj^r}^r} \prod_{k^r=2,3,4} [P_{iak^r}^r(\boldsymbol{\xi}_i) f_w(\boldsymbol{\xi}_i)]^{d_{iak^r}^r} \right]^{I(R_{ia}=r)} \quad (3.20)$$

and where  $\boldsymbol{\theta} \equiv (\boldsymbol{\alpha}' \quad \boldsymbol{\beta}' \quad \boldsymbol{\gamma}')$ ,  $I(A)$  is the indicator function that is equal to one if  $A$  is true and zero otherwise, and  $f_\xi(\cdot)$  is the pdf of  $\boldsymbol{\xi}$ . In the analysis that follows, we assume that  $\boldsymbol{\xi}$  has a standard multivariate normal distribution with identity covariance matrix. Finally, the variance of the estimated parameters is recovered as the inverse of the estimated Hessian matrix, which has the desirable asymptotic properties for maximum likelihood estimators. In practice, we use quadrature to approximate the integral of the likelihood function using Gaussian quadrature with seven points of support for each dimension of the integral.

### 3.5.3 Identification

Here we informally discuss the identification of our model's parameters. Identification of the parameters relating observed outcomes to observed characteristics is straightforward. These parameters are identified by variation in the observed characteristics under various assumptions about the distribution from which the transitory errors are drawn. Identification of the parameters associated with the person-specific unobservables is less straightforward and requires more discussion.

To identify the role of unobserved effects, we cannot rely on instrumental variable techniques. Instrumenting for all the previous choices in a person's career would not be feasible in this framework, as such valid instruments would not exist. This further motivates the need for our structural model.

As discussed in Section 3.5.1, we assume stochastic independence of  $\boldsymbol{\xi}_i$  and the distributions of the observables, so that  $\boldsymbol{\xi}_i$  are random effects with normalized mean and variance. The factor loadings  $\boldsymbol{\alpha}_\xi$  and  $\boldsymbol{\beta}_\xi$  represent the variance of the factors relative to their normalized values. Because  $\boldsymbol{\xi}_i$  is vector-valued and appears in both the utility and wage equations, we impose three exclusion restrictions in order to be able to identify the factor loading parameters: (i) the factor  $\xi_{i2}$  does not enter the ASVAB test score equations; (ii) the population covariance between  $\xi_{i1}$  and  $\xi_{i2}$  is zero; and (iii) the vector of family background characteristics  $\mathbf{f}_i$  does not enter the wage equation (see Willis and Rosen, 1979; Taber, 2001; Hotz et al., 2002).

In order to aid the interpretation of the factor loadings, we measure the first factor (representing cognitive skills) by utilizing ASVAB test scores. The intuition proceeds as follows: for a given vector of observables and outcomes (observed decisions, wages, and ASVAB subject test scores), the factor loading for  $\xi_{i1}$  in each of these equations measures the permanent covariance in the residuals among these alternatives, net of the other factor  $\xi_{i2}$ . The second factor  $\xi_{i2}$  (representing non-cognitive skills) is

identified in a similar way, but because we have no common measures of non-cognitive skills across the two NLSY datasets, we use the panel nature of each dataset to identify this parameter vector. In this case, the factor loading on  $\xi_{i2}$  is identified from permanent covariance among outcome residuals holding fixed observables,  $\xi_{i1}$ , and transitory variation. Thus, individuals that have higher-than-expected outcomes over time conditional on observables and  $\xi_{i1}$  would have higher levels of  $\xi_{i2}$ . Because of our ability to find measurements of only  $\xi_{i1}$ , and not  $\xi_{i2}$ , we are unable to identify their covariance and thus instead restrict it to be zero.

We follow the previous literature (see Willis and Rosen, 1979; Taber, 2001; Hotz et al., 2002) in excluding family background characteristics  $\mathbf{f}_i$  from the wage equations. While not crucial to identification in our context (because our selection specification is dynamic and by definition relies on panel data, as opposed to static sample selection specifications), this exclusion restriction helps pin down the factor loadings by allowing the set of observables to differ between the choice equations and the wage equations.

### 3.6 Results

In this section we discuss the results of our estimation, which form the foundation of our decompositions. We begin by discussing the various model specifications that we use to evaluate our analysis. We then discuss the different components of the model, specifically the returns to degree completion in various specifications, the returns to experience with and without selection controls, and the returns to unobserved ability as measured by our factor loading estimates. Throughout we are interested in how the results differ across the three birth cohorts represented in our data.

### 3.6.1 *Specification of the model*

As discussed above, our full model allows us to estimate wage returns by accounting for the endogeneity of schooling and working choices early in the life cycle. The experience and graduation variables enter our model via  $g(\mathbf{x}_{ia}^r)$ , which is linear in educational attainment (high school and college completion), but non-linear in years of schooling and work experience. We compare this specification with other models, specifically the classic Mincerian model (see Mincer, 1974) and the flexible specifications introduced in Heckman et al. (2006a).

The classic Mincerian model allows for the wage to be a linear function of the number of years of schooling and a quadratic function of the number of years of potential experience, i.e. age. The biggest criticism of this model is the strict linearity in the schooling terms, especially when it comes to significantly higher wages that a college graduate receives versus someone who left college after three years. Heckman et al. (2006a) build on this fact extensively in their analysis. Specifically, they sequentially relax three key Mincerian assumptions. First, they relax the linearity assumption in schooling by estimating a model that includes indicators for each year of schooling. They show that doing this provides for very non-linear estimates in returns to education around degree completions, both high school and college. Second, they relax the quadratic assumption on potential experience in the same way, and they find that this does not produce significant changes. Third and finally, they relax the assumption of separability of earnings in schooling and experience by estimating the returns to potential experience separately for each schooling class (HS dropout, HS graduate, etc). They show that a significant bias exists if this assumption is not relaxed.

The specifications we consider rely on many of the underlying concerns that Heckman et al. (2006a) express. We first examine a Mincerian specification that

relaxes the linearity in schooling, similar to the second method proposed by Heckman et al. (2006a). However, rather than include indicators for each year of schooling, we only use indicators for high school and college completion, and use a cubic in schooling for the rest of the observations. Including these graduation indicators captures many of the non-linearities in schooling, and a flexible cubic is sufficient for capturing the rest.

We next consider a specification that relaxes all the Mincerian assumptions, although in a slightly different way than in Heckman et al. (2006a). First, we use *actual* work experience rather than *potential* work experience. In addition to indicators on graduation events and a cubic in schooling experience, we add a cubic in each type of work experience (in-school, part-time, full-time) and we add an interaction of schooling experience with each type of work experience. These steps allow for more flexibility than a simple quadratic in age and address the separability assumption without requiring separate estimation procedures for each type of schooling class. Finally, to account for some selection on observable characteristics, we add in background characteristics such as race, nativity and birth year.<sup>18</sup>

Using this as our baseline model, we go one important step further by controlling for *unobservable* selection. As mentioned previously, we account for this by including random factors representing unobserved cognitive and non-cognitive ability and jointly estimating our structural model, which includes the wage, choice and ability equations. Instrumenting for all the previous choices in a person's career would not be feasible in this framework, as finding such valid instruments would be difficult. This motivates the need for our structural model.

We examine the results of these different specifications in Tables 3.9 and 3.10. These results are in the form of marginal effects. For the graduation dummies, these

---

<sup>18</sup> We cannot include ASVAB nor interact it with educational attainment because ASVAB is our factor measure and doing so would eliminate our identification of the factor.

are simply the estimate coefficients, since they enter the model linearly. However, for the accumulated experience variables  $\mathbf{x}_{ia}^r$  (schooling, work, military and other) that enter the model linearly, we calculate the marginal effect on the full-time wage of an additional unit of experience  $k$ :

$$g'_k(\mathbf{x}_{ia}^r) = \frac{\partial w_{ia4}(\mathbf{x}_{ia}^r)}{\partial x_{kia}^r}, \quad (3.21)$$

where wage is subscripted by  $j' = 4$  to denote we are examining full-time wages. Further, since this is a function of the experience terms, we need to choose a point of evaluation, which for this analysis is the average experience vector at age 29,  $\bar{\mathbf{x}}_{29}^r$ .<sup>19</sup> We use this age because (i) it is an age by which most people have completed schooling, and (ii) it is the last observation in our panel.<sup>20</sup>

### 3.6.2 Sheepskin effects

We start by discussing the sensitivity of the so-called “sheepskin effect” of degree completion to our different specifications, listed in Table 3.9. What we see in general are hump shaped patterns, with college sheepskin being highest for NLSY79 young and lowest for NLSY79 old. Also, we see high school sheepskin being lowest for NLSY79 young and highest for NLSY79 old. Further, these sheepskin effects become smaller and more similar across cohorts as we introduce more controls.

We now discuss the different specifications. In the first two columns of the table, we compare our baseline Mincerian approach to the unadjusted raw premiums for finishing high school and college. Unsurprisingly, we immediately see a decrease in all sheepskin effects as we introduce controls for experience. This decrease is strongest for the NLSY97.

---

<sup>19</sup> The full estimation results are available from the authors upon request.

<sup>20</sup> Estimating the returns even later in life would be interesting, but is not feasible for us given the data limitations.

The third and fourth columns of Table 3.9 report our results from specifications that are similar to Heckman et al. (2006a). As mentioned previously, these specifications include actual (instead of potential) experience, cubics in each type of experience, and linear interactions between years of schooling and each type of work experience. Column 4 further adds personal background characteristics: birth year, nativity status, and race/ethnicity. In each of these subsequent specifications, the sheepskin effects diminish markedly. Specifically, the college sheepskin effect decreases by 5 to 10 log points when introducing these controls. This is equivalent to a 15% to 35% decrease in the effect. Further, adding these controls consistently reduces the cross-cohort gap in the sheepskin effects.

Finally, our last column includes adds unobserved heterogeneity by jointly estimating the wage equation with the choice and ability equations. Foremost among the results in this column is the fact that the college sheepskin effect fell from 12 log points to 4 log points within the NLSY79. This finding highlights the importance of accounting for unobserved heterogeneity in estimating these effects. This role of unobservables is weaker in the more recent cohort, indicating that unobservables may be becoming less important in describing the college sheepskin effect.

In summary, as has been shown, and as can be expected, the more one controls for selection, the lower the sheepskin effects become. Thus, while the sheepskin effects are still large and significant, much of the evolution in wage returns to skills is due to other factors, which we discuss next.

### *3.6.3 Returns to experience*

We continue by discussing the evolution across NLSY cohorts in the returns to various forms of human capital, under varying assumptions about selection on unobservables. As mentioned, we evaluate the return at the marginal effect for average experience levels by age 29,  $g'_j(\bar{\mathbf{x}}_{29}^r)$ . These returns can be found in Table 3.10. These esti-



mates are calculated from our full model specification, which regresses the log wage on background characteristics, local labor market conditions, and demographic variables, as well as the experience terms. Panel (a) shows the results from a wage equation specification with no selection on unobservables, whereas panel (b) shows the estimates after controlling for selection by jointly estimating the wage equation with our choice model and ability equation. Note that the last two variables in each panel are the educational attainment variables from the last two columns of Table 3.9 and are reproduced here for completeness.

The first three elements involve the returns to schooling, where the first line refers to the returns to any type of schooling, and the next two refer to the *additional* returns to working while in high school and college, respectively. After controlling for selection, we see downward trends in the wage returns over time for schooling experience, as well as in the additional returns to working while in college, with a highly variable U-shaped trend for the additional returns to working while in high school.

Comparing these trends to each other allows us to answer one of our main research questions, specifically what are the trends in the wage returns to in-school work experience. At one extreme is the NLSY79 young cohort, where the additional returns to working while in high school are negative, thus a year of high school work gives only a 1% increase in future wages, versus almost a 4% increase for pure schooling. However, for the other cohorts the additional returns to an additional year of work in high school are positive, resulting in a 7% and 5% increase in later wages for the NLSY79 old and NLSY97 cohorts, respectively.

Performing a similar analysis for working or not in college, we see a different trend. All additional returns are positive, with very similar returns between the NLSY79 young and NLSY97. However, the returns to an additional year of college work are very large for the NLSY79 old at about 6.5%, which is on top of the

more than 5% that all forms of schooling result in. These results are somewhat consistent with previous literature. Specifically, Hotz et al. (2002) examine only the NLSY79 young cohort and find the same results for working while in high school, namely a smaller return than pure schooling. However, while they also find a smaller return for working while in college than pure schooling, we find a larger effect. This can be explained by our model having two factors and having a slightly different specification.

The other experience terms we are interested in are the full-time and part-time work experience variables. Not surprisingly, the returns to full-time work experience are always higher than those to part-time work experience. For full-time work experience, there is little change over time, with all cohorts having about a 3% return to an additional year of full-time work, with or without selection controls. However, one item of interest is the importance of controlling for selection for part-time work experience. In panel (a) of Table 3.10, we find negative returns to part-time experience on full-time wages if we ignore selection. However, recognizing the role of selection in panel (b) shows that, while still negative, the effects are diminished. This finding is important, especially in light of recent discussion surrounding the detrimental impact of underemployment in the Great Recession. While these workers are still worse off than had they remained fully employed, the losses are not as profound as previously thought.

Recall that in section 3.3.2, we showed that the level of in-school work and time spent in school has increased at the expense of full-time work experience, and that the incidence of high school and college graduation has also increased. In this section, we have shown that the wage returns to these skills have decreased. In order to isolate the specific sources of these trends, we need to do a decomposition. The method and results of our decomposition are discussed in the next section.

### 3.6.4 *Factor loadings*

Before discussing the wage decompositions, it is useful to examine the contribution of the unobserved factors to the wages of young men. Table 3.11 contains both the cognitive and non-cognitive factor loading estimates for the full-time wage equation in each cohort. Recall that the distribution of the factors is multivariate normal with mean zero and identity variance. Thus the interpretation of the estimates is the change in log wages (percent change in wage) due to a one standard deviation increase in the unobserved ability, holding fixed all observable characteristics and the other dimension of unobserved ability.

Our main finding is that the cognitive loading, at about 14-17 log points, provides a higher return to skill than the non-cognitive loading, which is around 11-12 log points. Across cohorts, the returns for each factor are lowest for the NLSY79 young cohort. The cognitive factor loading is highest for the NLSY79 old. Our results differ from Castex and Dechter (2014), who also look at the returns to ability between the NLSY79 and NLSY97 but find that the returns to AFQT have diminished greatly between the two. Our wage specification controls for selection in the wage equation that theirs does not account for. Our finding that the returns to unobserved ability (as measured by ASVAB components) has increased between the NLSY79 young and the NLSY97 can be explained by selection.

## 3.7 Decompositions

In this section we use the parameter estimates of the model to estimate a set of counterfactual analyses. We assess the relative importance of the changes in prices of skills versus changes in the composition of skills across the three NLSY cohorts in accounting for the observed differences in the wage premia to these skills. The key feature of our decomposition approach is that, unlike previous studies, the estimates

our model produces allow us to account for the endogenous nature of the changes in educational attainment and work experience for each of the cohorts.

### 3.7.1 Setup

This section describes the intuition of our approach. Our decomposition is a more detailed version of Oaxaca (1973) and fits under the classification of decompositions discussed in Fortin et al. (2011). Our approach differs from Oaxaca (1973) in two important ways: (i) we decompose the difference in wage *premia* to skills across different groups (e.g. birth cohorts) rather than the average difference in wage *levels* across groups (e.g. males and females); and (ii) we allow for a more flexible form of wages and impose more structure on the joint distribution of unobservables. This approach allows us to decompose our differences of interest into both observed and unobserved components, and direct and indirect component.

In accounting for the endogeneity of skill accumulation, we are also able to explain the evolution in wage premia in terms of skill accumulation *directly* related to the wage premium of interest versus skill accumulation *indirectly* related. For instance, it may be the case that students who are more likely to graduate college also accumulate more in-school work experience, so that some of the wage premium to college graduation reflects the premium to in-school work experience. Our definition of the wage premium allows us to incorporate both of these components. Further, we define the wage premium as the difference in the conditional expectation of wages, where we evaluate the expectation at different discrete points of the observable characteristics (i.e. the mean of the skill plus one unit versus the mean of the skill), rather than at the traditional marginal effect in equation (3.21).

For simplicity, we introduce some reduced notation to help motivate the analysis. Formal mathematical notation can be found in Section B.2. Let  $w$  be our outcome of interest (e.g. log wage),  $s$  be our covariate of interest (e.g. full-time work experience,

HS graduation status),  $x$  be a representative correlated covariate (e.g. schooling experience, local employment rates) and  $\beta$  be the vector of estimated parameters of the structural model.

By definition, the overall wage premium is simply the difference in the expected wage evaluated at different points:

$$\Delta_s E(w | s) = \begin{cases} E(w | s = \tilde{s} + 1) - E(w | s = \tilde{s}) & \text{if } s \text{ is continuous} \\ E(w | s = 1) - E(w | s = 0) & \text{if } s \text{ is binary,} \end{cases} \quad (3.22)$$

where  $\tilde{s}$  is a specific value of our variable of interest. These expected wages can be calculated directly from the data. The intuition of our writing of the wage premium into direct and indirect effects follows from the Law of Iterated Expectations:  $E(w | s) = E(E(w | s, x) | s)$ . However, to perform our decompositions, we use structural wage parameters, which means that the expected wages should be interpreted as predicted wages. Further, adding  $\beta$  into the conditioning allows us to more easily show the nested expectations:

$$\begin{aligned} E(w | s; \beta) &= E(E(w | s, x; \beta) | s) \\ &= E(s\beta_s + x\beta_x | s) \\ &= s\beta_s + E(x | s)\beta_x \\ &= \underbrace{\hat{w}(s; \beta)}_{\text{direct effect}} + \underbrace{\hat{w}(\bar{x}(s); \beta)}_{\text{indirect effect}}, \end{aligned} \quad (3.23)$$

where  $\hat{w}$  is the predicted wage as a function of the data and the estimated structural parameters, and  $\bar{x}(s) \equiv E(x | s)$  is the expected level of the correlated covariate conditional on  $s$ . The last two terms are the direct and indirect effects of interest, and we will use the  $\hat{w}$  notation going forward representing a wage that is predicted (i.e. a function of structural parameters) rather than expected (i.e. calculated from the raw data).

The *direct* effect of  $s$  on the predicted wage is:

$$\Delta_s \hat{w}(s; \beta) = \begin{cases} \hat{w}(s = \tilde{s} + 1; \beta) - \hat{w}(s = \tilde{s}; \beta) & \text{if } s \text{ is continuous} \\ \hat{w}(s = 1; \beta) - \hat{w}(s = 0; \beta) & \text{if } s \text{ is binary.} \end{cases} \quad (3.24)$$

If  $s$  is continuous (school, work, military, and other experience), this is generally the average experience evaluated at a certain point, for example a certain age.<sup>21</sup> If  $s$  is discrete,  $\tilde{s}$  is just the value at that same point. For example, if our variable of interest were full-time work experience, then  $\Delta_s \hat{w}$  would be the difference in predicted selection-corrected wages at a certain age given  $\tilde{s} + 1$  years of full-time work experience versus  $\tilde{s}$  years of experience. Likewise, if our variable of interest were college graduation, the direct wage premium would be the difference in predicted selection-corrected wages between those who have graduated college by a certain age and those who have not.

The evolution in the wage premia related to *indirect* skill accumulation is measured by conditioning the skill correlates on the variable of interest, as shown in equation (3.23):

$$\Delta_s \hat{w}(\bar{x}(s); \beta) = \begin{cases} \hat{w}(\bar{x}(s = \tilde{s} + 1); \beta) - \hat{w}(\bar{x}(s = \tilde{s}); \beta) & \text{if } s \text{ is continuous} \\ \hat{w}(\bar{x}(s = 1); \beta) - \hat{w}(\bar{x}(s = 0); \beta) & \text{if } s \text{ is binary,} \end{cases} \quad (3.25)$$

Continuing with the example of college graduation, we consider all other components of the wage equation to be covariates correlated with college graduation. The indirect college wage premium is then the wage premium arising from the fact that, for example, those who have graduated from college at a given age have higher levels of in-college work experience but lower levels of full-time work experience. The indirect premium is simply the difference in the the predicted selection-corrected wages between the correlated skills of college graduates and non-college graduates.

The *total* observed wage premium can be thought of as either the difference between the total predicted wage premium at  $\tilde{s} + 1$  and  $\tilde{s}$ , or, more simply, the sum

<sup>21</sup> As explained in Section 3.7.2, the age we use is 29, which is the last observation in the data.

of the direct and indirect effects:

$$\Delta_s \hat{w}(s, \bar{x}(s); \beta) = \Delta_s \hat{w}(s; \beta) + \Delta_s \hat{w}(\bar{x}(s); \beta). \quad (3.26)$$

Finally, in addition to the observed skills above, we also have unobserved skills,  $\xi$ , assumed to be correlated with  $s$  through the selection equations in the choice model. Because we don't observe  $\xi$ , we take the difference between the wage premia estimated with and without controlling for selection to pin down the effects of the unobserved skill. Thus, the *total* premium attributable to unobservable correlated covariates is a double difference:

$$\Delta_\beta \Delta_s \hat{w}(\xi(s), s, \bar{x}(s); \beta_r, \beta) = \Delta_s \hat{w}(s, \bar{x}(s); \beta_r) - \Delta_s \hat{w}(s, \bar{x}(s); \beta), \quad (3.27)$$

where  $\beta_r$  refers to the parameter estimates that appear in our reduced-form specification (i.e. the specification in the second to last column of Table 3.9).

We now discuss how to use these terms to create counterfactual wage premia that enter our decomposition using the covariates from one cohort and the parameters for another. For arbitrary cohorts  $A$  and  $B$ , this would be:

$$\Delta_s \hat{w}(s^A, \bar{x}^A(s^A); \beta^B) = \Delta_s \hat{w}(s^A; \beta^B) + \Delta_s \hat{w}(\bar{x}^A(s^A); \beta^B). \quad (3.28)$$

This notation allows us to express mathematically the different components of the decomposition.

Following the discussion in Fortin et al. (2011), we decompose the overall difference across cohorts in wage premia to various skills into four main components, one of which is divided into a direct and indirect effect as explained above:

1. Observed price effect: holds fixed everything in the two cohorts except the observable return to the skill of interest

$$= \Delta_s \hat{w}(s^A, \bar{x}^A(s^A); \beta^B) - \Delta_s \hat{w}(s^A, \bar{x}^A(s^A); \beta^A),$$

which is divided into two components:

- (a) Direct effect: holds fixed everything in the two cohorts except the observable *direct* return to the skill of interest

$$=\Delta_s \hat{w}(s^A; \beta^B) - \Delta_s \hat{w}(s^A; \beta^A)$$

- (b) Indirect effect: holds fixed everything in the two cohorts except the observable *indirect* return to the skill of interest (e.g. through returns to correlated skills)

$$=\Delta_s \hat{w}(\bar{x}^A(s^A); \beta^B) - \Delta_s \hat{w}(\bar{x}^A(s^A); \beta^A)$$

2. Observed composition effect: holds fixed everything in the two cohorts except the observable composition of the skill of interest and correlated skills

$$=\Delta_s \hat{w}(s^B, \bar{x}^B(s^B); \beta^B) - \Delta_s \hat{w}(s^A, \bar{x}^A(s^A); \beta^B)$$

3. Unobserved price effect: holds fixed everything in the two cohorts except the unobservable return to all skills

$$=\Delta_\beta \Delta_s \hat{w}(\xi(s^A), s^A, \bar{x}(s^A); \beta_r^B, \beta^B) - \Delta_\beta \Delta_s \hat{w}(\xi(s^A), s^A, \bar{x}(s^A); \beta_r^A, \beta^A)$$

4. Unobserved composition effect: holds fixed everything in the two cohorts except the unobservable composition of all skills

$$=\Delta_\beta \Delta_s \hat{w}(\xi(s^B), s^B, \bar{x}(s^B); \beta_r^B, \beta^B) - \Delta_\beta \Delta_s \hat{w}(\xi(s^A), s^A, \bar{x}(s^A); \beta_r^B, \beta^B)$$

With these terms, we then compute the decomposition. The observed components of the decomposition are calculated using the structural wage parameter estimates,  $\beta$ . The unobserved components are then calculated as the difference in the observed components evaluated at the structural,  $\beta$ , and the reduced-form wage parameter estimates,  $\beta_r$ . These components are contrasted with the overall change in the wage premia in equation (3.22), in the spirit of Oaxaca (1973).<sup>22</sup>

<sup>22</sup> As mentioned in Oaxaca (1973), inherent in this decomposition is the “index number problem,” where the results of the decomposition depend on the baseline group. Following Oaxaca (1973), we average across both groups when presenting our final decomposition results. That is, we calculate the counterfactual wages in two ways: (i) using the characteristics of cohort B and the parameters of cohort A; and (ii) using the characteristics of cohort A and the parameters of cohort B. We then average across the two.



### *Example*

To build intuition for our decomposition, we discuss a specific example (the college wage premium) and identify each component of our decomposition, where we decompose the evolution in the college wage premium between cohort B and cohort A. The observed direct price effect is what would traditionally be termed as the return to college graduation. If this return were positive, then the direct price effect states that a college degree by itself (i.e. holding everything else fixed) is rewarded more in cohort B than cohort A. However, if the *indirect* price effect were negative, then this would mean that the skills that college graduates have (aside from the college degree) are rewarded by less in cohort B than cohort A.

The unobserved price effect is calculated in a similar way to the observed price effect. However, instead of just looking at the difference in the prices (both direct and indirect), we consider how this difference compares between estimates corrected for selection and estimates not corrected for selection. Thus, the unobserved price effect is the predicted wage evaluated at the difference in the differences of the parameter vectors.

The observed composition effect looks at the evolution across cohorts in the difference in the skills that college graduates invested in versus the skills that non-college graduates invested in. To evaluate how much these skills are worth, we use one price across both cohorts and multiply the cross-cohort difference in observable characteristics by this price. The composition effect then states that cohort B chose skills that, on average, were worth 3% more than those chosen by cohort A.

To evaluate the role of unobserved composition, we compare the observed composition effect evaluated at two different parameter vectors: structural and reduced form. The difference between these two measures the role of unobservable composition.

### 3.7.2 Results

In this section we discuss the results of the decompositions and their relation to the research questions posed at the beginning of this paper. Tables 3.12 through 3.14 present these results in six columns. The far right hand column reports the observed evolution in the wage premium found in the raw data, and the five other columns show the five components of our counterfactual exercise as explained in section 3.7.1 and calculated in equation (3.28). Each component of the table can be understood as the counterfactual wage premium for the given component of our decomposition. We emphasize that the second column, “direct observed price,” is equivalent to the wage return to the corresponding skill. As such, we henceforth use the term “wage return” to reference this object. Additionally, because our decomposition uses a model, the sum of each component does not generally equal the premium in the data. We emphasize that this discrepancy arises from differences in statistical power between the data and the model, and *not* because the model does not fit the data.

We now discuss our pertinent findings for the three major research questions we consider: What is the relative importance of changes in skill price versus skill composition? What are the changes across cohorts in the wage returns to in-school work experience? How much of the college wage premium evolution actually reflects the increase in the in-school, and more generally early work experience?

We first discuss how the relationship between skill prices and skill composition has evolved. The objective of our decomposition analysis is to recover the size of both price and composition effects for each skill of interest.<sup>23</sup> Our primary finding is that both price and composition effects matter, but that their relative importance varies

---

<sup>23</sup> In his seminal work, Oaxaca (1973) finds price and composition effects that are always the same sign as each other and as the gender wage gap and as such simply presents each effects’ percentage of the gap. However, as is evident in Tables 3.12, the effects that we find both frequently differ in sign as well as have very large magnitudes. This happens both along the price and composition dimension as well as the observed and unobserved dimension. This high level of variation underscores the importance of accounting for more than just price effects or more than just observed effects.

depending on the skill. For instance, we find that, for college graduation, composition effects were small relative to price effects. In contrast, composition effects are much larger relative to price effects in explaining the evolution of the wage premium for working while in college between the NLSY79 young and the NLSY97.<sup>24</sup>

We now discuss how the the wage returns to in-school work experience have evolved. These results are listed in Table 3.12.<sup>25</sup> While the evolution in the overall wage premia for working in school have exhibited a hump shape, the wage returns to these skills have actually evolved quite differently. Comparing the direct price column for working in high school in panels (a) and (b) of Table 3.12 shows that the returns have actually evolved in a U shape, with a sharp increase between the NLSY79 young and the NLSY97 (6.8 log points versus -2.9 log points). For working in college, the return has decreased monotonically, but at a decreasing rate (-3.4 log points within the NLSY79 cohorts and -1.3 between the NLSY79 young and the NLSY97). In contrast, the gross wage returns to schooling (either with or without simultaneous work) have increased over time, but at a decreasing rate (6.0 log points within the NLSY79 cohorts and 2.3 between the NLSY79 young and the NLSY97).

The differences in the evolution between the observed premium and wage return are heterogeneous for in-school work. We find substantial magnitudes in the other components of our decomposition that overshadow the evolution in the wage returns. For example, the evolution in the wage premium for working while in high school exhibited large positive observed and unobserved composition effects in the NLSY79,

---

<sup>24</sup> Other findings of importance relate to the role of in-school work experience as well as the evolution of the college wage premium. For instance, we find negative indirect price effects over time for all skills except full- and part-time work. In contrast, we find that the composition effects for schooling-related activities were generally positive within the NLSY79 cohorts, but negative between the NLSY79 young and NLSY97.

<sup>25</sup> Similar results, without using our Law of Iterated Expectations approach, are discussed in Section 3.6.3. The overall trends between the two sections are the same with the exception of the returns to years of schooling because of the non-linearities of our model induced by the interaction between schooling and work experience.

but large negative unobserved price effects between the later two cohorts. The returns to working while in college are characterized by a large negative indirect price effect between the NLSY79 cohorts which was offset by even larger unobserved price and unobserved composition effects. The negative evolution in this premium across the latter two cohorts was driven almost exclusively by observed composition effects.

We now assess how much of the evolution in the college wage premium actually reflects the evolution in the in-school, and more generally early work experience. We find very small observed composition effects, as shown in Table 3.14.<sup>26</sup> Although the levels of in-college work experience increased substantially over this time period, the levels of full-time work experience decreased (because the two are substitutes; see Figure 3.1). The return to working in college decreased over time (see Table 3.12), which by itself would have led to a negative composition effect. However, the changes in full-time work would bring about a slightly positive composition effect. On the whole, after all skill correlates are considered, the negative impact of the rising level of in-college work experience is canceled out.

Finally, we discuss other findings from our decompositions. One interesting finding surrounding the college wage premium relates to the unobserved effects. Specifically, there is a very large increase in the college wage premium between the NLSY79

---

<sup>26</sup> The raw differences in the data shown in the final column are the overall changes in the “sheep-skin” effects for high school and college. Note, that these are slightly different than those found in Table 3.5, where we explicitly separate high school graduates with no college from those with some college. In this table and in our estimation, we group all high school graduates without a bachelor’s degree together. However, while the magnitudes are slightly different, the trends are very similar, both following U-shaped paths. Specifically, the raw high school wage premium went down between the NLSY79 cohorts, and then went back up for the NLSY97, while the overall college wage premium went up between the NLSY79 cohorts, and then back down for the NLSY97. Further, since these variables only enter the model linearly, there is no difference between the discrete calculation we performed here to calculate the observed price effect compared with that we presented in Table 3.10. Thus, the direct returns here in Table 3.14 are exactly the same as the returns found in the last two rows of the last two columns of Table 3.10. Finally, we see that these direct returns follow the same path as the raw differences in the data, albeit of differing magnitude. For example, the evolution among the NLSY79 cohort shows a change in the return to high school and college of -4.8% and 3.8%, respectively. These mirror the overall raw differences of -3.5% and 12.3%.

cohorts, due to the large, positive unobserved effects. In fact, the observed effects in aggregate are negative, although the direct return is positive. This finding of a large, positive role of both unobserved price and composition is very consistent with Taber (2001), who finds that the increase in the demand for unobserved ability was an important component of the increase in the college wage premium. Taber (2001), however, does not allow for composition effects.

For the high school wage premium, the observed price effects are both negative and large between the NLSY79 cohorts, with a 4.8% decrease in the return to the degree as well as a 4.2% decrease in the indirect return from the correlated skills. Offsetting this, however, is a 4.7% *increase* in the overall composition effect. This seems to indicate that individuals found themselves in “higher-value” activities. Further, comparing the NLSY79 young and NLSY97 cohorts reveals a similar trend in the indirect price and total composition effects, although the direct return is 2.6%. Finally, the net result of the unobserved effects is quite small for both, although each type of unobserved effect is fairly large.

### 3.8 Conclusion

In this paper, we examine what the changes are across cohorts of young men in the wage returns to various activities and attributes. Specifically, we examine the evolution of the returns to in-school work experience and to obtaining a college education, and how early-life choices influence later life outcomes, such as graduation rates and wages. We do this by separating the overall returns to experiences into price and composition effects.

Using a dynamic model of schooling and work decisions, we estimate the returns to various forms of experience, separately for three different cohorts of two NLSY panels. We find that the relative importance of the price and composition effects varies dramatically across skills. Regarding in-school work experience, we find that

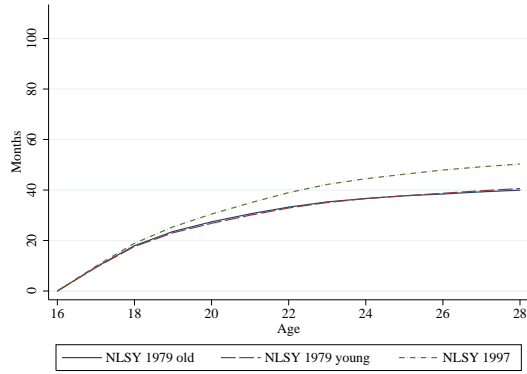
the direct returns to working while in college have decreased over time, with the decrease early on (between NLSY79 cohorts) due mostly to price effects, and the latter decrease (between NLSY79 young and NLSY97 cohorts) due mostly to composition effects. Further, composition effects explain little in the evolution of the college wage premium. Considering that we find both a significant increase in the incidence of in-college work over time and a decreasing direct wage return of in-college work, the negative impact this change had on the composition effect is offset by a positive net impact of the remaining skill correlates. Finally, and consistent with other studies (e.g. Taber, 2001), we find that almost all the increase in the college wage premium in the 1980s is due to a change in the returns to and composition of unobserved skills.

Our findings underscore the importance of the role of composition effects in the long-term evolution of the U.S. labor market. Our study also further supports the need to appropriately account for dynamic selection in analyzing wage returns to early-career schooling and work experience (Cameron and Heckman, 1998, 2001).

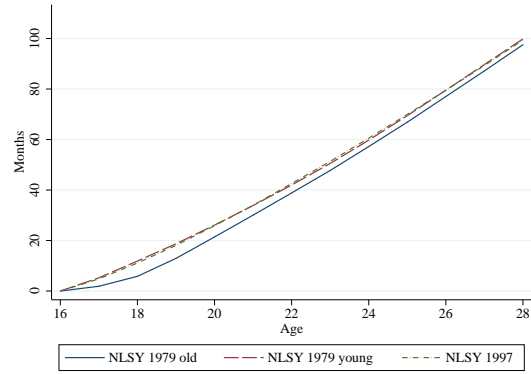
## Figures and Tables

Table 3.1: Risk sets and activities

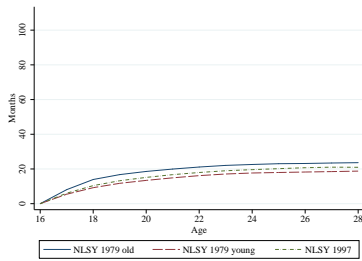
Activity ( $j^r$ )	Description
$R_{ia} = 1$ (Pre-High School Graduate):	
1	School only, no HS diploma or GED
2	Work in school, no HS diploma or GED
3	Work PT (no school), no HS diploma or GED
4	Work FT (no school), no HS diploma or GED
5	Military, no HS diploma or GED
6	Other, no HS diploma or GED
7	Graduate from HS at age $a$ (Attainment Activity)
$R_{ia} = 2$ (High School Graduate):	
1	School only, HS diploma or GED
2	Work in school, HS diploma or GED
3	Work PT (no school), HS diploma or GED
4	Work FT (no school), HS diploma or GED
5	Military, HS diploma or GED
6	Other, HS diploma or GED
7	Graduate with bachelor's degree at age $a$ (Attainment Activity)
$R_{ia} = 3$ (College Graduate):	
1	School only, bachelor's degree
2	Work in school, bachelor's degree
3	Work PT (no school), bachelor's degree
4	Work FT (no school), bachelor's degree
5	Military, bachelor's degree
6	Other, bachelor's degree



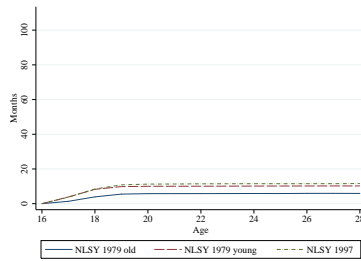
(a) Total school experience



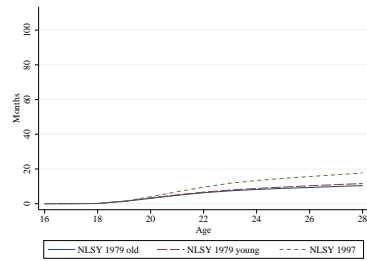
(b) Total work experience



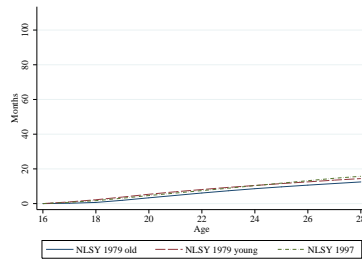
(c) School only



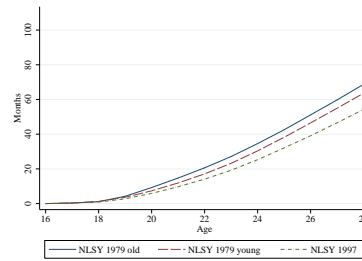
(d) Work in HS



(e) Work in college



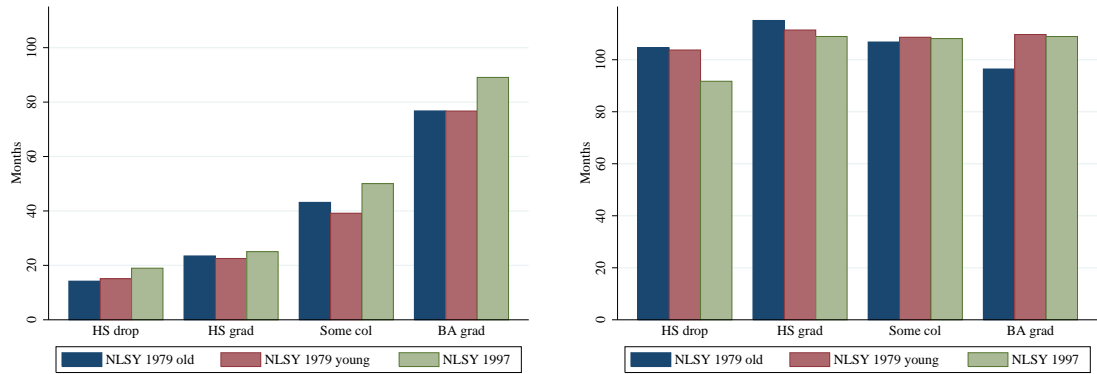
(f) Work PT (no school)



(g) Work FT (no school)

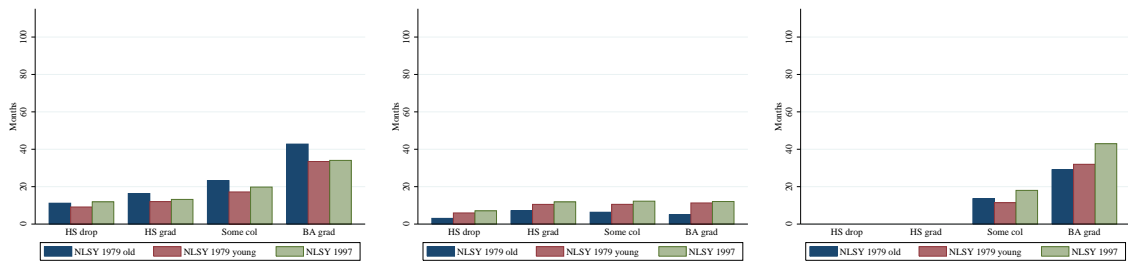
FIGURE 3.1: Distributions of average experience by age





(a) Total school experience

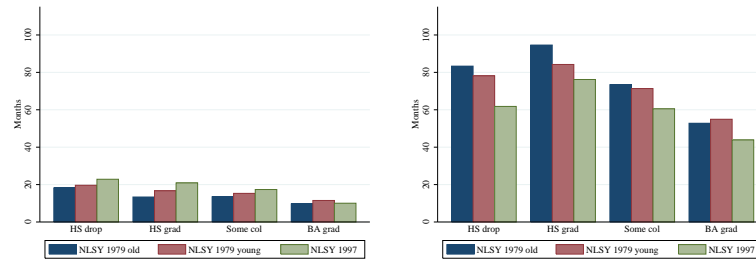
(b) Total work experience



(c) School only

(d) Work in HS

(e) Work in college



(f) Work PT (no school)

(g) Work FT (no school)

FIGURE 3.2: Distributions of average end-of-panel experience by final educational attainment

Table 3.2: Graduation probabilities by age

(a) Age 26					
Variable	79o	79y	97	$\Delta$ 79y-79o	$\Delta$ 97-79y
Pr(grad HS)	0.87	0.88	0.90	0.00	0.02**
Pr(start col)	0.55	0.59	0.64	0.04**	0.05***
Pr(grad BA)	0.19	0.22	0.22	0.03*	0.01
Pr(grad BA   start col)	0.35	0.37	0.35	0.02	-0.02
<i>N</i>	1,099	2,456	3,607		
(b) Age 29					
Variable	79o	79y	97	$\Delta$ 79y-79o	$\Delta$ 97-79y
Pr(grad HS)	0.87	0.89	0.91	0.01	0.02**
Pr(start col)	0.56	0.61	0.65	0.04**	0.05***
Pr(grad BA)	0.20	0.23	0.26	0.03**	0.02*
Pr(grad BA   start col)	0.36	0.38	0.39	0.02	0.01
<i>N</i>	1,064	2,400	1,930		

Notes: High school graduation includes earning either a GED or a diploma. Starting college refers to enrolling in either a 2- or 4-year institution. Significance reported at the 1% (\*\*\*), 5% (\*\*), and 10% (\*) levels

Table 3.3: Average growth in full-time wages due to various experiences by final educational attainment

(a) High school dropouts					
Experience	79o	79y	97	$\Delta$ 79y-79o	$\Delta$ 97-79y
work in HS	0.092	0.081	0.046	-0.011	-0.035**
work part time	-0.044	0.019	-0.036	0.062***	-0.055***
work full time	0.032	0.048	0.047	0.016***	-0.001
<i>N</i>	1,140	2,026	1,617		

(b) High school graduates					
Experience	79o	79y	97	$\Delta$ 79y-79o	$\Delta$ 97-79y
work in HS	0.043	0.032	0.061	-0.011	0.030***
work part time	0.003	-0.026	-0.011	-0.030**	0.016*
work full time	0.033	0.054	0.052	0.021***	-0.001
<i>N</i>	2,459	4,733	4,817		

(c) Some college					
Experience	79o	79y	97	$\Delta$ 79y-79o	$\Delta$ 97-79y
work in HS	-0.005	0.073	0.024	0.078***	-0.049***
work in college	0.071	0.072	0.074	0.001	0.002
work part time	-0.032	0.031	-0.028	0.064***	-0.059***
work full time	0.054	0.063	0.062	0.009**	-0.001
<i>N</i>	2,195	5,060	5,952		

(d) College graduates					
Experience	79o	79y	97	$\Delta$ 79y-79o	$\Delta$ 97-79y
work in HS	-0.010	0.061	-0.003	0.071***	-0.063***
work in college	0.077	0.034	0.036	-0.043***	0.002
work part time	0.014	0.043	-0.084	0.029	-0.127***
work full time	0.098	0.097	0.090	-0.001	-0.007
<i>N</i>	817	2,013	2,469		

(e) All individuals					
Experience	79o	79y	97	$\Delta$ 79y-79o	$\Delta$ 97-79y
work in HS	0.029	0.074	0.041	0.046***	-0.033***
work in college	0.087	0.111	0.094	0.024***	-0.017***
work part time	-0.021	-0.005	-0.056	0.015*	-0.051***
work full time	0.037	0.049	0.044	0.012***	-0.004**
<i>N</i>	6,611	13,832	14,855		

Notes: Estimates weighted by NLSY sampling weights. Estimates are coefficients from regressing log wage on each cumulative experience term separately. One monthly observation per year per individual is included in *N*. HS graduates included in this table are those who never attended college. "Some college" are those who attended college (either 2- or 4-year) but did not graduate with a 4-year degree. College graduates are those who graduated with a 4-year degree but who never attended graduate school. Significance reported at the 1% (\*\*\*), 5% (\*\*), and 10% (\*) levels.

Table 3.4: Average growth in full-time wages due to various experiences at age 29 by final educational attainment

(a) High school dropouts					
Experience	79o	79y	97	$\Delta$ 79y-79o	$\Delta$ 97-79y
work in HS	0.162	0.077	0.033	-0.084***	-0.044***
work part time	-0.072	-0.019	-0.112	0.053***	-0.093***
work full time	0.060	0.053	0.050	-0.007	-0.002
<i>N</i>	1,205	2,154	1,188		

(b) High school graduates					
Experience	79o	79y	97	$\Delta$ 79y-79o	$\Delta$ 97-79y
work in HS	0.051	0.045	0.029	-0.006	-0.017*
work part time	-0.039	-0.091	-0.058	-0.052***	0.033***
work full time	0.039	0.064	0.054	0.025***	-0.010**
<i>N</i>	2,799	5,636	3,439		

(c) Some college					
Experience	79o	79y	97	$\Delta$ 79y-79o	$\Delta$ 97-79y
work in HS	-0.019	0.062	-0.004	0.082***	-0.066***
work in college	0.077	0.056	0.036	-0.021**	-0.020***
work part time	-0.100	-0.036	-0.087	0.064***	-0.051***
work full time	0.054	0.037	0.034	-0.018***	-0.002
<i>N</i>	2,820	6,528	5,317		

(d) College graduates					
Experience	79o	79y	97	$\Delta$ 79y-79o	$\Delta$ 97-79y
work in HS	0.025	0.051	-0.007	0.026*	-0.058***
work in college	0.045	-0.025	-0.004	-0.070***	0.022***
work part time	-0.097	-0.093	-0.120	0.004	-0.026**
work full time	0.019	0.063	0.071	0.044***	0.008
<i>N</i>	1,316	3,648	3,545		

(e) All individuals					
Experience	79o	79y	97	$\Delta$ 79y-79o	$\Delta$ 97-79y
work in HS	0.029	0.076	0.014	0.047***	-0.062***
work in college	0.100	0.086	0.067	-0.013***	-0.020***
work part time	-0.085	-0.100	-0.124	-0.015**	-0.024***
work full time	0.006	0.006	-0.002	0.000	-0.008***
<i>N</i>	8,140	17,966	13,489		

Notes: Estimates weighted by NLSY sampling weights. Estimates are coefficients from regressing log wage on each cumulative experience term separately. One monthly observation per year per individual is included in *N*. HS graduates included in this table are those who never attended college. "Some college" are those who attended college (either 2- or 4-year) but did not graduate with a 4-year degree. College graduates are those who graduated with a 4-year degree but who never attended graduate school. Significance reported at the 1% (\*\*\*), 5% (\*\*), and 10% (\*) levels.

Table 3.5: College and HS wage premium and dispersion at age 29 for full-time workers

Variable	79o	79y	97	$\Delta$ 79y-79o	$\Delta$ 97-79y
<i>Average log wages</i>					
HS dropouts	1.86	1.81	1.75	-0.05***	-0.05***
HS graduates	2.00	1.92	1.91	-0.09***	-0.01
Some college	2.14	2.05	2.01	-0.09***	-0.04***
College graduates	2.31	2.35	2.28	0.04***	-0.08***
<i>Average wage premia</i>					
High school wage premium	0.14	0.11	0.16	-0.04**	0.05***
Some college wage premium	0.13	0.13	0.09	0.00	-0.04***
College wage premium	0.31	0.44	0.37	0.13***	-0.07***
<i>Standard deviation of log wages</i>					
HS dropouts	0.39	0.38	0.35		
HS graduates	0.37	0.39	0.39		
Some college	0.44	0.40	0.42		
College graduates	0.39	0.37	0.43		
<i>N</i> HS dropouts	1,205	2,154	1,188		
<i>N</i> HS graduates	2,727	5,452	3,403		
<i>N</i> Some college	2,820	6,528	5,317		
<i>N</i> College graduates	1,296	3,578	3,526		

Notes: Summary statistics weighted by NLSY sampling weights. All monthly log wage observations during the last year of the panel are included in *N*. HS graduates included in this table are those who never attended college. “Some college” are those who attended college (either 2- or 4-year) but did not graduate with a 4-year degree. College graduates are those who graduated with a 4-year degree but who never attended graduate school. “High school wage premium” refers to the log wage difference between HS graduates and HS dropouts. “Some college wage premium” refers to the log wage difference between “Some college” and HS graduates. “College wage premium” refers to the log wage difference between College graduates and HS graduates. Significance reported at the 1% (\*\*\*), 5% (\*\*), and 10% (\*) levels.

Table 3.6: Median AFQT score and dispersion by final educational attainment

	79o	79y	97	$\Delta$ 79y-79o	$\Delta$ 97-79y
<i>Median AFQT score</i>					
HS dropouts	-0.97	-0.97	-0.77	0.00	0.19**
HS graduates	0.05	-0.13	-0.14	-0.17**	-0.02
Some college	0.43	0.38	0.45	-0.05	0.07
College graduates	1.22	1.18	1.05	-0.04	-0.12***
<i>Standard deviation of AFQT score</i>					
HS dropouts	0.68	0.78	0.94		
HS graduates	0.79	0.85	0.89		
Some college	0.81	0.83	0.84		
College graduates	0.52	0.56	0.62		
<i>N</i> HS dropouts	179	379	416		
<i>N</i> HS graduates	338	774	923		
<i>N</i> Some college	391	939	1,358		
<i>N</i> College graduates	188	453	748		

Notes: AFQT distribution normalized so that the distribution including all cohorts is mean-zero, variance one. Significance reported at the 1% (\*\*\*), 5% (\*\*), and 10% (\*) levels using bootstrapped standard errors of the median (500 replications).

Table 3.7: Family background characteristics by final educational attainment

(a) High school dropouts					
Variable	79o	79y	97	$\Delta$ 79y-79o	$\Delta$ 97-79y
Mother's education	9.78	10.17	11.22	0.39	1.05***
Father's education	8.88	9.89	11.08	1.01***	1.19***
Family Income	20.04	20.58	19.64	0.54	-0.94
% live in female-headed HH	0.18	0.19	0.31	0.01	0.12***
<i>N</i>	206	409	603		
(b) High school graduates					
Variable	79o	79y	97	$\Delta$ 79y-79o	$\Delta$ 97-79y
Mother's education	11.12	10.91	11.93	-0.21	1.02***
Father's education	11.02	10.87	11.79	-0.15	0.92***
Family Income	31.42	26.58	25.88	-4.84***	-0.70
% live in female-headed HH	0.13	0.14	0.25	0.01	0.11***
<i>N</i>	373	800	1,202		
(c) Some college					
Variable	79o	79y	97	$\Delta$ 79y-79o	$\Delta$ 97-79y
Mother's education	11.95	11.82	12.97	-0.13	1.15***
Father's education	12.60	12.21	12.94	-0.39*	0.73***
Family Income	35.45	31.46	33.98	-3.99***	2.52**
% live in female-headed HH	0.11	0.14	0.24	0.03	0.10***
<i>N</i>	420	978	1,696		
(d) College graduates					
Variable	79o	79y	97	$\Delta$ 79y-79o	$\Delta$ 97-79y
Mother's education	13.34	13.38	14.52	0.04	1.14***
Father's education	14.10	14.38	14.98	0.28	0.60***
Family Income	47.04	45.16	49.44	-1.88	4.28**
% live in female-headed HH	0.08	0.10	0.12	0.02	0.02
<i>N</i>	197	462	873		
(e) All individuals					
Variable	79o	79y	97	$\Delta$ 79y-79o	$\Delta$ 97-79y
Mother's education	13.34	13.38	14.52	0.04	1.14***
Father's education	14.10	14.38	14.98	0.28	0.60***
Family Income	47.04	45.16	49.44	-1.88	4.28**
% live in female-headed HH	0.08	0.10	0.12	0.02	0.02
<i>N</i>	197	462	873		

Notes: Family income is in 1,000's of 1982-84\$. Education is highest grade of the respondent's biological parents. Female-headed household is from survey round 1 in NLSY79 and age 14 in NLSY97. Significance reported at the 1% (\*\*\*), 5% (\*\*), and 10% (\*) levels

Table 3.8: Local labor market conditions at various ages

(a) “employment rate”

Timing	79o	79y	97	$\Delta$ 79y-79o	$\Delta$ 97-79y
At age 16	0.72	0.75	0.88	0.03***	0.13***
At age 22	0.76	0.79	0.88	0.03***	0.09***
At age 26	0.81	0.84	0.88	0.03***	0.04***
At age 29	0.85	0.85	0.85	0.00	0.00

(b) income per worker

Timing	79o	79y	97	$\Delta$ 79y-79o	$\Delta$ 97-79y
At age 16	12.04	12.40	16.54	0.36***	4.14***
At age 22	12.53	13.71	18.13	1.18***	4.42***
At age 26	13.96	14.83	18.65	0.87***	3.82***
At age 29	14.94	14.98	18.52	0.04	3.54***

Notes: “Employment rate” in the respondent’s county of residence at each age is the number of employees reported by employers divided by population. Income per worker is the total wage and salary income of the county (in 1,000’s of 1982-84\$) divided by the number of workers. Significance reported at the 1% (\*\*\*), 5% (\*\*), and 10% (\*) levels.



Table 3.9: Measures of returns to schooling across specifications

(a) Any school

Specification	NLSY79 old	NLSY79 young	NLSY97	$\Delta$ 79y-79o	$\Delta$ 97-79y
(i) Raw	–	–	–	–	–
(ii) Mincer	0.016***	0.033***	0.049***	0.017***	0.016***
(iii) HLT (2006)	0.107***	0.073***	0.058***	-0.035***	-0.015***
(iv) +Actual Exp	-0.002	0.012***	0.026***	0.014***	0.015***
(v) +Background	0.026***	0.024***	0.007***	-0.002	-0.018***
(vi) +Unobserved	0.053***	0.038***	0.026***	-0.015***	-0.012***

(b) HS graduation

Specification	NLSY79 old	NLSY79 young	NLSY97	$\Delta$ 79y-79o	$\Delta$ 97-79y
(i) Raw	0.160***	0.156***	0.175***	-0.004	0.018***
(ii) Mincer	0.114***	0.089***	0.062***	-0.025***	-0.027***
(iii) HLT (2006)	0.115***	0.091***	0.060***	-0.024***	-0.031***
(iv) +Actual Exp	0.062***	0.061***	0.048***	-0.001	-0.012***
(v) +Background	0.081***	0.068***	0.037***	-0.014***	-0.030***
(vi) +Unobserved	0.060***	0.012***	0.038***	-0.048***	0.026***

(c) College graduation

Specification	NLSY79 old	NLSY79 young	NLSY97	$\Delta$ 79y-79o	$\Delta$ 97-79y
(i) Raw	0.245***	0.420***	0.367***	0.175***	-0.053***
(ii) Mincer	0.203***	0.354***	0.252***	0.151***	-0.102***
(iii) HLT (2006)	0.148***	0.319***	0.235***	0.171***	-0.084***
(iv) +Actual Exp	0.145***	0.304***	0.222***	0.159***	-0.082***
(v) +Background	0.136***	0.253***	0.209***	0.117***	-0.044***
(vi) +Unobserved	0.167***	0.205***	0.173***	0.038***	-0.033***

Notes:

Panel (a) is the wage return at age 29 of one extra year of schooling.

Panel (b) is the wage premium of earning a high school diploma relative to not earning a diploma.

Panel (c) is the wage premium of earning a bachelor's degree relative to a high school diploma.

(i) Indicates raw premium without any controls.

(ii) Includes a quadratic in age, a cubic in years of schooling, and dummies for type of work (in-school, part-time, full-time) as the only set of controls.

(iii) Increases flexibility similar to Heckman et al. (2006a). Replaces the quadratic in age with a cubic in age, adds a linear interaction between schooling experience and age, and adds race/ethnicity indicators.

(iv) Replaces potential experience (age) with actual work experience type (in-school, part-time, full-time), military experience, and other experience.

(v) Adds personal background characteristics.

(vi) Adds discrete choice estimation and person-specific random factors for dynamic selection.

Significance reported at the 1% (\*\*\*), 5% (\*\*), and 10% (\*) levels.

Table 3.10: Select wage equation marginal effects (age 29)

(a) No unobserved heterogeneity

Variable	NLSY79 old	NLSY79 young	NLSY97	$\Delta$ 79y-79o	$\Delta$ 97-79y
years of school	0.026*** (0.003)	0.024*** (0.002)	0.007*** (0.002)	-0.002*** (0.003)	-0.018*** (0.002)
work in HS	0.025*** (0.003)	0.024*** (0.002)	-0.003** (0.001)	-0.001*** (0.004)	-0.027*** (0.002)
work in college	0.067*** (0.003)	0.043*** (0.002)	0.040*** (0.001)	-0.024*** (0.004)	-0.003*** (0.002)
work PT only	-0.041*** (0.002)	-0.022*** (0.002)	-0.045*** (0.001)	0.019*** (0.003)	-0.023*** (0.002)
work FT only	0.034*** (0.001)	0.035*** (0.001)	0.035*** (0.001)	0.001*** (0.002)	0.001*** (0.001)
HS graduate	0.081*** (0.004)	0.068*** (0.002)	0.037*** (0.002)	-0.014*** (0.004)	-0.030*** (0.003)
College graduate	0.136*** (0.006)	0.253*** (0.004)	0.209*** (0.003)	0.117*** (0.007)	-0.044*** (0.005)

(b) With unobserved heterogeneity

Variable	NLSY79 old	NLSY79 young	NLSY97	$\Delta$ 79y-79o	$\Delta$ 97-79y
years of school	0.053*** (0.003)	0.038*** (0.002)	0.026*** (0.002)	-0.015*** (0.003)	-0.012*** (0.002)
work in HS	0.017*** (0.004)	-0.028*** (0.002)	0.027*** (0.001)	-0.045*** (0.004)	0.055*** (0.002)
work in college	0.065*** (0.004)	0.026*** (0.002)	0.021*** (0.002)	-0.040*** (0.004)	-0.005*** (0.003)
work PT only	-0.004** (0.002)	0.009*** (0.002)	-0.013*** (0.001)	0.013 (0.003)	-0.022*** (0.002)
work FT only	0.030*** (0.001)	0.029*** (0.001)	0.033*** (0.001)	-0.001*** (0.001)	0.004*** (0.001)
HS graduate	0.060*** (0.004)	0.012*** (0.003)	0.038*** (0.002)	-0.048*** (0.005)	0.026*** (0.004)
College graduate	0.167*** (0.005)	0.205*** (0.004)	0.173*** (0.003)	0.038*** (0.006)	-0.033*** (0.005)

Notes:

Panel (a) refers to wage equation marginal effects without correcting for selection on unobservables. This is the same specification as the “+Background” column in Table 3.9.

Panel (b) refers to wage equation marginal effects correcting for selection on unobservables. This is the same specification as the “Unobs Het” column in Table 3.9.

Marginal effects are evaluated at the cohort-specific sample averages at age 29 for each component of experience. Significance reported at the 1% (\*\*\*), 5% (\*\*), and 10% (\*) levels.

Table 3.11: Full-time wage factor loading estimates

Loading	NLSY79 old	NLSY79 young	NLSY97	$\Delta$ 79y-79o	$\Delta$ 97-79y
cognitive	0.174*** (0.0010)	0.145*** (0.0009)	0.163*** (0.0008)	-0.030*** (0.0014)	0.018*** (0.0012)
non-cognitive	0.114*** (0.0011)	0.108*** (0.0007)	0.117*** (0.0007)	-0.006*** (0.0013)	0.009*** (0.0010)

Notes: Factor loading estimates are from the specification found in the “Unobs het” column in Table 3.9. Significance reported at the 1% (\*\*\*), 5% (\*\*), and 10% (\*) levels.

Table 3.12: Full-time log wage decompositions at age 29 for an additional unit of schooling experience

(a)  $\Delta 79y-79o$

Experience Type	Observed				Unobserved		Difference in Data
	Total Price	Direct Price	Indirect Price	Total Composition	Total Price	Total Composition	
years of school	0.1	6.0	-5.8	-5.6	1.4	4.0	-3.7
work in HS	-3.7	-2.9	-0.8	5.7	5.5	2.3	15.7
work in college	-9.7	-3.4	-6.3	1.0	6.9	4.5	7.8

(b)  $\Delta 97-79y$

Experience Type	Observed				Unobserved		Difference in Data
	Total Price	Direct Price	Indirect Price	Total Composition	Total Price	Total Composition	
years of school	-3.4	2.3	-5.7	-0.3	2.4	0.8	-3.6
work in HS	5.4	6.8	-1.4	-1.8	-10.1	-1.5	-13.9
work in college	-2.6	-1.3	-1.3	-10.4	-0.9	-0.6	-18.5

Notes: Figures presented in wage percentage points (log points). Each column isolates a different component of the evolution in log wages across the two cohorts. See text for explanation of each component. The last column is the overall wage premium, which is the difference in expected wages for a one year difference in experience at age 29.

Table 3.13: Full-time log wage decompositions at age 29 for an additional unit of work experience

(a)  $\Delta$  79y-79o

Experience Type	Observed				Unobserved		Difference in Data
	Total Price	Direct Price	Indirect Price	Total Composition	Total Price	Total Composition	
work PT only	1.6	0.9	0.7	-3.4	1.6	0.6	-1.6
work FT only	-1.9	-0.4	-1.5	-4.1	0.2	1.0	-2.8

(b)  $\Delta$  97-79y

Experience Type	Observed				Unobserved		Difference in Data
	Total Price	Direct Price	Indirect Price	Total Composition	Total Price	Total Composition	
work PT only	1.3	-1.0	2.4	-0.1	-0.5	-2.0	-1.1
work FT only	1.7	-0.1	1.8	0.3	-1.2	-0.5	0.6

Notes: Figures presented in wage percentage points (log points). Each column isolates a different component of the evolution in log wages across the two cohorts. See text for explanation of each component. The last column is the overall wage premium, which is the difference in expected wages for a one year difference in experience at age 29.

Table 3.14: Full-time log wage decompositions at age 29 for educational attainment

(a)  $\Delta$  79y-79o

Educational Attainment	Observed				Unobserved		Difference in Data
	Total Price	Direct Price	Indirect Price	Total Composition	Total Price	Total Composition	
HS graduation	-8.9	-4.8	-4.2	4.7	3.6	-3.4	-3.5
College graduation	-6.5	3.8	-10.3	0.1	13.8	6.7	12.3

(b)  $\Delta$  97-79y

Educational Attainment	Observed				Unobserved		Difference in Data
	Total Price	Direct Price	Indirect Price	Total Composition	Total Price	Total Composition	
HS graduation	-1.4	2.6	-4.1	2.3	-6.9	4.8	3.9
College graduation	-8.3	-3.3	-5.1	0.1	-4.9	-2.6	-5.6

Notes: Figures presented in wage percentage points (log points). Each column isolates a different component of the evolution in log wages across the two cohorts. See text for explanation of each component. The last column is the overall wage premium, which is the difference in expected wages for a graduate vs a non-graduate at age 29. These numbers are constructed similarly to those in Table 3.5, though in this Table HS graduates include those who also have some college.

# Appendix A

## Data and Estimation Notes for Chapter 2

### A.1 Sample Selection

Table A.1 breaks down the sample selection into more detail. The first big group dropped are teachers who received their master's degree before beginning to teach. I do not include them because I do not have the opportunity to observe them at the time they make this choice. Also, these teachers are quite different from teachers who receive their master's degree after teaching. Two metrics differentiating these two groups is that those who received their MA before very rarely attend for-profit colleges and have much higher Praxis scores.

I also drop censored observations. This primarily takes into account situations where teachers transfer into non-teaching occupations within the school system or they leave for multiple periods and then return to teaching. Censored observations have teacher characteristics that are quite similar to the full sample.

The next group dropped are those with invalid history and/or experience. The primary culprit in this situation is one where the teacher does not have a start date, or the start date is earlier than the college graduation date. These teachers look

Table A.1: Sample Selection

	N	N*T
Teachers who began on or after 1995	157,533	788,150
Drop if received MA before teaching	16,869	76,742
Drop right- and left-censored observations	641	92,529
Drop if invalid history and/or experience	28,503	93,902
Drop if missing demographic data	10,454	22,761
Final	101,066	502,216

Sources: NCERDC and IPEDS

Note: Includes all teachers in the sample and their eventual master's degree status.

similar to the rest of the sample in terms of demographic characteristics. However, they do generally score lower on the Praxis exams, though they are also less likely to report taking the exam.

Finally, I drop individuals with missing demographic data. These are persons who never appear in certain personnel files, and as Table A.1 shows, they are only in the data on average for two periods. These teachers are disproportionately more likely to have higher Praxis scores, implying that perhaps some of the best and brightest are only testing the waters.

## A.2 Tuition

As mentioned, I supplement the IPEDS data with information from colleges websites. Specifically, I looked at a few years of tuition posted by North Carolina colleges and compared it with the IPEDS tuition. For universities that had a high match, I used the IPEDS tuition. For those without, I researched thoroughly to find as many actual data points of tuition as possible, and then imputed the missing values by interpolation, using the average trend over time of the other colleges of the same type.



### A.3 Identification

The assumption on the error term allows the *Emax* term in equation 2.7 to have a closed form solution:

$$E_t \max_{k \in \mathcal{J}^r} \{v_{ik,t+1}(x_{i,t+1}) + \varepsilon_{ik,t+1} | d_{it} = j\} = \ln \left( \sum_{k \in \mathcal{J}^r} \exp v_{ik,t+1}(x_{i,t+1}) \right) + \gamma \quad (\text{A.1})$$

where  $\gamma$  is the Euler-Mascheroni constant (see McFadden (1974) and Rust (1987)). Below I use new notation for the outside option, similar to my notation for  $\mathcal{J}^r$ . Specifically,  $O^r = O^{rit}$  indicates the outside option for choice set  $r_{it}$  of teacher  $i$  in period  $t$ . By adding and subtracting the value function for this outside option, I can represent equation (A.1) as

$$\begin{aligned} & \ln \left( \sum_{k \in \mathcal{J}^r} \exp v_{ik,t+1}(x_{i,t+1}) \right) + \gamma & (\text{A.2}) \\ &= \ln \left( \frac{\exp(v_{i,O^r,t+1}(x_{i,t+1}))}{\exp(v_{i,O^r,t+1}(x_{i,t+1}))} \sum_{k \in \mathcal{J}^r} \exp(v_{ikt+1}(x_{i,t+1})) \right) + \gamma \\ &= v_{i,O^r,t+1}(x_{i,t+1}) + \ln \left( \frac{\sum_{k \in \mathcal{J}^r} \exp(v_{ikt+1}(x_{i,t+1}))}{\exp(v_{i,O^r,t+1}(x_{i,t+1}))} \right) + \gamma \\ &= v_{i,O^r,t+1}(x_{i,t+1}) - \ln \left( \frac{\exp(v_{i,O^r,t+1}(x_{i,t+1}))}{\sum_{k \in \mathcal{J}^r} \exp(v_{ikt+1}(x_{i,t+1}))} \right) + \gamma \\ &= v_{i,O^r,t+1}(x_{i,t+1}) - \ln(p_{O^r,t+1}(x_{i,t+1})) + \gamma, & (\text{A.3}) \end{aligned}$$

where  $p_{O^r,t+1}(x_{i,t+1})$  is the conditional choice probability of choosing the outside option in period  $t+1$ , conditional on  $d_{it} = j$  and  $x_{i,t+1}$ .

I can now express the conditional value function as

$$v_{ijt}(x_{it}) = u_{ijt}(x_{it}) + \beta \sum_{x_{i,t+1}} [v_{i,O^r,t+1}(x_{i,t+1}) - \ln(p_{O^r,t+1}(x_{i,t+1})) + \gamma] f_j(x_{i,t+1} | x_{it}) \quad (\text{A.4})$$

where the only future terms are the one-period ahead conditional value function and choice probability for the outside option. The assumptions on the outside option are

important here. Since the outside option is assumed to be a terminal state, there are no more choices to make, thus the utility functions are undefined. I can define a conditional value function for this outside option, namely the discounted sum of future wages taken to the expected age of retirement. However, the other assumption on the outside option is that regardless of  $r$ , the wage for the outside option is the average county wage. This term is thus a constant and will cancel out in estimation. Relaxing this assumption by using different county wages for different degree levels does not significantly change the results. As such,  $v_{i,O^r,t+1}(x_{i,t+1})$  can be ignored and equation (2.9) can be written as:

$$v_{ijt}(x_{it}) = u_{ijt}(x_{it}) + \beta \sum_{x_{i,t+1}} [-\ln(p_{O^r,t+1}(x_{i,t+1})) + \gamma] f_j(x_{i,t+1}|x_{it}). \quad (\text{A.5})$$

Using more properties of the multinomial logit, the probability of teacher  $i$  choosing option  $j$  in period  $t$  can be expressed as

$$\begin{aligned} p_{jt}(x_{it}) &= Pr(d_{it} = j) \\ &= Pr(v_{ijt}(x_{it}) + \varepsilon_{ijt} \geq v_{ikt}(x_{it}) + \varepsilon_{ikt}), \forall k \neq j, k \in \mathcal{J}^r \\ &= \frac{\exp(v_{ijt}(x_{it}))}{\sum_{k \in \mathcal{J}^r} \exp(v_{ikt}(x_{it}))} \end{aligned} \quad (\text{A.6})$$

The likelihood for individual  $i$  in state  $s$  can be expressed as the product of all periods, choices, and choice sets

$$\mathcal{L}_{is}(\theta) = \prod_t \prod_j \prod_r [p_{jt}(x_{it}, s; \theta)^{d_{it}=j}]^{\mathbf{1}\{j \in \mathcal{J}^r\}} \phi_i(x_{it}, s; \theta)$$

where  $\theta = [\alpha \kappa \psi \delta \sigma]$ ,  $\alpha$  is a vector of all the  $\alpha_j$ 's and  $\kappa$  is a vector of all the  $\kappa_j$ 's. Additionally,  $\mathbf{1}\{j \in \mathcal{J}^r\}$  is an indicator for whether choice  $j$  is in choice set  $\mathcal{J}^r$ , and  $\phi_i(\cdot)$  is the contribution to the likelihood function from observed teacher ability from equation (2.1).

## A.4 Estimation

Equation (A.5) contains three components that need to be estimated: the transition probabilities, the one period ahead CCPs, and the period flow utility. A fourth component of the estimation is ability, equation (2.1). Estimation occurs in two stages. In the first stage I estimate the transition probabilities and the one period ahead CCPs. I then use these to estimate the utility and ability parameters in equations (2.9) and (2.1).

The transition probabilities from equation (2.2) are represented as a Markov transition matrix. This is calculated from the raw data as the probability of going from state  $x_t$  to  $x_{t+1}$  and is only estimated from the elements of  $x_{it}$  that evolve stochastically. SES quartiles, student test score quartiles, and teacher experience are the state variables with stochastic transitions. Experience is stochastic in the model because it does not always increase every period a teacher works, mainly due to time away from teaching within a year.

The one-period-ahead CCPs are estimated using a flexible multinomial logit

$$\tilde{d}_{it} = k = \arg \max_{j \in J} \{ \tilde{v}_{ijt}(x_{it}) + \tilde{\varepsilon}_{ijt} \}, \quad (\text{A.7})$$

where  $\tilde{v}_{ijt}(x_{it})$  differs from equation (A.5) in that  $u_{ijt}(x_{ijt})$  is augmented with a more flexible form, including higher-order terms and interactions. The error term  $\tilde{\varepsilon}_{ijt}$  is IID type-I extreme value. The CCPs are calculated for each choice for each possible realization of  $x_{t+1}$ .

Once the transition probabilities and (initial) conditional choice probabilities are calculated, all that remains to estimate are the coefficients on the observed and unobserved variables from the period flow utility,  $u_{ijt}(x_{ijt})$ . This is done in a second stage.

As mentioned, the unobserved terms are assumed to follow a mixture distribution. Let  $\pi_s$  be the sample probability of being type  $s$ , and let there be  $S$  possible types.

Then the full likelihood and log-likelihood functions can be written as

$$\mathcal{L}(\theta) = \prod_i \left( \sum_s^S \pi_s \mathcal{L}_{is}(\theta) \right) \quad (\text{A.8})$$

$$\ell(\theta) = \sum_i \ln \left( \sum_s^S \pi_s \mathcal{L}_{is}(\theta) \right). \quad (\text{A.9})$$

Next, define  $q_{is}$  as the probability of individual  $i$  being unobserved type  $s$ , conditional on the observed data:

$$q_{is} = \frac{\pi_s \mathcal{L}_{is}(\theta)}{\mathcal{L}_i(\theta)} \quad (\text{A.10})$$

$$= \frac{\pi_s \mathcal{L}_{is}(\theta)}{\sum_{s'}^S \pi_{s'} \mathcal{L}_{is'}(\theta)} \quad (\text{A.11})$$

where  $\mathcal{L}_i(\theta) = \sum_s^S \pi_s \mathcal{L}_{is}(\theta)$  is individual  $i$ 's full contribution to the likelihood. Then by definition

$$\pi_s = \frac{1}{N} \sum_i q_{is}. \quad (\text{A.12})$$

The definition of  $q_{is}$  allows for a different likelihood than equation (A.9):

$$\begin{aligned} \ell(\theta) &= \sum_i \sum_s^S q_{is} \ln(\mathcal{L}_{is}(\theta)) \\ \ell(\theta) &= \sum_i \sum_s^S q_{is} \sum_j^{\mathcal{J}^r} \sum_t^T (d_{it} = j) \ln(p_{jt}(x_{it}, s; \theta)), \end{aligned} \quad (\text{A.13})$$

which gives the same first-order conditions as equation (A.9).

Estimation then proceeds by using the Estimation-Maximization algorithm, as introduced in Arcidiacono and Miller (2011). Iteration  $m + 1$  will proceed as follows

1. Given  $\theta^{(m)}$  and  $\pi^{(m)}$ , calculate  $q^{(m+1)}$  from equation (A.10).
2. Using  $q^{(m+1)}$ , update  $\pi^{(m+1)}$  from equation (A.12).

3. Using  $q^{(m+1)}$ , update  $p_{O_r}^{(m+1)}$  from the parameters estimated in equation (A.7).
4. Using  $q^{(m+1)}$  and  $p_{O_r}^{(m+1)}$ , update  $\theta^{(m+1)}$  via maximum likelihood on equation (A.13).

Regarding initial values,  $\theta^{(0)}$  is estimated from (A.13) setting  $S = 1$ , and  $\pi^{(0)}$  is either the inverse of the number of types,  $\frac{1}{S}$ , or an average of random values of  $q_{is}$ .

# Appendix B

## Data and Estimation Notes for Chapter 3

### B.1 Sample Selection

The details of our sample selection can be found in Tables B.1 and B.2

Table B.1: Choice Sample Selection

Category	NLSY79 old <sup>a</sup>	NLSY79 young <sup>b</sup>	NLSY97
Starting persons	6,741	5,945	8,984
Drop females	3,355	2,928	4,599
Drop older birth cohorts <sup>c</sup>	1,698	0	0
Drop non-race oversamples <sup>d</sup>	492	251	0
Drop other race	0	0	40
Resulting persons	1,196	2,666	4,559
Survey rounds	15	15	15
Survey person-years <sup>e</sup>	12,628	33,983	57,522
Add retrospective data years <sup>f</sup>	2,920	675	843
Potential person-years	15,548	34,658	58,365
Potential person-months	186,576	415,896	688,903
Drop missing interview months <sup>g</sup>	8,250	19,638	101,853
Final persons	1,196	2,656	4,443
Final person-months	178,326	396,258	587,050
Final $\bar{T}$ (months)	149.1	149.2	132.1
Final max $T$ (months)	156	156	156

<sup>a</sup> Birth years 1957-1960.

<sup>b</sup> Birth years 1961-1964.

<sup>c</sup> Birth years 1957 and 1958.

<sup>d</sup> Oversamples of military personnel and disadvantage white individuals are both excluded from the analysis.

<sup>e</sup> This refers to the number of survey rounds available before an individual turns 28.

<sup>f</sup> This refers to adding retrospective data for the years 1974-1978 or 1993-1996 (if applicable).

<sup>g</sup> This refers to dropping any right-censored missing interview spells or any observations during or after a spell of 3+ missed interviews.

Table B.2: Wage Sample Selection

Category	NLSY79 old	NLSY79 young	NLSY97
Potential wage observations <sup>a</sup>	117,559	264,547	386,461
Drop self-employed wages	6,502	13,278	23,699
Drop outlying wages <sup>b</sup>	1,693	4,669	27,581
Drop non-reported wages	9,071	18,420	42,742
Final wage observations	100,293	228,180	292,529

<sup>a</sup> Potential wage observations refers to the the number of person-months choosing a work alternative.

<sup>b</sup> We drop wages below \$2 and above \$50 (in 1982-84\$).

## B.2 Formal definitions of decomposition components

As discussed previously, we estimate expected wage returns for specified values of some or all endogenous school and work experience variables, e.g.  $\mathbf{x}_{ia}^{r,c}$ , for birth cohort  $c \in \{79o, 79y, 97\}$ , and at specified ages,  $a = a'$ , where the expectations are taken over a cohort-specific population and assume that the model developed in Section 3.4 characterizes the choice process governing the conditioning endogenous variables.

In the following two sections, we formally define the expected wage returns (see equation (B.3)) and show how we compute counterfactual wage returns (see equation (B.6)) using the same formulas.

### *Wage returns*

Consider first the conditional expectation of log wages of cohort  $c$ , conditioned on one element of the experience profile, e.g.  $x_{kia}^{r,c} = \tilde{x}_{ka'}$ :

$$\begin{aligned}
& \mathbb{E} \left( w_{iaj'}^c \mid a = a', x_{kia}^{r,c} = \tilde{x}_{ka'}, \mathbf{q}_{ia}^c, \boldsymbol{\theta}^c \right) \\
&= \mathbb{E} \left( \beta_{0j'}^{r,c} + \beta_{\mathbf{m}}^c \mathbf{m}_{ia}^c + \beta_{\mathbf{z}}^c \mathbf{z}_i^c + \beta_{\mathbf{x}}^c g(\mathbf{x}_{ia}^{r,c}) + \beta_{\boldsymbol{\xi}j'}^c \boldsymbol{\xi}_i + \varepsilon_{iaj'} \mid \cdot \right) \\
&= \beta_{0j'}^{r,c} + \beta_{\mathbf{m}}^c \mathbb{E}(\mathbf{m}_{ia}^c \mid \cdot) + \beta_{\mathbf{z}}^c \mathbb{E}(\mathbf{z}_i^c \mid \cdot) \\
&\quad + \beta_{\mathbf{x}}^c \mathbb{E}(g(\mathbf{x}_{ia}^{r,c}) \mid \cdot) + \beta_{\boldsymbol{\xi}j'}^c \mathbb{E}(\boldsymbol{\xi}_i \mid \cdot) \\
&\quad + \mathbb{E}(\varepsilon_{iaj'} \mid \cdot) \tag{B.1} \\
&= \beta_{0j'}^{r,c} + \beta_{\mathbf{m}}^c \left[ \sum_i \mathbf{m}_{ia}^c \Pr(x_{kia}^{r,c} = \tilde{x}_{ka'} \mid a = a', \mathbf{q}_{ia}^c, \boldsymbol{\theta}^c) \right] \\
&\quad + \beta_{\mathbf{z}}^c \left[ \sum_i \mathbf{z}_{ia}^c \Pr(x_{kia}^{r,c} = \tilde{x}_{ka'} \mid a = a', \mathbf{q}_{ia}^c, \boldsymbol{\theta}^c) \right] \\
&\quad + \beta_{\mathbf{x}}^c \left[ \sum_i g(\mathbf{x}_{ia}^{r,c}) \Pr(x_{kia}^{r,c} = \tilde{x}_{ka'} \mid a = a', \mathbf{q}_{ia}^c, \boldsymbol{\theta}^c) \right] \\
&\quad + \beta_{\boldsymbol{\xi}j'}^c \mathbb{E}(\boldsymbol{\xi}_i \mid \cdot),
\end{aligned}$$



where “.” in the conditional expectation is used to denote the original condition on the left hand side of the equation. The notation  $\mathbf{q}_{ia}^c$  is meant to represent the matrix  $(\mathbf{m}_{ia}^{c'} \quad \mathbf{z}_i^{c'} \quad \mathbf{x}_{\sim k, ia}^{r, c'})'$ , and the notation  $\mathbf{x}_{\sim k, ia}^{r, c}$  refers to all experience variables except for  $x_{kia}^{r, c}$ . Further,  $\Pr(x_{kia}^{r, c} = \tilde{x}_{ka'} \mid a = a', \mathbf{q}_{ia}^c, \boldsymbol{\theta}^c)$  is the probability of the event  $x_{kia}^{r, c} = \tilde{x}_{ka'}$  occurring for individual  $i$ , e.g., that individual  $i$  is a college graduate by age  $a'$  with  $\tilde{x}_{kia}$  years of experience  $k$ , and  $\mathbb{E}(\boldsymbol{\xi}_i \mid a = a', x_{kia}^{r, c} = \tilde{x}_{ka'}, \mathbf{q}_{ia}^c, \boldsymbol{\theta}^c)$  is the truncated mean of  $\boldsymbol{\xi}_i$ , conditioned on the event  $x_{kia}^{r, c} = \tilde{x}_{ka'}$  occurring for individual  $i$ .

It follows that we can represent the expected wage return to wages for cohort  $c$  of a one-unit change in  $x_{ka}^r$ . In order to separate this into direct and indirect effects, we represent the experience function as:

$$g(\mathbf{x}_{ia}^{r, c}) = g_k(x_{kia}^{r, c}, \mathbf{x}_{\sim k, ia}^{r, c}) + g_{\sim k}(\mathbf{x}_{\sim k, ia}^{r, c}), \quad (\text{B.2})$$

where  $g_k(x_{kia}^{r, c}, \mathbf{x}_{\sim k, ia}^{r, c})$  are all the components, including linear, quadratic and cubic terms in experience, as well as any and all interactions between the experience variable of interest and the other experience terms, and  $g_{\sim k}(\mathbf{x}_{\sim k, ia}^{r, c})$  are all the other components of  $g(\mathbf{x}_{ia}^{r, c})$ , namely, those without any  $x_{kia}^{r, c}$  terms. The expected wage return can thus be represented as:

$$\begin{aligned} & \Delta_{x_{ka}^{r, c}} \mathbb{E}(w_{iaj'} \mid a = a', x_{kia}^{r, c} = \tilde{x}_{ka'}, \mathbf{q}_{ia}^c, \boldsymbol{\theta}^c) \\ &= \beta_{\mathbf{x}_k}^c \left[ g_k(\tilde{x}_{ka'} + 1, \mathbf{x}_{\sim k, ia}^{r, c}) - g_k(\tilde{x}_{ka'}, \mathbf{x}_{\sim k, ia}) \right] \\ &+ \beta_{\mathbf{m}}^c \left[ \sum_i \mathbf{m}_{ia}^c \Delta_{x_{ka}^{r, c}} \Pr(x_{kia}^{r, c} = \tilde{x}_{ka'} \mid a = a', \mathbf{q}_{ia}^c, \boldsymbol{\theta}^c) \right] \\ &+ \beta_{\mathbf{z}}^c \left[ \sum_i \mathbf{z}_{ia}^c \Delta_{x_{ka}^{r, c}} \Pr(x_{kia}^{r, c} = \tilde{x}_{ka'} \mid a = a', \mathbf{q}_{ia}^c, \boldsymbol{\theta}^c) \right] \\ &+ \beta_{\mathbf{x}_{\sim k}}^c \left[ \sum_i g_{\sim k}(\mathbf{x}_{\sim k, ia}^{r, c}) \Delta_{x_{ka}^{r, c}} \Pr(x_{kia}^{r, c} = \tilde{x}_{ka'} \mid a = a', \mathbf{q}_{ia}^c, \boldsymbol{\theta}^c) \right] \\ &+ \beta_{\boldsymbol{\xi}_{j'}}^c \Delta_{x_{ka}^{r, c}} \mathbb{E}(\boldsymbol{\xi}_i \mid a = a', x_{kia}^{r, c} = \tilde{x}_{ka'}, \mathbf{q}_{ia}^c, \boldsymbol{\theta}^c), \end{aligned} \quad (\text{B.3})$$

where

$$\begin{aligned} \Delta_{x_{ka}^r} \Pr(x_{kia'}^{r,c} = \tilde{x}_{ka'} | a = a', \mathbf{q}_{ia}^c, \boldsymbol{\theta}^c) \\ \equiv \Pr(x_{kia'}^r = \tilde{x}_{ka'} + 1 | a = a', \mathbf{q}_{ia}^c, \boldsymbol{\theta}^c) - \Pr(x_{kia'}^r = \tilde{x}_{ka'} | a = a', \mathbf{q}_{ia}^c, \boldsymbol{\theta}^c), \end{aligned} \quad (\text{B.4})$$

and

$$\begin{aligned} \Delta_{x_{ka}^r} \text{E}(\boldsymbol{\xi}_i | a = a', x_a^{r,c} = \tilde{x}_{a'}, \mathbf{q}_{ia}^c, \boldsymbol{\theta}^c) \\ \equiv \text{E}(\boldsymbol{\xi}_i | a = a', x_{kia'}^r = \tilde{x}_{ka'} + 1, \mathbf{q}_{ia}^c, \boldsymbol{\theta}^c) - \text{E}(\boldsymbol{\xi}_i | a = a', x_{kia'}^r = \tilde{x}_{ka'}, \mathbf{q}_{ia}^c, \boldsymbol{\theta}^c). \end{aligned} \quad (\text{B.5})$$

The second line (B.3) represents the *direct* observed component of the expected wage returns due to a marginal change in  $x_{ka}^{r,c}$  and the next three lines and the last line represent the *indirect* observed and unobserved components, respectively, of the expected wage returns due to the influence of a marginal change in  $x_{ka}^{r,c}$  on the composition of the population.

It follows that estimates of the expected wage returns in (B.3) are formed using  $\hat{\boldsymbol{\theta}}^c$  in place of  $\boldsymbol{\theta}^c$  and the expectations of the various functions formed by predicting in the sample for cohort  $c$  who will meet the condition  $x_{kia}^{r,c} = \tilde{x}_{ka'}$  at age  $a'$ , which determines the derivatives of the sample mean functions for  $\mathbf{m}_{ia}^c$ ,  $\mathbf{z}_i^c$ ,  $\mathbf{x}_{\sim k,ia}^{r,c}$  and  $\boldsymbol{\xi}_i$  with respect to  $x_{ka}^{r,c}$ .

### *Across-Cohort Counterfactual Evaluations*

We provide estimates of across-cohort counterfactual expected returns in order to examine the extent to which differences in unadjusted wage returns across cohorts are the result of changes in the composition of the cohorts, i.e., the differences across the cohorts in the observed,  $\mathbf{m}_{ia}$  ( $\mathbf{z}_i$ ) and unobserved ( $\boldsymbol{\xi}_i$ ) characteristics, versus changes in the structure of the wage function, i.e., changes in the  $\boldsymbol{\beta}^c$ 's. As discussed previously, this is an adaptation of the Oaxaca-Blinder decomposition method allowing for the endogeneity of the source of wage changes.<sup>1</sup> So, for example, we can

<sup>1</sup> See Oaxaca (1973) and Blinder (1973). Also see DiNardo (2002), who establishes the equivalence of the non-parametric Oaxaca-Blinder estimator with propensity score methods for the estimation

evaluate the counterfactual expected wage return to wages for cohort  $c = 97$  using their observed and unobserved characteristics,  $\mathbf{m}_{ia}^{97}$ ,  $\mathbf{z}_i^{97}$ , and  $\beta_\xi^{97} \boldsymbol{\xi}_i$  from a change in an element of  $\mathbf{x}_a^r$ ,  $x_{ka}^r$ , assuming that the wage and choice parameters are those that hold for cohort  $c = 79y$ , i.e.,  $\boldsymbol{\theta}^c = \boldsymbol{\theta}^{79y}$ :

$$\begin{aligned}
& \Delta_{x_{ka}^{r,97}} E \left( w_{ia,j'}^{97} \mid a = a', x_{kia}^{r,97} = \tilde{x}_{ka'}, \mathbf{q}_{ia}^{97}, \boldsymbol{\theta}^{79y} \right) \\
&= \beta_{\mathbf{x}_k}^{79y} \left[ g_k(\tilde{x}_{ka'} + 1, \mathbf{x}_{\sim k, ia}^{r,97}) - g_k(\tilde{x}_{ka'}, \mathbf{x}_{\sim k, ia}) \right] \\
&+ \beta_{\mathbf{m}}^{79y} \left[ \sum_i \mathbf{m}_{ia}^{97} \Delta_{x_{ka}^{r,97}} \Pr \left( x_{kia}^{r,97} = \tilde{x}_{ka'} \mid a = a', \mathbf{q}_{ia}^{97}, \boldsymbol{\theta}^{79y} \right) \right] \quad (\text{B.6}) \\
&+ \beta_{\mathbf{z}}^{79y} \left[ \sum_i \mathbf{z}_{ia}^{97} \Delta_{x_{ka}^{r,97}} \Pr \left( x_{kia}^{r,97} = \tilde{x}_{ka'} \mid a = a', \mathbf{q}_{ia}^{97}, \boldsymbol{\theta}^{79y} \right) \right] \\
&+ \beta_{\mathbf{x}_{\sim k}}^{79y} \left[ \sum_i g_{\sim k} \left( \mathbf{x}_{\sim k, ia}^{r,97} \right) \Delta_{x_{ka}^{r,97}} \Pr \left( x_{kia}^{r,97} = \tilde{x}_{ka'} \mid a = a', \mathbf{q}_{ia}^{97}, \boldsymbol{\theta}^{79y} \right) \right] \\
&+ \beta_{\boldsymbol{\xi}_{j'}}^{97} \Delta_{x_{ka}^{r,97}} E \left( \boldsymbol{\xi}_i \mid a = a', x_{kia}^{r,97} = \tilde{x}_{ka'}, \mathbf{q}_{ia}^{97}, \boldsymbol{\theta}^{79y} \right),
\end{aligned}$$

An analogous counterfactual set of effects can be obtained using the characteristics of the 1979 young cohort but assuming that the direct returns are those for the 1997 cohort. And finally, entire procedure can be duplicated to compare 1979 old with 1997 or the 1979 old with the 1979 young. We present results from each of these comparisons in the next sections.

---

of average treatment effects, and Kline (2011), who shows that the regression-based parametric Oaxaca-Blinder method is also a propensity score reweighting estimator when using a linear model to characterize the conditional odds of treatment.

# Bibliography

- Altonji, J. G., Bharadwaj, P., and Lange, F. (2009), “Constructing AFQT Scores that are Comparable Across the NLSY79 and the NLSY97,” Working paper, Yale University.
- Altonji, J. G., Bharadwaj, P., and Lange, F. (2012), “Changes in the Characteristics of American Youth: Implications for Adult Outcomes,” *Journal of Labor Economics*, 30, 783–828.
- Arcidiacono, P. and Miller, R. A. (2011), “Conditional Choice Probability Estimation of Dynamic Discrete Choice Models with Unobserved Heterogeneity,” *Econometrica*, 79, 1823–1867.
- Arcidiacono, P., Cooley, J., and Hussey, A. (2008), “The Economic Returns to an MBA,” *International Economic Review*, 49, 873–899.
- Arcidiacono, P., Hotz, V. J., and Kang, S. (2012), “Modeling College Major Choices Using Elicited Measures of Expectations and Counterfactuals,” *Journal of Econometrics*, 166, 3–16.
- Atrostic, B. K., Bates, N., Burt, G., and Silberstein, A. (2001), “Nonresponse in US Government Household Surveys: Consistent Measures, Recent Trends, and New Insights,” *Journal of Official Statistics*, 17, 209–226.
- Aucejo, E. (2014), “Explaining Cross-Racial Differences in the Educational Gender Gap,” Working paper, London School of Economics and Political Science.
- Aughinbaugh, A. and Gardecki, R. M. (2008), “Attrition in the National Longitudinal Survey of Youth 1997,” Working paper, Bureau of Labor Statistics.
- Bacolod, M. and Hotz, V. J. (2006), “Cohort Changes in the Transition from School to Work: Evidence from three NLS surveys,” *Economics of Education Review*, 25, 351–373.
- Blinder, A. S. (1973), “Wage Discrimination: Reduced Form and Structural Estimates,” *Journal of Human Resources*, 8, 436–455.

- Boehm, M. J. (2013), “Has Job Polarization Squeezed the Middle Class? Evidence from the Allocation of Talents,” Discussion Paper 1215, CEP.
- Bound, J., Lovenheim, M. F., and Turner, S. (2012), “Increasing Time to Baccalaureate Degree in The United States,” *Education Finance and Policy*, 7, 375–424.
- Cameron, S. V. and Heckman, J. J. (1998), “Life Cycle Schooling and Dynamic Selection Bias: Models and Evidence for Five Cohorts of American Males,” *Journal of Political Economy*, 106, 262–333.
- Cameron, S. V. and Heckman, J. J. (2001), “The Dynamics of Educational Attainment for Black, Hispanic, and White Males,” *Journal of Political Economy*, 109.
- Card, D. (1999), “The Causal Effect of Education on Earnings,” in *Handbook of Labor Economics*, vol. 3, chap. 30, pp. 1801–1863, Amsterdam: Elsevier.
- Card, D. and Lemieux, T. (2001), “Can Falling Supply Explain the Rising Return to College for Younger Men?” *Quarterly Journal of Economics*, 116, 705–746.
- Carneiro, P. and Lee, S. (2011), “Trends in Quality-Adjusted Skill Premia in the United States, 1960-2000,” *American Economic Review*, 101, 2309–2349.
- Castex, G. and Dechter, E. K. (2014), “The Changing Roles of Education and Ability in Wage Determination,” *Journal of Labor Economics*, 32, 685–710.
- Cellini, S. R. and Chaudhary, L. (2011), “The Labor Market Returns to a Private Two-Year College Education,” *George Washington University, working paper*.
- Clotfelter, C. T., Ladd, H. F., and Vigdor, J. L. (2007), “Teacher Credentials and Student Achievement: Longitudinal Analysis with Student Fixed Effects,” *Economics of Education Review*, 26, 673–682.
- Clotfelter, C. T., Ladd, H. F., and Vigdor, J. L. (2010), “Teacher Credentials and Student Achievement: A Cross-Subject Analysis with Student Fixed Effects,” *The Journal of Human Resources*, 45, 655–681.
- Cunha, F., Karahan, F., and Soares, I. (2011), “Returns to Skills and the College Premium,” *Journal of Money, Credit and Banking*, 43, 39–86.
- Deming, D. J., Goldin, C., and Katz, L. F. (2012), “The For-Profit Postsecondary School Sector: Nimble Critters or Agile Predators?” *Journal of Economic Perspectives*, 26, 139–164.
- DiNardo, J. (2002), “Propensity Score Reweighting and Changes in Wage Distributions,” Working paper, University of Michigan.

- Dolton, P. J., Makepeace, G. H., and Van der Klaauw, W. (1989), "Occupational Choice and Earnings Determination: The Role of Sample Selection and Non-pecuniary Factors," *Oxford Economic Papers*, pp. 573–594.
- Eckstein, Z. and Wolpin, K. I. (1989), "The Specification and Estimation of Dynamic Stochastic Discrete Choice Models: A Survey," *Journal of Human Resources*, 24, 562–598.
- Fang, H. (2006), "Disentangling the College Wage Premium: Estimating a Model with Endogenous Education Choices," *International Economic Review*, 47, 1151–1185.
- Fortin, N., Lemieux, T., and Firpo, S. (2011), "Decomposition Methods in Economics," in *Handbook of Labor Economics*, eds. O. Ashenfelter and D. Card, vol. 4A, pp. 1–102, Elsevier.
- Fortin, N. M. (2006), "Higher-education Policies and the College Wage Premium: Cross-state Evidence from the 1990s," *American Economic Review*, 96, 959–987.
- Goldhaber, D. D. and Brewer, D. J. (2000), "Does Teacher Certification Matter? High School Teacher Certification Status and Student Achievement," *Educational Evaluation and Policy Analysis*, 22, 129–145.
- Goldin, C. and Katz, L. F. (2007), "Long-Run Changes in the US Wage Structure: Narrowing, Widening, Polarizing," Working Paper 13568, National Bureau of Economic Research.
- Harris, D. N. and Sass, T. R. (2009), "The Effects of NBPTS-Certified Teachers on Student Achievement," *Journal of Policy Analysis and Management*, 28, 55–80.
- Heckman, J. J. and Sedlacek, G. (1985), "Heterogeneity, Aggregation, and Market Wage Functions: An Empirical Model of Self-selection in the Labor Market," *The Journal of Political Economy*, pp. 1077–1125.
- Heckman, J. J., Lochner, L. J., and Todd, P. E. (2006a), "Earnings Functions, Rates of Return and Treatment Effects: The Mincer Equation and Beyond," in *Handbook of the Economics of Education*, eds. E. A. Hanushek and F. Welch, vol. 1, pp. 307–458, Elsevier.
- Heckman, J. J., Stixrud, J., and Urzúa, S. (2006b), "The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior," *Journal of Labor Economics*, 24, 411–482.
- Hotz, V. J. and Miller, R. A. (1993), "Conditional Choice Probabilities and the Estimation of Dynamic Models," *The Review of Economic Studies*, 60, 497–529.

- Hotz, V. J., Xu, L. C., Tienda, M., and Ahituv, A. (2002), “Are There Returns to the Wages of Young Men from Working While in School?” *The Review of Economics and Statistics*, 84, 221–236.
- Juhn, C., Murphy, K. M., and Pierce, B. (1993), “Wage Inequality and the Rise in Returns to Skill,” *Journal of Political Economy*, 101, 410–442.
- Katz, L. F. and Murphy, K. M. (1992), “Changes in Relative Wages, 1963-1987: Supply and Demand Factors,” *Quarterly Journal of Economics*, 107, 35–78.
- Keane, M. and Wolpin, K. (1997), “The Career Decisions of Young Men,” *Journal of Political Economy*, 105, 473–522.
- Kline, P. (2011), “Oaxaca-Blinder as a Reweighting Estimator,” *American Economic Review*, 101, 532–37.
- Lee, D. and Wolpin, K. I. (2010), “Accounting for Wage and Employment Changes in the US from 1968–2000: A Dynamic Model of Labor Market Equilibrium,” *Journal of Econometrics*, 156, 68–85.
- Manski, C. F. (1987), “Academic Ability, Earnings, and the Decision to Become a Teacher: Evidence from the National Longitudinal Study of the High School Class of 1972,” in *Public sector payrolls*, ed. D. A. Wise, pp. 291–316, Chicago: University of Chicago Press.
- McFadden, D. (1974), “Conditional Logit Analysis of Qualitative Choice Behavior,” in *Frontiers in Econometrics*, ed. P. Zarembka, pp. 105–142, New York: Academic Press.
- Miller, R. A. (1984), “Job Matching and Occupational Choice,” *The Journal of Political Economy*, pp. 1086–1120.
- Mincer, J. (1974), *Schooling, Experience and Earnings*, Columbia University Press for National Bureau of Economic Research, New York.
- Moretti, E. (2011), “Local Labor Markets,” in *Handbook of Labor Economics*, eds. O. Ashenfelter and D. Card, vol. 4B, pp. 1237–1313, Elsevier.
- Oaxaca, R. (1973), “Male-Female Wage Differentials in Urban Labor Markets,” *International Economic Review*, 14, 693–709.
- Roy, A. D. (1951), “Some Thoughts on the Distribution of Earnings,” *Oxford Economic Papers*, 3, 135–146.
- Rust, J. (1987), “Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher,” *Econometrica: Journal of the Econometric Society*, pp. 999–1033.

- Scott-Clayton, J. (2012), “What Explains Trends in Labor Supply Among U.S. Undergraduates?” *National Tax Journal*, 65, 181–210.
- Siow, A. (1984), “Occupational choice under uncertainty,” *Econometrica: Journal of the Econometric Society*, pp. 631–645.
- Stinebrickner, T. R. (2001a), “Compensation Policies and Teacher Decisions,” *International Economic Review*, 42, 751–780.
- Stinebrickner, T. R. (2001b), “A Dynamic Model of Teacher Labor Supply,” *Journal of Labor Economics*, 19, 196–230.
- Sullivan, P. (2010), “A Dynamic Analysis of Educational Attainment, Occupational Choices, and Job Search,” *International Economic Review*, 51, 289–317.
- Taber, C. R. (2001), “The Rising College Premium in the Eighties: Return to College or Return to Unobserved Ability?” *The Review of Economic Studies*, 68, 665–691.
- U.S. Department of Education (1996), *Digest of Education Statistics 1995*, Washington, DC: National Center for Education Statistics.
- U.S. Department of Education (2003), *Digest of Education Statistics 2002*, Washington, DC: National Center for Education Statistics.
- U.S. Department of Education (2009), *Digest of Education Statistics 2008*, Washington, DC: National Center for Education Statistics.
- U.S. Department of Education (2013), *Digest of Education Statistics 2012*, Washington, DC: National Center for Education Statistics.
- U.S. Department of Education (2014), *Digest of Education Statistics 2013*, Washington, DC: National Center for Education Statistics.
- Willis, R. J. and Rosen, S. (1979), “Education and Self-Selection,” *Journal of Political Economy*, 87, S7–S36.
- Wiswall, M. (2007), “Licensing and Occupational Sorting in the Market for Teachers,” *New York University, Working Paper*.



# Biography

Jared Ashworth was born on August 11, 1981 in Provo, UT, where he spent all of his formative years. He was the youngest of seven, son to a Spanish professor and a music teacher.

After graduating with his diploma from Timpview High School in Provo, UT in May 1999, he enrolled full-time at Brigham Young University, excited to become an engineer. After his freshman year, he took a break from college and moved to Argentina to serve as a full time missionary for his church for two years. Upon his return, he discovered his love for economics, and jumped in with both feet forward. While an upperclassman, he met and married his wife Kristie, and graduated from BYU in April 2005 with a Bachelor's of Arts in Economics.

His first job after college was as a Financial Services Professional with MassMutual Financial Group in Salt Lake City, UT. After more than a year working there, he decided to look for a job with more of a research focus, and as such began working for LECG in Los Angeles, CA, where he and Kristie lived until for the next three years. While at LECG, he decided to return to school to get his Ph.D. and follow in his father's footsteps.

In August 2009 he enrolled at Duke University in Durham, NC to pursue a doctoral degree in Economics. He obtained a Master's of Arts from the Economics department in August 2010. Under the direction of his advisor Joe Hotz, he performed research in the fields of Labor Economics and the Economics of Education,

looking at the education decisions and work outcomes of individuals, with a focus on the career and education decisions of teachers. At Duke he received the Graduate School Tuition Scholarship every year. Further, in Summer 2012 he received the Duke Graduate School Summer Research Fellowship, and in Fall 2014 he received a Graduate Student award from the Southern Economic Association. Most importantly, during his time as a graduate student, he and his wife welcomed their two beautiful children into the world, Alia and Daxton. After graduation, he is taking a full-time position as an Assistant Professor in the Graziadio School of Business at Pepperdine University.