4-29-2020 10:00 AM

Relaxing the Rational Expectations Assumption: Data-based and Model-based Approaches

Yifan Gong, *The University of Western Ontario*

Supervisor: Stinebrickner, Todd R., *The University of Western Ontario*

A thesis submitted in partial fulfillment of the requirements for the Doctor of Philosophy degree in Economics

Abstract

The fundamental importance of beliefs about future outcomes in decision-making suggests that an accurate characterization of these beliefs is important for understanding individuals' behavior and for evaluating the counterfactuals typically needed for policy analysis. Traditionally, many researchers have been using some form of Rational Expectations (RE) assumptions to characterize these beliefs. However, empirical evidence suggests that the RE assumption might not hold in many contexts, and that incorrectly imposing the RE assumption can lead to biased policy predictions. Motivated by these findings, I explore alternative approaches to conducting economic analysis without imposing the RE assumption.

Chapters 2 and 3 of my thesis, which are co-authored with Todd and Ralph Stinebrickner, utilize unique survey expectations data from the Berea Panel Study (BPS) to characterize college students' beliefs about various future outcomes. Specifically, in Chapter 2, we characterize how much uncertainty about post-college income is present for students at college entrance and how quickly this uncertainty is resolved. Measuring an individual's income uncertainty by the variance of the distribution describing her beliefs about earnings at age 28, we find that, on average, students resolve roughly one-third of the income uncertainty present at the time of entrance during college. Consistent with the finding that the majority of initial income uncertainty remains at the end of college, We find that uncertainty about college GPA and field of study, which are the two primary income-influencing factors that are realized in college, can only account for about 19% to 27% of students' initial income uncertainty.

Chapter 3 provides a concrete example that illustrates the importance of quantifying the resolution of students' (income) uncertainty during college. By entering college, students have the option to decide whether to remain in college after receiving relevant new information. We show that the value of this option of receiving new information is determined by a student's dropout probability and how much uncertainty is resolved before the decision is made. Taking advantage of longitudinal expectations data from the BPS, we find that students have accurate perceptions about the amount of income uncertainty that is resolved during college but vastly underestimate the probability of dropping out of school. Consequently, on average, they underestimate this option value by 65%.

Chapter 4 proposes an alternative, model-based approach to jointly nonparametrically identify individuals' beliefs and the decision rule, which is a function that maps beliefs to decisions. My method can be applied to signal-based learning models, where individuals use signals to update their beliefs about an unknown permanent factor and repeatedly make decisions based on these beliefs. The econometrician observes individuals' decisions and the signals they receive at each period. Using data from the BPS, I apply my method to estimate the relationship between college students' study time and their beliefs about academic productivity as measured by the ratio of semester GPA to study time. I find that expectations about own academic productivity have a negative effect on study time. The RE assumption is rejected at a 10% level for a subgroup of students. Incorrectly imposing the RE assumption would lead to a substantially larger estimate of the effect of expectations about academic productivity on college study time.

Keywords: Rational Expectations assumption, expectations data, learning models, college education, income uncertainty, uncertainty resolution, option value

Summary for Lay Audience

People's decisions often depend on their beliefs about various future outcomes. For example, high school graduates make college attendance decisions partly based on their beliefs about the return to college education. Consequently, accurate characterization of these beliefs is of fundamental importance for understanding how decisions are made. The most commonly used approach is to impose the Rational Expectations (RE) assumption. However, empirical evidence suggests that the RE assumption might not hold in many contexts. Motivated by these findings, my thesis explores alternative approaches to conducting economic analysis without imposing the RE assumption.

Chapters 2 and 3 (co-authored with Todd Stinebrickner and Ralph Stinebrickner) utilize unique survey expectations data from the Berea Panel Study (BPS) to characterize college students' beliefs about various future outcomes. In Chapter 2, we find that students are quite uncertain about post-college income at the time of entrance. The majority of this initial income uncertainty remains unresolved by the end of college. A large fraction of the amount of uncertainty that is resolved during college can be attributed to learning about academic outcomes such as final GPA and final major.

Chapter 3 provides a concrete example that illustrates the importance of quantifying the resolution of students' (income) uncertainty during college. By entering college, students have the option to decide whether to remain in college after receiving relevant new information. We show that the value of this option is determined by a student's dropout probability and how much uncertainty is resolved before the decision is made. Taking advantage of longitudinal expectations data from the BPS, we find, on average, students underestimate this option value by 65%.

Recognizing the rarity of expectations data, Chapter 4 proposes an alternative, model-based approach to characterizing beliefs and investigating how these beliefs influence decisions. My method can be applied to environments where people repeatedly make decisions based on their beliefs about an unknown factor and update their beliefs using signals. Such environments are commonly studied in the existing literature. In the college education example, students learn about their return to college education from realized grade performance and repeatedly make dropout decisions.

Co-Authorship Statement

This thesis contains co-authored material. Chapters 2 and 3 are co-authored with Todd Stinebrickner and Ralph Stinebrickner. All authors are equally responsible for the work.

A version of Chapter 2 has been published in the Quantitative Economics with the following citation: "Yifan Gong, Todd Stinebrickner, and Ralph Stinebrickner. Uncertainty about future income: Initial beliefs and resolution during college. Quantitative Economics, 10(2):607–641, 2019." A version of Chapter 3 has been conditionally accepted for publication at the Journal of Applied Econometrics. Chapter 4 has not been submitted for publication.

Acknowlegements

I have been blessed with the opportunities to work with and learn from my advisor and co-author Todd Stinebrickner. As an advisor, his invaluable guidance, support, and encouragement have helped me through many difficult times. As a co-author, he is an ideal academic role model for me to follow. The experience that I gained from working on our joint projects has tremendously helped me grow as a researcher. I would like to take this opportunity to express my deepest gratitude.

I am greatly indebted to my thesis committee members, Nirav Mehta, Lance Lochner, and Tim Conley. I have tremendously benefited from their suggestions and support in many ways throughout my graduate study. I would like to also give special thanks to Ralph Stinbrickner for his invaluable input in the collection of the dataset used in this thesis (Berea Panel Study). The completion of this thesis would not have been possible without them.

My thesis has greatly benefited from participants of the Applied Economics/Econometrics workshop at Western, including, but certainly not limited to, Victor Aguiar, Roy Allen, Nail Kashaev, Salvador Navarro, David Rivers, and Chris Robinson. I would like to thank fellow graduate students, Cecilia Diaz Campo and Tian Liu, for taking time to write detailed and insightful reports for the last chapter of my thesis.

I owe a great deal of thanks to my fellow students, particularly to my cohort, Samantha Black, Matthew Carew, Fidel Perez Macal and Yuki Onozuka, and to the supportive Chinese community, Zhuang Liu, Sha Wang, Jin Zhou, Wenya Wang, Qian Liu, Ziyu Zhang, Mingshi Kang, and Zijian Wang. I am also grateful for all the support from the economics staff, Sandra Augustine, Morgan Bender, Karin Feulgen, Alexandra Houston, Jennifer Hope, Gary Kim, Leslie Kostal, Debra Merrifield, Maureen O'Connell, and Sharon Phillips. They have made my life as a PhD student at Western an enjoyable one.

I also wish to thank Charles Ka Yui Leung, who is my co-author and former advisor during my master study at the City University of Hong Kong. He exposed me to frontier economic research and encouraged me to pursue an academic career in economics.

I am deeply grateful to my parents and grandparents. It is regrettable that I have not been able to be around much in the past ten years, even when they needed me the most. They have never complained and always embrace me with their unconditional love and support. Finally, I would like to extend my sincere gratitude and appreciation to my fellow student and beloved wife, Yuxi Yao. We have been through many happy and hard times together and will go through many more together in the future. I cannot express how fortunate I feel that I received that phone call from my mother nine years ago which led me to go to Hong Kong for my master study. Without that call, I would not have met Yuxi and had the happy marriage that supported me through my PhD study.

Contents

List of Figures

List of Tables

List of Appendices

Chapter 1

Introduction

From a conceptual point of view, when an individual is making decisions, she often knows that the utility or value she will receive from each decision will depend on some factors that she does not have perfect information about at the time of decision-making. Consequently, her decisions depend on her beliefs about these factors. For example, the value of attending college largely depends on a student's potential income after obtaining a degree. Hence, a high school graduate's schooling decisions will tend to be influenced by her beliefs about future income in the scenario where she graduates from college.

The fundamental importance of beliefs about future outcomes in decision-making suggests that an accurate characterization of these beliefs is important for understanding individuals' behavior and for evaluating the counterfactuals typically needed for policy analysis. However, as is well known, it is difficult to separately identify individuals' beliefs and the decision rule, i.e., the function that maps beliefs to decisions (Manski, 2004). A common solution is to assume individuals have Rational Expectations (RE) so that their beliefs can be constructed from data on the realizations of future outcomes. However, empirical evidence suggests that the RE assumption might not hold in many contexts (see e.g., Pesaran and Weale, 2006, for a survey of tests for the RE assumption). Perhaps more importantly, a few recent articles find that, if the RE assumption is incorrectly imposed in the estimation of a structural model, the estimated structural parameters and policy predictions made based on the estimated model can be substantially biased (e.g., Gan et al., 2015, de Bresser, 2019). Motivated by these findings, I explore alternative approaches to conducting economic analysis without imposing the RE assumption. Specifically, I adopt both data-based and model-based approaches and focus on applications of these approaches in the context of higher education.

The data-based approach is built on the notion that maybe beliefs are best described as data that can be elicited using carefully worded survey expectations questions. This approach has become increasingly widely used in the literature (see Zafar, 2011, 2013, Wiswall and Zafar, 2015, 2016, Arcidiacono, Hotz, Maurel and Romano, 2019, Stinebrickner and Stinebrickner, 2014a/b for recent examples). Chapters 2 and 3 of my thesis, which are co-authored with Todd Stinebrickner and Ralph Stinebrickner, adopt this data-based approach. We utilize unique survey expectations data from the Berea Panel Study (BPS) to (1) characterize college students' beliefs about future outcomes, such as college completion and post-college income; (2) examine the relationship between these beliefs; and (3) investigate how these beliefs are related to various decisions made during college, such as whether to remain in college after receiving new information relevant for the return to college education.

Specifically, in Chapter 2, we characterize how much uncertainty about post-college income is present for students at college entrance and how quickly this uncertainty is resolved.[1] Measuring an individual's income uncertainty by the variance of the distribution describing her beliefs about earnings at age 28, we find that, on average, students resolve roughly one-third of the income uncertainty present at the time of entrance during college. Consistent with the finding that the majority of initial income uncertainty remains at the end of college, We find that uncertainty about college GPA and field of study, which are the two primary income-influencing factors that are realized in college, can only account for about 19% to 27% of students' initial income uncertainty. In addition, we find evidence that transitory factors, such as search frictions, are likely to play an important role in creating the income uncertainty remained at the end of college.

Chapter 3 of my thesis provides a concrete example that illustrates the importance of quantifying the resolution of students' (income) uncertainty during college. By entering college, students have the option to decide whether to remain in college after receiving relevant new information. We show that the value of this option of receiving new information is determined by a student's dropout probability and how much uncertainty is resolved before the decision is made. Taking advantage of unique longitudinal expectations data characterizing beliefs about dropout and future income, we find that students have accurate perceptions about the amount of income uncertainty that is resolved during college but vastly underestimate the probability of dropping out of school. Consequently, on average, they underestimate this option value by 65%. However, the fact that students underestimate the dropout probability despite having accurate perceptions about uncertainty resolution suggests that, at the time of entrance, they overestimate the value of college completion relative to the dropout alternative. We find that, considering the implications of our findings for college entrance decisions, the underappreciation of the value of new information is more than offset by over-optimism about the ex ante returns to college completion. Thus, once one takes into account both components of the over-

[1] A version of this chapter has been published in Quantitative Economics (Gong, Stinebrickner, and Stinebrickner, 2019).

all value of college, concerns that too few students enter college because of misperceptions tend to dissipate.

While the data-based approach is appealing in many contexts, expectations data are costly to collect, hence are not available in many large-scale datasets. Moreover, as found in Chapters 2 and 3, responses to survey expectations questions might contain measurement error that can potentially bias the results. Recognizing the rarity of survey expectations data and the measurement error issue, Chapter 4 of my thesis proposes an alternative, model-based approach to jointly nonparametrically identify individuals' beliefs and decision rules in signal-based learning models, where individuals use signals to update their beliefs about an unknown permanent factor and repeatedly make decisions based on these beliefs. The econometrician observes individuals' decisions and the signals they receive at each period. Identification builds on an assumption that is both intuitively appealing and standard in the literature: The posterior mean of the distribution describing beliefs is the same as the prior mean whenever the signal equals the prior mean. If an individual's decision only depends on the subjective mean in a time-invariant fashion, this assumption implies that the prior mean for an individual who does not change decisions in two consecutive periods equals the signal she receives between periods. My method can be applied to many models that are of interest to economists and policy-makers, including a firm's input choice problem under productivity/price uncertainty, and a student's dropout decision under uncertainty about the return to education. Using data from the Berea Panel Study, I demonstrate the empirical importance of relaxing the RE assumption by applying my method to estimate the relationship between college students' study time (analogous to a firm's input) and their beliefs about academic productivity (analogous to a firm's productivity) as measured by the ratio of semester GPA to study time. I find that high expectations about own academic productivity have a negative effect on students' study time. The RE assumption is rejected at a 10% level for students who spent less than 2 hours per day studying in high school. These students over-estimate their academic productivity in college. Incorrectly imposing the RE assumption would lead to a much more negative estimate of the effect of expectations about academic productivity on college study time, suggesting the importance of relaxing the RE assumption in this context.

1.1 Berea Panel Study

The empirical investigations conducted in Chapters 2-4 are based on the Berea Panel Study (BPS). While I will leave the detail of how to take advantage of this dataset for later chapters, here I briefly describe this survey project and discuss its importance for our empirical analysis.

Designed and administered by Todd Stinebrickner and Ralph Stinebrickner, the BPS is a multipurpose longitudinal survey project that followed two cohorts of students at the Berea College from their entrance in 2000 and 2001, until 2014. It collected detailed information of relevance for understanding a wide variety of issues in higher education, including those related to dropout, college major, time-use, social networks, peer effects, and transitions to the labor market.

Located in central Kentucky, Berea College has some unique features that have been documented in previous work. For example, it operates under the objective of providing educational opportunities to "students of great promise, but limited economics resources," and, as part of this objective, provides a full tuition subsidy to all students. Thus, as always, it is necessary to be appropriately cautious about the exact extent to which results from one school would generalize to other institutions. However, important for the notion that the basic lessons from this thesis are likely to be useful for thinking about what takes place elsewhere, Berea operates under a standard liberal arts curriculum and students at Berea are similar in academic quality, for example, to students at the University of Kentucky (Stinebrickner and Stinebrickner, 2008). Further, academic decisions and outcomes at Berea are similar to those found elsewhere (Stinebrickner and Stinebrickner, 2014a). For example, dropout rates are similar to the dropout rates at other schools (for students from similar backgrounds) and patterns of major choice and major-switching are similar to those found in the NLS by Arcidiacono (2004).

Most relevant for Chapters 2 and 3, the BPS had a specific focus on the collection of students' beliefs about various academic and labor market outcomes. Baseline surveys were administered to the first cohort (the 2000 cohort) immediately before it began its freshman year in the fall of 2000 and baseline surveys were administered to the second cohort (the 2001 cohort) immediately before it began its freshman year in the fall of 2001. An important aspect of the BPS in our context is that substantial follow-up surveys, which were administered at the beginning and end of each subsequent semester, documented how beliefs change over time.[2] Baseline surveys were completed in the presence of Todd Stinebrickner and/or Ralph Stinebrickner after students received classroom training. Subsequent in-school surveys were distributed through the campus mail system. Students returned completed surveys to Ralph Stinebrickner, who, after ensuring that surveys were completed in a conscientious manner, immediately provided compensation. This survey approach led to, not only high response rates, but also to, for example, virtually no item non-response.[3]

Much of the existing work using the BPS contributed to an early expectations literature that was interested in the quality of answers to expectations questions. As one example, Stine-

[2]The BPS is unique in its frequency of contact; each student was surveyed approximately 12 times each year while in school.

[3]BPS response rates were very high. Approximately 90% of all students who entered Berea College in 2001 responded to the baseline survey, and response rates were around 85% for subsequent in-school surveys.

brickner and Stinebrickner (2012) find that a simple theoretical implication related to college dropout - that the dropout decision should depend on both a student's cumulative GPA and beliefs about future GPA - is satisfied when beliefs are directly elicited through survey questions, but is not satisfied when beliefs are constructed under a version of Rational Expectations. As a second example, in Chapter 2, we propose and implement a method for characterizing the amount of measurement error in responses to expectations questions, which takes advantage of the fact that the BPS data often allow the unconditional subjective distribution of a particular outcome to be characterized using two different sets of expectations questions.

Of particular importance for Chapter 4, the BPS contains multiple 24-hour time diaries each semester, which allows me to construct a reasonably accurate measure of average daily study time for each semester. The BPS is linked with administrative data so I can observe semester GPA, and construct a measure of realized academic productivity for each student.

Bibliography

[1] Peter Arcidiacono. Ability sorting and the returns to college major. *Journal of Econometrics*, 121:343–375, 2004.

[2] Peter Arcidiacono, V Joseph Hotz, Arnaud Maurel, and Teresa Romano. Ex ante returns and occupational choice. *Working Paper*, 2019.

[3] Jochem de Bresser. Evaluating the accuracy of counterfactuals: the role of heterogeneous expectations in life cycle models. *working paper*, 2019.

[4] Li Gan, Guan Gong, Michael Hurd, and Daniel McFadden. Subjective mortality risk and bequests. *Journal of econometrics*, 188(2):514–525, 2015.

[5] Yifan Gong, Todd Stinebrickner, and Ralph Stinebrickner. Uncertainty about future income: Initial beliefs and resolution during college. *Quantitative Economics*, 10(2):607–641, 2019.

[6] Charles Manski. Measuring expectations. *Econometrica*, 72(5):1329–1376, 2004.

[7] Hashem Pesaran and Martin Weale. Survey expectations. *Handbook of economic forecasting*, 1:715–776, 2006.

[8] Ralph Stinebrickner and Todd Stinebrickner. Learning about academic ability and the college dropout decision. *Journal of Labor Economics*, 30:707–748, 2012.

[9] Ralph Stinebrickner and Todd Stinebrickner. Academic performance and college dropout: Using longitudinal expectations data to estimate a learning model. *Journal of Labor Economics*, 32(3):601–644, 2014.

[10] Ralph Stinebrickner and Todd Stinebrickner. A major in science? initial beliefs and final outcomes for college major and dropout. *Review of Economic Studies*, 81(1):426–472, 2014.

[11] Todd Stinebrickner and Ralph Stinebrickner. The causal effect of studying on academic performance. *The BE Journal of Economic Analysis & Policy*, 8(1), 2008.

[12] Matthew Wiswall and Basit Zafar. Determinants of college major choice: Identification using an information experiment. *The Review of Economic Studies*, 82(2):791–824, 2015.

[13] Matthew Wiswall and Basit Zafar. Human capital investments and expectations about career and family. *NBER Working Paper*, (w22543), 2016.

[14] Basit Zafar. How do college students form expectations? *Journal of Labor Economics*, 29(2):301–348, 2011.

[15] Basit Zafar. College major choice and the gender gap. *Journal of Human Resources*, 48(3):545–595, 2013.

Chapter 2

Uncertainty about Future Income: Initial Beliefs and Resolution During College

2.1 Introduction

From a conceptual standpoint, it is clear that the decision to enter or not enter college, as well as other college decisions, will depend on the amount of uncertainty about future income that is present at the time of college entrance.[1] However, college decisions will also be influenced by how quickly this initial uncertainty about future income is resolved. As one example, the option value of entering college will typically be higher when initial uncertainty is resolved more quickly. Further, the speed at which uncertainty is resolved is closely related to the important question of whether initial uncertainty is due to, for example, academic ability, college major, labor market frictions, future aggregate labor market conditions, or other factors.

A natural first step towards understanding how income uncertainty influences college decisions involves characterizing how much income uncertainty is present for students at the time of college entrance and how quickly (and why) this uncertainty is resolved.[2] Unfortunately, taking this first step has proven to be difficult (Cunha, Heckman, Navarro, 2005). This chapter takes advantage of unique expectations data from the Berea Panel Study (BPS), which is described in Section 2.2, to provide new evidence.[3] From the standpoint of characterizing uncertainty, the general benefit of the expectations approach is that survey questions can be de-

[1]More generally, Friedman(1953) suggests the importance of understanding the relative role of labor market uncertainty in determining distributions of wealth.

[2]Throughout the chapter our focus is on labor market income, and we use the terms earnings and income interchangeably.

[3]This approach is motivated by a recognition that individual beliefs about earnings (and other outcomes) are perhaps best viewed as data that can potentially be elicited using carefully worded survey questions (Manski, 1993, 2004, Dominitz and Manski, 1997a/b).

signed to elicit the entire distribution describing a student's beliefs about future income, which, for convenience, we often refer to as the student's subjective income distribution. Given our need to characterize income uncertainty throughout a student's entire time in college, a particular virtue of the BPS is that earnings expectations were collected longitudinally during college, with the first survey collection taking place at an ideal time – immediately before students began their first year courses. Our analysis also takes advantage of other unique expectations data available in the BPS. For example, information characterizing a student's beliefs about college grade performance and college major helps us understand why uncertainty is resolved.

In Section 2.3, we use beliefs elicited at the time of college entrance to characterize each student's initial amount of uncertainty about future earnings. The appeal of our direct, expectations-elicitation approach is in its simplicity. In contrast, traditional investigations require that an individual's beliefs about future earnings be ascertained from an observed distribution of realized earnings. This involves the challenge of decomposing the total amount of dispersion in realized earnings across workers into the portion due to individual-level uncertainty and the portion due to heterogeneity in ability and other income-influencing factors that are known by individuals. One tempting possibility might be to equate individual-level uncertainty with the amount of dispersion in earnings present within groups that are homogeneous in terms of observable earnings-influencing characteristics. However, when unobserved heterogeneity is prevalent (i.e., when many earnings-influencing characteristics are known to individuals but are not observed by the econometrician), this approach will tend to substantially overstate the amount of income variation that should be attributed to uncertainty.

In the schooling context, Carneiro, Hansen, and Heckman (2003) and Cunha, Heckman, and Navarro (2004, 2005) develop methods for separating uncertainty from heterogeneity that do not require the econometrician to observe all relevant characteristics that influence earning capabilities.[4] Specifically, they take advantage of situations where economic theory implies that the realization of uncertainty was unanticipated at the moment of decision making, and, therefore, was independent of the choices that economic agents made.[5] The general conclusion from these papers is that a substantial part of the variability in the ex post returns to schooling is predictable and acted on by agents. That is, "variability cannot be equated with uncertainty and this has important empirical consequences" (Cunha, Heckman, and Navarro, 2005).

Our results in Section 2.3 strongly reinforce this general message. At entrance, our measure of uncertainty, the standard deviation of the distribution describing a student's beliefs about her earnings at age 28, ranges from an average of $9,600 a year to an average of $13,900 a year,

[4] Cunha and Heckman (2007) provides a survey on this series of articles. See also Browning and Carro (2007) for a further discussion of the difficulties of separating uncertainty from heterogeneity.

[5] See also Blundell and Preston (1998) for early work using similar methods in a somewhat different substantive context.

across the different computational approaches that we take to ensure robustness. To characterize the relative importance of uncertainty and heterogeneity, we compute an expectations analog to the realized earnings distribution used in other papers by aggregating individual beliefs across the sample. The percentage of the total variation in this analog that should be attributed to (observed and unobserved) heterogeneity is always above 50% and is as high as 77%, depending on which computational approach is employed. We find that results do not change substantially when we correct for classical measurement error that might arise in the responses to the survey questions. This measurement error correction is made possible by the fact that there are two different sets of survey questions in the BPS that can be used to construct beliefs about future earnings.

In Section 2.4, we turn to examining issues related to the resolution of income uncertainty, with a particular focus on what happens during college. Given that empirical work has not typically examined these issues, it is an open question whether individuals believe that uncertainty will be resolved quickly after college entrance.[6] This issue is directly linked to the question of why uncertainty exists. For example, one particularly prominent potential source of uncertainty is college grade point average (GPA), which is widely viewed as the best available proxy for human capital at the time of college graduation. By definition, all uncertainty about final college GPA will be resolved by the end of college. Thus, if uncertainty about GPA is an important contributor to the initial uncertainty about earnings, then students will expect much of the uncertainty about earnings to be resolved at some point during college and that this resolution will take place early in college if learning about academic ability tends to happen quickly.[7] We are able to provide evidence about the importance of grade uncertainty in determining initial earnings uncertainty by taking advantage of survey questions eliciting beliefs about grade performance and survey questions eliciting beliefs about future earnings conditional on grade performance. We find that, on average, between 16% and 19% of the variance representing (age 28) earnings uncertainty at the time of college entrance can be attributed to uncertainty about grade performance at the time of college entrance. A related analysis finds that between 11% and 17% of the earnings uncertainty at the time of college entrance can be attributed to uncertainty about college major at the time of college entrance. Moreover, when combined, uncertainty about these two factors together can account for about 19% to 27% of overall initial uncertainty about future income.

[6]An exception is Navarro and Zhou (2017) who develop a model that identifies the path of uncertainty resolution over multiple periods. With each period having a length of six years, their first period (age 18-24) corresponds to the time that our sample spends in college and the first two years in the workforce.

[7]See Stinebrickner and Stinebrickner (2012, 2014b) and Zafar (2011) for research that uses expectations data to examine updating of beliefs about grade performance. See Altonji (1993) for early work recognizing the role that grade updating may play in schooling decisions.

The finding that students expect much uncertainty about earnings to remain even after resolving uncertainty about grade performance and college major raises the possibility that much uncertainty about earnings remains even at the end of college. The longitudinal nature of our expectations data allow us to examine this issue. We find that, on average, about 65% of a student's initial uncertainty about future earnings remains at the end of college. Further, this result, combined with the results in the end of the previous paragraph, suggests that the portion of uncertainty that is resolved during school can be largely attributed to what one learns about her academic ability and her college major during school.

It is worth considering why much of the initial uncertainty about earnings at age 28 is unresolved during college. We consider two broad explanations that may have different policy implications. The first explanation is that individuals might be unsure about what kinds of job offers they will receive at age 28 because of, for example, the existence of search frictions. The second explanation is that individuals might know the kinds of job offers they would receive at age 28, but might be unsure about which kinds of available job offers they will prefer/choose at this age.

2.2 Data

Our primary sample consists of the 650 students who answered the baseline surveys of the Berea Panel Study, which were collected immediately after students entering college.[8] While observable characteristics are not the primary focus of this chapter, we note that approximately 41% of the students in the sample are male, 15% of the students in the sample are black, and the average American College Test (ACT) score in the sample is approximately 25. In addition to collecting detailed background information, the baseline surveys were designed to take advantage of recent advances in survey methodology to collect beliefs (expectations) about future outcomes. An important aspect of the BPS in our context is that substantial follow-up surveys, which were administered at the beginning and end of each subsequent semester, documented how beliefs change over time.[9]

Our primary survey questions eliciting beliefs about future earnings are of the form of baseline Survey Question 1A, which is shown in Appendix A.[10] Specifically, Survey Question 1A elicited the minimum, the maximum, and the three quartiles of the subjective income distribu-

[8]Approximately 85% of all students who entered Berea in the fall of 2000 and the fall of 2001 completed the baseline surveys and, in part because surveys were reviewed before students left the survey site, the amount of item non-response was trivial.

[9]The BPS is unique in its frequency of contact; each student was surveyed approximately 12 times each year while in school.

[10]For another example of research that uses an expectations-based approach to elicit information about the entire distribution of future income, see Attanasio and Kaufmann (2014).

tion at three different ages (first year after graduation, age 28, and age 38), under a scenario in which the student graduates from college. Students received detailed classroom instruction related specifically to these questions, with the spirit of the discussion being similar to written instructions that were included with the survey (see Appendix A for these instructions). An almost identical set of questions (not shown) was used to elicit beliefs under the scenario in which the student does not graduate from college. A baseline survey question also elicited beliefs about earnings conditional on graduating with three particular levels of GPA (2.00, 3.00, 3.75). Question 1B in Appendix A shows the portion of this question related to graduating with a 2.00 GPA.

Table 2.1 shows descriptive statistics related to Question 1. The entries in the first row show the median (the second quartile) of the subjective income distribution, averaged over the sample, for several different age and academic performance scenarios. The first three columns show that, on average, the median increases with age. The second three columns show that, on average, the median increases with final grade point average. To provide some descriptive evidence about uncertainty, the entries in the second row show the interquartile range (the difference between the third quartile and the first quartile) of the subjective income distribution, averaged over the sample, for the same age and academic performance scenarios. The first three rows show that, on average, the interquartile range increases with age. The second three columns show that, on average, the interquartile range increases with final grade point average.

Table 2.1: Descriptive Statistics of Earnings Beliefs at Entrance

	1 Year Out	Age 28	Age 38	Age 28 GPA = 2.00	Age 28 GPA = 3.00	Age 28 GPA = 3.75
Median	39.5480 (18.3900)	49.1923 (21.9922)	60.5161 (36.7525)	41.8088 (21.7551)	48.1623 (23.7830)	54.7238 (26.3292)
Interquartile Range	12.6773 (10.3599)	15.3221 (12.8526)	19.2754 (30.7070)	12.3756 (10.6551)	13.8969 (12.1135)	15.9806 (13.6923)

Note: The unit of measurement for all entries is one thousand dollars. A particular entry in the table shows the sample mean and the sample standard deviation of the corresponding variable. For example, row 1, column 1 shows a sample mean of $39548.00 and a sample standard deviation of $18390.00 for the median of the distribution describing a student's beliefs about income in the first year out of college. Similarly, row 1, column 4 shows a sample mean of $41808.80 and a sample standard deviation of $21755.10 for the median of the distribution describing a student's beliefs about income at age 28 given that her final GPA is equal to 2.00.

Baseline Survey Question 2, which characterizes beliefs about future grade performance by eliciting the probabilities that a student's future semester grade point average will fall in the intervals [3.5, 4.00], [3.0, 3.49], [2.5, 2.99], [2.0, 2.49], [1.0, 1.99] and [0.0, .99], is also shown in Appendix A. In terms of other baseline information, this chapter takes advantage of survey questions eliciting each student's subjective probability of completing a degree in different possible major groups (Question 5, Appendix A), and each student's belief about how much noise exists in the grade process (Question 3, Appendix A).

2.3 Uncertainty about Future Income at College Entrance

This section examines uncertainty about future income at the time of college entrance. In Section 2.3.1, we characterize the amount of uncertainty that exists at college entrance. In Section 2.3.2, we construct an expectations analog to the realized earnings distribution and examine the relative importance of uncertainty and heterogeneity in determining the variance of this distribution.

2.3.1 Characterizing Uncertainty at Time of College Entrance

When measuring earnings uncertainty, we focus on earnings under the scenario in which a student graduates from college and, unless otherwise noted, examine beliefs about earnings at the age of 28.[11] The general object of interest is the distribution describing a student's subjective beliefs about her future income, which, as noted earlier, we often refer to as the student's subjective income distribution. While this entire section focuses on beliefs at the time of entrance, which we often refer to as "initial" beliefs, we include a time subscript in our notation for use in subsequent sections. We let w_i denote the earnings of person i at age 28, W_{it} denote the random variable describing student i's subjective beliefs at time t about w_i, and $f_{W_{it}}(w_{it})$ denote the density of W_{it}. Then, the standard deviation and variance of W_{it} are natural measures of a student's uncertainty about w_i at time t. Our objectives related to the issue of uncertainty motivate a focus on measures of dispersion, although it is necessary for parts of our analysis to also characterize measures of central tendency (e.g., the mean of W_{it}), which have received substantial attention in other previous work.

Our data allow us to take two different approaches for computing the standard deviation (and mean) of W_{it} from survey information. The first approach takes advantage of Survey Question 1A (Appendix A), which directly elicited the minimum, maximum, and three quartiles of the subjective income distribution. The standard deviation can be computed directly from this information given a distributional assumption for W_{it}. The second approach takes advantage of Survey Question 1B (Appendix A), which elicited the minimum, maximum, and

[11] We focus on the graduation scenario because this is the outcome that students overwhelmingly believe is most likely. Specifically, at the time of entrance, students believe, on average, that the probability of dropping out is only 0.14 (Question 4, Appendix A).

three quartiles of the subjective income distribution conditional on various levels of grade performance, and Survey Questions 2 and 3 (Appendix A), which provide information about a student's subjective grade distribution. While the second approach has the appeal of explicitly taking into account one particularly prominent source of income uncertainty – uncertainty about grade performance – it also requires additional survey questions and additional assumptions. Given the trade-offs between the two approaches, examining whether they yield similar results is valuable as a robustness check. In addition, the comparison is valuable because each of these approaches is utilized in other parts of our analysis.

Approach 1 for characterizing the standard deviation of W_{it}

Our first approach for characterizing income uncertainty takes advantage of information that was elicited by Question 1A about the unconditional distribution of W_{it}. We denote the elicited minimum, first quartile, second quartile, third quartile, and maximum of the distribution of W_{it} as $C^1_{it}, C^2_{it}, C^3_{it}, C^4_{it}$ and C^5_{it}, respectively. Characterizing the mean and standard deviation of W_{it} from this information requires a distributional assumption for W_{it}. We examine the robustness of our results to three different distributional assumptions.

a. Log-normal. We first consider the use of a log-normal distribution, following the suggestions in Manski (2004). The mean and standard deviation for the log-normal distribution are given by $E(W_{it}) = C_3 e^{\sigma^2/2}$ and $std(W_{it}) = E(W_{it})\sqrt{e^{\sigma^2} - 1}$, where $\sigma = log(\frac{C^4_{it}}{C^2_{it}})/2\Phi^{-1}(0.75)$ and Φ is the standard normal cumulative distribution function.

b. Normal. The log-normal distribution imposes an asymmetry that may or may not be present in the data. While the log-normal does have the appealing feature of ruling out negative income, the probability of negative income will tend to be small for the normal distribution when, as we find in our data, the mean is relatively large compared to the standard deviation. As described in Appendix B.1, we find that the fit of the two distributions is quite similar with, if anything, the normal having a slightly better fit. Then, given that these two distributions can potentially have quite different implications for characterizing the mean and variance, it seems worthwhile for robustness reasons to consider each of them. The mean and standard deviation of the normal distribution are given by $E(W_{it}) = C^3_{it}$ and $std(W_{it}) = (C^4_{it} - C^2_{it})/2\Phi^{-1}(0.75)$.

c. Stepwise Uniform. The log-normal and normal distributions do not utilize information about the minimum, C^1_{it}, or the maximum, C^5_{it}, because the supports of the distributions are R_{++} and R, respectively. To allow for a specification that uses these values along with the quartiles, we assume that W_{it} has the stepwise uniform pdf given by:

$$f_{W_{it}}(w_{it}) = \frac{0.25}{C_{it}^{n+1} - C_{it}^{n}}, \text{ if } w_{it} \in [C_{it}^{n}, C_{it}^{n+1}], \text{ for } n \in \{1, 2, 3, 4\}. \quad (2.1)$$

The mean and standard deviation are given by $E(W_{it}) = \sum_{n=1}^{4} \frac{C_{it}^{n+1}+C_{it}^{n}}{8}$ and $std(W_{it}) = \sqrt{\sum_{n=1}^{4} \frac{(C_{it}^{n+1})^2+C_{it}^{n+1}C_{it}^{n}+(C_{it}^{n})^2}{12} - (E(W_{it}))^2}$.

We examine the magnitude of earnings uncertainty at the time of college entrance ($t = 0$) for our sample of 650 students. The first three rows of Table 2.2 summarize the results for Approach 1. Depending on which distributional assumption is made (log-normal, normal, stepwise uniform), the average standard deviation of W_{i0} for the sample varies between \$9,653 and \$13,064 per year and the average standard deviation to mean ratio in the sample varies between 18.95% and 24.17% per year.[12] Thus, the results are generally quite similar across the three distributional assumptions. The numbers in parentheses in the standard deviation column of Table 2.2 indicate that there is substantial heterogeneity in uncertainty across students.

Approach 2 for characterizing the standard deviation of W_{it}

Letting g_i denote the final (cumulative) college GPA of person i and letting G_{it} denote the random variable describing student i's subjective beliefs at time t about g_i, our second approach for characterizing income uncertainty takes advantage of information that was elicited about the distribution of G_{it} and about the distribution of W_{it} conditional on G_{it}. The relationship between these distributions and the unconditional income distribution is given by:

$$f_{W_{it}}(w_{it}) = \int f_{W_{it}|G_{it}=g_{it}}(w_{it}) dF_{G_{it}}(g_{it}), \quad (2.2)$$

where g_{it} is a realization of G_{it} and where $F_{G_{it}}(g_{it})$ and $f_{W_{it}|G_{it}=g_{it}}(w_{it})$ denote the cdf of G_{it} and the pdf of $W_{it}|G_{it} = g_{it}$, respectively.

The analysis in this chapter mostly utilizes the mean, $E(W_{it})$, and the standard deviation, $std(W_{it})$, of W_{it}. We first consider $E(W_{it})$, which can be written as the expected value of

[12]Using log-normal distributions leads to the largest mean and standard deviation approximations and using stepwise uniform distributions leads to the smallest. Note that the distributions constructed using each of these two distributional assumptions share the same median. Hence, loosely speaking, log-normal distributions tend to have larger expectations because they are more left-skewed than the stepwise uniform distributions. While log-normal density functions have wider supports than stepwise uniform density functions, they also have different shapes which, all else equal, can lead to smaller standard deviations. Hence, the relative size of the standard deviations implied by the two distributions is theoretically ambiguous. In our case, the wider-support effect dominates the other effect.

Table 2.2: Earnings Beliefs at Entrance

# of Observations: 650	$E(W_{i0})$	$std(W_{i0})$	$\frac{std(W_{i0})}{E(W_{i0})}$
Approach 1, Log-normal	51.1742 (23.2062)	13.0641 (15.5580)	0.2417 (0.2055)
Approach 1, Normal	49.1524 (21.9879)	11.3152 (9.4768)	0.2295 (0.1617)
Approach 1, Stepwise Uniform	49.7633 (22.1799)	9.6529 (8.0391)	0.1895 (0.1165)
Approach 2, Log-normal	52.8264 (25.4735)	13.9455 (14.4897)	0.2557 (0.1736)
Approach 2, Normal	50.9118 (24.3631)	12.2488 (9.7170)	0.2427 (0.1435)
Approach 2, Stepwise Uniform	51.3497 (24.3550)	10.6537 (8.1362)	0.2061 (0.1142)

Note: The unit of measurement for W_{i0} is one thousand dollars. A particular entry in the table shows the sample mean and the sample standard deviation of the corresponding variable. For example, row 1, column 1 shows a sample mean of $51,174.20 and a sample standard deviation of $23,206.20 for $E(W_{i0})$. Similarly, row 1, column 2 shows a sample mean of $13,064.10 and a sample standard deviation of $15,558.00 for $std(W_{i0})$.

$E(W_{it}|G_{it})$ with respect to G_{it}. In cases like this, where an expresssion of interest involves iterated expectations (or variances), it is often useful for reasons of clarity to be explicit about the random variable on which the outer expectation (or variance) operates. Using this notational device,

$$E(W_{it}) = E_{G_{it}}(E(W_{it}|G_{it})).^{13} \tag{2.3}$$

We use a standard simulation-based method to approximate this integral, which requires repeatedly drawing from the distribution of G_{it} and evaluating $E(W_{it}|G_{it})$ at each of these draws. The complication that arises, in practice, is that $E(W_{it}|G_{it})$ and $F_{G_{it}}(g_{it})$ are not fully observed.

With respect to $E(W_{it}|G_{it})$, the complication arises because, as discussed in Section 2.2, a student reports information about her subjective conditional income distribution for only three different realizations of G_{it}: 3.75, 3.00, and 2.00. For these three g_{it} values, $E(W_{it}|G_{it})$ can be computed by assuming one of the three distributions. As described in detail in Appendix B.2.1, we interpolate the value of $E(W_{it}|G_{it})$ conditional on other realizations of G_{it} using an approach adopted in Stinebrickner and Stinebrickner (2014b).

With respect to $F_{G_{it}}(g_{it})$, the complication arises because the BPS did not directly elicit G_{it}, a student's beliefs at time t about final cumulative GPA, G_i. Given that a student's grades

[13] $E_{G_{it}}(E(W_{it}|G_{it})) = \int E(W_{it}|G_{it} = g_{it})dF_{G_{it}}(g_{it})$, with $E(W_{it}|G_{it} = g_{it}) = \int w_{it} f_{W_{it}|G_{it}=g_{it}}(w_{it})dw_{it}$.

before time t are observed in administrative data, the challenge in determining G_{it} comes from the need to characterize the student's beliefs at t about the average GPA (i.e., the cumulative GPA) she will receive over all remaining (future) semesters in school. The primary source of information used to construct these beliefs is Survey Question 2 (Appendix A), which elicits beliefs about semester GPA. However, even making the natural assumption that Question 2 represents a student's beliefs about semester GPA in each future semester, Question 2 alone is not enough to determine how uncertain a student is about the average GPA she will receive over all remaining semesters. This is the case because one's uncertainty about average GPA over multiple semesters will depend on beliefs about the correlation in semester GPA across semesters. For example, if uncertainty about semester GPA arises because of uncertainty about a factor such as ability that is permanent in nature, and, therefore, will tend to influence grades in each semester, then the uncertainty about semester GPA expressed in Question 2 will tend to be a good indicator of the student's uncertainty about average GPA over multiple semesters. On the other hand, if uncertainty about semester GPA arises because of semester-specific randomness in grades which is transitory in nature, and, therefore, will tend to average out to some extent over multiple semesters, then the uncertainty about semester GPA expressed in Question 2 might substantially overstate the student's uncertainty about average GPA over multiple semesters.[14] Our approach for characterizing a student's subjective beliefs about the cumulative GPA she will receive over all remaining semesters differentiates between these two types of possibilities by taking advantage of a novel survey question (Question 3 in Appendix A), which elicited beliefs about the importance of the semester-specific randomness. Appendix B.2.2 describes this approach in detail, focusing, for illustrative purposes, on the case of $t = 0$, which is of relevance in this section.

We now turn our attention to the measure of dispersion, $std(W_{it})$, which is given by:

$$std(W_{it}) = \sqrt{var_{G_{it}}(E(W_{it}|G_{it})) + E_{G_{it}}(var(W_{it}|G_{it}))}.^{15} \tag{2.4}$$

The value of $std(W_{it})$ can be approximated in a manner very similar to that described in the previous paragraphs for the approximation of $E(W_{it})$. Equation (2.4) shows that, in addition to using an interpolation approach to deal with the issue that $E(W_{it}|G_{it})$ and $F_{G_{it}}(g_{it})$ are not fully observed, it is also necessary to interpolate the value of $var(W_{it}|G_{it})$ at realizations of G_{it} other than 2.00, 3.00 or 3.75. The details of our interpolation approach are described in Appendix B.2.1.

---

[14] This randomness might be due to, for example, bad matches with instructors, sicknesses at inopportune times, or temporary personal problems.

[15] $Var_{G_{it}}(E(W_{it}|G_{it})) = \int (E(W_{it}|G_{it} = g_{it})) - E_{G_{it}}(E(W_{it}|G_{it} = g_{it}))^2 dF_{G_{it}}(g_{it})$ and $E_{G_{it}}(Var(W_{it}|G_{it})) = \int Var(W_{it}|G_{it} = g_{it}) dF_{G_{it}}(g_{it})$, with $Var(W_{it}|G_{it} = g_{it}) = \int (w_{it} - E(W_{it}|G_{it} = g_{it}))^2 f_{W_{it}|G_{it}=g_{it}}(w_{it}) dw_{it}$.

Using Approach 2, we examine the magnitude of earnings uncertainty for the same sample of 650 students. Results are summarized in the last three rows of Table 2.2. Depending on which distributional assumption is made, the average standard deviation of W_{i0} for the sample varies between \$10,654 and \$13,946 per year and the average standard deviation to mean ratio in the sample varies between 20.61% and 25.57% per year. Thus, we find that the results are reasonably robust to two computation approaches. In fact, results change more due to the choice of distribution than to the choice of computational approach.

Demographic Variables

It is worth examining whether the amount of uncertainty that is present at the time of entrance varies systematically with demographic information. To examine this issue, we regress $std(W_{i0})$ on Black, Male and ACT score for each of the six different distribution-approach combinations in Table 2.2. We find a seemingly important role for race. While full regression results are not shown, taking the average of estimated coefficients over the six different combinations, we find that black students have a standard deviation that is approximately \$1,536 higher than non-blacks. Further, the Black coefficient has a t-statistic greater than 1.5 in four of the six distribution-approach combinations, with the maximum t-statistic having a value of 2.6. Comparing these findings to those for our other binary variable, Male, we find that the coefficient for Male also has a t-statistic greater than 1.5 for four of the six combinations, but that the average coefficient for Male over the six distribution-approach combinations is only approximately 62% of the average coefficient for Black.

We stress that understanding the exact interpretation of these results is beyond the scope of this chapter. Among other things, interpretation is complicated by the fact that uncertainty could be caused by a lack of information, but it could also be caused by potential access to a wide range of job opportunities. The possibility that these two effects may sometimes push in opposite directions may explain, for example, why we do not find evidence of a relationship between ACT score and uncertainty.

2.3.2 Heterogeneity vs. Uncertainty

Traditionally, estimating the amount of uncertainty about earnings that is present at college entrance requires separating the importance of this uncertainty from the importance of heterogeneity - differences in ability and other income-influencing factors known by individuals - in determining a realized distribution of income. Thus, while characterizing the amount of uncertainty that is present at the time of college entrance is reasonably viewed as the primary goal, past work has found it natural to also report the percentage of the total variation in earnings that is due to this uncertainty. In Section 2.3.2 we compute an expectations analog to this percentage. We also examine the robustness of our results to a measurement error correction and describe how our expectations analog relates to the approach surveyed in Cunha and Heckman (2007). Given this discussion, we conclude that our results reinforce their findings.

Decomposition of heterogeneity and uncertainty

Suppose that a person's earnings in a future year (e.g., age 28) are determined by a vector of finitely many random variables X_i.[16] Further decompose X_i into factors that are observed by the students at t, X_i^{t-}, and those that are not, X_i^{t+}, and define $X_i \equiv (X_i^{t-}, X_i^{t+})$. Then, we can write the future income of student i, W_i, as:

$$W_i \equiv \mathcal{W}(X_i^{t-}, X_i^{t+}). \tag{2.5}$$

Although, a priori, individuals have identical distributions of X_i^{t-} and X_i^{t+}, realizations of these random variables vary across people. It is differences in these realizations that produce variation in the empirical earnings distribution. At the time t when individuals answer the survey, they have already observed X_i^{t-}. Heterogeneity in X_i^{t-} produces differences in the beliefs we observe as given by the distribution of W_{it}. To construct the expectations analog to the empirical earnings distribution, we take advantage of the fact that $var(W_i)$ can be written as a function of the conditional distributions that we observe:

$$var(W_i) = E_{X_i^{t-}}(var(W_i|X_i^{t-})) + var_{X_i^{t-}}(E(W_i|X_i^{t-})). \tag{2.6}$$

Under the assumption that X_i is independently distributed across students, taking an expectation with respect to X_i^{t-} is, in essence, averaging across individuals (whose beliefs about income at time t differ only through X_i^{t-}).[17] The first term on the right hand side of equation (2.6) shows, on average, how uncertain individuals are about earnings. Thus, this term represents the contribution of uncertainty to total variation. Using either of the two approaches in Section 2.3.1, we are able to compute the sample analog of this term as the sample mean of $var(W_{it})$. Similarly, taking a variance with respect to X_i^{t-} is, in essence, measuring dispersion across individuals. The second term on the right hand side shows how much dispersion exists in expected earnings across individuals, arising from the heterogeneity term X_i^{t-}. Therefore, this second term represents the contribution of heterogeneity to total variation. Using either of

[16] Note that these random variables represent both factors related to the worker and factors related to the labor market.

[17] In Section 2.3.2, we discuss scenarios under which the independence assumption would tend to be violated and the implications of these scenarios.

the two approaches in Section 2.3.1, we are able to compute the sample analog of this term as the sample variance of $E(W_{it})$.

Note that if beliefs are correct, i.e., if $W_{it} \equiv W_i | X_i^{t-}$, the sum of the two terms will correspond to the variance of the realized income distribution. If beliefs are not correct, the sum of the terms corresponds to what individuals believe about the the variance of the realized income distribution.

For each of our six approach-distribution combinations, the first column of Table 2.3 shows the first (uncertainty) term from equation (2.6), the second column shows the second (heterogeneity) term from equation (2.6), the third column shows the sum of the first two columns (the total variation), and the final column shows the ratio of the second column (heterogeneity) to the third column (total variation).

Table 2.3: Heterogeneity and Uncertainty

# of Observations: 650	Uncertainty: Sample Mean of $var(W_{i0})$	Heterogeneity: Sample Variance of $E(W_{i0})$	Total	Heterogeneity Ratio
Approach 1, Stepwise Uniform	157.7	491.9	649.7	75.72%
Approach 1, Log-normal	412.4	538.5	950.9	56.63%
Approach 1, Normal	217.7	483.5	701.2	68.95%
Approach 2, Stepwise Uniform	179.6	593.2	772.8	76.76%
Approach 2, Log-normal	404.1	648.9	1,053.0	61.62%
Approach 2, Normal	244.3	593.6	837.9	70.84%

Note: The unit of measurement for W_{i0} is one thousand dollars. The third column (Total) is the sum of the first two columns. The fourth column (Heterogeneity Ratio) is the ratio of column 2 (Heterogeneity) to column 3 (Total).

Consistent with what we found earlier, Approach 1 and Approach 2 deliver results that are quite similar. While larger differences in results are generated by the distributional assumption than by the choice of computational approach (Approach 1 and Approach 2 in Section 2.3.1, all three of the distributional assumptions suggest a large role for heterogeneity. For the stepwise uniform distribution, heterogeneity accounts for over 75% of overall variation. This percentage is approximately 60% and 70% for the log-normal distribution and the normal distribution, respectively.

Allowing for measurement error

While the conceptual virtues of expectations data are well-recognized, it is generally difficult to know the extent to which the benefits of this approach are mitigated by, for example, measurement error in responses to expectations questions. In our context, classical measurement error in the income expectations responses would tend to lead to an overstatement of the importance of heterogeneity relative to the importance of uncertainty. This is the case because, as can be seen in equation (2.6), the measured contribution of heterogeneity (the second term) is represented by a sample variance (which will tend to increase with the amount of classical measurement error), while the measured contribution of uncertainty (the first term) is represented by a sample mean (which will tend to be consistent even in the presence of classical measurement error). To provide some evidence about the quantitative importance of measurement error, we take advantage of the fact that our two computational approaches in Section 2.3.1 allow us to compute $E(W_{it})$ in two separate ways. We refer to the computed values from Approach 1 and Approach 2 as $\widetilde{E^1}(W_{it})$ and $\widetilde{E^2}(W_{it})$, respectively. The intuition underlying the measurement error correction is that, in an environment with no interpolation, the two computed values will be identical if the responses to the survey questions used to compute these values are not affected by measurement error. However, when the two computed values are different, the importance of measurement error can be ascertained if one specifies the manner in which measurement error affects the responses to the survey questions.

Starting with Approach 1, the computed value $\widetilde{E^1}(W_{it})$ comes directly from Question 1A (which elicits the unconditional subjective income distribution). We assume that measurement error enters the computed value $\widetilde{E^1}(W_{it})$ in a classical manner;

$$\widetilde{E^1}(W_{it}) = E(W_{it}) + \varsigma_i, \tag{2.7}$$

where ς_i is the classical measurement error attached to the true value $E(W_{it})$. Dispersion in the computed value, $\widetilde{E^1}(W_{it})$, across students originates from both dispersion in the true value, $E(W_{it})$, across students and randomness caused by measurement error, ς_i. This can be seen by taking the variance of both sides of equation (2.7):

$$var(\widetilde{E^1}(W_{it})) = var(E(W_{it})) + var(\varsigma_i). \tag{2.8}$$

Equation (2.8) reveals that the true contribution of heterogeneity, $var(E(W_{it}))$, can be obtained by subtracting the variance of the measurement error, ς_i, from the measured contribution of heterogeneity, $var(\widetilde{E^1}(W_{it}))$. Thus, the remainder of this section focuses on estimating the variance of ς_i.

Turning to Approach 2, the value $\widetilde{E^2}(W_{it})$ is computed from the responses to questions eliciting beliefs about income conditional on the three particular realizations of final GPA (questions such as 1B) as well as questions eliciting beliefs about grade performance (Questions 2 and 3). Similar to the assumption made in equation (2.7), we assume that measurement error influences the responses to questions such as 1B in a classical manner, that is,

$$\widetilde{E}(W_{it}|G_{it} = g_{it}) = E(W_{it}|G_{it} = g_{it}) + \varsigma_i^{g_{it}} \qquad g_{it} = 2.00, 3.00 \text{ or } 3.75, \tag{2.9}$$

where $\widetilde{E}(W_{it}|G_{it} = g_{it})$ is the measured value of the true value $E(W_{it}|G_{it} = g_{it})$ and $\varsigma_i^{g_{it}}$, g_{it} = 2.00, 3.00 or 3.75, are the corresponding classical measurement errors.

As discussed in Section 2.3.1, the computation of $\widetilde{E^2}(W_{it})$ requires information on $\widetilde{E}(W_{it}|G_{it})$ at all realizations of G_{it} and the distribution of G_{it}. However, because we only observe the measured value $\widetilde{E}(W_{it}|G_{it})$ for three specific realizations of G_{it}, we need to interpolate the value of $\widetilde{E}(W_{it}|G_{it})$ at other realizations. Under the interpolation approach that we adopted in Section 2.3.1, $\widetilde{E^2}(W_{it})$ can be written as a weighted sum of $\widetilde{E}(W_{it}|G_{it} = 2.0)$, $\widetilde{E}(W_{it}|G_{it} = 3.0)$, and $\widetilde{E}(W_{it}|G_{it} = 3.75)$:

$$\widetilde{E^2}(W_{it}) = \sum_{g_{it}} \lambda_i^{g_{it}} \widetilde{E}(W_{it}|G_{it} = g_{it}) \qquad g_{it} = 2.00, 3.00 \text{ or } 3.75, \tag{2.10}$$

where, as shown in Appendix B.3, the weights $\lambda_i^{2.0}$, $\lambda_i^{3.0}$, and $\lambda_i^{3.75}$ are integrals that depend on the distribution of G_{it}. Here, we assume that no errors are introduced by the interpolation approach. However, in Appendix F we discuss why our conclusion about the importance of heterogeneity in this section will tend to be conservative if this type of interpolation error exists or if error is introduced during the computation of G_{it}.

Combining equation (2.9) and equation (2.10), we obtain the following equation:

$$\begin{aligned}\widetilde{E^2}(W_{it}) &= \sum_{g_{it}} \lambda_i^{g_{it}} E(W_{it}|G_{it} = g_{it}) + \sum_{g_{it}} \lambda_i^{g_{it}} \varsigma_i^{g_{it}} \\ &= E(W_{it}) + \sum_{g_{it}} \lambda_i^{g_{it}} \varsigma_i^{g_{it}}.\end{aligned} \tag{2.11}$$

Taking the difference between the mean computed using Approach 1 and the mean computed using Approach 2, we obtain:

$$\widetilde{E^1}(W_{it}) - \widetilde{E^2}(W_{it}) = \varsigma_i - \sum_{g_{it}} \lambda_i^{g_{it}} \varsigma_i^{g_{it}}. \tag{2.12}$$

Using equation (2.12) to estimate $var(\varsigma_i)$ requires assumptions about the joint distribution of ς_i, $\varsigma^{2.0}$, $\varsigma^{3.0}$ and $\varsigma^{3.75}$. The prior assumption that ς_i and $\varsigma_i^{g_{it}}$s represent classical measurement error implies that they have mean zero and are independent of other factors. In addition, we assume that the four measurement error terms are independent and identically distributed.

Under these assumptions, as shown in Appendix B.4,

$$var(\varsigma_i) = \frac{var(\widetilde{E^1}(W_{it}) - \widetilde{E^2}(W_{it}))}{1 + \sum_{g_{it}} E((\lambda_i^{g_{it}})^2)}. \tag{2.13}$$

Note that we can compute the sample analogs of $var(\widetilde{E^1}(W_{it}) - \widetilde{E^2}(W_{it}))$ and $E((\lambda_i^{g_{it}})^2)$ from data available to us.[18] Hence, $var(\varsigma_i)$ can be estimated. The first column of Table 2.4 reports the estimates of $var(\varsigma_i)$. Subtracting the measurement error component from measured heterogeneity (column 2 in Table 2.4 for the three rows associated with Approach 1) yields the magnitude of true heterogeneity $var(E(W_{it}))$, which is reported in the second column. In the third column, we report the adjusted heterogeneity ratio, which is defined as the ratio of true heterogeneity (column 2 in Table 2.4) to the sum of true heterogeneity (column 2 in Table 2.4) and uncertainty (column 1 in Table 2.3).

We find that the magnitude of measurement error is relatively small compared to measured heterogeneity across all specifications so that the true contribution of heterogeneity to overall earnings dispersion remains large.

Table 2.4: Heterogeneity and Uncertainty (Measurement Error Adjusted)

# of Observations: 650	Measurement Error $var(\varsigma_i)$	Adjusted Heterogeneity	Adjusted Heterogeneity Ratio
Stepwise Uniform	83.6	408.3	72.14%
Log-normal	110.6	428.0	50.93%
Normal	93.6	390.0	64.17%

Note: The second column (Adjusted Heterogeneity) is found by subtracting column 1, Table 2.4 from column 2, Table 2.3. The third column (Adjusted Heterogeneity Ratio) is the ratio of column 2, Table 2.4, to the sum of column 2, Table 2.4 and column 1, Table 2.3.

Discussion

There are reasons that our results are not directly comparable to the results surveyed in Cunha and Heckman (2007), which are obtained using a realized income distribution. One particularly notable difference is that our analysis is based on a sample of relatively homogeneous students from one college. A second difference is that our survey questions (Question 1A/B) are able to take into account individual-level uncertainty due to a potentially important factor, the aggregate state of the economy in the future, which does not generate variation in the realized income distribution in a particular year. However, if we were to broaden our sample

[18]For example, the sample analog of $var(\widetilde{E^1}(W_{it}) - \widetilde{E^2}(W_{it}))$ involves finding the difference between the mean computed by Approach 1 and the mean computed by Approach 2 for each individual and then computing the variance of this difference across all individuals in the sample.

to include students who are likely to have systematically different views about future earnings (e.g., students who do not attend college) or if we were to remove any uncertainty that exists due to business cycles, then we would tend to find an even more prominent role for heterogeneity relative to uncertainty.[19] Thus, it is reasonable to conclude that our findings reinforce the strong message in Cunha and Heckman (2007) that taking into account heterogeneity is essential for characterizing the amount of uncertainty that exists about future earnings at the time of college entrance.

2.4 Uncertainty Resolution

In this section, we turn to examining when and why initial uncertainty about income is resolved. In Section 2.4.1, we examine one particularly prominent potential source of uncertainty, one's college grade point average. By definition, all uncertainty about final college GPA will be resolved by the end of college. Thus, if uncertainty about GPA is an important contributor to overall earnings uncertainty, then students will expect much earnings uncertainty to be resolved at some point during college, and much resolution may be expected to take place early in school if students tend to learn quickly about their academic ability (Stinebrickner and Stinebrickner, 2012, 2014b). In Section 2.4.2, we perform a related analysis to examine how much earnings uncertainty at the time of entrance can be attributed to uncertainty about college major. The findings in Section 2.4.1 and Section 2.4.2 raise the possibility that much uncertainty about earnings may remain unresolved at the end of college. Section 2.4.3 takes advantage of the longitudinal expectations data in the BPS to show that this is the case, and, finally, Section 2.4.4 explores the factors that could contribute to this finding.

2.4.1 How Much Does Grade Uncertainty Contribute to Earnings Uncertainty?

In addition to being useful for examining robustness and correcting for measurement error, our second computational approach (Section 2.3.1) provides a natural way to quantify the importance of uncertainty about final GPA in determining overall uncertainty about future income. Equation (2.4) yields a natural decomposition of income uncertainty. The first term in the

[19]The former is true if, e.g., the amount of uncertainty in other groups tends to be roughly similar to that of students in our sample. The latter statement holds if aggregate and individual income-influencing factors are multiplicatively separable. The proof is available upon request.

Another difference is that, unlike articles surveyed in Cunha and Heckman (2007), we do not control for observed characteristics before computing the relative importance of uncertainty and heterogeneity. However, this difference is unlikely to be important: we find that observable characteristics explain relatively little of the total variation in $E(W_{it})$.

square root shows the degree to which a student believes that the mean of W_{it} varies across different final GPA realizations. Thus, it measures the contribution of uncertainty about grade performance to income uncertainty. The second term is an average (across GPA realizations) of how much uncertainty is present conditional on a particular realization of final GPA. Thus, it measures the contribution of other factors to income uncertainty, including, for example, uncertainty about major choice, labor market frictions, and future labor market conditions.[20]

Formally, we define the contribution of grade uncertainty to income uncertainty as the fraction of overall uncertainty that can be attributed to the first term:

$$\begin{aligned} R^G_{it} &= \frac{var_{G_{it}}(E(W_{it}|G_{it}))}{var(W_{it})} \\ &= \frac{var_{G_{it}}(E(W_{it}|G_{it}))}{var_{G_{it}}(E(W_{it}|G_{it})) + E_{G_{it}}(var(W_{it}|G_{it}))}. \end{aligned} \tag{2.14}$$

Table 2.5: Contribution of R^G_{i0}: Mean and Quartiles

# of Observations: 650	Mean	25%	50%	75%
Stepwise Uniform	0.1879	0.0118	0.0850	0.2813
Log-normal	0.1609	0.0102	0.0648	0.2360
Normal	0.1655	0.0112	0.0687	0.2545

Note: The first column shows the mean of the sample distribution of R^G_{i0}. The final three columns show the three quartiles of the sample distribution of R^G_{i0}.

Table 2.5 summarizes the results for the time of entrance. The first column shows that, on average, 19% of income uncertainty is due to uncertainty about final GPA when we use the stepwise uniform assumption and that, on average, 16% of income uncertainty is due to uncertainty about final GPA when we use the log-normal or normal distributions. The final three columns show the three quartiles for the three distributional assumptions. For the log-normal and normal distributions, only roughly 25% of students believe that more than roughly 25% of overall income uncertainty is due to uncertainty about final GPA. For the stepwise uniform case, only 25% of students believe that more than 28% of income uncertainty is due to uncertainty about final GPA. Hence, we conclude that, while uncertainty about grade performance has a non-trivial effect on overall earnings uncertainty, the large majority of uncertainty exists for other reasons.

[20] Of course, it is desirable to directly investigate the importance of each of the "other" factors as thoroughly as possible. In Section 2.4.2 we do examine the contribution of major choice to overall earnings uncertainty, and in Section 2.4.4 we do investigate the relative importance of labor market frictions and future labor market conditions in determining the substantial uncertainty that is found to remain at the end of college.

We can also provide evidence about the determinants of the heterogeneity in the Table 2.5 fractions. While individuals with higher fractions do tend to have slightly less income uncertainty because of factors other than GPA, they have much more income uncertainty because of GPA. For example, splitting the sample based on the median in the second (Log-normal) row of Table 2.5, the first term in the denominator of equation (2.14) is 11 times larger for students above the median and the second term in the denominator is 32% smaller for students above the median. Differences in the amount of income uncertainty that is due to GPA could arise, not only because of differences in uncertainty about GPA, but also because of differences in beliefs about how GPA translates to income. We find evidence that, in practice, both of these sources of heterogeneity matter.[21]

2.4.2 How Much Does Major Uncertainty Contribute to Earnings Uncertainty?

Another important determinant of income that is fully realized during college is college major (Altonji, Blom, and Meghir, 2012, Stinebrickner and Stinebrickner, 2014a, Altonji, Arcidiacono, and Maurel, 2016). A decomposition relevant for investigating the role that uncertainty about major plays in determining total income uncertainty can be obtained in a way similar to the decomposition for GPA in equation (2.4):

$$var(W_{it}) = var_{M_{it}}(E(W_{it}|M_{it})) + E_{M_{it}}(var(W_{it}|M_{it})), \tag{2.15}$$

where M_{it} is a discrete random variable describing student i's beliefs about final major at time t, which takes on one of seven possible majors j with probability P_{ijt}.[22] The first term on the right side of equation (2.15) shows how the mean of W_{it} varies across different majors. Thus, it measures the contribution of uncertainty about major to income uncertainty. The second term is an average (across major realizations) of how much uncertainty is present conditional on a particular realization of final major. Thus, it measures the contribution of other factors to

[21] Evidence about the importance of the first source of heterogeneity can be seen by computing the sample interquartile range of $var_{G_{i0}}(E(W_{i0}|G_{i0}))$ assuming that, conditional on a given realization of GPA, all students have identical beliefs about the mean of the subjective conditional income distribution. In practice, we set these means equal to their sample averages. Evidence about the importance of the second source of heterogeneity can be seen by computing the sample interquartile range of $var_{G_{i0}}(E(W_{i0}|G_{i0}))$ assuming that all students have identical beliefs about final GPA. In practice, we set the parameters of the subjective GPA distribution equal to their sample averages. We find that, depending on which of the three distributional assumptions is used, the interquartile range for the first source of heterogneity is roughly 35% to 40% as large as the interquartile range for the second source of heterogeneity.

[22] The numbers $1, ..., 7$ correspond to the following eight major groups: 1. Agricultural and Physical Education; 2. Business; 3. Elementary Education; 4. Humanities; 5. Natural Sciences/Math; 6. Professional Programs; 7. Social Sciences, where Economics is included in Social Sciences and where, for convenience, we have grouped Agriculture and Physical Education together because of their small sizes.

income uncertainty. Then, analogous to our GPA analysis, the goal is to estimate the fraction of total income uncertainty that is due to major uncertainty using the following formula:

$$R_{it}^{M} = \frac{var_{M_{it}}(E(W_{it}|M_{it}))}{var_{M_{it}}(E(W_{it}|M_{it})) + E_{M_{it}}(var(W_{it}|M_{it}))}. \tag{2.16}$$

Unfortunately, unlike what was the case for our GPA analysis in Section 2.4.1, the data do not include all of the information that would allow us to directly compute the two terms, $var_{M_{it}}(E(W_{it}|M_{it}))$ and $E_{M_{it}}(var(W_{it}|M_{it}))$, that enter this fraction. Specifically, while our analysis in Section 2.4.1 took advantage of the fact that $var(W_{it}|G_{it})$ is available in the data, $var(W_{it}|M_{it})$ is not available. However, given information that is observed about $E(W_{it})$, $var(W_{it})$ and the probabilities P_{ijt}, $j = 1, ..., 7$, we are able to estimate the two terms if we make additional assumptions about how the mean and variance of the subjective income distribution conditional on a major varies across students.

Estimation

The objective of this section is to examine the fraction of income uncertainty that is due to uncertainty about major at the time of entrance ($t = 0$). With P_{ij0} observed from Survey Question 5 in Appendix A for $j = 1, ..., 7$, Equation (2.15) shows that estimating the two terms requires knowledge of $E(W_{i0}|M_{i0})$ and $var(W_{i0}|M_{i0})$. We estimate these conditional means and conditional variances under the assumption that they are homogeneous across students conditional on observable characteristics, X_i, that are known to the student at time $t = 0$,

$$\begin{aligned} E(W_{i0}|M_{i0} = j) &= \alpha_w + X_i\beta + \delta_j \\ var(W_{i0}|M_{i0} = j) &= \alpha_v + X_i\gamma + \theta_j, \end{aligned} \tag{2.17}$$

where δ_j, $j = 1, ..., 7$ and θ_j, $j = 1, .., 7$ represent differences in the conditional means and the conditional variances, respectively, across majors.[23]

The unconditional mean $E(W_{i0})$ can be written as $E_{M_{i0}}(E(W_{i0}|M_{i0}))$, and, therefore, is a function of $E(W_{i0}|M_{i0})$ and the random variable M_{i0}. Similarly, the unconditional variance $var(W_{i0})$ can be written as $var_{M_{i0}}(E(W_{i0}|M_{i0})) + E_{M_{i0}}(var(W_{i0}|M_{i0}))$, and, therefore is a function of $E(W_{i0}|M_{i0})$, $var(W_{i0}|M_{i0})$, and the random variable M_{i0}. Then, following the same assump-

[23]While the linear specification does not restrict the conditional means and variances in equation (2.17) to be positive, in practice we find that these objects are typically estimated to be positive. Nonetheless, we also estimated a specification in which we assumed that the conditional means and variances were exponential functions. This specification, in which the means and variances are restricted to be positive, produces results that are quite similar to those obtained for the linear case.

tion as in Section 2.3.2, the unconditional mean that is computed from Survey Question 1A using Approach 1, $\widetilde{E^1}(W_{i0})$, is determined by adding classical measurement error, ς_i, to the true unconditional mean, $E(W_{i0})$. Similarly, the unconditional variance, $\widetilde{Var}(W_{i0})$, that is computed from Survey Question 1A using Approach 1 is determined by adding classical measurement error, u_i, to the true unconditional variance, $var(W_{i0})$. This implies that

$$\begin{aligned}\widetilde{E^1}(W_{i0}) = E_{M_{i0}}(E(W_{i0}|M_{i0})) + \varsigma_i = \sum_{j=1}^{7} P_{ij0}E(W_{i0}|M_{i0} = j) + \varsigma_i \\ = \alpha_w + X_i\beta + \sum_{j=1}^{7} P_{ij0}\delta_j + \varsigma_i\end{aligned} \tag{2.18}$$

$$\begin{aligned}\widetilde{var}(W_{i0}) = var_{M_{i0}}(E(W_{i0}|M_{i0})) + E_{M_{i0}}(var(W_{i0}|M_{i0})) + u_i \\ = var_{M_{i0}}(\delta_j) + \alpha_v + X_i\gamma + \sum_{j=1}^{7} P_{ij0}\theta_j + u_i\end{aligned} \tag{2.19}$$

Normalizing the Social Science coefficients δ_7 and θ_7 to zero, we estimate the remaining parameters, α_w, β, δ_j, $j = 1, ..., 6$, α_v, γ, and θ_j, $j = 1, ..., 6$, which are needed to estimate $E(W_{i0}|M_{i0} = j)$, $j = 1, ..., 7$ and $var(W_{i0}|M_{i0} = j)$, $j = 1, ..., 7$ (equation 2.17), and, therefore, the two terms that appear in the fraction R^M_{i0} (equation 2.16). We obtain estimates by:

1. Regressing $\widetilde{E^1}(W_{i0})$ on X_i and P_{ij0}, $j = 1, ..., 7$ to obtain estimates of α_w, β and δ_j, $j = 1, ..., 6$.

2. Using the estimates $\widehat{\delta_j}$, $j = 1, ..., 6$ and the normalized value $\delta_7 = 0$ to compute an estimate of $var_{M_{i0}}(\delta_j)$, $j = 1, ..., 7$ for each person i.

3. Regressing $\widetilde{var}(W_{i0}) - \widehat{var}_{M_{i0}}(\delta_j)$ on X_i and P_{ij0}, $j = 1, ..., 7$ to obtain estimates of α_v, γ and θ_j, $j = 1, ..., 6$.

Results

Including Black, Male, and ACT score in X_i, Table 2.6 shows the results. The first column shows that, on average, 17% of income uncertainty is due to uncertainty about final major when we use the stepwise uniform assumption, on average, 12% of income uncertainty is due to uncertainty about final major when we use the log-normal assumption, and, on average, 11% of income uncertainty is due to uncertainty about final major when we use the normal assumption. Thus, the conclusions for major are fairly similar to the conclusions for GPA - while students believe that uncertainty about major plays non-trivial role in creating the overall uncertainty about income, much of the uncertainty about income is present for other reasons.

Table 2.6: Contribution of R^M_{i0}: Mean and Quartiles

# of Observations: 682	Mean	25%	50%	75%
Stepwise Uniform	0.1669	0.0419	0.1458	0.2508
Log-normal	0.1152	0.0333	0.0932	0.1645
Normal	0.1125	0.0407	0.0957	0.1672

Note: The first column shows the mean of the sample distribution of R^M_{i0}. The final three columns show the three quartiles of the sample distribution of R^M_{i0}.

Table 2.7: Estimates for δ_j and θ_j

		$j = 1$	$j = 2$	$j = 3$	$j = 4$	$j = 5$	$j = 6$	$j = 7$
δ_j	Stepwise Uniform	-1.3487 (0.7612) [3]	8.8563 (0.0380) [1]	-11.3784 (0.0176) [7]	-3.7377 (0.3542) [5]	-3.8968 (0.2932) [6]	-2.6666 (0.5168) [4]	0 N.A. [2]
	Log-normal	-2.6844 (0.5554) [4]	8.0725 (0.0720) [1]	-13.2090 (0.0098) [7]	-6.0781 (0.1498) [6]	-2.4133 (0.5350) [3]	-3.5936 (0.3994) [5]	0 N.A. [2]
	Normal	-2.8511 (0.5080) [4]	7.2972 (0.0780) [1]	-11.4803 (0.0156) [7]	-6.7927 (0.0926) [6]	-1.9434 (0.5954) [3]	-3.2801 (0.4164) [5]	0 N.A. [2]
θ_j	Stepwise Uniform	32.6860 (0.3368) [1]	12.2821 (0.7798) [3]	-86.5519 (0.0676) [7]	16.8320 (0.6150) [2]	-11.6128 (0.6676) [5]	-30.0232 (0.3176) [6]	0 N.A. [4]
	Log-normal	68.7708 (0.2612) [1]	24.4051 (0.7140) [3]	-139.0690 (0.0696) [7]	47.7077 (0.4082) [2]	18.9153 (0.7582) [4]	-41.2298 (0.4746) [6]	0 N.A. [5]
	Normal	50.8042 (0.2824) [1]	8.6432 (0.8916) [2]	-93.4664 (0.1060) [7]	1.4545 (0.9932) [4]	8.4128 (0.8628) [3]	-40.2733 (0.3514) [6]	0 N.A. [5]

List of majors: 1. Agricultural and Physical Education; 2. Business; 3. Elementary Education; 4. Humanities; 5. Natural Sciences/Math; 6. Professional Programs; 7. Social Sciences.
Note: Equal-tail bootstrap P-values are in the parenthesis. Ranks are in the brackets.

Table 2.7 reports the estimates for δ_j and θ_j. The first three rows indicate that students believe there are substantial differences in mean earnings across majors. For example, the Business major (j = 2) has a significantly higher mean than the Social Science major (j = 7), while the Education major (j = 3) has a significantly lower mean than the Social Science major. The last three rows indicate that there are also differences in uncertainty about income across majors. Most notably, consistent with the rigid pay scale that exists in public schools, the variance is estimated to be the smallest for Elementary Education.

2.4.3 Total Uncertainty Resolution

The findings in Section 2.4.1 and Section 2.4.2 raise the possibility that much uncertainty about earnings may remain unresolved at the end of college. However, while grade performance (academic ability) and college major are prominent income-influencing factors that a student could learn about during college, they are not the only possible factors of relevance. In this section, we examine the actual evolution of income uncertainty over time during school, by taking advantage of the fact that the BPS elicited information about subjective income distributions in each year of school (using questions such as Question 1A in Appendix A). We again focus on subjective beliefs about income at age 28 under the scenario in which a student graduates from college.

Table 2.8: Uncertainty Resolution

# of Observations: 246		Beginning	Year 1	Year 2	Year 3	End
Sample Average of $std(W_{it})$	Stepwise Uniform	10.1310	9.1084	8.3859	8.2887	8.2874
	Log-normal	13.4582	11.8160	11.0484	11.0632	10.7536
	Normal	11.7686	10.7123	10.0112	9.6912	9.6384
Percentage of Uncertainty Resolved	Stepwise Uniform	N.A.	0.1917	0.3148	0.3306	0.3308
	Log-normal	N.A.	0.2291	0.3261	0.3242	0.3615
	Normal	N.A.	0.1714	0.2764	0.3219	0.3292

Note: The unit of measurement for W_{it} is one thousand dollar. The percentage of initial uncertainty resolved by Year t (row 4-6) is obtained in the manner described in the text.

The first three rows of Table 2.8 report the average standard deviation of the subjective earnings distribution at five different points in college - the beginning of college, the end of the first year, the end of the second year, the end of the third year, and the time of graduation (End) - for each of our three distributional assumptions, using Approach 1.[24] We restrict our sample to students who answered income expectations questions at all five points. Looking across columns, as would be expected, students become increasingly certain about their future income as they progress through college.[25]

[24]For t greater than zero, computing $std(W_{it})$ using Approach 2 requires using a student's cumulative GPA at time t to construct the distribution describing subjective beliefs about final grades at time t. We avoid this complication by computing $std(W_{it})$ using only Approach 1.

[25]The only exception is a slight increase of sample average of $std(W_{it})$ from the end of Year 2 to the end of Year

In order to facilitate a comparison between total uncertainty resolution and the findings in Section 2.4.1 and 2.4.2, we define the percentage of uncertainty resolution as the percentage decrease in the variance of the subjective income distribution. Since the variance is simply the square of the standard deviation, we compute these percentages using entries in the first three rows of Table 2.8. As an example, the second column in the fourth row shows that $1 - \frac{9.1084^2}{10.1310^2} = 19.17\%$ of total income uncertainty was resolved during the first year of college, when we use the stepwise uniform distribution.

The last three rows of Table 2.8 show the percentage of uncertainty that is resolved as of the five different points. The results indicate that, depending on the distributional assumption that is made, between 33% and 36% of uncertainty is resolved by the end of college. Thus, the evidence indicates that much uncertainty does remain unresolved during college. Further, comparing the last three columns, we find that the majority of uncertainty resolution took place in the first two years of college, with little uncertainty resolved after the end of the third year. This finding suggests that learning about future income happens relatively quickly in college. Given evidence that uncertainty about grade performance and major is resolved relatively quickly, the finding is consistent with an environment where learning about grade performance (ability) and major contribute heavily to the total resolution of income uncertainty.

Further, comparing the sum of the contribution of GPA uncertainty (Table 2.5) and major uncertainty (Table 2.6) to the results in the last three rows of Table 2.8 provides some direct evidence about whether this is the case. However, this sum would give a biased view of the joint contribution of GPA and major if these two factors tend to be correlated. The joint contribution of GPA and major is determined by an equation analogous to Equation (2.4) and Equation (2.15):

$$var(W_{it}) = var_{G_{it},M_{it}}(E(W_{it}|G_{it}, M_{it})) + E_{G_{it},M_{it}}(var(W_{it}|G_{it}, M_{it})). \tag{2.20}$$

The first term on the right side of Equation (2.20), which is the variance of $E(W_{it}|G_{it}, M_{it})$ over the joint distribution of G_{it} and M_{it}, represents the joint contribution of uncertainty about final GPA and major to total income uncertainty. The second term on the right side of Equation (2.20), which is the mean of $var(W_{it}|G_{it}, M_{it})$ over the joint distribution of G_{it} and M_{it}, represents the contribution of other factors to total initial income uncertainty. Analogous to Equation (2.14) and Equation (2.16), we define the contribution of final GPA and major to total income uncertainty, R_{it}^{GM}, as the ratio of the first term to the sum of the two terms.

We compute R_{it}^{GM} for the time of entrance ($t = 0$) using a method described in Appendix B.6. Table 2.9 summarizes the results. The first column shows that, on average, 27% of initial

3 when using log-normal distribution. This increase, however, is quite small and can be reasonably attributed to measurement error.

income uncertainty is due to uncertainty about final GPA and major when we use the stepwise uniform assumption, on average, 19% of initial income uncertainty is due to uncertainty about final GPA and major when we use the log-normal assumption, and, on average, 23% of initial income uncertainty is due to uncertainty about final GPA and major when we use the normal assumption. Thus, the results in Table 2.9 along with the results in the last three rows of Table 2.8 do indicate a very substantial role for final GPA and major in the resolution of uncertainty.

Table 2.9: Contribution of R_{i0}^{GM}: Mean and Quartiles

# of Observations: 588	Mean	25%	50%	75%
Stepwise Uniform	0.2685	0.1549	0.2377	0.3604
Log-normal	0.1926	0.0952	0.1446	0.2298
Normal	0.2262	0.1254	0.1792	0.2757

Note: The first column shows the mean of the sample distribution of R_{i0}^{GM}. The final three columns show the three quartiles of the sample distribution of R_{i0}^{GM}.

Selection

In order to keep the sample constant across columns in Table 2.8, the sample used includes only students who graduated. A natural question is how the results in Table 2.8 would change if no selection issues were present, that is, if we could compute these numbers for the full sample of all students who entered college - both those who graduated and those who dropped out. Thinking about how the full sample might differ from the sample of graduates, it is not clear from a conceptual standpoint whether individuals who drop out of school would tend to resolve more uncertainty or less uncertainty than individuals who remain in school. This is the case because students who drop out could tend to be those that resolve a substantial amount of uncertainty or could be students who were very close to the margin of indifference at the time of entrance, and, therefore, could be induced to leave school even without resolving much uncertainty. As such, whether the amount of uncertainty that would be resolved for the full sample would tend to be higher or lower than the amount of uncertainty that is resolved for the sample of graduates is an empirical question. We are able to provide some evidence about this question by taking advantage of the fact that income expectations were elicited twice during the first year, before much dropout occurs. We find that, depending on the distributional assumption we use, individuals in the full sample resolve between 7% and 9% of initial uncertainty during this period, while individuals who graduate resolve between 15% and 17% of uncertainty during the first year. Thus, the amount of uncertainty that is resolved for students in the full sample seems to be, if anything, lower than the amount of uncertainty that is resolved for students who graduated. This suggests that our conclusion from Table 2.8 - that much uncertainty remains unresolved at the time of graduation - would be strengthened further if we were able to examine the resolution of earnings for our full sample of students who answered the baseline survey.

It is worth considering whether it seems generally plausible that much uncertainty may remain unresolved at the end of college. Of central relevance, it seems reasonable to believe that, during college, a student may be able to resolve uncertainty about her own ability or other permanent factors, but it may be, by definition, difficult to resolve uncertainty about transitory shocks that could occur in the labor market. Then, the notion that substantial uncertainty remains at the end of college may not be entirely surprising given that a broad literature finds that transitory components play an important role in the earnings process (Blundell and Preston, 1998, Meghir and Pistaferri, 2004). Consistent with these findings, using our post-college data to estimate a random effects model of earnings, we find that the transitory component has a standard deviation of approximately \$9,000.[26] While a variety of concerns could arise from comparing this standard deviation from the realized earnings data to standard deviations elicited using expectations questions, it does seem generally relevant that \$9,000 is non-trivial when viewed next to the standard deviations in Table 2.8.

Demographic Variables

In Section 2.3.1 we found that black students are particularly uncertain about income at the time of entrance. A natural question is whether these students resolve more uncertainty early in college, so that they ultimately end up with similar amounts of uncertainty as other students. Given that Table 2.8 found that the majority of resolution during college takes place during the first two years, we regress $std(W_{i2})$ on Black, Male and ACT score for the three different distributional assumptions associated with Approach 1. We find that black students are no longer more uncertain at the end of the second year; the estimated coefficient on Black in all three regressions is slightly negative.

The previous paragraph suggests that black students are resolving more uncertainty than other students. To provide more direct evidence, we regress the change in uncertainty, as measured by $std(W_{i2}) - std(W_{i0})$, on Black, as well as Male and ACT score for the three distributional assumptions associated with Approach 1. As expected, we find that the coefficient on Black is significant at a .1 level in all three regressions, with the largest t-statistic having a

[26]We estimate a random effects model with annual income as the dependent variable and Black, Male, ACT score, cohort dummy and year dummy as regressors. We use data during 2009-2012 for estimation because most students in our sample turn 28 around year 2010 or 2011.

value of 2.31. Averaging the coefficient for Black across the three regressions, we find that the decrease in uncertainty is $3088 larger for blacks than for non-blacks.

2.4.4 What Factors Account for End-of-College Income Uncertainty?

With the goal of providing a more concrete understanding of why a substantial amount of uncertainty about income at age 28 remains unresolved at the end of college, we consider two broad explanations. The first explanation is that individuals might be unsure about what kinds of job offers they will receive at age 28. The second explanation is that individuals might know the kinds of job offers they will receive, but might be unsure about what kinds of jobs they will prefer to hold/choose in the future. These two explanations may have different policy implications for a variety of reasons, including the fact that the latter represents variation in future income that is at least partially under the control of individuals.

We begin by considering the second explanation. Traditionally, especially for women, uncertainty about hours of work would have represented a particularly salient reason for this explanation, with uncertainty about hours of work having an obvious, direct link to uncertainty about income. However, Stinebrickner, Stinebrickner, and Sullivan (2018) find that this reason is unlikely to be of particular importance for our recent cohort of college graduates; the large majority of both men and women work full-time throughout their first decade in the labor force, with even departures for children tending to be short.

A second possible reason for the second explanation is that individuals may be uncertain about what types of work they will prefer to perform in the future, with uncertainty about types of work having a link to uncertainty about income because income varies substantially across different types of work (Gibbons and Katz, 1992, Heckman and Sedlacek, 1985, Acemoglu and Autor, 2011, Autor and Handel, 2013). We use Survey Question 7 to look for evidence of this type of uncertainty. Because it is not possible to elicit preferences about all types of work, the question stratifies the set of possible jobs into three broad categories: jobs that do not require a college degree (No-Degree-Needed), jobs that require a college degree in a student's specific area of study (Degree-My-Area), and jobs that do not require a college degree in a student's specific area of study (Degree-Any-Area).

Uncertainty about preferences towards the three categories in Question 7 would be particularly relevant for creating income uncertainty if individuals tend to be uncertain about whether they will wish to work in No-Degree-Needed jobs, because these jobs tend to pay substantially less than jobs that require a college degree. However, Survey Question 7 suggests that this is unlikely. Only between 2-3% of all students prefer No-Degree-Needed jobs to jobs that require a college degree and the preference for the types of work in college jobs is very strong, with the average respondent requiring an income premium of over 50% ($45,500 v.s. $30,000) to change from her preferred college job to a No-Degree-Needed job. Further, there seems to be relatively little uncertainty about what types of jobs students prefer even when we take a further step and differentiate between Degree-Any-Area jobs and Degree-My-Area jobs. More than 80% of students prefer Degree-My-Area jobs, and, on average, these individuals would have to be paid a roughly 47% income premium to accept Degree-Any-Area jobs instead.[27] Thus, Question 7 does not provide evidence that the second explanation is important. However, we can not rule out that the second explanation is important because it is possible that workers are uncertain about their preferences towards the different types of jobs that are present within each of the broad categories in Question 7.

We consider several possible reasons for the first explanation. The first reason we consider is that uncertainty may exist about the state of the economy at age 28. To examine this reason, we take advantage of the fact that, as students approached the end of college, the BPS elicited beliefs about not only earnings at the age of 28, but also about earnings in the first year out of college. As shown in the first column of Table 2.10, at the end of college ($t = 4$), the average standard deviation of the subjective distribution of earnings in the first post-college year is between six thousand and nine thousand dollars, depending on the distributional assumption that is employed. This standard deviation tends to be approximately 75% of the standard deviation associated with age 28 (second column) and approximately 60% of the standard deviation associated with age 38 (third column). The fact that much uncertainty exists for the first year out of school suggest that, at the very least, factors other than the state of the economy are influencing income uncertainty.

Roughly speaking, we could group the remaining reasons for the first explanation under the heading of frictions. One possibility is that information frictions are present. For example, students may begin school with uncertainty about the type of job opportunities that tend to be available for college graduates, and this uncertainty may not be entirely resolved even by the end of college (Betts, 1996). It is somewhat difficult to provide direct evidence about the importance of this type of friction. However, we are able to provide some evidence about a second potential type of frictions - labor market/search frictions. The first piece of evidence comes from Survey Question 6. Although we found that more than 80% of students prefer a Degree-My-Area job, Question 6 indicates that, on average, students believe there is only a 50% chance of ending up in such a job in the first year. Further, while almost no students prefer a No-Degree-Needed job, on average, students believe there is almost a 20% chance of being forced to accept this type of job. The second piece of evidence comes from Survey Question

[27] In addition, the 16% of students who prefer a Degree-Any-Area job also seem to be quite certain about their preferences. On average, these students would have to be paid around 44% more to accept Degree-My-Area jobs.

Table 2.10: Earnings Beliefs at the End of College

# of Observations: 359	$std(W_{i4}^{a,1})$		
	a = 1 Year Out	a = 28	a = 38
Stepwise Uniform	6.3281 (4.9854)	8.6233 (6.9139)	10.9518 (9.3353)
Log-normal	9.0638 (13.0912)	11.4029 (11.7530)	14.1175 (14.5838)
Normal	7.2301 (5.6186)	9.9188 (7.9017)	12.5319 (10.3720)

Note: For different ages a, the table shows the standard deviation of the subjective income distribution at the end of college ($t = 4$) for the graduation scenario ($s = 1$). The unit of measurement for $W_{i4}^{a,1}$ is one thousand dollars. A particular entry in the table shows the sample mean and standard deviation of $std(W_{i4}^{a,1})$ for a particular age a. For example, row 1, column 1 shows a sample mean of \$6,328.10 and a sample standard deviation of \$4,985.40 for $std(W_{i4}^{a,1})$ for the age a corresponding to the first post-college year.

8. On average, students believe that there is a 22% probability that it will take five or more months of search to find a job. Further, on average, students believe that there is only a 20% chance of obtaining a job with less than one month of search.[28] While we stress that it is not possible to determine the relative importance of the different reasons for the first explanation, the results suggest that search frictions are likely to be relevant.

2.5 Conclusion

Whether large amounts of uncertainty about future earnings tend to be resolved during college has been an open question. Large amounts would tend to be resolved if: 1) the substantial dispersion found in realized earnings is indicative of substantial amounts of uncertainty at the time of college entrance, and 2) much of this initial uncertainty is resolved during college as students learn about earnings-influencing factors.

Prior evidence about 1) is provided by research such as Cunha, Heckman, and Navarro (2005). They conclude that only a relatively small portion of the variation in realized earnings should be attributed to uncertainty, leaving a large role for heterogeneity. We find direct evidence in support of their conclusion when, taking advantage of expectations data collected at the time of college entrance, we decompose an expectations analog to the realized wage

[28]The survey question elicits beliefs about search frictions during school. The assumption in this discussion is that these beliefs are related to beliefs about search frictions in the post-schooling period. This assumption is consistent with the assumptions made, out of necessity, in a broader search literature.

distribution into the portion due to uncertainty and the portion due to heterogeneity.

Very little evidence about 2) is present in the literature. Taking advantage of the longitudinal nature of our expectations data, we find that much of the income uncertainty that is present at the time of entrance remains unresolved at the time of graduation. Further, taking advantage of a variety of unique data features, we provide evidence about the amount of initial income uncertainty that is and is not resolved. Our findings suggest that the portion of uncertainty that is resolved during school can be largely attributed to what one learns about her academic ability and her college major during school. As for why some uncertainty remains unresolved, we find evidence that transitory factors, such as search frictions, are likely to play an important role in creating initial uncertainty.

Bibliography

[1] Daron Acemoglu and David Autor. Skills, tasks and technologies: Implications for employment and earnings. In *Handbook of labor economics*, volume 4, pages 1043–1171. Elsevier, 2011.

[2] Joseph G Altonji. The demand for and return to education when education outcomes are uncertain. *Journal of Labor Economics*, 11(1, Part 1):48–83, 1993.

[3] Joseph G Altonji, Peter Arcidiacono, and Arnaud Maurel. The analysis of field choice in college and graduate school. *Handbook of the Economics of Education*, 5:305–396, 2016.

[4] Joseph G Altonji, Erica Blom, and Costas Meghir. Heterogeneity in human capital investments: High school curriculum, college major, and careers. *Annu. Rev. Econ.*, 4(1):185–223, 2012.

[5] Peter Arcidiacono. Ability sorting and the returns to college major. *Journal of Econometrics*, 121(1):343–375, 2004.

[6] Orazio P Attanasio and Katja M Kaufmann. Education choices and returns to schooling: Mothers' and youths' subjective expectations and their role by gender. *Journal of Development Economics*, 109:203–216, 2014.

[7] David H Autor and Michael J Handel. Putting tasks to the test: Human capital, job tasks, and wages. *Journal of Labor Economics*, 31(S1):S59–S96, 2013.

[8] Julian R Betts. What do students know about wages? evidence from a survey of undergraduates. *Journal of human resources*, pages 27–56, 1996.

[9] Richard Blundell and Ian Preston. Consumption inequality and income uncertainty. *The Quarterly Journal of Economics*, 113(2):603–640, 1998.

[10] Martin Browning and Jesus Carro. Heterogeneity and microeconometrics modeling. *Econometric Society Monograph*, 43:47, 2007.

[11] Pedro Carneiro, Karsten T Hansen, and James J Heckman. 2001 lawrence r. klein lecture estimating distributions of treatment effects with an application to the returns to schooling and measurement of the effects of uncertainty on college choice. *International Economic Review*, 44(2):361–422, 2003.

[12] Flavio Cunha and James J Heckman. Identifying and estimating the distributions of ex post and ex ante returns to schooling. *Labour Economics*, 14(6):870–893, 2007.

[13] Flavio Cunha, James J Heckman, and Salvador Navarro. Separating heterogeneity from uncertainty an aiyagari-laitner economy. In *Goldwater Conference on Labor Markets in Arizona*, 2004.

[14] Flavio Cunha, James J Heckman, and Salvador Navarro. Separating uncertainty from heterogeneity in life cycle earnings. *Oxford Economic Papers*, 57(2):191–261, 2005.

[15] Jeff Dominitz and Charles F Manski. Perceptions of economic insecurity: Evidence from the survey of economic expectations. *Public Opinion Quarterly*, pages 261–287, 1997a.

[16] Jeff Dominitz and Charles F Manski. Using expectations data to study subjective income expectations. *Journal of the American statistical Association*, 92(439):855–867, 1997b.

[17] Milton Friedman. Choice, chance, and the personal distribution of income. *Journal of Political Economy*, 61(4):277–290, 1953.

[18] Robert Gibbons and Lawrence Katz. Does unmeasured ability explain inter-industry wage differentials? *The Review of Economic Studies*, 59(3):515–535, 1992.

[19] James J Heckman and Guilherme Sedlacek. Heterogeneity, aggregation, and market wage functions: an empirical model of self-selection in the labor market. *Journal of political Economy*, 93(6):1077–1125, 1985.

[20] Charles F Manski. Adolescent econometricians: How do youth infer the returns to schooling? In *Studies of supply and demand in higher education*, pages 43–60. University of Chicago Press, 1993.

[21] Charles F Manski. Measuring expectations. *Econometrica*, 72(5):1329–1376, 2004.

[22] Costas Meghir and Luigi Pistaferri. Income variance dynamics and heterogeneity. *Econometrica*, 72(1):1–32, 2004.

[23] Salvador Navarro and Jin Zhou. Identifying agent's information sets: An application to a lifecycle model of schooling, consumption and labor supply. *Review of Economic Dynamics*, 25:58–92, 2017.

[24] Ralph Stinebrickner and Todd Stinebrickner. The effect of credit constraints on the college drop-out decision: A direct approach using a new panel study. *The American economic review*, 98(5):2163–2184, 2008.

[25] Ralph Stinebrickner and Todd Stinebrickner. Learning about academic ability and the college dropout decision. *Journal of Labor Economics*, 30(4):707–748, 2012.

[26] Ralph Stinebrickner and Todd Stinebrickner. A major in science? initial beliefs and final outcomes for college major and dropout. *Review of Economic Studies*, 81(1):426–472, 2014a.

[27] Ralph Stinebrickner and Todd Stinebrickner. Academic performance and college dropout: Using longitudinal expectations data to estimate a learning model. *Journal of Labor Economics*, 32(3):601–644, 2014b.

[28] Ralph Stinebrickner, Todd Stinebrickner, and Paul Sullivan. Job tasks and the gender wage gap among college graduates. *NBER Working Papers 24790*, 2018.

[29] Basit Zafar. How do college students form expectations? *Journal of Labor Economics*, 29(2):301–348, 2011.

Chapter 3

Perceived and Actual Option Values of College Enrollment

3.1 Introduction

An important feature of post-secondary schooling is the experimentation that accompanies sequential decision-making.[1] Specifically, by entering college, a student gains the option to decide at a future time ($t = t^*$) whether it is optimal to remain in college ($s = 1$) or to drop out ($s = 0$), after resolving uncertainty that existed at entrance ($t = t_0$) about academic ability or other factors that affect her return to college. This chapter uses data from the Berea Panel Study to contribute to a literature that has recognized the importance of quantifying the value of this option (Heckman, Lochner, and Todd, 2006, Heckman and Navarro, 2007, and Stange, 2012). The unique nature of the data allows an examination of whether students' perceptions about option values tend to be accurate by allowing, for the first time, a distinction to be made between "actual" option values and "perceived" option values.

For the purpose of illustration, consider a scenario where all that occurs between t_0 and t^* is that students resolve uncertainty that existed at entrance.[2] In this case, in the absence of the option to make decisions after receiving new information, the decision of whether to enter college after high school is equivalent to a decision of whether to commit to staying in school until college graduation. The value of the option quantifies how beneficial it is to be

[1]This notion that education can be considered as a sequential choice that is made under uncertainty has been widely accepted in the literature since the seminal work in Manski (1989) and Altonji (1993).

[2]If there are also direct net benefits/costs associated with staying in school between t_0 and t^* (e.g., tuition, utility or dis-utility of schooling, foregone earnings), students' entrance decisions would also depend on such direct benefits/costs, which could complicate the illustrative discussion in the introduction. However, we note that our formal approach for quantifying the option value does not rely on the assumption that there are no direct benefits/costs between t_0 and t^*.

able to delay the graduation decision until after some uncertainty is resolved during the early portion of college. For a student who would not enter college in the absence of the option, the expected lifetime utility at $t = 0$ of graduating, which we denote $E_{t=t_0} V_i^{s=1}$, is lower than the expected utility at $t = 0$ of not graduating, which we denote $E_{t=t_0} V_i^{s=0}$. Roughly speaking, the option value for this student tends to be substantial when, given the magnitude of the (negative) difference between these two expected utilities at t_0, the information she will obtain after entering college will often push her across the margin of indifference to a situation where the expected utility at t^* of graduating ($E_{t=t^*} V_i^{s=1}$) is non-trivially higher than the expected utility at t^* of not graduating ($E_{t=t^*} V_i^{s=0}$). Similarly, for a student who would enter college in the absence of the option, the expected utility at t_0 of graduating is higher than the expected utility at t_0 of not graduating. Roughly speaking, the option value for this student tends to be substantial when, given the size of the (positive) difference between these two expected utilities at t_0, the information she will obtain after entering college will often push her across the margin of indifference to a situation where the expected utility at t^* of graduating is non-trivially lower than the expected utility at t^* of not graduating.

The importance of quantifying the option value comes from its fundamental importance for understanding/interpreting college attendance and college dropout decisions; while policy discussion often suggests that college attendance rates are too low or college dropout rates are too high, it is difficult to reach an informed view of these rates without understanding the option value's importance.[3] In terms of college entrance, as implied by the discussion in the previous paragraph, the number of high school graduates who should find it optimal to enter will depend directly on the option value; when option values are close to zero, students will tend to enter college only if the expected utility at t_0 of graduating is greater than the expected utility at t_0 of not graduating, while substantially higher option values can induce entrance even for students for which the difference between these expected utilities (hereafter denoted $E_{t=t_0}(V_i^{s=1} - V_i^{s=0})$ and referred to as the "initial expectations gap" at $t = 0$) is substantially negative. Further, this effect on who attends college also leads to a very direct link between option values and dropout rates. Indeed, inconsistent with policy discussion that tends to view dropout as inherently bad, if high option values imply that students with substantial negative initial expectations gap find it useful to enter college, then a non-trivial amount of dropout would be a natural part of a healthy environment in which schools are providing useful information to students.

The well-recognized difficulty of characterizing option values can be viewed as arising, to a large extent, because of data issues. We illustrate these issues using a stylized model that cap-

[3]As one of many examples, a recent article in the Forbes (June 6, 2018) suggests that "The sad reality is that far too many students invest scarce time and money pursuing a degree they never finish, frequently winding up worse off than if they'd never set foot on campus in the first place".

tures the key components of learning in the college environment of interest. Consistent with the discussion in the second paragraph, we show that the option value is determined by the initial expectations gap $E_{t=t_0}(V_i^{s=1} - V_{it^*}^{s=0})$ and the amount of uncertainty about the expected gap that will be resolved before making the dropout decision at t=1, which we denote σ_i. Then, because $E_{t=t_0}(V_i^{s=1} - V_i^{s=0})$ and σ_i completely determine the dropout probability, which we denote $P_i^{s=0}$, what is needed to characterize the option value is any two of: $P_i^{s=0}$, $E_{t=t_0}(V_i^{s=1} - V_i^{s=0})$, and σ_i. Unfortunately, while administrative data sources can provide direct evidence about $P_i^{s=0}$, they are not well-suited for providing direct evidence about the other two objects.[4] As such, research characterizing option values typically has turned to fully specified models (often dynamic discrete choice models) to estimate the option value.[5] In contrast, the BPS data allow the option value to be computed in a more direct way; in addition to containing information about dropout, evidence about uncertainty resolution, which arises in our baseline model because of learning about pecuniary factors under the scenario in which a student graduates from college, comes from the fact that the distribution describing beliefs about future earnings is collected at multiple times during school.

A feature of the models traditionally used to estimate option values is that Rational Expectations (RE) assumptions are employed to link actual outcomes to choices that depend on students' subjective expectations. Consequently, these approaches do not make a distinction between students' perceptions about option values (hereafter referred to as "perceived" option values) and their values implied by rational expectations (hereafter referred to as "actual" option values); roughly speaking, the option values computed using these models are a mix of perceived and actual option values. Generally, the potential importance of this distinction is highlighted by a recent expectations literature, which has found that perceptions about objects of relevance for educational decisions are often inaccurate.[6] In the particular context of interest here, it seems quite possible that students may not entirely appreciate the benefits of experi-

[4]It is hard to provide information about the initial expectations gap because this gap includes not only the financial return to schooling but also non-pecuniary benefits of schooling. These benefits are inherently difficult to observe directly. Instead, many researchers have treated them as the "residual" in the contemporaneous utility function and have identified/estimated their values from the component of schooling attendance decisions that is not explained by pecuniary factors (e.g., Keane and Wolpin, 1997, Cunha, Heckman, and Navarro, 2005, Heckman, Lochner, and Todd, 2006, and Abbott, et al., forthcoming).

[5]Estimation of σ_i typically requires researchers to either impose or estimate the structure of agent information sets at college entrance and the end of college. As one example of the former, Stange (2012) assumes that students update their beliefs about the benefit of college mainly through observing grades as signals. As one example of the latter, Heckman and Navarro (2007) estimate students' information sets using a method developed by Cunha, Heckman, and Navarro (2005).

[6]The importance of whether perceptions tend to be accurate can be seen in recent research emphasizing the value of supplementing expectations data with data on actual outcomes (e.g., Arcidiacono, Hotz, Maurel and Romano, 2019, Stinebrickner and Stinebrickner, 2014a, Wiswall and Zafar, 2016, D'Haultfoeuille, Gaillac, and Maurel, 2018, and Giustinelli and Shapiro, 2019).

mentation. Indeed, the importance of learning models was not even widely recognized in the economics of education literature until quite recently, and policy discussion does not tend to extol the virtues of experimentation.[7] Our ability to differentiate between perceived and actual option values comes from the fact that 1) in addition to observing actual dropout rates, the BPS collected information about perceived dropout rates and 2) in addition to being able to characterize students' actual uncertainty resolution from longitudinal earnings expectations data, students' perceptions about how much uncertainty will be resolved can be estimated using a simple model describing the relationship between the perceived dropout probability, the initial expectations gap, and the perceived amount of uncertainty resolution.

We find that, on average, students' perceptions about the value of the option understate the actual value of the option substantially: The average perceived option value is $8,670, roughly 65% smaller than the average actual option value, $25,040. Importantly, our approach allows us to examine why this overstatement occurs. We find that it is not driven by an understatement of the amount of earnings uncertainty that is resolved in college - both the actual and the perceived fraction of initial earnings uncertainty that is resolved in college are 0.51. Instead, we find that students' perceptions tend to substantially overstate the initial expectations gap $E_{t=t_0}(V_i^{s=1} - V_i^{s=0})$. This result follows from our finding that perceptions about uncertainty resolution are accurate along with a finding that individuals are too optimistic about the probability of graduating (perceived probability 0.853, actual probability 0.647), since, as we show using a stylized model, the initial expectations gap is decreasing in the dropout probability holding constant uncertainty resolution.[8] As a robustness check, we examine the implications of allowing students to learn about non-pecuniary factors and also about their non-college option.

Our findings about the reason for misperceptions about the option value are important because, while it may seem at a first glance that an understatement of the option value would necessarily lead to too few students entering college, in reality whether this is true depends critically on why misperceptions exist.[9] This is the case because the overall value of college, which is the relevant object for the college entrance decision, is strongly related but not identical to the option value. Under the illustrative scenario in the second paragraph - where all that occurs between t_0 and t^* is that students resolve uncertainty that existed at entrance - the

[7]The Berea Panel Study was designed (in 1998) with the specific objective of understanding the importance of learning in educational decisions. At the time, Altonji (1993) and Manski (1989) represented perhaps the only research specifically focusing on the importance of learning models for understanding dropout.

[8]Our finding that, on average, students overstate the initial expectations gap implies that there must be some unexpected systematic downward changes in students' beliefs between t_0 and t^*. This is the case because, if they were anticipated, they should be incorporated into perceived initial expectations gaps, which would then be correct, on average. Therefore, we interpret these changes as unexpected corrections of systematic overoptimism.

[9]The relevance of this concern is apparent in related research which, for example, examines whether higher-education decisions are influenced by misperceptions about college costs (Bleemer and Zafar, 2018) or by misperceptions about available opportunities (Hoxby and Turner, 2013).

overall value of college is equal to the sum of the option value and the initial expectations gap under the most likely scenario where the initial expectations gap is positive. We find that the understatement of the option value is more than offset by the optimism about the initial expectations gap. Thus, once one takes into account both components of the overall value of college, concerns that too few students enter college tend to dissipate.

.

3.2 Data

In the context here, of particular importance are survey questions eliciting students' perceptions about the probability of dropping out and perceptions about future earnings under a scenario in which the student graduates ($s = 1$) and under a scenario in which the student drops out ($s = 0$). We focus on the 2001 cohort of the Berea Panel Study because the 2000 cohort did not answer the key survey question about perceived dropout probability in the baseline survey. Unless otherwise noted, the analyses in this chapter involves the 337 students (from the 2001 cohort) who provided complete answers to these questions on the baseline survey. Providing evidence in support of the notion that the elicited dropout probabilities contain useful content, we find that the null hypothesis that perceived dropout probabilities are unrelated to actual dropout outcomes is rejected at a .10 level of significance.[10]

3.3 Defining the Option Value in a Stylized Learning Model

In this section, we define the option value in the context of a stylized model that captures the key features of learning in the college environment of interest. When entering college at t_0, a student knows that she will have the option to choose between college completion ($s = 1$) and dropping out ($s = 0$) at a future time t^*, after resolving a certain fraction of her initial uncertainty (i.e., uncertainty at $t = t_0$) about the value of each alternative. We denote the value of the two alternatives as $V_i^{s=1}$ and $V_i^{s=0}$, respectively, and denote student i's expectations about the two values at $t = t^*$ as $V_{it^*}^{s=1} \equiv E_{t=t^*} V_i^{s=1}$ and $V_{it^*}^{s=0} \equiv E_{t=t^*} V_i^{s=0}$, respectively.

Formally, the option value can be defined as:

$$OV_i \equiv E_{t=t_0} \max(V_{it^*}^{s=1}, V_{it^*}^{s=0}) - \max(E_{t=t_0}(V_{it^*}^{s=1}), E_{t=t_0}(V_{it^*}^{s=0})). \tag{3.1}$$

[10]Of course, from a conceptual standpoint, a strong relationship between perceptions about an object of interest and the actual outcomes of that object are not necessary for expectations data to be useful. Indeed, much of the motivation for the direct elicitation of expectations comes from the possibility that beliefs may be incorrect. Nonetheless, given the difficulty of providing evidence in support of the quality of expectations data, much previous research has examined whether a relationship exists between perceptions and actual outcomes.

Let $\Delta_i = (V_{it^*}^{s=1} - V_{it^*}^{s=0}) - E_{t=t_0}(V_{it^*}^{s=1} - V_{it^*}^{s=0})$ represent the new information received between t_0 and t^*. We assume that Δ_i is normally distributed.[11] It has a mean of zero by construction, and we denote its variance as σ_i^2.

At time t^*, student i chooses to drop out if and only if $V_{it^*}^{s=0} > V_{it^*}^{s=1}$. Given the normality assumed for Δ_i, her dropout probability $P_i^{s=0}$ is given by:

$$P_i^{s=0} = \Phi(\frac{E_{t=t_0}(V_{it^*}^{s=0} - V_{it^*}^{s=1})}{\sigma_i}), \tag{3.2}$$

where $\Phi(\cdot)$ is the cdf of the standard normal distribution.

In Equation (3.1), $E_{t=t_0}\max(V_{it^*}^{s=1}, V_{it^*}^{s=0})$, which can be referred to as the continuation value of college enrollment, is given by:

$$E_{t=t_0}\max(V_{it^*}^{s=1}, V_{it^*}^{s=0}) = P_i^{s=1}E_{t=t_0}(V_{it^*}^{s=1}) + P_i^{s=0}E_{t=t_0}(V_{it^*}^{s=0}) + \sigma_i\phi(\frac{E_{t=t_0}(V_{it^*}^{s=1} - V_{it^*}^{s=0})}{\sigma_i}), \tag{3.3}$$

where $P_i^{s=1} \equiv 1 - P_i^{s=0} = \Phi(\frac{E_{t=t_0}(V_{it^*}^{s=1} - V_{it^*}^{s=0})}{\sigma_i})$ is the probability of completing college, and $\phi(\cdot)$ is the pdf of the standard normal distribution.[12]

Equation (3.3) allows us to express the option value (OV) as a function of σ_i and $P_i^{s=0}$:

$$\begin{aligned}
OV_i &\equiv E_{t=t_0}\max(V_{it^*}^{s=1}, V_{it^*}^{s=0}) - \max(E_{t=t_0}(V_{it^*}^{s=1}), E_{t=t_0}(V_{it^*}^{s=0})) \\
&= \begin{cases} P_i^{s=0}E_{t=t_0}(V_{it^*}^{s=0} - V_{it^*}^{s=1}) + \sigma_i\phi(\frac{E_{t=t_0}(V_{it^*}^{s=0} - V_{it^*}^{s=1})}{\sigma_i}) \text{ if } E_{t=t_0}(V_{it^*}^{s=1}) > E_{t=t_0}(V_{it^*}^{s=0}) \\ P_i^{s=1}E_{t=t_0}(V_{it^*}^{s=1} - V_{it^*}^{s=0}) + \sigma_i\phi(\frac{E_{t=t_0}(V_{it^*}^{s=1} - V_{it^*}^{s=0})}{\sigma_i}) \text{ if } E_{t=t_0}(V_{it^*}^{s=1}) \leq E_{t=t_0}(V_{it^*}^{s=0}) \end{cases} \\
&= \begin{cases} P_i^{s=0}\sigma_i\Phi^{-1}(P_i^{s=0}) + \sigma_i\phi(\Phi^{-1}(P_i^{s=0})) \text{ if } P_i^{s=0} < 0.5 \\ P_i^{s=1}\sigma_i\Phi^{-1}(P_i^{s=1}) + \sigma_i\phi(\Phi^{-1}(P_i^{s=1})) \text{ if } P_i^{s=0} \geq 0.5 \end{cases} \\
&\equiv \begin{cases} \sigma_i G(P_i^{s=0}) \text{ if } P_i^{s=0} < 0.5 \\ \sigma_i G(P_i^{s=1}) \text{ if } P_i^{s=0} \geq 0.5 \end{cases}, \\
&\equiv \begin{cases} \sigma_i G(P_i^{s=0}) & \text{if } P_i^{s=0} < 0.5 \\ \sigma_i G(1 - P_i^{s=0}) & \text{if } P_i^{s=0} \geq 0.5 \end{cases},
\end{aligned} \tag{3.4}$$

where $G(P) \equiv P\Phi^{-1}(P) + \phi(\Phi^{-1}(P))$ is a known function of P, which has the easily verifiable property:

[11] Later in Section 3.4.2, to obtain baseline results, we impose an assumption that uncertainty resolution in school is through learning about future earnings. In this case, the normality assumption for Δ_i can be motivated by the finding in Gong, Stinebrickner, and Stinebrickner (2019) that a normal distribution fits students' responses to earnings expectations question better than a log-normal distribution.

[12] Equation (3.3) is equivalent to a well-known alternative formulation: $E_{t=t_0}\max(V_{it^*}^{s=1}, V_{it^*}^{s=0}) = E_{t=t_0}(V_{it^*}^{s=1}|V_{it^*}^{s=1} \geq V_{it^*}^{s=0})P_i^{s=1} + E_{t=t_0}(V_{it^*}^{s=0}|V_{it^*}^{s=0} > V_{it^*}^{s=1})P_i^{s=0}$. A comparison between the two formulations reveals that the last term in Equation (3.3) captures the difference between the conditional and unconditional means of $V_{it^*}^{s=1}$ and $V_{it^*}^{s=0}$.

Lemma 3.3.1 *$G(P)$ is monotonically increasing in P for $P \in (0, 1)$.*

Lemma 3.3.1 implies the following propositions.

Proposition 3.3.2 *The option value, OV_i, has the following properties with respect to the amount of uncertainty resolved before t^*, σ_i, and the probability of dropping out, $P_i^{s=0}$.*

1. *The OV_i is uniquely determined by σ_i and $P_i^{s=0}$;*

2. *The OV_i is multiplicatively separable in σ_i and $P_i^{s=0}$;*

3. *The OV_i is linearly increasing in σ_i;*

4. *The OV_i is monotonically increasing in $P_i^{s=0}$ for $P_i^{s=0} \in (0, 0.5)$ and monotonically decreasing in $P_i^{s=0}$ for $P_i^{s=0} \in [0.5, 1)$.*

Proposition 3.3.2.1 shows that data on the dropout probability, $P_i^{s=0}$, and the amount of uncertainty resolved during college, σ_i, are sufficient for determining the OV, with Equation (3.2) detailing how the initial expectations gap is uniquely characterized by these two terms. Important for our analysis in Section 3.4, Proposition 3.3.2.2 shows that σ_i and $P_i^{s=0}$ enter the expression of OV_i in a multiplicatively separable fashion. Proposition 3.3.2.3 and Proposition 3.3.2.4 qualitatively describe how σ_i and $P_i^{s=0}$ affect the value of OV_i.

3.4 Characterizing the Option Value

Proposition 3.3.2 showed that the OV is uniquely determined by the dropout probability, $P_i^{s=0}$, and the amount of uncertainty that is resolved during college, σ_i. In Section 3.4.1, we describe the direct information available in the BPS about both actual and perceived values of $P_i^{s=0}$. In Section 3.4.2, we impose more structure on the general model described in Section 3.3 in order to estimate the actual and perceived values of σ_i. In Section 3.4.3, combining information about $P_i^{s=0}$ and σ_i, we compute both actual and perceived option values for each student. Comparing actual option values (obtained using actual $P_i^{s=0}$ and actual σ_i) to perceived option values (obtained using perceived $P_i^{s=0}$ and perceived σ_i) provides evidence about the accuracy of beliefs about option values at the time of entrance. Finally, in Section 3.4.4, we discuss the policy implications of potential misperceptions.

3.4.1 Actual and Perceived Dropout Probabilities

Both actual dropout outcomes and perceived dropout probabilities can be obtained directly from the BPS data. 218 out of the 337 students in the sample eventually graduated from Berea College, which implies a dropout rate, or equivalently an average actual dropout probability, of 0.353. Question 4 in Appendix A elicits a student's perceived probability of graduating from Berea College. Subtracting this number from 1 yields the perceived dropout probability, $P_i^{s=0}$. We find that the average perceived dropout probability of students in our sample is 0.147, 58% smaller than the average actual dropout probability.

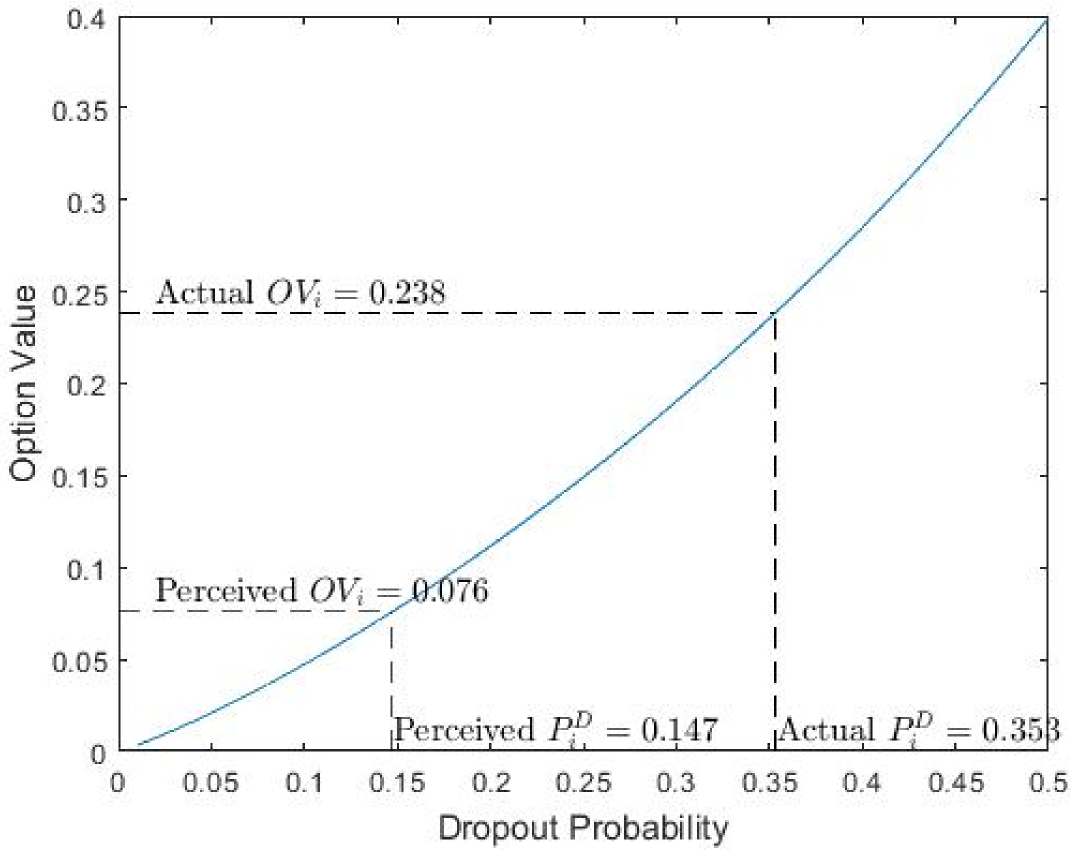
Proposition 3.3.2 is useful for examining how the underestimation of $P_i^{s=0}$ influences the size of the perceived OV relative to the size of the actual OV. Suppose students have rational expectations about σ_i. Since the OV is multiplicatively separable in $P_i^{s=0}$ and σ_i, without loss of generality, we set $\sigma_i = 1$. As implied by proposition 3.3.2.4, Figure 3.1 shows that the OV is increasing in the dropout probability over the range $(0, 0.5)$. Evaluating the OV at the average actual dropout probability leads to an actual OV of 0.238. Evaluating the OV at the average perceived dropout probability leads to a perceived OV of 0.076. Then, for a "representative" student, the perceived value of OV is 68% lower than the actual value of OV.

Of course, in reality there is no reason that individuals would necessarily have Rational Expectations about σ_i. Proposition 3.3.2.1 indicates that obtaining point estimates for the actual and perceived values of the OV requires knowledge of actual and perceived values of σ_i. In the next section, we discuss our approach for taking advantage of additional unique data to obtain these objects. Nonetheless, the evidence presented in the previous paragraph strongly suggests that we are likely to find that students at Berea College tend to underestimate the option value at the time of entrance. Indeed, using Proposition 3.3.2.3, we see that the representative student would need to overestimate σ_i by at least 214% in order to not underestimate the OV.

Before we turn to the characterization of σ_i, we note that, in order to compute the option value for each student, individual-specific measures of actual and perceived dropout probabilities are required. As mentioned earlier, individual-specific perceived dropout probabilities can be directly obtained from students' responses to Question 4 in Appendix A. The sample standard deviation of perceived dropout probabilities is 0.180. In contrast, individual-specific measures of actual dropout probabilities are not directly available. We allow for individual heterogeneity by assuming that a student's actual dropout probability is equal to the predicted probability from a probit regression of a dropout dummy on observables.[13]

[13]The observables in the probit regression include gender, race, high school GPA, ACT score, and a student's perceived dropout probability.

Figure 3.1: Option Value and Dropout Probability



3.4.2 Actual and Perceived Earnings Uncertainty Resolution

In this section, we describe the construction of the actual and perceived values of σ_i. We first show that, under the assumption that the learning of relevance during college is about future earnings associated with college completion, σ_i can be computed by combining: 1) data characterizing student i's uncertainty at the time of entrance (i.e., initial uncertainty) about future earnings under the scenario in which she graduates from college and 2) a parameter ρ capturing the fraction of this initial uncertainty that is resolved between t_0 and t^*. We then describe how we can construct measures of initial earnings uncertainty from survey questions eliciting subjective beliefs about future earnings. The actual fraction of uncertainty resolution, which we denote ρ_A, and therefore the actual σ_i, can be consistently estimated by taking advantage of the longitudinal feature of our expectations data. The perceived fraction of uncertainty resolution, which we denote ρ_P, and, therefore the perceived σ_i, can be consistently estimated by taking advantage of data on students' perceived dropout probabilities and students' initial subjective beliefs about future earnings.

Defining σ_i in a Fully Specified Model

We consider a model in which the value of alternative s, $V^s_{it^*}$, is equal to the expectation, at time t^*, of the sum of the discounted lifetime earnings associated with this alternative, Y^s_i, and an additional term γ^s_i summarizing student i's overall non-pecuniary benefit from s:

$$V^s_{it^*} = E_{t=t^*}(Y^s_i + \gamma^s_i). \tag{3.5}$$

We start by specifying the discounted lifetime earnings, Y^s_i, for each alternative. If a student chooses $s = 1$, the student stays in college until time $\bar{t}$, then starts to work. For ease of notation, we index time t by a student's age a. $Y^{s=1}_i$ is then given by $Y^{s=1}_i = \sum_{a=\bar{t}}^{\bar{A}} \beta^{a-t^*} w^{a,s=1}_i$, where $w^{a,s}_i$ represents the earnings that student i receives at age a given her choice of s, β is the discount factor and $\bar{A}$ is the age of retirement. Similarly, if the student chooses $s = 0$, she leaves college and starts working immediately. The discounted lifetime earnings associated with this alternative, $Y^{s=0}_i$, is given by $Y^{s=0}_i = \sum_{a=t^*}^{\bar{A}} \beta^{a-t^*} w^{a,s=0}_i$.

Turning to the non-pecuniary benefit/utility associated with the choice of s, the immediate exit from school that accompanies a choice of $s = 0$ implies that $\gamma^{s=0}_i$ will tend to capture a person's preferences about working in jobs that do not require a college degree. On the other hand, $\gamma^{s=1}_i$ will capture not only preferences for working in the types of jobs that are obtained with a college degree, but also a person's utility gain/loss from staying in college until graduation.

For our primary results, we make the simplifying assumption that the only updating that occurs during college is about the future earnings that would be received under the graduation scenario. That is, students learn only about $Y_i^{s=1}$ while in college. Abstracting away from learning about earnings under the dropout scenario, $Y_i^{s=0}$, allows for a more transparent discussion of identification, but is also consistent with the intuitively appealing notion that college is best suited for providing information about one's ability to perform high skilled jobs. Further, when relaxing this assumption as a robustness check in Appendix C.2, we find strong evidence in support of this notion; 1) Students resolve less substantially uncertainty about earnings under the dropout scenario than under the graduation scenario, and 2) our main results remain quantitatively similar when we relax this assumption.

Abstracting away from learning about the non-pecuniary benefits, γ_i^s, while obviously not literally correct, would tend to not be particularly problematic if students tend to have a good sense of how much they like school by the end of high school or if the overall non-pecuniary benefit of the graduation alternative ($s = 1$) arises largely because a college degree affects the non-wage aspects of one's work over her lifetime - since individuals presumably learn the most about these non-wage aspects when they actually hold these jobs after graduation.[14] Nonetheless, in Appendix C.3, we discuss how relaxing this assumption would affect our results. In particular, we show that, if, as in Stinebrickner and Stinebrickner (2012), a common set of signals (e.g., grades) influences what a student learns about both pecuniary and non-pecuniary benefits, our estimates of actual option values tend to be downward biased while our estimates of perceived option values remain consistent.

The assumptions in the previous paragraph imply that $E_{t=t_0}(\gamma_i^s) = E_{t=t^*}(\gamma_i^s)$ for $s = 0, 1$, and $E_{t=t_0}(Y_i^{s=0}) = E_{t=t^*}(Y_i^{s=0})$. Then, the relevant new information $\Delta_i \sim N(0, \sigma_i^2)$ is given by:

$$\begin{aligned}
\Delta_i &= (V_{it^*}^{s=1} - V_{it^*}^{s=0}) - E_{t=t_0}(V_{it^*}^{s=1} - V_{it^*}^{s=0}) \\
&= E_{t=t^*}[(Y_i^{s=1} + \gamma_i^{s=1}) - (Y_i^{s=0} + \gamma_i^{s=0})] - E_{t=t_0}[(Y_i^{s=1} + \gamma_i^{s=1}) - (Y_i^{s=0} + \gamma_i^{s=0})] \\
&= E_{t=t^*}(\sum_{a=\bar{t}}^{\bar{A}} \beta^{a-t^*} w_i^{a,s=1}) - E_{t=t_0}(\sum_{a=\bar{t}}^{\bar{A}} \beta^{a-t^*} w_i^{a,s=1}) \\
&= \sum_{a=\bar{t}}^{\bar{A}} \beta^{a-t^*} [E_{t=t^*}(w_i^{a,s=1}) - E_{t=t_0}(w_i^{a,s=1})].
\end{aligned} \tag{3.6}$$

With σ_i representing the standard deviation of Δ_i, Equation (3.6) reveals that σ_i is determined by how much a student updates her expectations about earnings under the graduation

[14]While students do likely learn something about how much they like school after entrance, this learning only affects utility for the short period of time between t^* and $\bar{t}$. In contrast, the non-wage aspects of one's future work would have a lifelong impact on her utility.

scenario, or equivalently, by how much initial earnings uncertainty is resolved between t_0 and t^*. We begin the process of characterizing this updating by writing $w_i^{a,s=1}$, without loss of generality, as the sum of three independently distributed factors, $\epsilon_{i\tau_1}^{a,s=1}$, $\epsilon_{i\tau_2}^{a,s=1}$, and $\epsilon_{i\tau_3}^{a,s=1}$, that are observed by the student in the period before t_0 (denoted τ_1), in the period between t_0 and t^* (denoted τ_2), and in the period after t^* (denoted τ_3), respectively:

$$w_i^{a,s=1} = \epsilon_{i\tau_1}^{a,s=1} + \epsilon_{i\tau_2}^{a,s=1} + \epsilon_{i\tau_3}^{a,s=1}. \tag{3.7}$$

At the time of entrance, there exists no uncertainty about $\epsilon_{i\tau_1}^{a,s=1}$ because, by definition, students have observed its realization. On the other hand, uncertainty does exist about $\epsilon_{i\tau_2}^{a,s=1}$ and $\epsilon_{i\tau_3}^{a,s=1}$. We assume that $\epsilon_{i\tau_2}^{a,s=1}$ and $\epsilon_{i\tau_3}^{a,s=1}$ are each normally distributed. We normalize each of their means to be zero and denote their standard deviations as $\sigma_{i\tau_2}^{a,s=1}$ and $\sigma_{i\tau_3}^{a,s=1}$, respectively.

Denote $\tilde{w}_{it}^{a,s}$, $s = 0, 1$, as the random variable describing student i's beliefs, at time t, about $w_i^{a,s}$. With $\epsilon_{i\tau_1}^{a,s=1}$ observed before t_0, $\tilde{w}_{it_0}^{a,s=1} \sim N(\epsilon_{i\tau_1}^{a,s=1}, (\sigma_{i\tau_2}^{a,s=1})^2 + (\sigma_{i\tau_3}^{a,s=1})^2)$. Similarly, with $\epsilon_{i\tau_2}^{a,s=1}$ observed between t_0 and t^*, $\tilde{w}_{it^*}^{a,s=1} \sim N(\epsilon_{i\tau_1}^{a,s=1} + \epsilon_{i\tau_2}^{a,s=1}, (\sigma_{i\tau_3}^{a,s=1})^2)$. Then, Equation (3.6) becomes:

$$\begin{aligned}\Delta_i &= \sum_{a=\bar{t}}^{\bar{A}} \beta^{a-t^*} [E_{t=t^*}(w_i^{a,s=1}) - E_{t=t_0}(w_i^{a,s=1})] \\ &= \sum_{a=\bar{t}}^{\bar{A}} \beta^{a-t^*} (\epsilon_{i\tau_2}^{a,s=1}).\end{aligned} \tag{3.8}$$

Motivated by the notion that, during college, learning about future earnings is mostly through permanent factors such as innate ability, we assume that the $\epsilon_{i\tau_2}^{a,s=1}$ are perfectly correlated across all future ages a. Under this assumption, computing the standard deviation of Δ_i from Equation (3.8) implies that σ_i is given by:

$$\sigma_i = \sum_{a=\bar{t}}^{\bar{A}} \beta^{a-t^*} (\sigma_{i\tau_2}^{a,s=1}). \tag{3.9}$$

Motivated by the obvious difficulties of writing survey questions that could directly elicit information about uncertainty resolution, which implies that $\sigma_{i\tau_2}^{a,s=1}$ is not directly available in the data, we proceed under the assumption that all students resolve the same fraction of their initial earnings uncertainty before t^*. Denoting this fraction ρ and recalling that the initial uncertainty about earnings at age a is given by $\sqrt{(\sigma_{i\tau_2}^{a,s=1})^2 + (\sigma_{i\tau_3}^{a,s=1})^2}$, Equation (3.9) becomes:

$$\sigma_i = \rho \sum_{a=\bar{t}}^{\bar{A}} \beta^{a-t^*} \left(\sqrt{(\sigma_{i\tau_2}^{a,s=1})^2 + (\sigma_{i\tau_3}^{a,s=1})^2} \right). \tag{3.10}$$

The computation of the components of Equation (3.10) is discussed in the remainder of Section 3.4.2.

Computing $\sum_{a=\bar{t}}^{\bar{A}} \beta^{a-t^*}(\sqrt{(\sigma_{i\tau_2}^{a,s=1})^2 + (\sigma_{i\tau_3}^{a,s=1})^2})$ from Survey Data

In this section, we describe the computation of $\sum_{a=\bar{t}}^{\bar{A}} \beta^{a-t^*}(\sqrt{(\sigma_{i\tau_2}^{a,s=1})^2 + (\sigma_{i\tau_3}^{a,s=1})^2})$. This term corresponds to the standard deviation of the random variable describing student i's beliefs about the discounted lifetime earnings associated with the graduation alternative, $Y_i^{s=1}$. As a result, we denote this term $\tilde{\sigma}_{it_0}^{Y,s=1}$. Similarly, because the term $\sqrt{(\sigma_{i\tau_2}^{a,s=1})^2 + (\sigma_{i\tau_3}^{a,s=1})^2}$ corresponds to the standard deviation of $\tilde{w}_{it_0}^{a,s=1}$, we denote it $\tilde{\sigma}_{it_0}^{a,s=1}$. With this notation, Equation (3.10) can be written as:

$$\begin{aligned} \sigma_i &= \rho \tilde{\sigma}_{it_0}^{Y,s=1} \\ &= \rho \sum_{a=\bar{t}}^{\bar{A}} \beta^{a-t^*} \tilde{\sigma}_{it_0}^{a,s=1}. \end{aligned} \tag{3.11}$$

Our approach for computing $\tilde{\sigma}_{it_0}^{a,s=1}$, and therefore $\tilde{\sigma}_{it_0}^{Y,s=1}$, takes advantage of a sequence of survey questions that elicits information about $\tilde{w}_{it_0}^{a,s=1}$. Specifically, following the format of Question 1 in Appendix A, a respondent reports, at a particular time t, the three quartiles, $Q_{it}^{k,a,s}$, $k = 1, 2, 3$, of the distribution describing her beliefs about what her earnings will be at a particular future age a under choice s. Maintaining the assumption that this distribution is normal, the standard deviation ($\tilde{\sigma}_{it}^{a,s=1}$) of the distribution is given by:

$$\tilde{\sigma}_{it}^{a,s} = (Q_{it}^{3,a,s} - Q_{it}^{1,a,s}) / [\Phi(0.75) - \Phi(0.25)], \tag{3.12}$$

where $\Phi(\cdot)$ is the standard normal cdf.

Equation (3.11) shows that the computation of $\tilde{\sigma}_{it_0}^{Y,s=1}$ requires taking into account a student's uncertainty about earnings, $\tilde{\sigma}_{it}^{a,s=1}$, for all future ages a. As can be seen in Question 1, the earnings expectations questions in the BPS were asked for three specific ages a: the first year after graduation (age 23), age 28, and age 38. Following Stinebrickner and Stinebrickner (2014b), we assume that $\tilde{\sigma}_{it}^{a,s=1}$ grows linearly between the first post-college year and age 28, grows linearly between ages 28 and 38, and does not change after age 38 (until the age of retirement, $\bar{A} = 65$). We operationalize our stylized model by assuming that a student enters college at age 19 ($t_0 = 19$), decides whether to drop out at the end of the third year ($t^* = t_0 + 3$), and graduates at age 23 ($\bar{t} = 23$) if she chooses to remain in school.[15] Focusing on the case

[15]Our choice of $t^* = t_0 + 3$ was informed by Gong, Stinebrickner and Stinebrickner (2019) who found that

where $t = t_0$ and $s = 1$, Equation (3.12), together with the interpolation and timing assumptions above, allows the computation of $\tilde{\sigma}^{Y,s=1}_{it_0}$.[16] We report all values in 2001 dollars. The first column of Table 3.1 shows that the average value of $\tilde{\sigma}^{Y,s=1}_{it_0}$ is \$226,000 for our primary sample.[17]

Table 3.1: Descriptive Statistics

# of Observations: 337	$\tilde{\sigma}^{Y,s=1}_{it_0}$	$\tilde{\sigma}^{Y,s=0}_{it_0}$	$\tilde{\mu}^{Y,s=1}_{it_0}$	$\tilde{\mu}^{Y,s=0}_{it_0}$
Sample Mean	226	163	954	680
Sample Std	201	145	436	333

Note: The unit of measurement is one thousand dollars.

Actual Uncertainty Resolution

In this section, we describe the estimation of the actual fraction of initial earnings uncertainty that is resolved before t^*, ρ_A. Our approach takes advantage of the fact that the longitudinal nature of the BPS expectations data provides direct evidence about the extent to which uncertainty decreases over time.

Earlier we have shown that $\tilde{\sigma}^{Y,s=1}_{it_0}$ can be constructed from the expectations data reported at the time of entrance. Using the same method, the expectations data collected at t^* allows us to also construct $\tilde{\sigma}^{Y,s=1}_{it^*}$, the standard deviation of student i's beliefs about $Y_i^{s=1}$ at t^*. Of interest here is the relationship between these values. Recall that:

$$\tilde{\sigma}^{Y,s=1}_{it_0} = \sum_{a=\bar{t}}^{\bar{A}} \beta^{a-t^*} \left(\sqrt{(\sigma^{a,s=1}_{i\tau_2})^2 + (\sigma^{a,s=1}_{i\tau_3})^2} \right). \tag{3.13}$$

Taking into account that the factor $\epsilon^{a,s=1}_{i\tau_2}$ is realized by t^*, so that no uncertainty about this

the vast majority of uncertainty resolution during college takes place before the end of the third year. However, perhaps more importantly, we find that, because uncertainty resolution tends to take place rather quickly, our results change little if we assume that dropout takes place at the end of the second year, i.e., $t^* = t_0 + 2$.

[16]We assume that the discount factor β is equal to 0.95.

[17]Using the same method, we can also compute $\tilde{\sigma}^{Y,s=0}_{it_0} \equiv \sum_{a=t^*}^{\bar{A}} \beta^{a-t^*} (\tilde{\sigma}^{a,s=0}_{it_0})$, the standard deviation of the random variable describing student i's beliefs about the discounted lifetime earnings associated with the dropout alternative. As reported in the second column of Table 3.1, the sample average of $\tilde{\sigma}^{Y,s=0}_{it_0}$ is \$163,000, implying that students on average are more uncertain about earnings associated with the graduation alternative.

factor remains at t^*,

$$\begin{aligned}
\tilde{\sigma}^{Y,s=1}_{it^*} &= \sum_{a=\bar{t}}^{\bar{A}} \beta^{a-t^*} (\sigma^{a,s=1}_{i\tau_3}) \\
&= \sum_{a=\bar{t}}^{\bar{A}} \beta^{a-t^*} \left(\sqrt{(\sigma^{a,s=1}_{i\tau_2})^2 + (\sigma^{a,s=1}_{i\tau_3})^2 - (\sigma^{a,s=1}_{i\tau_2})^2} \right) \\
&= \sqrt{1-\rho_A^2} \sum_{a=\bar{t}}^{\bar{A}} \beta^{a-t^*} \left(\sqrt{(\sigma^{a,s=1}_{i\tau_2})^2 + (\sigma^{a,s=1}_{i\tau_3})^2} \right),
\end{aligned} \tag{3.14}$$

where the last line in Equation (3.14) follows from the assumption that all students resolve the same fraction of initial earnings uncertainty: $\sigma^{a,s=1}_{i\tau_2} = \rho_A \sqrt{(\sigma^{a,s=1}_{i\tau_2})^2 + (\sigma^{a,s=1}_{i\tau_3})^2}$ for all i.

Equation (3.13) and (3.14) together show that $\sqrt{1-\rho_A^2}$ can be computed using the ratio of the average of $\tilde{\sigma}^{Y,s=1}_{it^*}$ to the average of $\tilde{\sigma}^{Y,s=1}_{it_0}$ for the same sample of students.[18] Using the sample of students who were still in school at $t^* = 3$, the estimated value of $\sqrt{1-\rho_A^2}$ is 0.86.[19] Hence, the estimated value of ρ_A is 0.51.[20] Then, Equation (3.11) can be used to compute the actual value of σ_i for each student in our sample.

Perceived Uncertainty Resolution

In this section, we describe how the perceived fraction of initial earnings uncertainty that is resolved before t^*, ρ_P, can be esimated using a simple model of dropout.

At the time of entrance (t_0), a student reports her perceived dropout probability, $P^{P,s=0}_i$.

[18] We choose to use the ratio of the average of $\tilde{\sigma}^{Y,s=1}_{it^*}$ to the average of $\tilde{\sigma}^{Y,s=1}_{it_0}$ rather than, for example, the average of the ratio of $\tilde{\sigma}^{Y,s=1}_{it^*}$ to $\tilde{\sigma}^{Y,s=1}_{it_0}$, because the former tends to be a consistent estimator of $\sqrt{1-\rho_A^2}$ even when individual uncertainty measures might potentially contain measurement error.

[19] In practice, some students dropped out of college before $t^* = 3$, and, therefore, were not included in the estimation of ρ_A. One might be concerned that those who dropped out before t^* might have resolved systematically different fractions of their initial uncertainty under the counterfactual in which they stayed until t^* than those who actually remained in our sample until t^*. As a robustness check, it would be desirable to add students who dropped out before t^* to our estimation sample. We do this by using a student's last observed earnings uncertainty as a proxy for what her earnings uncertainty would have been at t^*. Given that students who dropped out before t^* would have resolved additional uncertainty between the time of dropout and t^* if they had remained in school, the resulting estimator should produce a lower bound for ρ_A. We find that this lower bound is 0.41 and that the corresponding lower bound for the average actual option value is \$19,990. As we show later in Section 3.4.3, this lower bound is still substantially higher than the estimated average perceived option value, suggesting that our main conclusion that students vastly underestimate the option value is robust to the selection issue.

[20] Our results about actual earnings uncertainty resolution are comparable in magnitude to what was found in Gong, Stinebrickner, and Stinebrickner (2019), which also take advantage of the BPS dataset. Using data for both the 2000 and the 2001 cohorts, they find that the sample average of the standard deviation of the distribution describing students' beliefs about $w^{28,s=1}_i$ at the end of the third year ($t = t^*$) is roughly 82% of the sample average of the standard deviation of the distribution describing students' beliefs about $w^{28,s=1}_i$ at the beginning of college ($t = t_0$).

Equation (3.2) shows that this perceived probability depends on a student's beliefs about the distance from the margin of indifference at t_0, $E_{t=t_0}(V_{it^*}^{s=1} - V_{it^*}^{s=0})$, and her perceptions about σ_i. With the expression for $V_{it^*}^{s}$ coming from Equation (3.5) and the expression for the perceived value of σ_i coming from Equation (3.11) with ρ replaced by ρ_P, we obtain:

$$\begin{aligned} P_i^{P,s=0} &= \Phi(\frac{E_{t=t_0}[(Y_i^{s=0} + \gamma_i^{s=0}) - (Y_i^{s=1} + \gamma_i^{s=1})]}{\rho_P \tilde{\sigma}_{it_0}^{Y,s=1}}) \\ &= \Phi(\frac{E_{t=t_0}(Y_i^{s=0}) - E_{t=t_0}(Y_i^{s=1}) + E_{t=t_0}(\gamma_i^{s=0} - \gamma_i^{s=1})}{\rho_P \tilde{\sigma}_{it_0}^{Y,s=1}}) \\ &= \Phi(\frac{\tilde{\mu}_{it_0}^{Y,s=0} - \tilde{\mu}_{it_0}^{Y,s=1} + E_{t=t_0}(\gamma_i^{s=0} - \gamma_i^{s=1})}{\rho_P \tilde{\sigma}_{it_0}^{Y,s=1}}), \end{aligned} \tag{3.15}$$

where $\tilde{\mu}_{it_0}^{Y,s} \equiv E_{t=t_0}(Y_i^s)$ represents the mean of student i's beliefs about Y_i^s at t_0 and $\tilde{\gamma}_i \equiv E_{t=t_0}(\gamma_i^{s=0} - \gamma_i^{s=1})$ represents student i's expectation about the difference in the non-pecuniary benefits associated with the two alternatives.

The intuition underlying the role of ρ_P in Equation (3.15) is clear. The numerator in the probability expression is the difference between the expected utility of $s = 0$ and the expected utility of $s = 1$, at t_0. Thus, for example, a negative numerator represents the distance that a student is "above" the margin of dropping out at the time of entrance. A larger denominator implies that a student resolves more uncertainty about earnings between t_0 and t^*, thereby increasing the probability that the new information she receives will push her across the margin into a dropout decision; all else equal, in the seemingly most likely scenario in which the numerator is negative, the dropout probability will tend to be increasing in the denominator.[21] Roughly speaking, identification of ρ_P comes from the fact that the relationship between the amount of uncertainty at the time of entrance, $\tilde{\sigma}_{it_0}^{Y,s=1}$, and the perceived dropout probability, $P_i^{P,s=0}$, will tend to be stronger when ρ_P is high (than when ρ_P is low) because ρ_P maps the amount of initial uncertainty into the amount of uncertainty that the students believes will be resolved.

As described in previous sections, $P_i^{P,s=0}$ and $\tilde{\sigma}_{it_0}^{Y,s=1}$ can be obtained using students' responses to survey Questions 4 and 1, respectively. Appendix C.1 shows that $\tilde{\mu}_{it_0}^{Y,s}$ can also be computed using survey Question 1, in a manner similar to that used for the computation of $\tilde{\sigma}_{it_0}^{Y,s=1}$. As reported in the last two columns of Table 3.1, at t_0, the sample average of expected lifetime earnings associated with the graduation scenario ($\tilde{\mu}_{it_0}^{Y,s=1}$) and the dropout scenario ($\tilde{\mu}_{it_0}^{Y,s=0}$) are approximately \$954,000 and \$680,000, respectively.

[21] Of course, from a theoretical standpoint, when experimentation plays a role in the decision to enter school, a student might enter even if she has a positive numerator.

The only components in Equation (3.15) that are yet known to us are a common parameter ρ_P and individual-specific net non-pecuniary benefits $\tilde{\gamma}_i$. To estimate the value of ρ_P (and the distribution of $\tilde{\gamma}_i$), we rewrite Equation (3.15) as follows:

$$\Phi^{-1}(P_i^{P,s=0})\tilde{\sigma}_{it_0}^{Y,s=1} = \frac{\bar{\gamma}}{\rho_P} + [\tilde{\mu}_{it_0}^{Y,s=0} - \tilde{\mu}_{it_0}^{Y,s=1}]\frac{1}{\rho_P} + \frac{\tilde{\gamma}_i - \bar{\gamma}}{\rho_P}, \tag{3.16}$$

where $\bar{\gamma} = E(\tilde{\gamma}_i)$.

If all the expectations variables ($P_i^{P,s=0}$, $\tilde{\mu}_{it_0}^{Y,s=1}$, $\tilde{\mu}_{it_0}^{Y,s=0}$, and $\tilde{\sigma}_{it_0}^{Y,s=1}$) are measured perfectly, then $\frac{\bar{\gamma}}{\rho_P}$ and $\frac{1}{\rho_P}$ can be estimated via an easy-to-implement OLS regression of $\Phi^{-1}(P_i^{P,s=0})\tilde{\sigma}_{it_0}^{Y,s=1}$ on $[\tilde{\mu}_{it_0}^{Y,s=0} - \tilde{\mu}_{it_0}^{Y,s=1}]$. However, it is worthwhile to address the concern that responses to survey questions eliciting expectations may contain a non-trivial amount of measurement error (e.g., Manski and Molinari, 2010, Ameriks et al., 2019, Giustinelli, Manski, and Molinari, 2019, and Gong, Stinebrickner, and Stinebrickner, 2019), which can lead to well-known attenuation bias in the estimation of linear models such as Equation (3.16). We first modify Equation (3.16) to accommodate measurement error:

$$\Phi^{-1}(P_i^{P,s=0})\tilde{\sigma}_{it_0}^{Y,s=1} + \Delta y_i = \frac{\bar{\gamma}}{\rho_P} + [\tilde{\mu}_{it_0}^{Y,s=0} - \tilde{\mu}_{it_0}^{Y,s=1} + \Delta\mu_i^Y]\frac{1}{\rho_P} + \frac{\tilde{\gamma}_i - \bar{\gamma}}{\rho_P}. \tag{3.17}$$

In this specification, the observed measure of the pecuniary component of the initial expectations gap, $\tilde{\mu}_{it_0}^{Y,s=0} - \tilde{\mu}_{it_0}^{Y,s=1}$, contains classical measurement error $\Delta\mu_i^Y$. In addition, we also allow the computed value of $\Phi^{-1}(P_i^{P,s=0})\tilde{\sigma}_{it_0}^{Y,s=1}$ to contain classical measurement error Δy_i.[22] In Appendix C.4.2, we show that, under these assumptions, the attenuation bias in the estimation of $\frac{\bar{\gamma}}{\rho_P}$ and $\frac{1}{\rho_P}$ can be corrected if the variance of $\Delta\mu_i^Y$ is known.[23] In Appendix C.4.1, we describe how to utilize the method developed in Gong, Stinebrickner and Stinebrickner (2019) to estimate $var(\Delta\mu_i^Y)$.[24] We find that, after correcting for the attenuation bias, the estimate of ρ_P is 0.51, which is almost identical to the second decimal to the actual value of ρ.

[22] Δy_i may be relevant because either answers to survey questions eliciting perceived dropout probabilities or answers to survey questions eliciting earnings expectations (specifically, initial uncertainty) may be measured with error.

[23] The presence of classical measurement error in dependent variable (Δy_i) does not affect the consistency of the OLS estimator.

[24] The BPS contains two sets of survey questions that can be used to compute a student's unconditional expectation about earnings at age 28, $\tilde{\mu}_{it_0}^{28,s=1}$. Intuitively, differences in the unconditional expectations computed using these two sets of expectations questions are informative about the amount of measurement error, $\Delta\mu_i^{28,s=1}$, present in the observed measure of $\tilde{\mu}_{it_0}^{28,s=1}$. Gong, Stinebrickner and Stinebrickner (2019) formalize this intuition and develop a method to estimate $var(\Delta\mu_i^{28,s=1})$ under the assumption that the measurement error is classical. In Appendix C.4.1,, we adopt the same method to estimate $var(\Delta\mu_i^{28,s=1})$ using our sample (the 2001 cohort) and detail the assumptions that are required to compute $var(\Delta\mu_i^Y)$ using $var(\Delta\mu_i^{28,s=1})$

3.4.3 Actual and Perceived Option Values

Given individual-specific actual and perceived values for σ_i and $P_i^{s=0}$, we are able to compute the actual and perceived option value for each student using Equation (3.4). The solid line in Figure 3.2 shows the cdf for the estimated actual option values. The sample average and standard deviation of the actual option values are \$25,040 and \$28,440, respectively. Our finding about the average actual option value is generally similar to what has been found in the literature using very different methods. For example, estimating a schooling decision model under Rational Expectations assumptions, Stange (2012) finds that the OV is roughly \$19,000 (in 2001 dollars) for an average high school graduate in the United States.

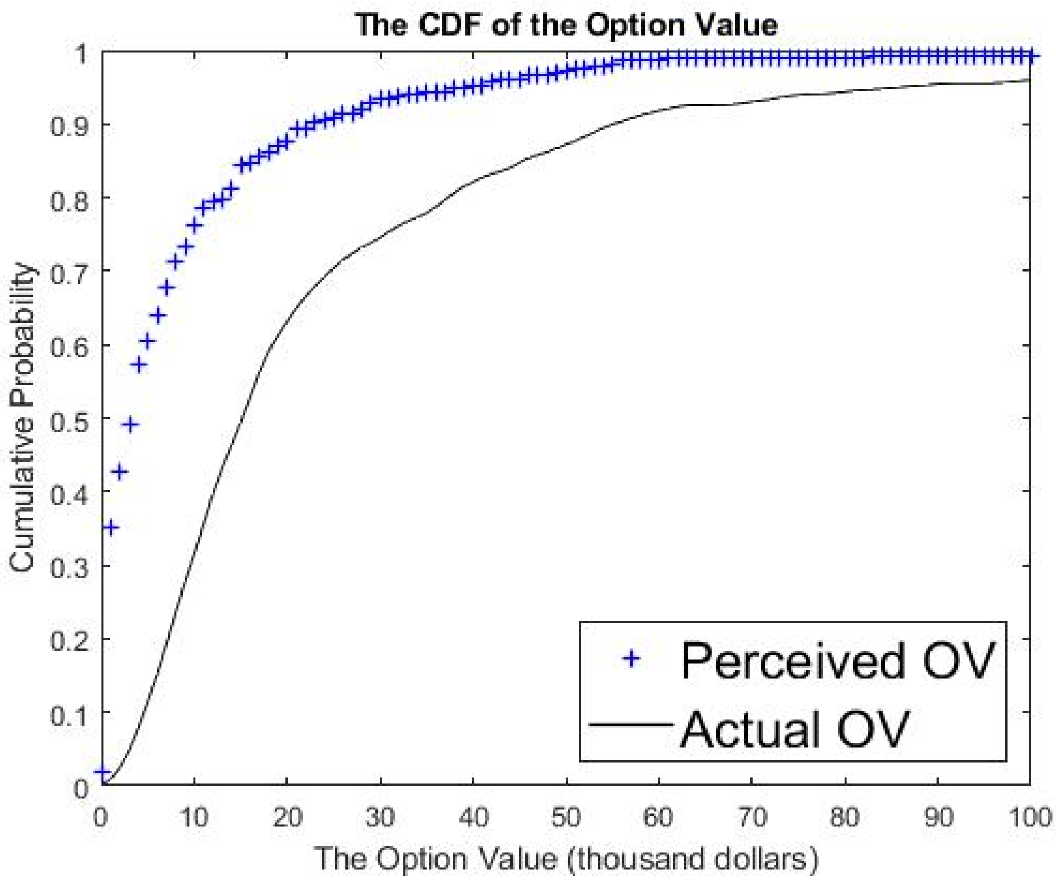
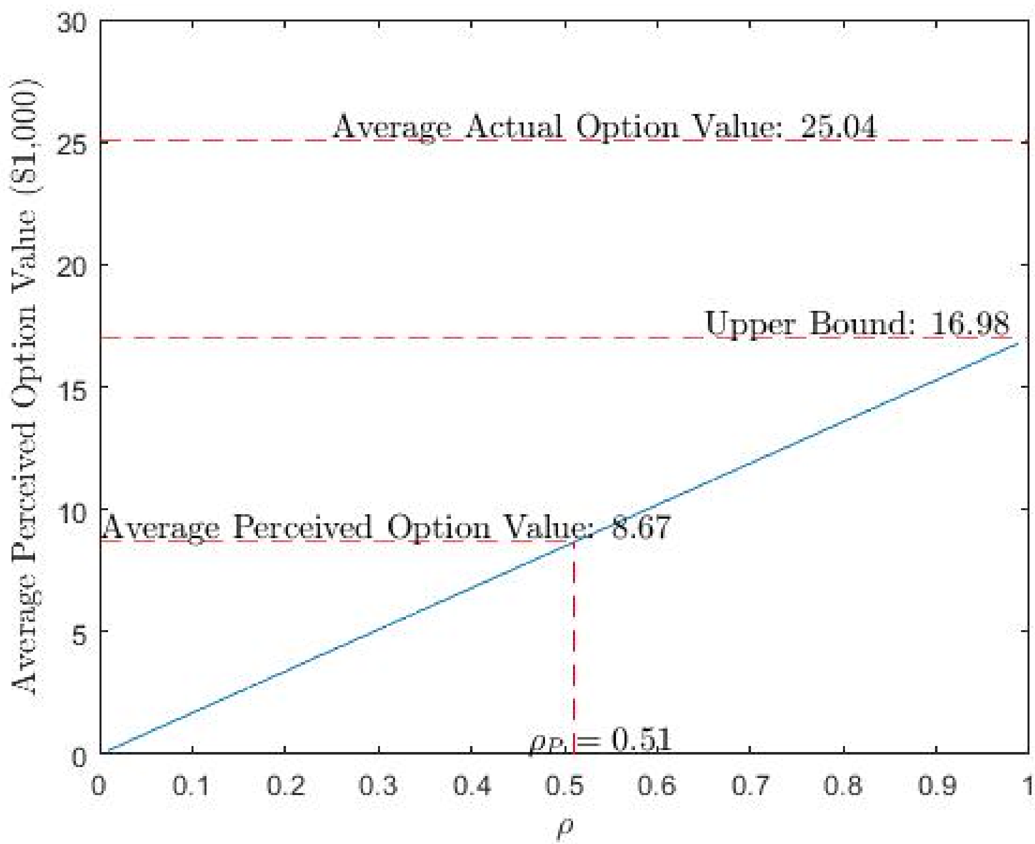


Figure 3.2: The CDF of the Option Value

The "+" line in Figure 3.2 shows the cdf for estimated perceived options values. The sample average and standard deviation of the perceived option values are \$8,670 and \$16,400, respectively. Consistent with what was suggested by a comparison between actual and perceived dropout probabilities in Section 3.4.1, students at Berea College do indeed vastly underestimate the option value of college enrollment.

Perhaps the most tenuous parameter to estimate is the perceived fraction of initial uncertainty that is resolved between t_0 and t^*, ρ_P, which in turn determines the perceived value of σ_i. However, the finding that students underestimate the option value is robust to the estimate of the perceived value of σ_i. An upper bound on the perceived value of σ_i can be obtained from Equation (3.11) by assuming that students believe they will fully resolve their initial uncertainty about lifetime earnings ($\rho_P = 1$). Combining this upper bound of σ_i and data on the perceived dropout probability, we can compute an upper bound for the perceived option value. As shown in Figure 3.3, the upper bound for the average perceived OV would still be roughly \$8,000 lower than the average actual OV, due to the considerable underestimation of the dropout probability.

Figure 3.3: Perceived Option Value and ρ_P



Our results about actual and perceived option values are obtained under a simplifying assumption that ρ_A and ρ_P are homogeneous across students. Motivated by Stinebrickner and Stinebrickner (2012), who show that the amount of learning during college tends to be different between males and females, we also redo our analysis separately for males and females.

We find that, while male students understate the amount of earnings uncertainty resolution ($\rho_A = 0.64$, $\rho_P = 0.40$), female students overstate the amount of earnings uncertainty resolution somewhat ($\rho_A = 0.40$, $\rho_P = 0.56$). We combine these gender-specific estimates of ρ_A and ρ_P with information on dropout probabilities and initial earnings uncertainty to compute actual and perceived option values. Our main result that students underestimate the option value holds for both males and females. However, we do find some gender differences; while, on average, perceptions about the option value are similar for males and females ($8,440 for males, $7,660 for females), there exists a substantial gender gap in the average actual option value ($39,690 for males, $15,200 for females). [25]

3.4.4 Policy Implications

Our finding that students' perceptions understate the actual option value of college enrollment raises a question fundamental to the general policy concern that informational problems may cause too few students to enter college: what would happen if misperceptions about the option value were corrected? Importantly, our approach allows us to examine not only whether misperceptions about the option value exist, but also why they exist. As highlighted by the simple conceptual model in the second paragraph of the introduction, the understatement of the option value could be caused by either an understatement of how much uncertainty is resolved during college or an overly optimistic view about the size of the initial expectations gap. Then, our finding that perceptions about uncertainty resolution are accurate implies that individuals overstate the size of the initial expectations gap. Correcting misperceptions about the option value would involve providing information about, for example, the returns to college.

These findings about the reason for misperceptions about the option value are important because, while it may seem at a first glance that an understatement of the option value would necessarily lead to too few students entering college, in reality whether this is true depends critically on why misperceptions exist. This is the case because the overall value of college, which is the relevant object for the college entrance decision, is strongly related but not identical to the option value. Under the illustrative scenario in the second paragraph of the introduction - where all that occurs between t_0 and t^* is that students resolve uncertainty that existed at entrance - the overall value of college is given by the net continuation value (Heckman, Lochner,

[25] We also conducted a similar analysis to examine whether ρ_A and ρ_P depend on other observed characteristics. For example, dividing students in our sample into two equal-sized subgroups based on their high school GPA (HSGPA), we find that, while students with high HSGPA expect to resolve a slightly larger fraction of initial earnings uncertainty ($\rho_P = 0.60$) than students with low HSGPA ($\rho_P = 0.46$), the actual fraction ρ_A is very similar for these two groups of students ($\rho_A = 0.51$ for high HSGPA, $\rho_A = 0.52$ for low HSGPA). For each group, we again find that, on average, perceptions about the option value ($9,680 for high HSGPA, $8,310 for low HSGPA) understate the actual option value ($14,500 for high HSGPA, $37,420 for low HSGPA).

and Todd, 2006, Heckman and Navarro, 2007, and Heckman and Urzua, 2008).[26] Defined as $E_{t=t_0}[\max(V_{it^*}^{s=1}, V_{it^*}^{s=0})] - E_{t=t_0}(V_{it^*}^{s=0})$, the net continuation value (NCV) captures the expected continuation value of college enrollment net of the value of the outside option (dropout). In the scenario where the initial expectations gap, $E_{t=t_0}(V_{it^*}^{s=1} - V_{it^*}^{s=0})$, is negative, the net continuation value and the option value are identical. However, in the more likely case where the initial expectations gap is positive, the net continuation value is equal to the sum of the option value and the initial expectations gap. Therefore, with the option value computed using methods described in previous sections and $E_{t=t_0}(V_{it^*}^{s=1} - V_{it^*}^{s=0})$ uniquely determined by data on σ_i and $P_i^{s=0}$, we can compute the actual and perceived NCV for each student.[27]

We start by computing the perceived and actual value of $E_{t=t_0}(V_{it^*}^{s=1} - V_{it^*}^{s=0})$ for each student. Equation (3.2) implies that $E_{t=t_0}(V_{it^*}^{s=1} - V_{it^*}^{s=0})$ is given by:

$$E_{t=t_0}(V_{it^*}^{s=1} - V_{it^*}^{s=0}) = -\Phi^{-1}(P_i^{s=0})\sigma_i. \tag{3.18}$$

Using the actual dropout probability, $P_i^{s=0}$, and the actual amount of uncertainty resolution, σ_i, we find that, on average, the actual value of $E_{t=t_0}(V_{it^*}^{s=1} - V_{it^*}^{s=0})$ is \$45,120. Similarly, using perceived values of $P_i^{s=0}$ and σ_i, we find that the average perceived value of $E_{t=t_0}(V_{it^*}^{s=1} - V_{it^*}^{s=0})$ is \$160,930. Thus, at the time of entrance, students overestimated the expected net benefit of college completion (i.e., the initial expectations gap) by more than \$115,000. This implies that misperceptions about the option value and misperceptions about the initial expectations gap work in an offsetting manner when computing the NCV. Taking both into account, we find that perceptions about the NCV somewhat overstate its actual value; the actual and perceived NCV are \$76,130 and \$173,110, respectively. Thus, while students underestimate the option value, once one considers the NCV, concerns that there might be too few students attending college tend to dissipate.

3.5 Conclusion

From a student's perspective, the return to college education is likely to be uncertain when she makes the college attendance decision. Having the option to decide whether to remain in college or to drop out after receiving relevant new information can potentially help students insure against this uncertainty. Complementing administrative data on college completion with data describing students' beliefs, at the time of entrance, about the probability of dropping out

[26]In the more general case, the overall value of college continues to take into account the NCV, but also takes into account the direct utility differences between the two options over the period t_0 to t^*.

[27]Alternatively, similar to what we did for the option value, we can directly express the net continuation value as a function of σ_i and $P_i^{s=0}$. We can show that the NCV is increasing in σ_i and decreasing in $P_i^{s=0}$.

and data describing students' beliefs, at multiple points in college, about future earnings allows us to pay careful attention to the distinction between perceived and actual option values.

We find strong evidence that students substantially underestimate the experimentation benefits of enrolling in college. However, importantly, we find that this underestimate is caused by an overly optimistic view about the size of the initial expectations gap, rather than an understatement of the amount of uncertainty that is resolved during college. This has important implications for whether inaccurate perceptions create a situation where too few students are entering college. In the calculation of the overall value of college, the underappreciation of the experimental benefit is more than offset by overoptimism about the initial expectations gap. Once one considers the overall value of college, concerns that there might be too few students attending college tend to dissipate.

As in our other work using the BPS, we feel it is important to be appropriately cautious when thinking about exactly how the results from our study would generalize to other institutions. Our results are perhaps most relevant for thinking about students from low income backgrounds, who are a primary focus of the educational mission at Berea College. This group is of particular policy interest, in part because they may be more likely to be affected by informational problems. In addition, from a methodological standpoint, we feel that this chapter provides a concrete example of how unique expectations data can be useful for characterizing difficult-to-identify objects of direct policy relevance.

Bibliography

[1] Brant Abbott, Giovanni Gallipoli, Costas Meghir, and Giovanni L Violante. Education policy and intergenerational transfers in equilibrium. *Journal of Political Economy*, forthcoming.

[2] Joseph G Altonji. The demand for and return to education when education outcomes are uncertain. *Journal of Labor Economics*, 11:48–83, 1993.

[3] John Ameriks, Gábor Kézdi, Minjoon Lee, and Matthew D Shapiro. Heterogeneity in expectations, risk tolerance, and household stock shares: The attenuation puzzle. *Journal of Business & Economic Statistics*, pages 1–27, 2019.

[4] Peter Arcidiacono. Ability sorting and the returns to college major. *Journal of Econometrics*, 121:343–375, 2004.

[5] Peter Arcidiacono, Esteban Aucejo, Arnaud Maurel, and Tyler Ransom. College attrition and the dynamics of information revelation. *NBER Working Paper*, (w22325), 2016.

[6] Peter Arcidiacono, V Joseph Hotz, Arnaud Maurel, and Teresa Romano. Ex ante returns and occupational choice. *Working Paper*, 2019.

[7] Zachary Bleemer and Basit Zafar. Intended college attendance: Evidence from an experiment on college returns and costs. *Journal of Public Economics*, 157:184–211, 2018.

[8] Flavio Cunha, James J Heckman, and Salvador Navarro. Separating uncertainty from heterogeneity in life cycle earnings. *Oxford Economic Papers*, 57:191–261, 2005.

[9] Xavier d'Haultfoeuille, Christophe Gaillac, and Arnaud Maurel. Rationalizing rational expectations? tests and deviations. *NBER Working Paper*, (w25274), 2018.

[10] Pamela Giustinelli, Charles F Manski, and Francesca Molinari. Precise or imprecise probabilities? evidence from survey response on late-onset dementia. *NBER Working Paper*, (w23636), 2019.

[11] Pamela Giustinelli and Matthew D Shapiro. Seate: Subjective ex ante treatment effect of health on retirement. *NBER Working Paper*, (w26087), 2019.

[12] Yifan Gong, Todd Stinebrickner, and Ralph Stinebrickner. Uncertainty about future income: Initial beliefs and resolution during college. *Quantitative Economics*, 10(2):607–641, 2019.

[13] James J Heckman, Lance J Lochner, and Petra E Todd. Earnings functions, rates of return and treatment effects: The mincer equation and beyond. *Handbook of the Economics of Education*, 1:307–458, 2006.

[14] James J Heckman and Salvador Navarro. Dynamic discrete choice and dynamic treatment effects. *Journal of Econometrics*, 136:341–396, 2007.

[15] James J Heckman and Sergio Urzua. The option value of educational choices and the rate of return to educational choices. In *Unpublished manuscript, University of Chicago. Presented at the Cowles Foundation Structural Conference, Yale University*, 2008.

[16] Caroline Hoxby, Sarah Turner, et al. Expanding college opportunities for high-achieving, low income students. *Stanford Institute for Economic Policy Research Discussion Paper*, (12-014), 2013.

[17] Michael P Keane and Kenneth I Wolpin. The career decisions of young men. *Journal of political Economy*, 105(3):473–522, 1997.

[18] Charles F Manski. Schooling as experimentation: a reappraisal of the postsecondary dropout phenomenon. *Economics of Education review*, 8(4):305–312, 1989.

[19] Charles F Manski and Francesca Molinari. Rounding probabilistic expectations in surveys. *Journal of Business & Economic Statistics*, 28(2):219–231, 2010.

[20] Kevin M Stange. An empirical investigation of the option value of college enrollment. *American Economic Journal: Applied Economics*, 4:49–84, 2012.

[21] Ralph Stinebrickner and Todd Stinebrickner. The effect of credit constraints on the college drop-out decision: A direct approach using a new panel study. *The American economic review*, 98:2163–2184, 2008.

[22] Ralph Stinebrickner and Todd Stinebrickner. Learning about academic ability and the college dropout decision. *Journal of Labor Economics*, 30:707–748, 2012.

[23] Ralph Stinebrickner and Todd Stinebrickner. A major in science? initial beliefs and final outcomes for college major and dropout. *Review of Economic Studies*, 81:426–472, 2014a.

[24] Ralph Stinebrickner and Todd Stinebrickner. Academic performance and college dropout: Using longitudinal expectations data to estimate a learning model. *Journal of Labor Economics*, 32:601–644, 2014b.

[25] Ralph Stinebrickner and Todd R Stinebrickner. Understanding educational outcomes of students from low-income families evidence from a liberal arts college with a full tuition subsidy program. *Journal of Human Resources*, 38:591–617, 2003.

[26] Matthew Wiswall and Basit Zafar. Human capital investments and expectations about career and family. *NBER Working Paper*, (w22543), 2016.

Chapter 4

Identification of Signal-based Learning Models without the Rational Expectations Assumption

4.1 Introduction

Motivated by the potential limitations of the data-based approach adopted in Chapter 2 and Chapter 3, in this chapter I provide constructive proofs for the nonparametric identification of individuals' beliefs and the decision rule, i.e., the function that maps beliefs to decisions, in signal-based learning models. Specifically, I consider a multi-period environment where individuals use signals to update their beliefs about an unknown permanent factor and repeatedly make decisions based on these beliefs. The econometrician observes individuals' decisions and signals as well as factors that determine individuals' initial beliefs. This environment nests many models that are of interest to researchers. In the context of higher education, college students use their semester GPA to update their beliefs about own academic ability, which influence their college attendance/dropout decisions (Stange, 2012; Stinebrickner and Stinebrickner, 2014a). Conley and Udry (2010) model the adoption of a new production technology as a learning process: Workers use realized output/profit as signals to update their beliefs about the production function and choose the level of input based on these beliefs.

This chapter focuses on the identification of the mean of the distribution describing individuals' beliefs, which, hereafter, I refer to as the subjective mean. The identification result leverages an assumption on the process governing the updating of an individual's subjective mean that is both intuitively appealing and standard in the literature: The posterior mean is (1) strictly increasing in the signal and (2) the same as the prior mean whenever the signal is equal to the prior mean.[1] The main identification results, presented in Section 4.3, are for the case where, conditional on other observables that affect their decisions, the average decision of individuals with identical subjective means is a time-invariant and strictly monotonic function of this subjective mean and does not depend on other moments of their subjective beliefs. Intuitively, in this environment, part (2) of the assumption on updating rules implies that, if a group of individuals with identical prior means receive a signal that is equal to this prior mean, then their average decision before receiving the signal is identical to their average decision after receiving the signal. The monotonicity assumptions guarantee that the converse of this statement is also true: The prior mean for this group of individuals is equal to the signal that induces identical average decisions in the two periods. Now consider the case where there are sufficiently many such groups. The prior mean and the average decision for each of these groups can be identified following the strategy above, which allows the decision rule to be pinned down.

I choose this environment as the benchmark primarily for three reasons. First, this environment nests many canonical models. For example, the input decision of a profit-maximizing firm with a linear production function and a convex cost function is an increasing function of the firm's subjective mean of the price for its output (or the subjective mean of its productivity), and does not depend on other moments of its subjective beliefs (Baron, 1970; Sandmo, 1971; Leland, 1972; Holthausen, 1976). Second, this environment can be considered as a limiting case of a fairly general class of models. For example, while the decision rules in a finite horizon dynamic model generally vary over time even when individuals' per-period utility functions and constraints are time-invariant, the decision rules in the first two periods may become arbitrarily close as the time horizon becomes sufficiently long. Third, this parsimonious environment highlights the most fundamental elements of my identification strategy. Moreover, as shown in Section 4.6.1, the identification strategy in this environment can be extended to more complicated, nonstationary environments such as dynamic discrete choice models.

As shown in Section 4.3, no additional parametric assumptions on either the decision rule or the updating rule are required for identification. Hence, in theory, individuals' prior means and the decision rule can be nonparametrically identified. However, as is well known, nonparametric estimators often suffer from the curse of dimensionality in practice. Motivated by this concern, in Section 4.4, I propose a feasible semiparametric estimator that is free of the curse of dimensionality and examine its performance through simulation exercises. Importantly, I impose parametric assumptions on the decision rule but still allow updating rules to be nonparametric and heterogeneous across individuals. The crucial assumption that facili-

[1]If an individual updates her beliefs in a Bayesian fashion, then the second part of this assumption is satisfied if both the prior distribution and the distribution of errors in the signal are symmetric (Chambers and Healy, 2012).

tates this semiparametric estimator is that an individual's misperception, as measured by the difference between her prior mean and the subjective mean that is implied by the RE assumption, is a function of a finite type variable that is observed by the econometrician. With the RE-consistent subjective means estimated from realized signals, this assumption reduces the estimation of prior means to the estimation of finitely many possible misperceptions.

It is established in the literature that students' study time is an important determinant of their academic achievements (Stinebrickner and Stinebrickner, 2008; De Fraja, Oliveira, and Zanchi, 2010). However, not much is known about how study time is determined.[2] In Section 4.5, I apply my method to provide evidence on how students' expectations about academic productivity affect their study time. Analogous to productivity in a linear production function, I define a student's academic productivity in each semester as the ratio of her semester GPA to her average daily study time. Hence, if students are solving a utility maximization problem that is similar to the canonical profit-maximizing problem mentioned above, expectations about academic productivity should have a positive effect on study time. Alternatively, if students' primary goal is to achieve certain grades, then students with high expectations would believe that they do not need to spend much time studying. In this case, the effect of expectations about academic productivity on study time should be negative.

The empirical investigation takes advantage of the Berea Panel Study, which contains detailed information about students' study time and GPA for multiple semesters. I estimate a negative effect of expectations about own academic productivity on study time for students in Berea College, suggesting that a student's study effort is likely to be induced by a desire to achieve a fixed grade. A particular focus of this empirical investigation is to test whether college students' prior means of their academic productivity deviate from RE-consistent subjective means. I find that this is the case for students who spent less than 2 hours per day studying in high school: On average, these students overestimate their academic productivity in college by over 15%, and the RE assumption is rejected at a 10% level for them. If I incorrectly impose the RE assumption, the estimated effect of expectations about own academic productivity on study time remains negative. However, the magnitude of this estimate is more than 75% larger than the estimate obtained without the RE assumption.

4.2 Literature Review

This chapter belongs to a relatively recent literature on the joint identification of beliefs and decision rules/preferences without the RE assumption. This identification issue is most exten-

[2]Adopting a field-experimental approach, Ersoy (2019) provides a recent investigation of the relationship between students' study effort and their beliefs about own academic productivity.

sively studied in macroeconomic models. Different from the setting in this chapter, macroeconomists are mostly interested in identifying economic agents' beliefs about aggregate equilibrium objects, such as inflation rates and equilibrium prices. Woodford (2013) provides a recent survey of a variety of approaches to specify and identify agents' beliefs in macroeconomic models. Also interested in identifying non-rational expectations about equilibrium objects, Aguirregabiria and Magesan (forthcoming) consider a game-theoretic setting and allow players' beliefs about the distribution of other players' actions to be biased. Other recent papers on the joint identification of beliefs and preferences in settings different from the one in this chapter include Caplin, Leahy, and Matejka (2016) and Olivi (2019).

Learning about individual-specific permanent factors through private signals is more commonly seen in dynamic discrete choice models where individuals learn about their choice-specific permanent factors through experimenting with different choices. Examples include students learning about (major-specific) academic ability (Arcidiacono, 2004; Stange, 2012; Stinebrickner and Stinebrickner, 2014a; Arcidiacono, Aucejo, et al., 2016) and consumers learning about product-specific taste (Erdem and Keane, 1996; Ackerberg, 2003; Crawford and Shum, 2005; Osborne, 2011). The majority of these papers assume that individuals' initial beliefs are rational conditional on their information set, hence they rule out the possibility of systematically biased initial beliefs. One exception is Ackerberg (2003), which identifies the systematic bias in consumers' initial beliefs about their taste by comparing the pre-experimentation distribution and the post-experimentation distribution of consumers' purchase behavior. Ackerberg (2003) does not assume the econometrician has access to direct signals of consumers' tastes but requires that each consumer's expectation about her taste eventually converges to its true value. In contrast, my method relies on the econometrician observing the signals but does not require an individual to ever fully recover the true value of the unknown permanent factor from the signals.

The empirical application in Section 4.5 contributes to the understanding of the determination of students' study effort. Based on an information experiment with a widely used online learning platform, Ersoy (2019) exogenously manipulates students' beliefs about the effort-performance relationship and finds that students change their study effort in the same direction with the shifts in their beliefs about returns to effort. The difference between my findings in Section 4.5 and the results in Ersoy (2019) suggests that the direction of the effect of expectations about academic productivity on study effort might be context-specific. Indeed, students who voluntarily choose to take classes on online learning platforms are more likely to be active learners whose study effort is chosen to balance the trade-off between the benefits and costs of studying.

The main identification results outlined above require the econometrician to find a group of individuals with identical prior mean. This implies that there is no unobserved (by the econometrician) heterogeneity in individuals' prior means. I show how to extend my method to allow for unobserved heterogeneity in prior means in Section 4.6.2. This extension is related to and motivated by a strand of literature interested in identifying unobserved heterogeneity in prior beliefs under the RE assumption (Carneiro, Hansen, and Heckman, 2003; Cunha, Heckman, and Navarro, 2004, 2005).

4.3 Identification in a Benchmark Environment

In this section, I provide constructive proofs for the identification of individuals' beliefs and the decision rule in a benchmark environment where an individual's decision as a function of her subjective mean of a permanent factor is time-invariant. I start by specifying the environment and notation in Section 4.3.1. The identification results are presented in Section 4.3.2. Finally, I discuss the empirical relevance of this benchmark environment in Section 4.3.3.

4.3.1 Environment and Notation

Throughout this chapter, I use capital letters and lowercase letters to represent random variables and their realizations, respectively. Let μ_{it} denote the mean of the distribution describing individual i's beliefs about an unknown factor A_i at the beginning of period t, and refer to it as the subjective mean of A_i.

Consider an environment where, at each period t, individual i makes decision d_{it} based on observed (by the econometrician) factors x_{it}, unobserved (by the econometrician) factors ϵ_{it}, and the subjective mean μ_{it}, according to a time-invariant decision rule. This decision rule is given by:

$$d_{it} = D(x_{it}, \epsilon_{it}, \mu_{it}). \tag{4.1}$$

Since factors ϵ_{it} are not observed by the econometrician, the identification of $D(x_{it}, \epsilon_{it}, \mu_{it})$ is theoretically impossible. Hence, I focus on the expectation of $D(x_{it}, \epsilon_{it}, \mu_{it})$ with respect to ϵ_{it} instead:

$$\bar{D}_t(x_{it}, \mu_{it}) \equiv E_{\epsilon_{it}|x_{it},\mu_{it}} D(x_{it}, \epsilon_{it}, \mu_{it}) \tag{4.2}$$

Throughout this section, I impose the following assumption unless otherwise specified:

Assumption 4.3.1 *ϵ_{it} is independent from any other factors and is identically distributed over time.*

Assumption 4.3.1 guarantees that the average decision function $\bar{D}_t(x_{it}, \mu_{it})$ is also time-invariant.[3] I therefore remove the time subscript from this function and define $\bar{D}(x, \mu) \equiv \bar{D}_t(x, \mu)$. Moreover, the independence part of this assumption also allows me to use $\bar{D}(x_{it}, \mu_{it})$ to conduct counterfactuals or characterize the causal effect of x_{it} and μ_{it} on the expectation of d_{it}.

I focus on the case in which the econometrician observes individuals' decisions d_{it} and decision-influencing factors x_{it} in two consecutive periods, $t = 0, 1$. Individuals' prior ($t = 0$) and posterior ($t = 1$) beliefs about A_i are not observed. But the econometrician knows that the prior mean μ_{i0} is fully determined by factors z_i that are also observed by the econometrician:

$$\mu_{i0} = B(z_i). \tag{4.3}$$

Between the two periods, each individual receives a noisy signal about A_i, which I denote as s_{i0}. Upon receiving the signal, individual i updates μ_{it} according to the following updating rule:

$$\mu_{i1} - \mu_{i0} \equiv \Gamma_{i0}(s_{i0} - \mu_{i0}) = \Gamma_0(s_{i0} - \mu_{i0}; z_i, \xi_i). \tag{4.4}$$

The function $\Gamma_0(s)$ maps the net signal $s_{i0} - \mu_{i0}$ to the difference between the posterior and prior means. I allow $\Gamma_0(s)$ to depend on beliefs-influencing factors z_i. This is motivated by the observation that, in a standard Bayesian updating setting, the difference between the posterior and prior means depends not only on the net signal, but also on, for example, how uncertain about A_i the individual was before receiving the signal.[4] I also allow $\Gamma_0(s)$ to depend on an unobserved independently distributed shock ξ_i. One interpretation of ξ_i is that it represents the error that individual i makes in the updating process. I assume that both the beliefs-influencing factors z_i and the signal s_{i0} are observed by the econometrician.

The primary objects of interest are the average decision function $\bar{D}(x_{it}, \mu_{it})$ and the prior mean function $B(z_i)$. In the next section, I discuss the additional assumptions that are sufficient for nonparametric identification of these two objects. I then provide constructive proofs for the nonparametric identification results.

[3] The requirement that ϵ_{it} is identically distributed over time is sufficient but not necessary for the time-invariance of $\bar{D}_t(x_{it}, \mu_{it})$. For example, if ϵ_{it} enters $D(x_{it}, \epsilon_{it}, \mu_{it})$ in an additively separable fashion, then the assumption that $E\epsilon_{i0} = E\epsilon_{i1}$ is sufficient. Such assumption is frequently used. For example, regression models commonly assume that ϵ_{it} is additively separable and has mean zero.

[4] The assumption that $\Gamma_0(s)$ depends on z_i but not other observables is not crucial for the identification results, but simplifies the presentation of proofs.

4.3.2 Nonparametric Identification Results

At period $t = 0$, the econometrician observes each individual's decision d_{i0}, decision-influencing factors x_{i0}, and beliefs-influencing factors z_i. At period $t = 1$, for each individual, the econometrician observes d_{i1}, x_{i1}, z_i, and a signal s_{i0}. I do not assume the econometrician can observe the same individuals in both periods. Hence, the identification of $\bar{D}(x_{it}, \mu_{it})$ and $B(z_i)$ builds on the relationship between d_{i0} and (x_{i0}, z_i), and the relationship between d_{i1} and (x_{i1}, z_i, s_{i0}).

Let $\tilde{D}_0(x_{i0}, z_i) \equiv E_{\epsilon_{i0}|x_{i0},z_i} D(x_{i0}, \epsilon_{i0}, B(z_i))$ denote the expectation of d_{i0} conditional on x_{i0} and z_i, and let $\tilde{D}_1(x_{i1}, z_i, s_{i0}) \equiv E_{(\epsilon_{i1},\xi_i)|x_{i1},z_i,s_{i0}} D(x_{i1}, \epsilon_{i1}, B(z_i) + \Gamma_0(s_{i0} - B(z_i); z_i, \xi_i))$ denote the expectation of d_{i1} conditional on x_{i1}, z_i, and s_{i0}. $\tilde{D}_0(x, z)$ is defined on the support of the joint distribution of X_{i0} and Z_i and $\tilde{D}_1(x, z, s)$ is defined on the support of the joint distribution of X_{i1}, Z_i, and S_{i0}. In this section, I first consider a hypothetical scenario where the true value of conditional mean functions $\tilde{D}_0(x, z)$ and $\tilde{D}_1(x, z, s)$ are known. I then discuss the complications that arise because $\tilde{D}_0(x, z)$ and $\tilde{D}_1(x, z, s)$ are not available and need to be estimated.

Identification with Known Conditional Mean Functions

I show that the average decision function $\bar{D}(x_{it}, \mu_{it})$ and the prior mean function $B(z_i)$ can be identified from conditional mean functions $\tilde{D}_0(x_{i0}, z_i)$ and $\tilde{D}_1(x_{i1}, z_i, s_{i0})$ if the following assumptions are satisfied.

Assumption 4.3.2 *For any (z, ξ), (i) $\Gamma_0(0; z, \xi) = 0$ and (ii) $\Gamma_0(s; z, \xi)$ is strictly monotonic in s.*

Assumption 4.3.3 *For any x, $\bar{D}(x, \mu)$ is strictly monotonic in μ.*

Assumption 4.3.4 *For any (x_{i1}, z_i), $B(z_i) \in supp(S_{i0}|(x_{i1}, z_i))$.*

Assumption 4.3.2(i) is most crucial for the identification results in this chapter. It requires that an individual who observes a signal that is equal to her prior mean will not update her subjective mean. This assumption is intuitively appealing and consistent with Bayesian learning about a permanent factor under fairly general conditions. Chambers and Healy (2012) show that if both the prior distribution and the distribution of errors in the signal are symmetric, then the Bayesian updating rule has this property.

My method requires the econometrician to observe the same group of individuals (i.e., individuals with the same value of observables x and z) in both periods. I stress the point that, given that decision-influencing factors x_{it} might be time-varying, the members of a certain group may change over time as well. For example, a student's study effort might directly depend on her expectations about own academic productivity (as will be discussed in Section

4.5) and her current health condition (healthy or unhealthy), which is time-varying. Hence, to parcel out the effect of time-varying health condition on study effort, my method is based on a comparison between decisions made by students who are healthy (unhealthy) in period 0 and decisions made by students who are healthy (unhealthy) in period 1 after they receive signals s_{i0}, despite that these might not be the same students.

Roughly speaking, Assumption 4.3.2(i) implies that, if this group of individuals receive a signal that is equal to their prior mean, then their average decision at $t = 1$ is the same as their average decision at $t = 0$. In other words, the value of the conditional mean function at $t = 1$, $\tilde{D}_1(x, z, s)$, is equal to the average decision made by this group of individuals at $t = 0$, $\tilde{D}_0(x, z)$, when the signal s is equal to the prior mean $B(z)$ for this group. The strict monotonicity of $\bar{D}(x, \mu)$ (Assumption 4.3.3) and $\Gamma_0(s; z, \xi)$ (Assumption 4.3.2(ii)) imply that $\tilde{D}_1(x, z, s)$ is also strictly monotonic, hence there is at most one signal s that can induce identical average decisions in two periods. This guarantees that the converse of the previous statement is also true. Finally, Assumption 4.3.4 guarantees the existence of such a signal. Formally, the following theorem holds.

Theorem 4.3.1 *For any $(x, z) \in supp((X_{i0}, Z_i)) \cap supp((X_{i1}, Z_i))$, if Assumption 4.3.1-4.3.4 are satisfied, then:*

1. *the average decision $\bar{D}(x, B(z)) = \tilde{D}_0(x, z)$;*

2. *$\tilde{D}_1(x, z, s)$ is invertible in s and the prior mean $B(z) = \tilde{D}_1^{-1}(x, z, \tilde{D}_0(x, z))$.*

Proof Fix $(x, z) \in \text{supp}((X_{i0}, Z_i)) \cap \text{supp}((X_{i1}, Z_i))$. By construction, $\tilde{D}_0(x, z)$ is a known constant and $\tilde{D}_1(x, z, s)$ is a known function defined on $\text{supp}(S_{i0}|(x, z))$. The value of $\tilde{D}_0(x, z)$ is given by:

$$\begin{aligned}
\tilde{D}_0(x, z) &= E_{\epsilon_{i0}|x,z} D(x, \epsilon_{i0}, B(z)) \\
&= E_{\epsilon_{i0}} D(x, \epsilon_{i0}, B(z)) \\
&= \bar{D}(x, B(z)),
\end{aligned} \tag{4.5}$$

where the second line follows from the independence of ϵ_{i0}. This proves the first part of the theorem.

Fix $s \in \text{supp}(S_{i0}|(x, z))$, similarly, the value of $\tilde{D}_1(x, z, s)$ is given by:

$$\begin{aligned}
\tilde{D}_1(x, z, s) &= E_{(\epsilon_{i1}, \xi_i)|x,z,s} D(x, \epsilon_{i1}, B(z) + \Gamma_0(s - B(z); z, \xi_i)) \\
&= E_{\xi_i} E_{\epsilon_{i1}} D(x, \epsilon_{i1}, B(z) + \Gamma_0(s - B(z); z, \xi_i)) \\
&= E_{\xi_i} \bar{D}(x, B(z) + \Gamma_0(s - B(z); z, \xi_i)),
\end{aligned} \tag{4.6}$$

where the second line follows from the independence of ϵ_{i1} and ξ_i.

Assumption 4.3.2 and 4.3.3 imply that, for any (x, z, ξ_i), $\bar{D}(x, B(z) + \Gamma_0(s - B(z); z, \xi_i))$ is strictly monotonic in s. I first consider the case where it is strictly increasing in s. Now fix $s_1 > s_2$. Then:

$$\tilde{D}_1(x, z, s_1) - \tilde{D}_1(x, z, s_2) = E_{\xi_i}(\bar{D}(x, B(z) + \Gamma_0(s_1 - B(z); z, \xi_i)) - \bar{D}(x, B(z) + \Gamma_0(s_2 - B(z); z, \xi_i)). \tag{4.7}$$

The right-hand-side of Equation (4.7) is the integral of a strictly positive function over a set that has probability measure 1. Hence, it is strictly positive which implies that $\tilde{D}_1(x, z, s_1) > \tilde{D}_1(x, z, s_2)$. Therefore, for any x, z, $\tilde{D}_1(x, z, s)$ is strictly increasing, hence invertible, in s. In the case where $\bar{D}(x, B(z) + \Gamma_0(s - B(z); z, \xi_i))$ is strictly decreasing in s, I can show that $\tilde{D}_1(x, z, s)$ is strictly decreasing, hence invertible, in s, following the same steps.

Given the invertibility of $\tilde{D}_1(x, z, s)$, to prove $B(z) = \tilde{D}_1^{-1}(x, z, \tilde{D}_0(x, z))$, I need to show (1) $\tilde{D}_1(x, z, B(z)) = \tilde{D}_0(x, z)$ and (2) $(x, z, B(z))$ is in the domain of $\tilde{D}_1(x, z, s)$. Assumption 4.3.2(i) implies that, when $s = B(z)$, $\Gamma_0(s - B(z); z, \xi_i)) = 0$ for all ξ_i. Hence, $\tilde{D}_1(x, z, B(z)) = \bar{D}(x, B(z)) = \tilde{D}_0(x, z)$. Assumption 4.3.4 guarantees that $(x, z, B(z))$ is in the domain of $\tilde{D}_1(x, z, s)$.

Remark The validity of Theorem 4.3.1 does not require any specific assumptions on the joint distribution of x_{it} and z_i. For example, they can each be binary, discrete, or continuous. They can be completely separate or completely overlapping sets of variables. However, in order to apply Theorem 4.3.1 to empirically conduct counterfactual analysis with respect to the prior mean μ_{i0}, some features of this joint distribution are desirable.

For a particular individual i with decision-influencing factors x_{i0} and prior mean $\mu_{i0} = B(z_i)$, this counterfactual analysis typically involves evaluating the identified average decision function $\bar{D}(x, \mu)$ at $x = x_{i0}$ and $\mu \neq B(z_i)$. Doing this requires that Z_i and X_{i0} are not perfectly dependent, i.e., Z_i is not a deterministic function of X_{i0}. Conceptually, this is the case because, analogous to the issue of perfect multicollinearity in a linear regression model, if Z_i and X_{i0} are perfectly dependent, then it is not possible to separate the effect of $\mu_{i0} = B(z_i)$ on d_{i0} from the effect of x_{i0} on d_{i0}.[5]

Intuitively, this implies that, in order to perform any meaningful counterfactual analysis, there must exist at least one observed (by the econometrician) factor that determines the prior mean μ_{i0}, but does not have a direct effect on the decision d_{i0}. Moreover, from an empirical standpoint, it would be helpful if this additional beliefs-influencing factor has a wide support.

[5] Note that Theorem 4.3.1 only allows for identification of the average decision $\bar{D}(x, B(z))$ if $(x, z) \in \text{supp}((X_{i0}, Z_i)) \cap \text{supp}((X_{i1}, Z_i))$. If Z_i and X_{i0} are perfectly dependent, then given $X_{i0} = x_{i0}$, $\text{supp}((X_{i0}, Z_i)) \cap \text{supp}((X_{i1}, Z_i))$ is either a singleton with (x_{i0}, z_i) as the only element or the empty set. Therefore, for any $x = x_{i0}$ and $\mu \neq B(z_i)$, $\bar{D}(x, \mu)$ cannot be identified using Theorem 4.3.1.

Then, by changing the value of this factor, we are able to identify the value of $\bar{D}(x,\mu)$ for a wide range of subjective mean $\mu_{i0} = B(z_i)$.

Identification with Consistently Estimated Conditional Mean Functions

Theorem 4.3.1 provides a strategy to identify the average decision function $\bar{D}(x,\mu)$ and the prior mean function $B(z)$ from *known* conditional mean functions $\tilde{D}_0(x,z)$ and $\tilde{D}_1(x,z,s)$. In practice, however, the econometrician needs to estimate these two conditional mean functions using data on d_{it}, x_{it}, z_i, and s_{i0}. I denote $\hat{D}_0(x,z)$ and $\hat{D}_1(x,z,s)$ as generic consistent estimators of the two functions and denote $\hat{D}_1^{-1}(x,z,d)$ as a generic consistent estimator of the inverse of $\tilde{D}_1(x,z,s)$.

A natural question is whether Theorem 4.3.1 holds if the conditional mean functions are replaced with their consistent estimators and equality is replaced with convergence in probability. The answer is "yes" to the first statement and "no" to the second statement in Theorem 4.3.1. Roughly speaking, this is because, under Assumption 4.3.1-4.3.4, $\tilde{D}_1^{-1}(x,z,d)$ is a generic strictly monotonic function of d, hence does not necessarily preserve limits. To validate the second statement, additional continuity assumptions are required.

Assumption 4.3.5 *For any (z,ξ), $\Gamma_0(s;z,\xi)$ is continuous in s and is bounded on any finite interval in $\mathbb{R}$.*

Assumption 4.3.6 *For any x, $\bar{D}(x,\mu)$ is continuous in μ and is bounded on any finite interval in $\mathbb{R}$.*

Assumption 4.3.7 *For any $(x,z) \in supp((X_{i1},Z_i))$, $supp(S_{i0}|(x,z))$ is a finite interval.*[6]

Assumption 4.3.5-4.3.7 help guarantee that $\tilde{D}_1^{-1}(x,z,d)$ is continuous in d.

Lemma 4.3.2 *For any $(x,z) \in supp((X_{i0},Z_i)) \cap supp((X_{i1},Z_i))$, if Assumption 4.3.1-4.3.7 are satisfied, then $\tilde{D}_1(x,z,s)$ is continuous and invertible in s and $\tilde{D}_1^{-1}(x,z,d)$ is continuous in d.*

Proof See Appendix D.1.

Under the continuity assumptions listed above, the average decision function $\bar{D}(x,\mu)$ and the prior mean function $B(z)$ can be consistently estimated using consistent estimators of $\tilde{D}_0(x,z)$ and $\tilde{D}_1(x,z,s)$ ($\tilde{D}_1^{-1}(x,z,d)$).

[6] This assumption is primarily required to ensure the continuity of the conditional mean function at $t = 1$, $\hat{D}_1(x,z,d)$, in the presence of unobserved heterogeneity in updating rules. Technically it assumes aways many canonical signal distributions considered in the literature, such as normal and log-normal distributions. However, in many applications, the support of signals are bounded empirically. For example, a person's semester grade is typically bounded by 0 and 4; realized annual income is also bounded above since it would be impossible for someone to earn more than the world can physically produce in a year.

Assumption 4.3.8 *For any* $(x, z) \in supp((X_{i0}, Z_i)) \cap supp((X_{i1}, Z_i))$*,*

1. $\hat{D}_0(x, z) \underset{p}{\rightarrow} \tilde{D}_0(x, z)$*;*

2. $\sup_{d \in \tilde{D}_1(x,z,supp(S_{i0}|(x,z)))} |\hat{D}_1^{-1}(x, z, d) - \tilde{D}_1^{-1}(x, z, d)| \underset{p}{\rightarrow} 0$*, i.e.,* $\hat{D}_1^{-1}(x, z, d)$ *uniformly converges in probability to* $\tilde{D}_1^{-1}(x, z, d)$*.*

Theorem 4.3.3 *For any* $(x, z) \in supp((X_{i0}, Z_i)) \cap supp((X_{i1}, Z_i))$*, if Assumption 4.3.1-4.3.8 are satisfied, then:*

1. $\hat{D}_0(x, z) \underset{p}{\rightarrow} \bar{D}(x, B(z))$*;*

2. $\hat{D}_1^{-1}(x, z, \hat{D}_0(x, z)) \underset{p}{\rightarrow} B(z)$*.*

Proof The first statement in Theorem 4.3.3 is trivially implied by the first statement in Theorem 4.3.1 and the first part of Assumption 4.3.8.

To prove the second statement in Theorem 4.3.3, first rewrite $\hat{D}_1^{-1}(x, z, \hat{D}_0(x, z))$ as the sum of several terms:

$$\begin{aligned} \hat{D}_1^{-1}(x, z, \hat{D}_0(x, z)) = \tilde{D}_1^{-1}(x, z, \tilde{D}_0(x, z)) + (\tilde{D}_1^{-1}(x, z, \hat{D}_0(x, z)) - \tilde{D}_1^{-1}(x, z, \tilde{D}_0(x, z))) \\ + (\hat{D}_1^{-1}(x, z, \hat{D}_0(x, z)) - \tilde{D}_1^{-1}(x, z, \hat{D}_0(x, z))). \end{aligned} \tag{4.8}$$

By Lemma 4.3.2, $\tilde{D}_1^{-1}(x, z, d)$ is continuous in d. Since $\hat{D}_0(x, z) \underset{\mathrm{p}}{\rightarrow} \tilde{D}_0(x, z)$, the continuous mapping theorem states that $\tilde{D}_1^{-1}(x, z, \hat{D}_0(x, z)) \underset{\mathrm{p}}{\rightarrow} \tilde{D}_1^{-1}(x, z, \tilde{D}_0(x, z))$. The second part of Assumption 4.3.8 implies that $\hat{D}_1^{-1}(x, z, \hat{D}_0(x, z)) - \tilde{D}_1^{-1}(x, z, \hat{D}_0(x, z)) \underset{\mathrm{p}}{\rightarrow} 0$. Therefore, the right-hand-side of Equation (4.8) converges in probability to $\tilde{D}_1^{-1}(x, z, \tilde{D}_0(x, z))$, which is equal to $B(z)$ by Theorem 4.3.1.

4.3.3 Discussion

This environment has two defining features. First, an individual repeatedly makes decisions based on her subjective mean of a permanent unknown factor. Second, the average decision as a function of the subjective mean is time-invariant, strictly monotonic, and continuous. I choose this environment as the benchmark primarily for three reasons.

First, this environment nests many models that are of interest to researchers, including the canonical model of input choices under price/demand/productivity uncertainty (e.g., Baron, 1970; Sandmo, 1971; Leland, 1972; Holthausen, 1976). In Appendix D.2.1, I provide a concrete example of this model. Note that the continuity assumption in Theorem 4.3.3 is imposed on the average decision function, not the decision rule. Therefore, even in models where the decision d_{it} is discrete, I may still be able to apply Theorem 4.3.3. In Appendix D.2.2, I show this using a model of an individual's rural-urban migration decisions as an example.

Second, this benchmark environment also provides a good approximation to many commonly used models that might not fit the description of the benchmark environment exactly. For example, in most finite horizon dynamic models, an individual's decision rules are typically not time-invariant even when preferences and constraints in each period remain the same. However, under fairly general conditions, the contraction mapping theorem implies that if the time-horizon is sufficiently long, then individuals' decision rules in the first two periods can be arbitrarily close. In the extreme case, the model becomes an infinite horizon problem with stationary per-period environment and has a time-invariant decision rule. Similarly, if the length of a period is sufficiently short (e.g., weekly quizzes followed by dropout decisions), an individual's decision rules in two consecutive periods can be approximately identical as well. If an individual receives a signal immediately after making a decision and is asked to make decisions again, it would be natural to expect that her new decision rule should be very similar to her decision rule before receiving the signal.[7]

Third, the identification strategy highlighted in this parsimonious environment can be naturally extended to achieve identification in more complicated, nonstationary environments that cannot be well approximated by the benchmark environment. In Section 4.6.1, I show this using a simple dynamic discrete choice model that captures the key elements of the dropout model in Stange (2012) as an example.

4.4 Feasible Semiparametric Estimator

In Section 4.3, I showed that the average decision function $\bar{D}(x,\mu)$ and the prior mean function $B(z)$ can be jointly nonparametrically estimated if consistent nonparametric estimators of conditional mean functions $\tilde{D}_0(x,z)$ and $\tilde{D}_1(x,z,s)$ are available. In practice, their nonparametric estimation suffers from the standard curse of dimensionality. This issue might be particularly severe in my context because, given the conditional homogeneity assumption on individuals' prior means, it is desirable to include a large number of observables in the vector of beliefs-influencing factors, z_i.

[7] Another type of deviation from the benchmark environment is to allow an individual's decision to depend on not only her subjective mean of A_i, but also higher moments of her subjective distribution. Since the assumptions on the updating rules do not impose any restrictions on how individuals update higher moments of their subjective distributions, these higher moments are generally unidentified. However, if the magnitude of learning is small such that individuals' subjective distributions of A_i do not change much over the two periods, an individual's decision rule as a function of the subjective mean may be approximately time-invariant as well.

Note that, as long as no parametric and/or homogeneity assumptions are imposed on the updating rules, $\tilde{D}_1(x, z, s)$ is a nonparametric function of x, z, and s, even if both $\bar{D}(x, \mu)$ and $B(z)$ are fully parametric. In other words, parametric assumptions on the average decision function and the prior mean function alone do not help alleviate the curse of dimensionality in the estimation of $\tilde{D}_1(x, z, s)$. Additional assumptions are required.

In this section, I show that, if individual i's misperception, as measured by the difference between her prior mean $\mu_{i0} = B(z_i)$ and the subjective mean that is implied by the Rational Expectations assumption, $E(A_i|z_i)$, is determined by a finite type variable k_i, then parametric functions $\bar{D}(x, \mu)$ and $B(z)$ can be consistently estimated using a curse-of-dimensionality-free approach even if no parametric and homogeneity assumptions are imposed on the updating rules. I then examine the performance of this feasible semiparametric estimator through a simulation exercise. In particular, I compare its performance to a fully parametric estimator, where, as is often the case in the empirical literature on Bayesian learning, linearity and homogeneity assumptions are imposed on the updating rule.

4.4.1 Method

I keep the general environment in Section 4.3 but impose parametric assumptions on the decision rule $D(x_{it}, \epsilon_{it}, \mu_{it})$, the prior mean function $B(z_i)$, and the true conditional expectation $A(z_i) = E(A_i|z_i)$. I still allow the updating rule to be nonparametric and heterogeneous across individuals. For the purpose of illustration, I consider below the case where both $D(x_{it}, \epsilon_{it}, \mu_{it})$ and $A(z_i)$ are linear. As shown in Appendix D.3, the method can be applied to cases where $D(x_{it}, \epsilon_{it}, \mu_{it})$ and $A(z_i)$ are arbitrary parametric functions.[8]

Most importantly, I also make the additional assumption that $B(z_i) - A(z_i) = \tilde{\pi}_B(k_i)$, where k_i only takes finitely many values and is known to both the individual and the econometrician.[9] $\tilde{\pi}_B(k_i)$ measures individual i's misperception about A_i at period $t = 0$. Within each group k_i, individuals may have different prior means $B(z_i)$, but their misperceptions $\tilde{\pi}_B(k_i)$ are identical. This restricts the amount of heterogeneity in prior means that need to be estimated nonparametrically.

Individual i's decision d_{it}, prior mean μ_{i0}, and signal s_{i0} are given by:

$$d_{it} = D(x_{it}, \epsilon_{it}, \mu_{it}) = x'_{i0}\alpha + \beta\mu_{it} + \epsilon_{it}, \tag{4.9}$$

$$\mu_{i0} = A(z_i) + \tilde{\pi}_B(k_i) = z'_i\pi_R + \tilde{\pi}_B(k_i), \tag{4.10}$$

$$s_{i0} = A_i + v_{i0} = A(z_i) + (A_i - A(z_i)) + v_{i0} \equiv z'_i\pi_R + \eta_i + v_{i0}, \tag{4.11}$$

[8] $D(x_{it}, \epsilon_{it}, \mu_{it})$ needs to satisfy Assumption 4.3.3 and 4.3.6.

[9] Since k_i is also a beliefs-influencing factor, by construction, it is perfectly determined by z_i.

where ϵ_{it} is a mean zero independent shock in decision d_{it}, v_{i0} is the independent noise term in signal s_{i0} and $\eta_i \equiv A - A(z_i)$ summarizes the beliefs-influencing factors that are not in individual i's information set at $t = 0$. η_i is orthogonal to $A(z_i)$ by construction. Here I impose a stronger assumption that it is independent of z_i.

Equation (4.11) shows that a consistent estimator of $A(z_i) = z_i'\pi_R$ can be obtained through a linear regression of s_{i0} on z_i. I denote this consistent estimator as $\hat{A}(z_i)$.

Combining Equation (4.9) and (4.10), I obtain the following equations for individual i's decisions at period $t = 0$ and $t = 1$:

$$d_{i0} = x_{i0}'\alpha + z_i'(\pi_R\beta) + \beta\tilde{\pi}_B(k_i) + \epsilon_{i0} \equiv \tilde{D}_0(x_{i0}, (z_i, k_i)) + \epsilon_{i0}, \tag{4.12}$$

$$d_{i1} = \tilde{D}_0(x_{i1}, (z_i, k_i)) + \beta\Gamma_{i0}(s_{i0} - z_i'\pi_R - \tilde{\pi}_B(k_i)) + \epsilon_{i1}. \tag{4.13}$$

Since k_i only takes finitely many values, $\tilde{\pi}_B(k_i)$ can be written as a linear function of a series of dummy variables. Hence, $\tilde{D}_0(x_{i0}, (z_i, k_i))$ can be consistently estimated through a linear regression of d_{i0} on x_{i0}, z_i, and these dummy variables. I denote this consistent estimator as $\hat{D}_0(x_{i0}, (z_i, k_i))$.

Let $\tilde{d}_{i1} \equiv d_{i1} - \hat{D}_0(x_{i1}, (z_i, k_i))$ and $\tilde{s}_{i0} \equiv s_{i0} - \hat{A}(z_i)$. Then,

$$\tilde{d}_{i1} = \beta\Gamma_{i0}(\tilde{s}_{i0} - \tilde{\pi}_B(k_i)) + \epsilon_{i1} + \delta_i, \tag{4.14}$$

where δ_i represents the estimation error and converges in probability to zero as the number of observations goes to infinity.[10]

Since k_i is a finite type variable, it is feasible to stratify individuals by k_i. For each type k_i, $\tilde{\pi}_B(k_i)$ is a constant and $\tilde{d}_{i1}$ only depends on one "observable" $\tilde{s}_{i0}$ (and the unobserved error term $\epsilon_{i1} + \delta_i$).

In the case where no parametric and/or homogeneity assumptions are imposed on the updating rules $\Gamma_{i0}(s)$, the probability limit of $E(\tilde{d}_{i1}|\tilde{s}_{i0}, k_i)$ is a continuous and strictly monotonic function of $\tilde{s}_{i0}$ and equals zero when $\tilde{s}_{i0} = \tilde{\pi}_B(k_i)$ given Assumption 4.3.2 and 4.3.5. Hence, for each k_i, $\tilde{\pi}_B(k_i)$ can be consistently estimated following a two-step procedure:

1. Nonparametrically estimate $E(\tilde{d}_{i1}|\tilde{s}_{i0}, k_i)$ as a continuous and strictly monotonic function of $\tilde{s}_{i0}$ using the isotonic regression (Barlow et al., 1972);[11]

[10]Formally, δ_i is given by $\delta_i \equiv \tilde{D}_0(x_{i0}, (z_i, k_i)) - \hat{D}_0(x_{i0}, (z_i, k_i)) + \beta[\Gamma_{i0}(s_{i0} - A(z_i) - \tilde{\pi}_B(k_i)) - \Gamma_{i0}(\tilde{s}_{i0} - \tilde{\pi}_B(k_i))]$.

[11]The out-of-the-box isotonic regression typically yields a weakly monotonic function. Hence, in theory, the root of the estimator of $E(\tilde{d}_{i1}|\tilde{s}_{i0}, k_i)$ may not be unique. In practice, I introduce a tie-breaking rule to ensure the uniqueness of the root.

There are also many other existing methods for monotonic nonparametric estimation of conditional mean functions. Examples include spline regression with shape constraints (e.g., Ramsay, 1988) and kernel regression with shape constraints (e.g., Hall and Huang, 2001).

2. Solve for the unique root of this function which consistently estimates $\tilde{\pi}_B(k_i)$.

Hereafter I refer to this estimator as the feasible semiparametric estimator.

Alternatively, if the updating rule $\Gamma_{i0}(s)$ is linear in s and homogeneous across i, i.e., $\Gamma_{i0}(s) = \theta s$, Equation (4.14) becomes:

$$\tilde{d}_{i1} = -\beta\theta\tilde{\pi}_B(k_i) + \beta\theta\tilde{s}_{i0} + \epsilon_{i1} + \delta_i. \tag{4.15}$$

For each type variable k_i, consistent estimators of $-\beta\theta\tilde{\pi}_B(k_i)$ and $\beta\theta$ can be obtained by a linear regression of $\tilde{d}_{i1}$ on $\tilde{s}_{i0}$. The negative of the ratio of the former to the latter consistently estimates $\tilde{\pi}_B(k_i)$. Hereafter I refer to this estimator as the linear homogeneous updating rule (LHU) estimator.

After obtaining consistent estimators of $A(z_i)$ and $\tilde{\pi}_B(k_i)$ for each k_i, I can consistently estimate the the prior mean $\mu_{i0} = A(z_i) + \tilde{\pi}_B(k_i)$ for each individual, which I denote $\hat{\mu}_{i0}$. If $z_i|x_{i0}$ is not degenerate, then the structural parameters in the decision rule, α and β, can be consistently estimated through a linear regression of d_{i0} on decision-influencing factors x_{i0} and the generated prior mean $\hat{\mu}_{i0}$.

4.4.2 Simulation Exercise

I use a numerical example to examine the performance of this feasible semiparametric estimator and compare it to the LHU estimator. Consider the linear model in Section 4.4.1 where x_{i0} contains a constant, z_i contains a constant and a dummy variable q_i that takes value 1 with 50% probability, and k_i is the same as q_i:

$$\begin{aligned} d_{it} &= \alpha + \beta\mu_{it} + \epsilon_{it}, \\ s_{i0} &= \pi_{R,0} + \pi_{R,1}q_i + \eta_i + v_{i0}, \\ \mu_{i0} &= \pi_{R,0} + \pi_{R,1}q_i + \tilde{\pi}_B(q_i). \end{aligned} \tag{4.16}$$

In the numerical exercise, I set $\alpha = 0, \beta = 1, \pi_{R,0} = 2, \pi_{R,1} = 2, \tilde{\pi}_B(0) = -1$, and $\tilde{\pi}_B(1) = 1$. Type $q_i = 0$ individuals' prior mean of A_i is 1 while their Rational-Expectations-consistent subjective mean is 2. Type $q_i = 1$ individuals' prior mean of A_i is 5 while their Rational-Expectations-consistent subjective mean is 4.

To close the model, I also need to specify the updating rule $\Gamma_{i0}(s)$ for each individual. To facilitate comparisons between the LHU estimator and the semiparametric estimator, it would be desirable if $\Gamma_{i0}(s)$ has a form that is generally nonlinear in s and heterogeneous across individuals, but also has the linear homogeneous updating rule as a special case. The following class of updating rules has this feature:

$$\Gamma_{i0}(s) = sgn(s)\theta_i|s|^{\rho}, \tag{4.17}$$

where $\theta_i = \frac{e^{\xi_i}}{1+e^{\xi_i}}$ and $\xi_i \sim N(0, \sigma_\xi^2)$.

When $\sigma_\xi^2 > 0$, $\Gamma_{i0}(s)$ is heterogeneous across individuals. When $\rho \neq 1$, $\Gamma_{i0}(s)$ is nonlinear in s. In the special case where $\rho = 1$ and $\sigma_\xi^2 = 0$, the updating rule is linear in s and homogeneous across individuals: $\Gamma_{i0}(s) = 0.5s$.

Table 4.1: Comparison of Two Estimators

		$\alpha = 0$	$\beta = 1$	$\mu_{i0} = 1$ ($Q_i = 0$)	$\mu_{i0} = 5$ ($Q_i = 1$)
$N = 200$	Semiparametric	-0.170	1.064	1.083	4.873
		(0.355)	(0.101)	(0.261)	(0.249)
	Linear	0.017	0.994	0.977	5.024
		(0.277)	(0.077)	(0.250)	(0.246)
$N = 2,000$	Semiparametric	-0.003	1.003	0.993	4.994
		(0.199)	(0.055)	(0.162)	(0.159)
	Linear	0.002	1.000	0.997	5.001
		(0.084)	(0.023)	(0.072)	(0.072)
$N = 20,000$	Semiparametric	0.013	1.002	0.984	4.978
		(0.091)	(0.026)	(0.073)	(0.076)
	Linear	-0.000	1.000	1.000	4.998
		(0.027)	(0.007)	(0.023)	(0.022)

Notes: Numbers without parentheses are the averages of the estimates for 1,000 random samples. Numbers in parentheses are the standard deviations of the estimates for 1,000 random samples.

I first consider the case where $\Gamma_{i0}(s) = 0.5s$. In this case, both the feasible semiparametric estimator and the LHU estimator are consistent. However, they might have different small-sample properties and different convergence rates. To make comparisons, I generate 1,000 random samples using the above data generating process and estimate prior means (μ_{i0}) and model parameters (α and β) using both the feasible semiparametric estimator and the LHU estimator.

Table 4.1 reports the mean and standard deviation of both estimators. The first panel ($N = 200$) shows that, overall the LHU estimator exhibits smaller small-sample bias than the semiparametric estimator. For example, the average semiparametric estimate of the slope parameter β is 0.064 higher than the true value $\beta = 1$ while the average LHU estimate is only 0.006 lower than the true value. Similar comparisons are found for the intercept parameter α and prior means μ_{i0} as well. Additionally, the LHU estimator also converges faster than the semiparametric estimator. Again consider the slope parameter β as an example. Comparing the second column of the first and the last panels, I find that increasing the sample size from 200 to 20,000 leads to a roughly 94% reduction in the standard deviation of the LHU estimator

(0.077 to 0.007) and a roughly 74% reduction in the standard deviation of the semiparametric estimator (0.101 to 0.026). These results suggest that when the updating rule is correctly specified, the LHU estimator has better properties than the semiparametric estimator.

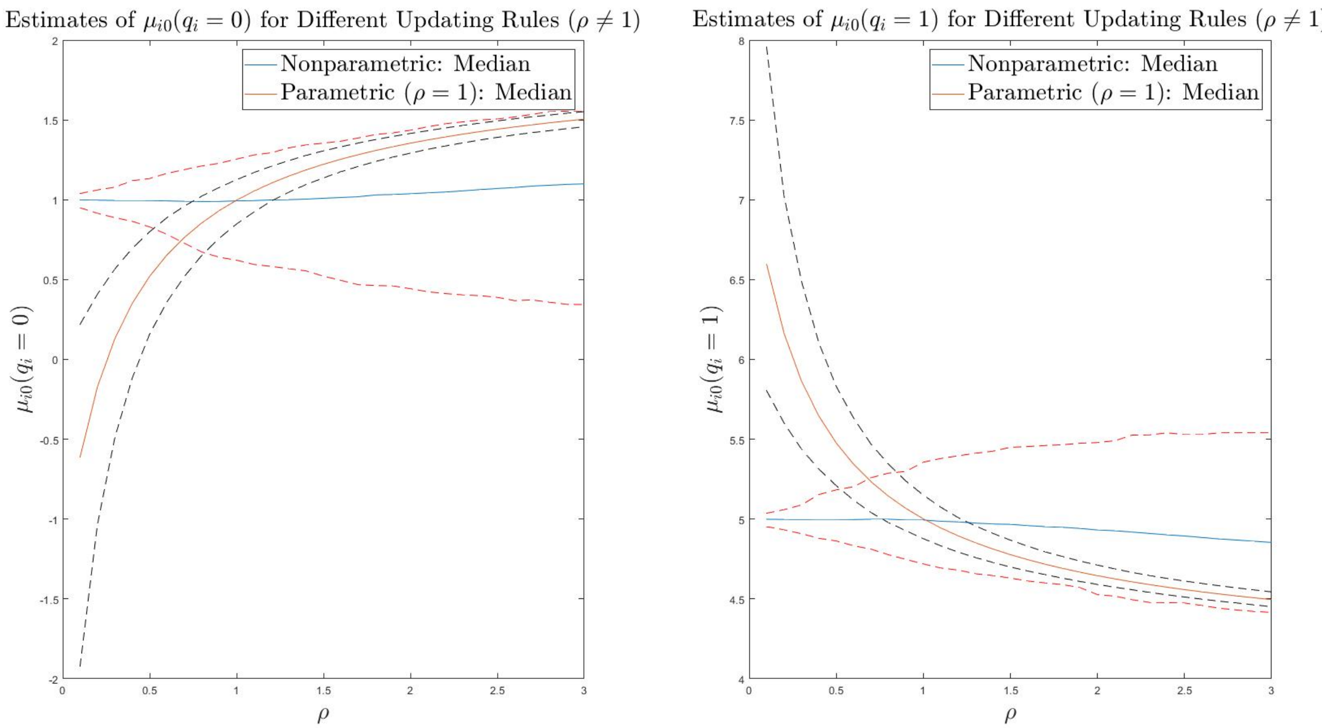


Figure 4.1: Comparison of Two Estimators for $\rho \neq 1$: Priors

Next I examine how nonlinearity of the updating rule influences the performance of the two estimators. I keep the assumption that θ_i is homogeneous across individuals ($\theta_i = 0.5$) but allow ρ to be an arbitrary positive number. For each ρ, I generate 1,000 random samples and estimate μ_{i0}, α, and β using both estimators. Each random sample consists of 2,000 observations.

Figure 4.1 and 4.2 depict the median and 95% interval of these estimators for $\rho \in (0, 3)$. An immediate observation is that, unlike the feasible semiparametric estimator, the LHU estimators are not consistent when $\rho \neq 1$. For any $\rho \in (0, 3)$, the 95% interval of the feasible semiparametric estimators of μ_{i0}, α, and β always contain their true values. On the contrary, only when ρ is close to zero does the 95% interval of the LHU estimator contain the true value of the object of interest. For example, the right panel of Figure 4.2 shows that the true value of the slope parameter $\beta = 1$ is in the 95% interval of its LHU estimator when $\rho \in [0.9, 1.1]$, and falls outside of the 95% interval of its LHU estimator for any other value of ρ.

Interestingly, the feasible semiparametric estimator is not always more volatile than the LHU estimator: When ρ is very small, the 95% intervals of feasible semiparametric estimators

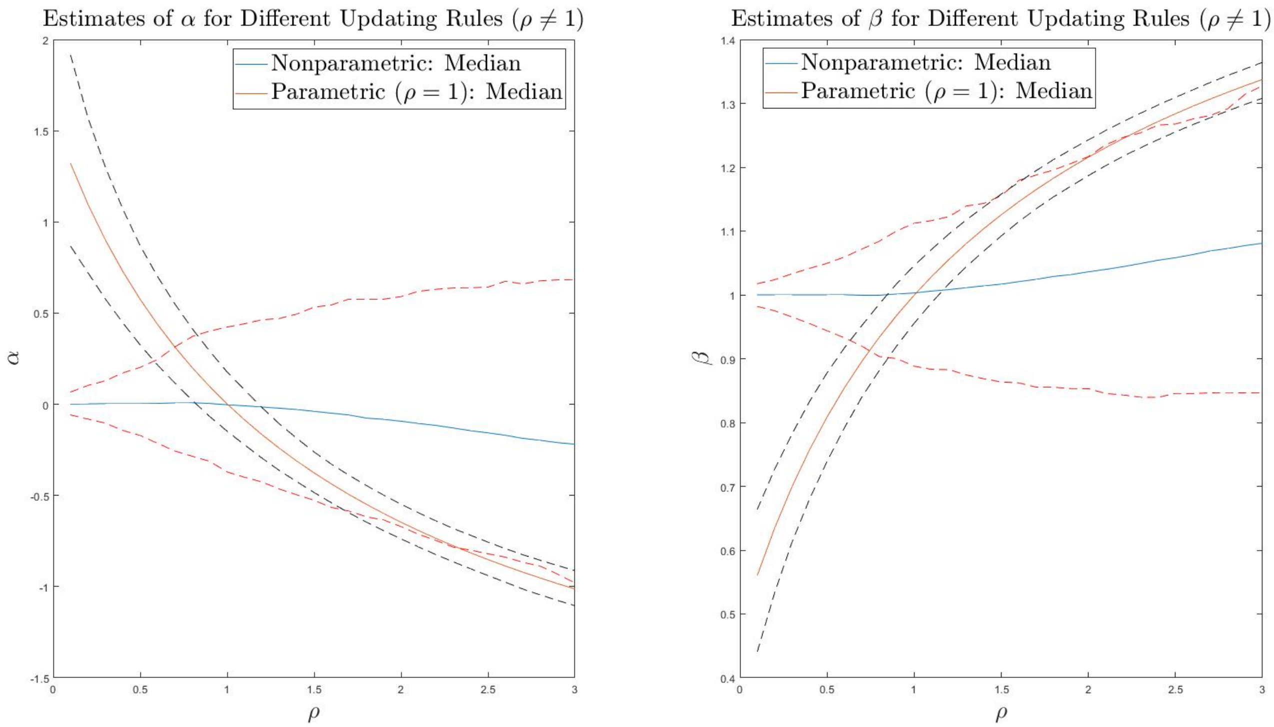


Figure 4.2: Comparison of Two Estimators for $\rho \neq 1$: Parameters

are narrower than the 95% intervals of LHU estimators. Intuitively, this is because the accuracy of the feasible semiparametric estimator is determined by how responsive individual i's decision d_{i1} is to the signal s_{i0} when the signal is close to her prior mean μ_{i0}. If $\rho < 1$, as s_{i0} converges to μ_{i0}, the derivative of posterior mean μ_{i1} with respect to signal s_{i0} diverges to infinity. In this case, a small change in s_{i0} around $s_{i0} = \mu_{i0}$ can lead to a substantial change in the posterior mean μ_{i1} and the expected value of d_{i1}. Therefore, the impact of sampling variations in the estimated average decision functions on the estimates of μ_{i0} should be small.

Lastly, I examine the performance of the two estimators when the updating rules are heterogeneous across individuals. I assume that the updating rule is linear for each individual ($\rho = 1$) , but the parameters θ_i differ across people, i.e., $\sigma_\xi^2 > 0$. Similar as before, for each σ_ξ^2, I generate 1,000 random samples, each of which contains 2,000 observations, and estimate μ_{i0}, α, and β using both estimators.

Figure 4.3 and 4.4 plot the median and 95% interval of these estimators against the sample interquartile range of θ_i. The main finding is similar as above: When there is heterogeneity in updating rules, the LHU estimator "precisely" produces inconsistent estimates. Consequently, it rarely produces estimates that are close to the true value, especially when the heterogeneity in updating rules is substantial. As shown in the right panel of Figure 4.4, the true value of the

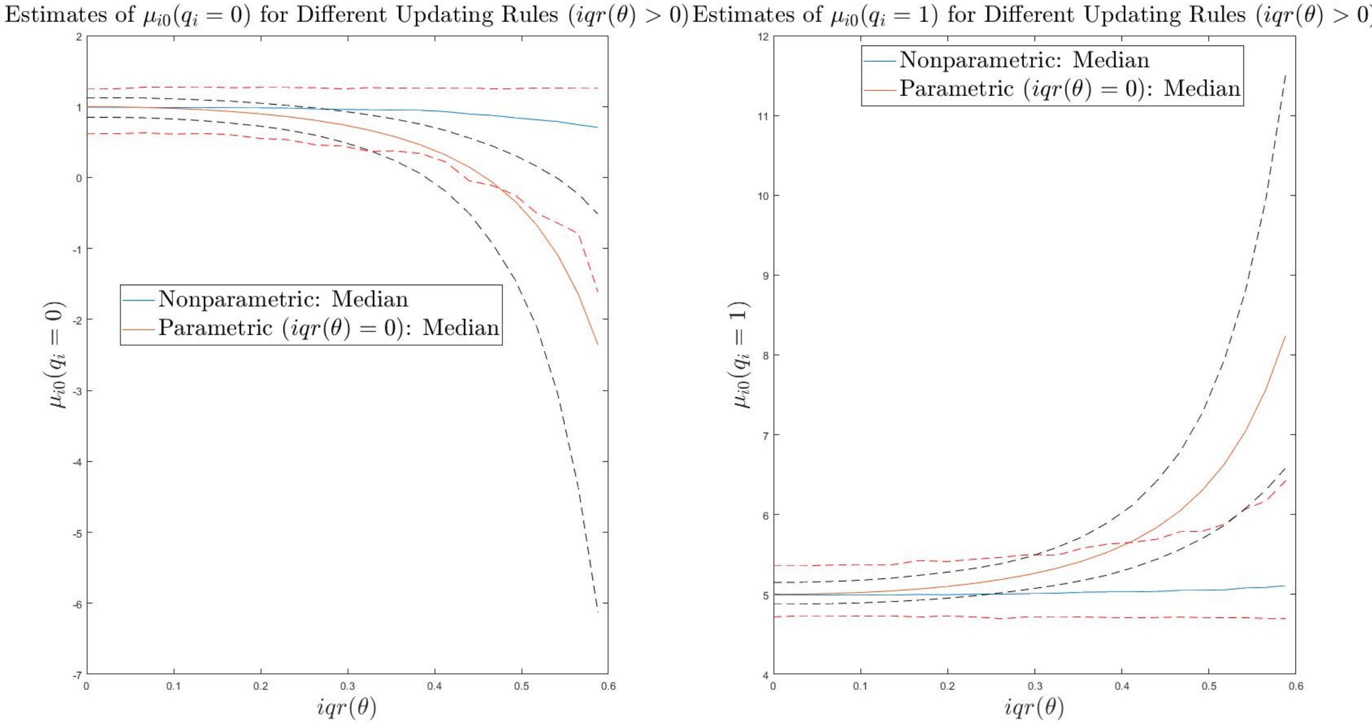


Figure 4.3: Comparison of Two Estimators for $iqr(\theta) > 0$: Priors

slope parameter $\beta = 1$ is in the 95% interval of its LHU estimator when the interqualtile range of θ_i is smaller than 0.2, and falls outside of the 95% interval of its LHU estimator otherwise. As a comparison, the 95% interval of the feasible semiparametric estimator contains the true value $\beta = 1$ regardless of how much heterogeneity there is in updating rules. Moreover, when the magnitude of heterogeneity is large, the 95% intervals of the LHU estimator and the feasible semiparametric estimator do not overlap, suggesting that despite being volatile, the feasible semiparametric estimator is much less likely to produce estimates that are too far away from the true value. Similar results can be found for α and μ_{i0} as well.

4.5 Empirical Application

Students' study effort might be affected by their expectations about how good they are at "producing" grades. Analogous to the productivity of a firm, I measure the academic productivity of a student in a given semester by the ratio of her semester GPA to average daily study time. In this section, I use data from the Berea Panel Study to empirically estimate college students' expectations about their academic productivity and estimate the effect of these expectations on students' study time. I first estimate these objects under the Rational Expectations (RE)

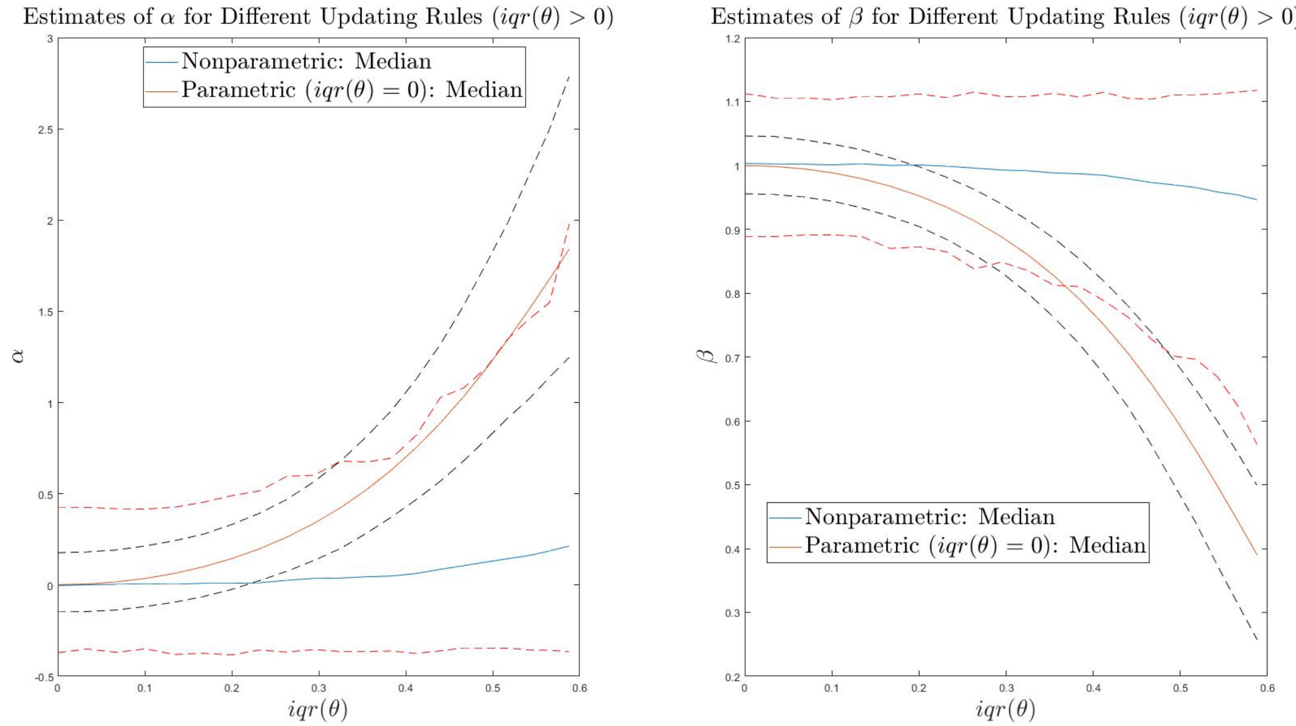


Figure 4.4: Comparison of Two Estimators for $iqr(\theta) > 0$: Parameters

assumption, then use both the LHU estimator and the feasible semiparametric estimator to estimate these objects without the RE assumption. A particular focus of this section is to examine whether these students have rational expectations about own academic productivity and whether relaxing the RE assumption leads to different estimates of the effect of expectations about academic productivity on study time.

4.5.1 Empirical Model

Consider the first two semesters in college ($t = 0, 1$). In each semester t, student i chooses average daily study time d_{it} based on factors x_{it} and the subjective mean of her academic productivity, s_{it}. The decision rule is given by:

$$d_{it} = x_{it}'\alpha + \beta E^b(s_{it}) + \epsilon_{it}, \tag{4.18}$$

where student i's academic productivity in semester t is defined as the ratio of her GPA g_{it}, to her study time d_{it}, i.e., $s_{it} \equiv \frac{g_{it}}{d_{it}}$.

Academic productivity s_{it} is the sum of a permanent factor A_i and a mean zero, independently distributed shock v_{it}, i.e., $s_{it} = A_i + v_{it}$.[12] Let μ_{it} denote student i's subjective mean of A_i in semester t. Then $E^b(s_{it}) = E_t^b(A_i) = \mu_{it}$. Equation (4.18) can be rewritten as:

$$d_{it} = x_{it}'\alpha + \beta\mu_{it} + \epsilon_{it}. \tag{4.19}$$

The objects of primary interest are students' prior means μ_{i0}, and the parameter β, which represents the effect of students' subjective mean μ_{it} on study time d_{it}. Conceptually, it is not clear whether β is positive or negative. For example, if students are solving a utility maximization problem that is similar to the canonical input choice problem faced by profit-maximizing firms, expectations about academic productivity would have a positive effect on study time because of the complementarity between academic productivity and study time in producing grades. Alternatively, if students' primary goal is to achieve certain grades, then students with high expectations about own academic productivity would believe that they do not need to spend much time studying. In this case, the effect of expectations about academic productivity on study time would be negative.

Under the RE assumption, student i's prior mean μ_{i0} is given by $\mu_{i0} = E(A_i|z_i) = E(s_{i0}|z_i)$, where z_i are the beliefs-influencing factors observed by individual i before the first semester. Provided that I observe both $s_{i0} \equiv \frac{g_{i0}}{d_{i0}}$ and z_i, I can consistently estimate $E(s_{i0}|z_i)$. In this application, I assume that $E(s_{i0}|z_i) \equiv A(z_i) = z_i'\pi_R$ and estimate π_R through linearly regressing s_{i0} on z_i. This provides consistent estimates of $\mu_{i0} = E(s_{i0}|z_i) = z_i'\pi_R$ for each student.

Without the RE assumption, however, student i's prior mean μ_{i0} is generally different from $z_i'\pi_R$. Following Section 4.4.1, I assume that student i's misperception about A_i in the first semester, $\mu_{i0} - z_i'\pi_R$, is determined by a finite type variable k_i. At the end of the first semester, student i observes her semester GPA g_{i0} and uses $s_{i0} \equiv \frac{g_{i0}}{d_{i0}}$ to update her subjective mean μ_{it}. Under this setting, I can use the LHU estimator and the feasible semiparamatric estimator developed in Section 4.4.1 to consistently estimate μ_{i0} for each student.

Finally, if beliefs-influencing factors Z_i and decision influencing factors X_{i0} are linearly independent, then X_{i0} and the estimated value of μ_{i0} are also linearly independent. Hence, consistent estimators of β can be obtained by linearly regressing d_{i0} on x_{i0} and the estimated value of μ_{i0}.

[12]To facilitate identification, I assume away the possibility that A_i can change between period $t = 0$ and $t = 1$. While I consider this to be a reasonable approximation because of the short time span, it is certainly possible that A_i is time-varying and may depend on past investments and achievements within a longer time frame (Cunha, Heckman, Lochner, et al., 2006; Cunha and Heckman, 2007, 2008; Cunha, Heckman, and Schennach, 2010).

4.5.2 Data

The empirical analysis is based on the Berea Panel Study. Of particular importance here, the BPS contains multiple 24-hour time diaries each semester, which allows me to construct a reasonably accurate measure of average daily study time d_{it} for $t = 0, 1$. The BPS is linked with administrative data so I observe semester GPA, g_{i0}, for each student.

The BPS also contains data on various decision-influencing factors x_{it} and beliefs-influencing factors z_i. In this empirical investigation, I let x_{it} include a constant, gender, and race. Beliefs-influencing factors z_i need to contain at least one variable that is not in x_{it}. I assume that in addition to all the variables in x_{it}, student i's prior mean is also determined by her high school study time. Conceptually, since a student's study time depends on her subjective mean of academic productivity, the student's high school study time should be informative about her subjective mean of academic productivity shortly before she enters college.

Lastly, I specify the type variable k_i that determines student i's misperception about A_i. Motivated by the link between a student's high school study time and prior mean, I assume that $k_i = 0$ ($k_i = 1$) if student i's study time in high school is (no) less than 2 hours per day.

Table 4.2: Descriptive Statistics

	Male	Black	HSTU	d_{i0}	d_{i1}	g_{i0}	s_{i0}
Full Sample: N = 475	0.398	0.181	1.607	2.969	3.014	2.964	1.272
	(0.490)	(0.386)	(1.234)	(1.350)	(1.440)	(0.733)	(0.983)
k_i = 1: N = 151	0.411	0.245	3.120	3.612	3.518	2.983	0.962
	(0.494)	(0.432)	(0.851)	(1.397)	(1.502)	(0.778)	(0.536)
k_i = 0: N = 324	0.392	0.151	0.902	2.669	2.779	2.955	1.416
	(0.489)	(0.359)	(0.575)	(1.218)	(1.350)	(0.712)	(1.104)

Notes: Standard deviations are in the parentheses.

I restrict the sample to the 475 students for whom I have complete information about d_{it}, g_{i0}, x_{it}, z_i, and k_i. Around two-thirds of these 475 students spend less than 2 hours per day studying in high school ($k_i = 0$). Table 4.2 summarizes the descriptive statistics for this sample. The first row shows that, approximately 40% of the students are male, 18% of the students are black, and the average high school daily study time is approximately 1.6 hours. On average, students in the sample spend around 3 hours per day studying in both the first and the second semester. The average semester GPA and academic productivity for this sample are 2.96 (on a scale from 0-4) and 1.27 per hour, respectively.

Rows 2 and 3 report descriptive statistics for type $k_i = 1$ students and type $k_i = 0$ students, respectively. As shown in the "HSTU" column, by construction, type $k_i = 1$ students spend much more time studying in high school compared to type $k_i = 0$ students (3.12 hours v.s. 0.90 hours). The "d_{it}" columns show that type $k_i = 1$ students spend more time studying in college as well. Interestingly, these two types of students do not obtain systematically different GPAs in the first semester as shown in the "g_{i0}" column. Consequently, the average actual academic productivity s_{i0} is much higher for type $k_i = 0$ students (1.42 per hour) than for type $k_i = 1$ students (0.96 per hour).

Results in Table 4.2 provide suggestive evidence about the signs of β and the misperception $\tilde{\pi}_B(k_i)$. A comparison of the "d_{i0}" column and the "s_{i0}" column shows that a group of students who have higher academic productivity spend less time studying in college on average. This suggests that if the students' expectations about academic productivity are rational, then the parameter β, which represents the effect of these expectations on study time, should be negative.

Comparing the "d_{i0}" column and the "d_{i1}" column, I find that, while the average study time for the full sample does not change much between the first two semesters, the change in average study time is more substantial for each of the two sub-samples. From the first semester to the second semester, on average, type $k_i = 1$ students adjust downward their daily study time by 0.1 hour while type $k_i = 0$ students adjust upward their daily study time by 0.11 hour. This provides suggestive evidence that, at the time of college entrance, both types of students might have non-rational expectations about their academic productivity. Moreover, the misperceptions for the two types of students should have opposite signs.

4.5.3 Estimation Results

Table 4.3 reports the estimates of β and $\tilde{\pi}_B(k_i)$ using three different estimation methods. The first row reports the estimates under the RE assumption. In this case, the misperception $\tilde{\pi}_B(k_i)$ is equal to zero, by construction. Consistent with the pattern found in Table 4.2, the estimate of β is negative and significant at the 1% level: one unit increase in students' expectation about academic productivity leads to 2.28 hours reduction in average daily study time.

Table 4.3: Estimation Results

$N = 475$	β	$\tilde{\pi}_B(k_i = 0)$	$\tilde{\pi}_B(k_i = 1)$
RE Assumption	-2.276	0	0
	(0.00)	N.A.	N.A.
LHU	-1.293	0.216	-0.167
	(0.00)	(0.09)	(0.39)
Semiparametric	-1.548	0.143	-0.103
	(0.00)	(0.26)	(0.88)

Notes: Two-sided bootstrap p-values are in parentheses

The second and third row report the LHU and feasible semiparametric estimates, respectively. In both cases, the point estimates of $\tilde{\pi}_B(k_i = 0)$ are positive while the point estimates of

$\tilde{\pi}_B(k_i = 1)$ are negative, suggesting that students with low/high high school study time overestimate/underestimate their academic productivity. Due to the small sample size, these estimates are somewhat imprecise. Nonetheless, I can still reject the null hypothesis that type $k_i = 0$ students have rational expectations about their academic productivity at the 10% level under the assumption that the updating rule for μ_{it} is linear and homogeneous across students.

Since students with low/high high school study time are the ones who have high/low academic productivity, the previous results about $\tilde{\pi}_B(k_i = 0)$ imply that the difference in expectations about academic productivity between these two types of students is larger than the difference in average actual academic productivity between these two types of students. Consequently, as shown in the first column of Table 4.3, while the LHU and feasible semiparametric estimates of β are also negative, they are much smaller in magnitude compared to the estimate of β obtained under the RE assumption.

4.6 Alternative Environments

4.6.1 A Simple Dynamic Discrete Choice Model

There are many finite horizon dynamic models with signal-based learning that are of interest to economists. For example, in the context of higher education, many researchers have modelled students' dropout decisions and major choices as results of students learning about their (major-specific) innate ability (Arcidiacono, 2004; Stange, 2012; Stinebrickner and Stinebrickner, 2014a; Arcidiacono, Aucejo, et al., 2016). However, since the environment specified in Section 4.3 requires the decision rule to be a time-invariant function of the subjective mean of the unknown factor, Theorem 4.3.3 cannot be directly applied to identify prior means and decision rules in most finite horizon dynamic models. Nonetheless, combined with additional assumptions on higher moments of the distribution describing an individual's beliefs about the unknown factor, the intuition underlying Theorem 4.3.3 can be extended to achieve identification in finite horizon dynamic models with a stationary per-period environment. In this section, I show this using a simple dynamic discrete choice model that captures the key elements in Stange (2012).

Environment and Notation

Consider a standard dynamic discrete choice model (DDCM) with signal-based learning. Time is discrete and indexed by t. At each period t, each individual i maximizes the present value of her total expected per-period utility over the rest of her lifecycle (from t to $t = T$) by choosing an action d_{it} from a non-empty discrete choice set $C_{it} \subseteq \{0,1\}$. For the purpose of illustration, here I consider the case where $T = 1$ and assume $C_{i0} = \{0,1\}$.

The choice set at period $t = 1$, C_{i1}, depends on the action d_{i0} at period $t = 0$. In this section, I consider a special case where action 0 is an absorbing state. This allows me to highlight the most salient difficulty in the identification of DDCM while keeping the presentation of theorems and proofs succinct. The existence of an absorbing state is commonly assumed in models of schooling decisions. For example, when studying college dropout decisions, researchers often assume that once a student chooses to drop out of college, she cannot choose to go back to college later (e.g., Stange, 2012; Stinebrickner and Stinebrickner, 2014a). Formally, C_{i1} is given by:

$$C_{i1} = \begin{cases} \{0\} & \text{if } d_{i0} = 0, \\ \{0,1\} & \text{if } d_{i0} = 1. \end{cases} \tag{4.20}$$

Individual i's per-period utility from choosing action j is determined by a set of state variables, including observed (by the econometrician) factors x_{ijt}, unobserved (by the econometrician) factors ϵ_{ijt}, and the subjective mean μ_{it} of an unknown permanent factor A_i.

The per-period utility function is separable in (x_{ijt}, μ_{it}) and ϵ_{ijt}, and is linear in μ_{it} given x_{ijt}:

$$u_{ijt} = f_j^{\alpha}(x_{ijt}) + f_j^{\beta}(x_{ijt})\mu_{it} + \epsilon_{ijt}. \tag{4.21}$$

I normalize $f_j^{\alpha}(x_{i0t}) = f_j^{\beta}(x_{i0t}) = 0$ for all x_{i0t} following the majority of the DDCM literature.

Similar to the setting in Section 4.3, individual i's prior mean of A_i is given by $B(z_i)$. Between the two periods, if individual i chooses $d_{i0} = 1$ at period $t = 0$, she receives a signal $s_{i0} = A_i + v_{i0}$ and updates her beliefs about A_i. The updating rule for the subjective mean is given by:

$$\mu_{i1} - \mu_{i0} = \Gamma_0(s_{i0} - \mu_{i0}; z_i). \tag{4.22}$$

I do not specify and restrict how individuals update higher moments of their beliefs about A_i.

The econometrician observes decisions d_{it}, signals s_{i0}, utility-influencing factors x_{ijt} and beliefs-influencing factors z_i. The objects of interest are utility functions $f_j^{\alpha}(x_{ijt})$ and $f_j^{\beta}(x_{ijt})$, and prior mean function $B(z_i)$.

Identification Results

Analogous to Section 4.3.2, in this section, I assume that the (joint) distributions of the observables are known by the econometrician and abstract away from complications that arise in estimation. I leave these complications for future research.

Most of the assumptions required to identify the model are identical or similar to those in

Section 4.3. Furthermore, these assumptions are also consistent with the ones made in Stange (2012). Specifically, I impose the following assumptions.

Assumption 4.6.1 *ϵ_{ijt} is independent from any other factors and is identically distributed over time for all j.*

Assumption 4.6.2 *$\epsilon_{i0t} - \epsilon_{i1t}$ has a known strictly monotonic CDF $F_{\Delta\epsilon}(\epsilon)$ and its support is $\mathbb{R}$.*

As will be clearer later, the strict monotonicity of $F_{\Delta\epsilon}(\epsilon)$ helps guarantee that the probability of choosing $d_{it} = 1$ (analogous to the average decision in Section 4.3) is strictly monotonic in the subjective mean μ_{it}. The assumption that $F_{\Delta\epsilon}(\epsilon)$ is known is a standard normalization assumption. The second part of this assumption ensures that the probability of choosing either of the two actions is strictly positive. It guarantees that there are always individuals who receive a signal between the two periods and are allowed to choose between the two actions at period $t = 1$.

Assumption 4.6.3 *For any z, $\Gamma_0(0; z) = 0$ and $\Gamma_0(s; z)$ is strictly monotonic in s.*

The assumptions on the updating rules are essentially the same as those in Section 4.3.2. The only difference is that now I do not allow the updating rule to depend on the unobserved factor ξ_i. This restriction allows me to recover each individual's subjective distribution of S_{i0} from the joint distribution of Z_i and S_{i0}.

Assumption 4.6.4 *Conditional on any set of other observables, the support of the error term in signals, v_{it}, is $\mathbb{R}$.*

Assumption 4.6.5 *At period $t = 0$, individual i has rational expectations about the distributions of X_{ij1} and ϵ_{ij1}.*

I abstract away from non-rational expectations about other factors because my method requires one signal for each factor about which individuals might have non-rational expectations.

Assumption 4.6.6 *Let S^b_{i0} describe individual i's beliefs about S_{i0}. Then $S^b_{i0} = S_{i0} + (B(z_i) - E(S_{i0}|z_i))$.*

Since individuals have biased initial beliefs about the mean of A_i, their beliefs about the mean of signal s_{i0} have to be biased by the same amount as well. For identification, I assume that they have correct beliefs about higher moments of the distribution of S_{i0}. Note that this does not necessarily imply that individuals have correct beliefs about higher moments of the distribution of A_i. For example, they may be over-confident about how precise their perceptions about A_i are and (incorrectly) think the majority of the variations in the signal are due to the shock term v_{i0}.

Theorem 4.6.1 *Under Assumption 4.6.1-4.6.6, for any (x, z), $f_1^{\alpha}(x) + f_1^{\beta}(x)B(z)$, and $B(z)$ are identified if*

1. $f_1^{\beta}(x) \neq 0$;

2. $(x, x, z) \in supp((X_{i10}, X_{i11}, Z_i)|d_{i0} = 1)$.

Proof For ease of presentation, I prove Theorem 4.6.1 for the case where x_{i1t} and z_i are constant. Allowing x_{i1t} and z_i to be more flexible will not change the structure of the proof. In this case, $f_1^{\alpha}(x_{i1t})$, $f_1^{\beta}(x_{i1t})$, $B(z_i)$, and $E(S_{i0}|z_i)$ are all constant as well, which I denote $\alpha_1, \beta_1, \bar{\mu}$, and $\bar{s}$, respectively. The objects of interest are $\alpha_1 + \beta_1\bar{\mu}$ and $\bar{\mu}$.

At period $t = 1$, consider an individual i who chose $d_{i0} = 1$ at period $t = 0$. She chooses $d_{i1} = 1$ if and only if her subjective expectation of u_{i11} is higher than her subjective expectation of u_{i01}, i.e., $E_1^b(u_{i11}) > E_1^b(u_{i01})$. Hence, the probability of choosing $d_{i1} = 1$, which is also the expectation of d_{i1}, is given by:

$$\begin{aligned}
\tilde{D}_1(s) \equiv \text{Prob}(d_{i1} = 1) &= \text{Prob}(E_1^b(u_{i11}) > E_1^b(u_{i01})) \\
&= \text{Prob}(\alpha_1 + \beta_1\mu_{i1} + \epsilon_{i11} > \epsilon_{i01}) \\
&= F_{\Delta\epsilon}(\alpha_1 + \beta_1\mu_{i1}) \\
&= F_{\Delta\epsilon}(\alpha_1 + \beta_1(\bar{\mu} + \Gamma_0(s - \bar{\mu}))). \qquad (4.23)
\end{aligned}$$

Since $F_{\Delta\epsilon}(u)$ is strictly monotonic (Assumption 4.6.2), I can invert this function to obtain:

$$h(s) \equiv F_{\Delta\epsilon}^{-1}(\tilde{D}_1(s)) = \alpha_1 + \beta_1\bar{\mu} + \beta\Gamma_0(s - \bar{\mu}). \qquad (4.24)$$

Equation (4.24) shows that $h(s)$ is equal to $\alpha_1 + \beta_1\bar{\mu}$ when $s = \bar{\mu}$. Under Assumption 4.6.2 and 4.6.3, $h(s)$ and $\tilde{D}_1(s)$ are strictly monotonic. Specifically, $h(s)$ is strictly increasing (decreasing) in s when $\tilde{D}_1(s)$ is strictly increasing (decreasing) in s. Hence, $\bar{\mu}$ is identified if $\alpha_1 + \beta_1\bar{\mu}$ can be identified from individuals' decisions at period $t = 0$. However, complications arise because their decisions at $t = 0$ depend on not only the per-period utility $\alpha_1 + \beta_1\bar{\mu}$, but also the expected future value of choosing each of the two actions. Hence, roughly speaking, the identification strategy involves computing, then eliminating the expected future values from the overall value of the two actions at $t = 0$.

In Appendix D.4, I show that individual i's (subjective) expectation of the future value of choosing action 1 is a known function of individual i's prior mean $\bar{\mu}$.[13] Furthermore, it is

[13] The computation of this expected future value involves converting conditional choice probabilities to future values. See Hotz and Miller (1993) for a seminal work on this technique.

strictly increasing (decreasing) in $\bar{\mu}$ if $\tilde{D}_1(s)$ is strictly increasing (decreasing) in s. Let $\bar{V}_1(\bar{\mu})$ denote this expected future value.

Consider individual i's utility maximization problem at period $t = 0$. Let δ denote the discount factor. Following the majority of the literature, I assume that δ is set outside of the model. The subjective expected value of choosing $d_{i0} = j$ at period $t = 0$, V_{ij0}, is given by:

$$\begin{aligned} V_{i10} &= \alpha_1 + \beta_1 \bar{\mu} + \epsilon_{i10} + \delta \bar{V}_1(\bar{\mu}), \\ V_{i00} &= \epsilon_{i00} + 0 = \epsilon_{i00}. \end{aligned} \tag{4.25}$$

Individual i chooses $d_{i0} = 1$ if and only if $V_{i10} > V_{i00}$. Hence, the probability of choosing $d_{i0} = 1$, which is also the expectation of d_{i0}, is given by:

$$\begin{aligned} \tilde{D}_0 \equiv \text{Prob}(d_{i0} = 1) &= \text{Prob}(V_{i10} > V_{i00}) \\ &= \text{Prob}(\alpha_1 + \beta_1 \bar{\mu} + \epsilon_{i10} + \delta \bar{V}_1(\bar{\mu}) > \epsilon_{i00}) \\ &= F_{\Delta\epsilon}(\alpha_1 + \beta_1 \bar{\mu} + \delta \bar{V}_1(\bar{\mu})). \end{aligned} \tag{4.26}$$

Inverting $F_{\Delta\epsilon}(\mu)$, I obtain:

$$F^{-1}_{\Delta\epsilon}(\tilde{D}_0) = \alpha_1 + \beta_1 \bar{\mu} + \delta \bar{V}_1(\bar{\mu}). \tag{4.27}$$

Define $g(s) \equiv F^{-1}_{\Delta\epsilon}(\tilde{D}_0) - \delta \bar{V}_1(s)$. Since $F^{-1}_{\Delta X}(d)$, $\tilde{D}_0$, δ, and $\bar{V}_1(s)$ are all known to the econometrician, $g(s)$ is a known function of s. Equation (4.27) shows that $g(s) = \alpha_1 + \beta_1 \bar{\mu}$ when $s = \bar{\mu}$. The strict monotonicity of $\bar{V}_1(s)$ implies that if $\tilde{D}_1(s)$ is strictly increasing (decreasing) in s, then $g(s)$ is strictly decreasing (increasing) in $\bar{\mu}$.

Combining the properties of $h(s)$ and $g(s)$, I can show that $\bar{\mu}$ is equal to the unique root of $h(s) - g(s)$. I first show that $\bar{\mu}$ solves $h(s) - g(s) = 0$. This is the case because $h(\bar{\mu}) = g(\bar{\mu}) = \alpha_1 + \beta_1 \bar{\mu}$. To prove it is the only solution, I need to show that $h(s) - g(s)$ is strictly monotonic. Consider the case where $\tilde{D}_1(s)$ is strictly increasing in s. It implies that $h(s)$ is strictly increasing in s and $g(s)$ is strictly decreasing in s. Therefore, $h(s) - g(s)$ is strictly increasing in s. Similarly, in the alternative case where $\tilde{D}_1(s)$ is strictly decreasing in s, $h(s) - g(s)$ is strictly decreasing in s as well.

Finally, with $\bar{\mu}$ identified following the steps above, $\alpha_1 + \beta_1 \bar{\mu}$ can be obtained using $h(\bar{\mu}) = g(\bar{\mu}) = \alpha_1 + \beta_1 \bar{\mu}$.

4.6.2 Unobserved Heterogeneity in Prior Means

All non/semi-parametric identification results in this chapter (Section 4.3 and Section 4.6.1) rely on the assumption that there is no unobserved heterogeneity in prior means, i.e., beliefs-influencing factors z_i are fully observed by the econometrician. This assumption is necessary primarily for two reasons. First, when the decision rule or the updating rule is nonlinear, it is generally the case that the average decision of a group of individuals with heterogeneous prior means will change when they receive a signal that is equal to the average of their prior means. Hence, without additional parametric restrictions, it is not even possible to identify the average prior mean for a group of ex ante heterogeneous individuals. Second, if there is an unobserved (by the econometrician) factor that determines A_i, conceptually it should influence both prior mean μ_{i0} and signal s_{i0}, which generates a correlation between the unobserved components in μ_{i0} and s_{i0}. This is problematic because it leads to an omitted variable bias when estimating decisions at period $t = 1$ as a function of signals.

Motivated by a literature interested in identifying unobserved heterogeneity in prior beliefs under the RE assumption (Carneiro, Hansen, and Heckman, 2003; Cunha, Heckman, and Navarro, 2004, 2005), I investigate this identification problem without the RE assumption in this section. Naturally, to allow for unobserved heterogeneity in prior means, these two issues have to be addressed. In this section, I show that with linearity assumptions on the decision rule, updating rule, and prior mean function, and one more period of information on signal s_{it}, I can identify both the decision rule and the average prior mean for each observed group (indexed by z_i) of individuals in the presence of unobserved heterogeneity in prior means. The linearity assumptions allow me to solve the first issue. With two observations on s_{it} for each individual, I can difference out the permanent unobserved component in s_{it} and construct an instrumental variable for signal s_{i0}.

Environment

Consider the case where the econometrician does not observe some factors that are in individuals' information set at $t = 0$. Denote the actual and perceived effects of these factors on A_i as v_i^R and v_i^B, respectively. I maintain the same linearity assumptions as in Section 4.4.1. The full model is given by:

$$\begin{aligned} d_{it} &= x_{i0}'\alpha + \beta\mu_{it} + \epsilon_{it}, \\ \mu_{i0} &= z_i'\pi_B + v_i^B, \\ \mu_{i1} &= \mu_{i0} + \theta(s_{i0} - \mu_{i0}), \\ s_{it} &= z_i'\pi_R + v_i^R + \eta_i + v_{it}. \end{aligned} \tag{4.28}$$

I rewrite this linear model to eliminate unobserved subjective means μ_{it}:

$$d_{i0} = x'_{i0}\alpha + z'_i(\pi_B\beta) + \beta v_i^B + \epsilon_{i0}, \tag{4.29}$$

$$d_{i1} = x'_{i1}\alpha + z'_i[\pi_B\beta(1-\theta)] + \beta\theta s_{i0} + \beta(1-\theta)v_i^B + \epsilon_{i1}, \tag{4.30}$$

$$s_{it} = z'_i\pi_R + v_i^R + \eta_i + v_{it}. \tag{4.31}$$

The parameters of primary interest are β and π_B.

Identification Results

For illustration purpose, here I consider a special case where none of the random variables contained in Z_i is a linear combination of X_{it}. In Appendix D.5, I show that the method proposed in this section is valid as long as X_{it} and Z_i are linearly independent.

Uncorrelated Unobserved Components

Equation (4.29) shows that α and $\pi_B\beta$ can be estimated by regressing d_{i0} on x_{i0} and z_i. Similarly, if v_i^R and v_i^B are uncorrelated, OLS regression of d_{i1} on x_{i1}, z_i, and s_{i0} gives estimators of α, $\pi_B\beta(1-\theta)$, and $\beta\theta$. Hence, θ can be consistently estimated by $1 - \frac{\widehat{\pi_B\beta(1-\theta)}}{\widehat{\pi_B\beta}}$, β can be consistently estimated by $\frac{\widehat{\beta\theta}}{\hat{\theta}}$, and π_B can be consistently estimated by $\frac{\widehat{\pi_B\beta}}{\hat{\beta}}$.

Intuitively, the effect of beliefs-influencing factor z_i on decision d_{it} is the product of the effect of z_i on the subjective mean μ_{it} and the effect of μ_{it} on d_{it}. Given the linear updating rule, the effect of z_i on μ_{it} diminishes at the rate of $1-\theta$. Since the effect of μ_{it} on d_{it} is a constant (β), the effect of z_i on d_{it} also diminishes at the rate of $1-\theta$. This allows for the identification of θ.[14]

Similarly, the effect of the signal s_{i0} on the decision d_{i1} is the product of the effect of s_{i0} on the subjective mean μ_{i1} (θ) and the effect of μ_{i1} on d_{i1} (β). Hence, with θ identified as above, β can be identified as well. Finally, π_B, which represents the effect of z_i on the prior mean μ_{i0}, can be estimated by the ratio of the effect of z_i on the first period decision d_{i0} to β.

Correlated Unobserved Components

Conceptually, however, it is natural to expect v_i^R and v_i^B to be correlated since they originate from the same factors. In this case, the OLS estimator of Equation (4.30) is biased because s_{i0} is correlated with $\beta\theta v_i^B$. To deal with this endogeneity issue, I construct an instrumental variable for s_{i0} by taking the difference between individual i's first two signals. This IV is

[14]The idea that a linear homogeneous updating rule can be estimated from the speed at which the effect of beliefs-influencing factors on decisions diminishes over time is also used in the employer learning literature (Farber and Gibbons, 1996, Altonji and Pierret, 2001, Lange, 2007) to estimate the speed of learning.

given by:

$$\Delta s_{i0} = s_{i0} - s_{i1} = v_{i0} - v_{i1}. \tag{4.32}$$

Δs_{i0} is a valid instrument for s_{i0} because (1) both v_{i0} and v_{i1} are error components in the signals, hence are independent of the unobserved heterogeneity in prior means; and (2) Δs_{i0} is correlated with s_{i0} since they share a common component v_{i0}. Using Δs_{i0} as an IV, I can consistently estimate α, $\pi_B\beta(1-\theta)$, $\beta\theta$ based on Equation (4.30) and consistently estimate β and π_B following the steps above.

4.7 Conclusion

In this chapter, I have developed a novel method to jointly identify and estimate individuals' prior means and decision rules without the Rational Expectations assumption in a multi-period environment where individuals use signals to update their beliefs about an individual-specific unknown permanent factor and repeatedly make decisions based on such beliefs. My method requires the econometrician to observe individuals' decisions and signals as well as factors that determine individuals' initial beliefs. The identification follows a two-step procedure. First, building on a crucial assumption on how individuals update their subjective means, I identify the prior mean for each group of individuals who have identical beliefs-influencing factors based on how their average decision changes after receiving a signal. Second, when the support of beliefs-influencing factors is sufficiently large, I can identify the prior mean and the average decision for each of these groups and effectively pin down the decision rule.

Using data from the Berea Panel Study, I apply my method to estimate the relationship between college students' study time and their expectations about academic productivity. This empirical exercise contributes to a recent literature interested in understanding the determinants of students' study effort. I find that high expectations about own academic productivity have a negative effect on students' study time. A particular focus of this application is to demonstrate the empirical importance of relaxing the RE assumption. I find that students who spent less than 2 hours per day studying in high school over-estimate their academic productivity in college. The RE assumption is rejected at the 10% level for these students. Incorrectly imposing the RE assumption leads to a much more negative estimate of the effect of expectations about academic productivity on college study time.

I would like to propose several potential avenues for future research. First, the identification results, presented in Section 4.6.1, suggest that my method can be applied to estimate college attendance/dropout models without the RE assumption using large-scale datasets such as the NLSY97. This can be considered as a cost-effective alternative to the direct elicitation of students' expectations, for investigating students' misperceptions about the return to college education and the impact of these misperceptions on their schooling decisions. Second, my method only imposes restrictions on how an individual updates her subjective mean, which is a specific moment of the distribution describing her beliefs about a permanent factor. Consequently, other moments (e.g., the variance) of the distribution cannot be identified using my method. It would be interesting to investigate whether stronger/alternative assumptions on updating rules can help identify these moments. For example, the standard assumption of Bayesian learning imposes restrictions on how an individual updates the entire distribution describing her beliefs upon receiving a signal. Finally, the feasible semiparametric estimator developed in Section 4.4.1 requires the econometrician to know how to group individuals by their misperceptions. A natural and desirable way to relax this requirement is to assume the econometrician knows how many types of misperceptions there are, but does not know each individual's type ex ante and has to infer their types from observables. In this case, one might be able to develop a feasible semiparametric estimator based on results from the literature on nonparametric clustering algorithms.

Bibliography

[1] Daniel Ackerberg. Advertising, learning, and consumer choice in experience good markets: an empirical examination. *International Economic Review*, 44(3):1007–1040, 2003.

[2] Victor Aguirregabiria and Arvind Magesan. Identification and estimation of dynamic discrete games when players' beliefs are not in equilibrium. *Review of Economic Studies*, forthcoming.

[3] Joseph Altonji and Charles Pierret. Employer learning and statistical discrimination. *The Quarterly Journal of Economics*, 116(1):313–350, 2001.

[4] Peter Arcidiacono. Ability sorting and the returns to college major. *Journal of Econometrics*, 121(1-2):343–375, 2004.

[5] Peter Arcidiacono, Esteban Aucejo, Arnaud Maurel, and Tyler Ransom. College attrition and the dynamics of information revelation. *NBER Working Paper*, (w22325), 2016.

[6] Richard Barlow, David Bartholomew, James Bremner, and Daniel Brunk. *Statistical inference under order restrictions: The theory and application of isotonic regression*. 1972.

[7] David Baron. Price uncertainty, utility, and industry equilibrium in pure competition. *International Economic Review*, pages 463–480, 1970.

[8] Andrew Caplin, John Leahy, and Filip Matejka. Rational inattention and inference from market share data. *URL: http://home. cergeei. cz/matejka/Market Share Final. pdf*, 2016.

[9] Pedro Carneiro, Karsten Hansen, and James Heckman. 2001 lawrence r. klein lecture estimating distributions of treatment effects with an application to the returns to schooling and measurement of the effects of uncertainty on college choice. *International Economic Review*, 44(2):361–422, 2003.

[10] Christopher Chambers and Paul Healy. Updating toward the signal. *Economic Theory*, 50(3):765–786, 2012.

[11] Timothy Conley and Christopher Udry. Learning about a new technology: Pineapple in ghana. *American economic review*, 100(1):35–69, 2010.

[12] Gregory Crawford and Matthew Shum. Uncertainty and learning in pharmaceutical demand. *Econometrica*, 73(4):1137–1173, 2005.

[13] Flavio Cunha and James Heckman. The technology of skill formation. *American Economic Review*, 97(2):31–47, 2007.

[14] Flavio Cunha and James Heckman. Formulating, identifying and estimating the technology of cognitive and noncognitive skill formation. *Journal of human resources*, 43(4):738–782, 2008.

[15] Flavio Cunha, James Heckman, Lance Lochner, and Dimitriy Masterov. Interpreting the evidence on life cycle skill formation. *Handbook of the Economics of Education*, 1:697–812, 2006.

[16] Flavio Cunha, James Heckman, and Salvador Navarro. Separating heterogeneity from uncertainty an aiyagari-laitner economy. In *Goldwater Conference on Labor Markets in Arizona*, volume 608, 2004.

[17] Flavio Cunha, James Heckman, and Salvador Navarro. Separating uncertainty from heterogeneity in life cycle earnings. *oxford Economic papers*, 57(2):191–261, 2005.

[18] Flavio Cunha, James Heckman, and Susanne Schennach. Estimating the technology of cognitive and noncognitive skill formation. *Econometrica*, 78(3):883–931, 2010.

[19] Gianni De Fraja, Tania Oliveira, and Luisa Zanchi. Must try harder: Evaluating the role of effort in educational attainment. *The Review of Economics and Statistics*, 92(3):577–597, 2010.

[20] Tülin Erdem and Michael Keane. Decision-making under uncertainty: Capturing dynamic brand choice processes in turbulent consumer goods markets. *Marketing Science*, 15(1):1–20, 1996.

[21] Fulya Ersoy. Effects of perceived productivity on study effort: Evidence from a field experiment. *mimeo*, 2019.

[22] Henry Farber and Robert Gibbons. Learning and wage dynamics. *The Quarterly Journal of Economics*, 111(4):1007–1047, 1996.

[23] Peter Hall and Li-Shan Huang. Nonparametric kernel regression subject to monotonicity constraints. *The Annals of Statistics*, 29(3):624–647, 2001.

[24] Duncan Holthausen. Input choices and uncertain demand. *The American Economic Review*, 66(1):94–103, 1976.

[25] V Joseph Hotz and Robert A Miller. Conditional choice probabilities and the estimation of dynamic models. *The Review of Economic Studies*, 60(3):497–529, 1993.

[26] Fabian Lange. The speed of employer learning. *Journal of Labor Economics*, 25(1):1–35, 2007.

[27] Hayne Leland. Theory of the firm facing uncertain demand. *The American Economic Review*, 62(3):278–291, 1972.

[28] Alan Olivi. Revealed preferences and beliefs from consumption-savings decisions. *mimeo*, 2019.

[29] Matthew Osborne. Consumer learning, switching costs, and heterogeneity: A structural examination. *Quantitative Marketing and Economics*, 9(1):25–70, 2011.

[30] James Ramsay. Monotone regression splines in action. *Statistical science*, 3(4):425–441, 1988.

[31] Agnar Sandmo. On the theory of the competitive firm under price uncertainty. *The American Economic Review*, 61(1):65–73, 1971.

[32] Kevin Stange. An empirical investigation of the option value of college enrollment. *American Economic Journal: Applied Economics*, 4(1):49–84, 2012.

[33] Ralph Stinebrickner and Todd Stinebrickner. Academic performance and college dropout: Using longitudinal expectations data to estimate a learning model. *Journal of Labor Economics*, 32(3):601–644, 2014.

[34] Todd Stinebrickner and Ralph Stinebrickner. The causal effect of studying on academic performance. *The BE Journal of Economic Analysis & Policy*, 8(1), 2008.

[35] Michael Woodford. Macroeconomic analysis without the rational expectations hypothesis. *Annu. Rev. Econ.*, 5(1):303–346, 2013.

Appendix A

Survey Questions

Question 1. The following questions will ask you about the income you might earn in the future at different ages under several hypothetical scenarios. We realize that you will not know exactly how much money you would make at a particular point in time. However, you may believe that some amounts of money are quite likely while others are quite unlikely. We would like to know what you think. We first ask you to indicate the lowest possible amount of money you might make and the highest amount of money you might make. We then ask you to divide the values between the lowest and the highest into four intervals. Please mark the intervals so that there is a 25% chance that your income will be in each of the intervals. When reporting incomes, take into account the possibility that you will work full-time, the possibility that you will work part-time, the possibility that you will not be working, and (for the hypothetical scenarios which involve graduation) the possibility that you will attend graduate or professional school. When reporting income you should ignore the effects of price inflation. (NOTE TO READER: Before answering Question 1, students received classroom training related to these specific questions. The written instructions/example shown in this appendix after Question 1 are strongly related to the classroom training.)

Question 1A. For ALL of question 1A, **assume that you graduate from Berea**. Think about the kinds of jobs that will be available for you and those that you would accept. Please write the **FIVE NUMBERS** that describe the income which you would expect to earn at the following ages or times under this hypothetical scenario.

I. Your income during the first full year after you leave school

|____|

lowest highest

II. Your income at age 28 (note: if you are 20 years of age or older, give your income 10 years from now)

| |
lowest highest

III. Your income at age 38 (note: if you are 20 years of age or older, give your income 20 years from now)

| |
lowest highest

Question 1B. For ALL of question 1B, assume that you graduate from Berea. Question 1A did not make any assumptions about your final grade average. For this question, **assume that you graduate with a grade point average of 2.0** (a C average). Please describe the income which you would expect to earn at the following ages or times under this hypothetical scenario.

I. Your income during the first full year after you leave school

| |
lowest highest

II. Your income at age 28

| |
lowest highest

III. Your income at age 38

| |
lowest highest

NOTE TO READER: In the paper, we also use close variants of Question 1, in which students were asked to consider scenarios in which they leave Berea after three years of study or graduate with other grade point averages (GPA) (3.00 and 3.75).

INSTRUCTIONS AND EXAMPLE To illustrate what we are asking you to do, consider the following example. A student is asked to describe what she thinks about how well she will do on an exam before taking it. Before the exam the person will not know exactly what grade she will receive. However, she will have some idea of what grade she will receive. Suppose that the person believes that the lowest possible grade she will receive is a 14 and the highest possible grade is 100 (so she believes that there is no chance that she will receive less than a 14 and some chance she will earn as high as 100).

1) The above person would begin by indicating the lowest and highest value on the line. (We will provide the lines for you whenever they are needed.)

14 100

lowest highest

2) The person would then divide the values between 14 and 100 into four intervals so that she thinks that there is a 25% chance that her grade will be in each interval. For example, suppose that the person marked three points between 14 and 100 and labeled them 52, 80 and 92.

14 54 80 92 100

lowest highest

This would mean that the person thinks there is a 25% chance she will get a grade between 14 and 52. Similarly, the person thinks there is a 25% chance she will get a grade between 52 and 80, a 25% chance she will get a grade between 80 and 92, and there is a 25% chance she will get a grade between 92 and 100. (This also means that the person thinks that there is a 50% chance she will get a grade less than 80 and a 50% chance that she will get a grade higher than 80.)

NOTE that the intervals o not have to have the same widths. For example, the interval between 14 and 52 is wider than the other intervals. This suggests that the student believes that she has a smaller chance of receiving a particular grade in this interval than a particular grade in the higher intervals. For example, the person may think that she is less likely to receive a 30 than 82.

A different person taking the exam might have very different views about how he might do on the exam. For example, a student might fill in the line to look like

0 32 51 63 90

lowest highest

This student thinks that the smallest possible grade is 0 and the highest possible grade he will receive is 90. When compared to the other student, this student thinks he is more likely to get a lower grade. For example, he thinks that there is a 25% chance he will get a grade less than 32. There is a 25% chance he will get a grade between 32 and 51. The chance that he gets a grade higher than 63 is only 25%. This person thinks there is a 50% chance he will get less than 51 and a 50% chance he will get more than 51.

We will be asking you questions about income instead of grades. However, the process will be the same as above. **For each question, please do the following:**

1) **Write the lowest and highest possible incomes above the words lowest and highest on the line.** Give the salary in thousands of dollars. If you write 15, you will mean $15,000. If you write 120, you will mean $120,000.

2) **Mark three points on the line between the lowest and highest values and write an income above each point.** These income values should divide the line into four intervals. As in the previous example, the numbers should be chosen so that there is a 25% chance that your income will be in each interval. The middle value you write should be the number such that there is a 50% chance that you will make more money and a 50% chance you will make less money.

Note: For each line you should enter five numbers.

The following questions will ask you about the income you would expect to earn under several hypothetical scenarios. Each of the questions will have the same format. In particular, each question will be divided into three parts. Each part will ask you the income that you will earn at a particular time in your life. The questions will differ in their assumptions about how far you go in school an how well you do in classes. In the first three questions, we will ask you about your income under several scenarios in which you do not graduate. In the last four questions, we ask you about your income under several scenarios in which you graduate with different grade point averages.

Question 2. We realize that you do not know exactly how well you will do in classes. However, we would like to have you describe your beliefs about the grade point average that you expect to receive in the first semester. Given the amount of study-time you indicated, please tell us the percent chance that your grade point average will be in each of the following intervals. That is, for each interval, write the number of chances out of 100 that your final grade point average will be in that interval.

Note: The numbers on the six lines must add up to 100.

Interval	**Percent Chance(number of chances out of 100)**
[3.5,4.00]	________________________

[3.0,3.49] ________________________

[2.5,2.99] ________________________

[2.0,2.49] ________________________

[1.0,1.99] ________________________

[0.0,0.99] ________________________

Note: A=4.0, B=3.0, C=2.0, D=1.0, F=0.0

Question 3. Your grades are influenced by your academic ability/preparation and how much you decide to study. However, your grades may also be influenced to some extent by good or bad luck which may vary from term to term and may be out of your control. Examples of "luck" may include 1) The quality of the teachers you happen to get and how hard or easy they grade; 2) Whether you happened to get sick (or didn't get sick) before important exams; 3) Whether a noisy dorm kept you from sleeping before an important exam; 4) Whether you happened to study the wrong material for exams; 5) Whether unexpected personal problems or problems with your friends and family made it hard to concentrate on classes.

We would like to know how important you think "luck" is in determining your grades in a particular semester. We'll have you make comparisons relative to a semester in which you have "average" luck. Average luck means that a usual number of things go right and wrong during the semester. Assume you took classes at Berea for many semesters.

BAD LUCK IN A TERM MEANS THAT YOU HAVE WORSE THAN AVERAGE LUCK IN THAT TERM

Assume for this section that you are in a semester in which you have bad luck

In what percentage of semesters that you have bad luck would bad luck lower your grade point average (GPA) by between 0.00 points and 0.25 points? ____________
(If you are taking four courses, bad luck would lower your GPA by 0.25 points if bad luck led to a full letter grade reduction in one of your courses.)

In what percentage of semesters that you have bad luck would bad luck lower your grade point average (GPA) by between 0.26 points and 0.50 points? ____________
(If you are taking four courses, bad luck would lower your GPA by 0.50 points if bad luck led to a full letter grade reduction in two of your courses or a two letter grade reduction in one of your courses.)

In what percentage of semesters that you have bad luck would bad luck lower your grade point average (GPA) by 0.51 or more points? ____________

(For a student taking four courses, this would mean that bad luck would lead to a full letter grade reduction in three or more courses.)

The numbers in the three spaces above should add up to 100(because if you are in a semester where you have bad luck, bad luck must lower your grades by between 0 and 0.25 points, or by between 0.25 and 0.5 points, or by more than 0.5 points).

GOOD LUCK IN A TERM MEANS THAT YOU HAVE BETTER THAN AVERAGE LUCK IN THAT TERM

Assume for this section that you are in a semester in which you have good luck

In what percentage of semesters that you have good luck would good luck raise your grade point average (GPA) by between 0.00 points and 0.25 points compared to a semester in which you received "average" luck? ____________
(If you are taking four courses, good luck would raise your GPA by 0.25 points if good luck led to a full letter grade increase in one of your courses.)

In what percentage of semesters that you have good luck would good luck raise your grade point average (GPA) by between 0.26 points and 0.50 points compared to a semester in which you received "average" luck? ____________
(If you are taking four courses, good luck would raise your GPA by 0.50 points if good luck led to a full letter grade increase in two of your courses or a two letter grade increase in one of your courses.)

In what percentage of semesters that you have good luck would good luck raise your grade point average (GPA) by 0.51 or more points compared to a semester in which you received "average" luck? ____________
(For a student taking four courses, this would mean that good luck would lead to a full letter grade increase in three or more courses.)

The numbers in the three spaces above in the good luck section should add up to 100(because if you are in a semester where you have good luck, good luck must increase your grades by between 0 and 0.25 points, or by between 0.25 and 0.5 points, or by more than 0.5 points).

Question 4. What is the percent chance that you will eventually graduate from Berea College? ____________Note: Number should be between 0 and 100 (could be 0 or 100).

Question 5. We realize that you may not be sure exactly what area of study you will eventually choose. In this first column below are listed possible areas of study. In the second column write down the percent chance that you will have this area of study (note: the percent chance of each particular area of study should be between 0 and 100 and the numbers in the percent chance column should add ip to 100). In the third column, please write down the grade point average (GPA) you would expect ot receive in a typical semester in the future if you had each of these areas of study.

Humanities include Art, English, Foreign Languages, History, Music, Philosophy, Religion, and Theatre.
Natural Science and Math includes Biology, Chemistry, Computer Science, Physics and Mathematics.
Professional Programs include Industrial Arts, Industrial Technology, Child Development, Dietetics, Home Economics, Nutrition, and Nursing.
Social Sciences include Economics, Political Science, Psychology and Sociology.

Area of Study	Percent Chance	Expected GPA
1. Agricultural (and Natural Resources)	____________	____________
2. Business	____________	____________
3. Elementary Education	____________	____________
4. Humanities	____________	____________
5. Natural Science & Math	____________	____________
6. Physical Education	____________	____________
7. Professional Programs	____________	____________
8. Social Sciences	____________	____________

Question 6. After graduating there are different types of jobs that you may hold. For Question 6 and 7, **NO-DEGREE-NEEDED** means all jobs that do not require a college degree. **DEGREE-ANYAREA** means all jobs that require a college degree of any type. **DEGREE-MYAREA** means all jobs that require a college degree specifically in your area of study. Please tell us the percent chance that your first job after graduating will be in each of these types of jobs.

Job-Type	Percent Chance
NO-DEGREE-NEEDED	____________
DEGREE-ANYAREA	____________
DEGREE-MYAREA	____________

Note: The numbers should add up to 100 and all numbers should be between 0 and 100. Write

0 if there is no chance that you will have a particular type of job. Write 100 if you know for sure that you will have a particular type of job.

Question 7. It is possible that how happy you will be in your job will depend on what type of job you have since different types of jobs require different types of work. Suppose you were offered the **same pay** to work in a NO-DEGREE-NEEDED job, a DEGREE-ANYAREA job, and a DEGREE-MYAREA job. Which would you choose? Circle one.

NO-DEGREE-NEEDED DEGREE-ANYAREA DEGREE-MYAREA

If NO-DEGREE-NEEDED, skip to 7.1. If DEGREE-ANYAREA, skip to 7.2. If DEGREE-MYAREA, skip to 7.3.

7.1 IF you circled NO-DEGREE-NEEDED

You have indicated that you would enjoy working in a NO-DEGREE-NEEDED job more than in either a DEGREE-ANYAREA job or a DEGREE-MYAREA job if all the jobs had the same pay. Therefore, in order to be convinced to choose a DEGREE-ANYAREA job or a DEGREE-MYAREA job, you would have to receive a job offer which paid more money than the job offer in your NO-DEGREE-NEEDED job.

If the NO-DEGREE-NEEDED job paid $30,000, how much would you have to be paid by the DEGREE-ANYAREA job to convince you to choose the DEGREE-ANYAREA job instead?
____________Note: should be more than $30,000.

If the NO-DEGREE-NEEDED job paid $30,000, how much would you have to be paid by the DEGREE-MYAREA job to convince you to choose the DEGREE-MYAREA job instead?
____________Note: should be more than $30,000.

7.2 IF you circled DEGREE-ANYAREA

You have indicated that you would enjoy working in a DEGREE-ANYAREA job more than in either a NO-DEGREE-NEEDED job or a DEGREE-MYAREA job if all the jobs had the same pay. Therefore, in order to be convinced to choose a NO-DEGREE-NEEDED job or a DEGREE-MYAREA job, you would have to receive a job offer which paid more money than the job offer in your DEGREE-ANYAREA job.

If the DEGREE-ANYAREA job paid $30,000, how much would you have to be paid by the NO-DEGREE-NEEDED job to convince you to choose the NO-DEGREE-NEEDED job instead?

____________Note: should be more than $30,000.

If the DEGREE-ANYAREA job paid $30,000, how much would you have to be paid by the DEGREE-MYAREA job to convince you to choose the DEGREE-MYAREA job instead?

____________Note: should be more than $30,000.

7.3 IF you circled DEGREE-MYAREA

You have indicated that you would enjoy working in a DEGREE-MYAREA job more than in either a NO-DEGREE-NEEDED job or a DEGREE-ANYAREA job if all the jobs had the same pay. Therefore, in order to be convinced to choose a NO-DEGREE-NEEDED job or a DEGREE-ANYAREA job, you would have to receive a job offer which paid more money than the job offer in your DEGREE-MYAREA job.

If the DEGREE-MYAREA job paid $30,000, how much would you have to be paid by the NO-DEGREE-NEEDED job to convince you to choose the NO-DEGREE-NEEDED job instead?
____________ Note: should be more than $30,000.

If the DEGREE-MYAREA job paid $30,000, how much would you have to be paid by the DEGREE-ANYAREA job to convince you to choose the DEGREE-ANYAREA job instead?
____________ Note: should be more than $30,000.

Question 8. Suppose during this school year that you searched seriously for a job. You may not know exactly how long it would take to find a job. What is the percent chance that it would take the following amounts of time to receive a job offer from the time you start searching seriously?
Note: A serious job search is one that involves actively looking for a job by participating in activities such as on-campus interviewing, reading and responding to want-ads, or contacting potential employees even if they have not posted want ads.

Amount of time to find a job-Interval	Percent Chance
[0,1) months	____________
[1,2) months	____________
[2,3) months	____________
[3,5) months	____________
[5,6) months	____________
6 months or more	____________

Appendix B

Appendices for Chapter 2

B.1 Approximation Error: Normal Versus Log-normal

When computing subjective income distributions using either normal or log-normal distributions, we have only used data on the median (C_{it}^3) and the difference between first and third quartiles ($C_{it}^4 - C_{it}^2$ or C_{it}^4/C_{it}^2). Hence, for either the normal and log-normal distributions, the three quartiles reported in the data (C_{it}^2, C_{it}^3, C_{it}^4) will not partition the support of the subjective income distribution into four segments that each have a probability of .25, unless the distributional assumption is exactly correct. Therefore, we evaluate the validity of a particular distributional assumption using the loss function:

$$AE(D) = \frac{1}{N}\sum_{i=1}^{N}[(F(C_{it}^3; D) - F(C_{it}^2; D) - 0.25)^2 + (F(C_{it}^4; D) - F(C_{it}^3; D) - 0.25)^2], \quad \text{(B.1)}$$

where $F(w; D)$ is the cdf of the distribution computed using distributional assumption D.

Using the same sample as in Section 2.3, we compute the value of $AE(D)$ for D = normal and D = log-normal. We find that $AE(\text{normal}) = 0.0101$ and $AE(\text{log-normal}) = 0.0103$. Hence, we conclude that the fit of the two distributions is quite similar with, if anything, the normal having a slightly better fit.

B.2 Approach 2: Computation Details

B.2.1 Construction of $E(W_{it}|G_{it} = g_{it})$ and $std(W_{it}|G_{it} = g_{it})$ (or, equivalently, $var(W_{it}|G_{it} = g_{it})$) at Realizations of G_{it} Other than 2.00, 3.00 or 3.75

Survey questions eliciting subjective income distributions conditional on final GPA are in the same form as the survey questions eliciting unconditional subjective income distributions shown in Question 1 of Appendix A. Hence, assuming either a log-normal, normal, or stepwise uniform distribution, Approach 1 can be used to compute $E(W_{it}|G_{it} = g_{it})$ (henceforth, $E(W_{it}|g_{it})$, for the ease of notation) and $std(W_{it}|G_{it} = g_{it})$ (henceforth, $std(W_{it}|g_{it})$) for g_{it} = 2.00, 3.00, or 3.75. However, we need to approximate $E(W_{it}|g_{it})$ and $std(W_{it}|g_{it})$ for all other possible values of g_{it}. Following a straightforward interpolation approach adopted in Stinebrickner and Stinebrickner (2014b), we assume that both $E(W_{it}|g_{it})$ and $std(W_{it}|g_{it})$ are linear between $g_{it} = 2.00$ and $g_{it} = 3.00$. We also assume that $E(W_{it}|g_{it})$ and $std(W_{it}|g_{it})$ are linear between $g_{it} = 3.00$ and $g_{it} = 4.00$, with the slope being identified by the observed values at $g_{it} = 3.00$ and $g_{it} = 3.75$ (i.e., we extrapolate values of $E(W_{it}|g_{it})$ and $std(W_{it}|g_{it})$ between $g_{it} = 3.75$ and $g_{it} = 4.00$).

B.2.2 Construction of the subjective final GPA distribution, $F_{G_{it}}(g_{it})$

In this subsection we discuss how we construct the subjective distribution G_{i0} describing beliefs, at the time of college entrance, about final cumulative GPA. A student's final GPA, G_i, is the average of the student's semester GPA over her eight semesters, k=1,...,8, subject to the constraint that the student obtains the 2.0 average that is needed to graduate. Thus, G_{i0} is given by:

$$G_{i0} = \sum_{k=1}^{8} G_{i0}^{k}/8, \text{ if } \sum_{k=1}^{8} G_{i0}^{k}/8 \geq 2, \tag{B.2}$$

where G_{i0}^{k} is the subjective distribution describing beliefs, at time $t = 0$, about semester GPA in semester k.

We view Question 2 in Appendix A as eliciting a student's subjective distribution about GPA in a typical future semester. That is, it elicits the marginal distributions of G_{i0}^{k}, $k = 1, ..., 8$. The fact that G_{i0} is the average of the G_{i0}^{k}'s implies that the mean of G_{i0} is given by the mean of the distribution elicited by Question 2. However, computing the variance of G_{i0} requires additional information describing beliefs about how the G_{i0}^{k}'s are correlated across semesters. For example, if students believe that grades are independent across time, then the variance of

G_{i0} would be found by dividing the variance elicited by Question 2 by the number of semesters (eight). On the other hand, this type of "averaging out" would not occur and the variance of G_{i0} would tend to be considerably larger if a student believes that grade performance is highly (positively) correlated across time. To formalize this notion, we denote a latent grade belief variable:

$$\tilde{G}^k_{i0} = a_{i0} + \xi^k_{i0}, \text{ where } G^k_{i0} = 0 \text{ if } \tilde{G}^k_{i0} < 0, G^k_{i0} = 4 \text{ if } \tilde{G}^k_{i0} > 4, \text{ and } G^k_{i0} = \tilde{G}^k_{i0}, \text{ otherwise.} \quad \text{(B.3)}$$

a_{i0} represents student i's ($t = 0$) beliefs about permanent (academic) ability and ξ^k_{i0} describes i's ($t = 0$) beliefs about the mean-zero transitory shock component of grades which is independent across semesters k. Thus, the G^k_{i0}'s will tend to be highly correlated if uncertainty in Survey Question 2 reflects uncertainty about ability and will have a smaller correlation if uncertainty in Survey Question 2 reflects a belief that there exists substantial transitory variation. Survey Question 2 alone provides only information about the total amount of uncertainty about grade performance. To differentiate between the two sources of uncertainty, we take advantage of Survey Question 3, which quantifies the importance of uncertainty due to the transitory shock component by asking students to report the probability that their grades in a semester would turn out to be 0.25 points and 0.5 points higher than expected due to good luck (and also bad luck).

In terms of implementation, we assume that a_{i0} and ξ^k_{i0} are normally distributed: $a_{i0} \sim N(\mu^a_{i0}, \sigma^a_{i0})$ and $\xi^k_{i0} \sim N(0, \sigma^\xi_{i0})$. For each student, we numerically search for the set of parameters $\{\mu^a_{i0}, \sigma^a_{i0}, \sigma^\xi_{i0}\}$ that minimizes a weighted sum of the discrepancies between observed and model implied probabilities. We weight each category by its associating probability to account for the fact that errors in categories with lower probability have less impact on the computation of unconditional moments of subjective income distribution.[1] Formally, we have:

$$\begin{aligned}\{\widehat{\mu^a_{i0}}, \widehat{\sigma^a_{i0}}, \widehat{\sigma^\xi_{i0}}\} = argmin &\sum_{cat^g_j \in CAT^g} Pr_{model}(G^k_{i0} \in cat^g_j)(Pr_{obs}(G^k_{i0} \in cat^g_j) - Pr_{model}(G^k_{i0} \in cat^g_j))^2 \\ &+ \sum_{cat^\epsilon_j \in CAT^\xi} Pr_{model}(\xi^k_{i0} \in cat^\xi_j)(Pr_{obs}(\xi^k_{i0} \in cat^\xi_j) - Pr_{model}(\xi^k_{i0} \in cat^\xi_j))^2, \quad \text{(B.4)}\end{aligned}$$

where $CAT^g = \{[3.5, 4.00], [3.0, 3.49], [2.5, 2.99], [2.0, 2.49], [1.0, 1.99], [0.0, .99]\}$ and $CAT^\xi = \{(-\infty, -0.5], (-0.5, -0.25], (-0.25, 0], (0, 0.25], (0.25, 0.5], (0.5, \infty)\}$.

Once parameters $\{\mu^a_{i0}, \sigma^a_{i0}, \sigma^\xi_{i0}\}$ are estimated, we can approximate the distribution of G_{i0} by simulation using equation (B.2) and (B.3).

[1] We have also estimated a non-weighted version. The results are similar.

B.3 Expressing $E(W_{it})$ as a weighted sum of $E(W_{it}|G_{it} = 2.00)$, $E(W_{it}|G_{it} = 3.00)$, and $E(W_{it}|G_{it} = 3.75)$

We show that $E(W_{it})$ can be expressed as a weighted sum of $E(W_{it}|G_{it} = 2.00)$, $E(W_{it}|G_{it} = 3.00)$, and $E(W_{it}|G_{it} = 3.75)$. For the ease of notation, we write $E(W_{it}|G_{it} = g_{it})$ as $E(W_{it}|g_{it})$. Hence,

$$
\begin{aligned}
E(W_{it}) = E_{G_{it}}(E(W_{it}|G_{it})) &= \int_2^4 E(W_{it}|g_{it})dF_{G_{it}}(g_{it}) \\
&= \int_2^3 [E(W_{it}|2.00) + \frac{E(W_{it}|3.00) - E(W_{it}|2.00)}{3.00 - 2.00}(g_{it} - 2)]dF_{G_{it}}(g_{it}) \\
&+ \int_3^4 [E(W_{it}|3.00) + \frac{E(W_{it}|3.75) - E(W_{it}|3.00)}{3.75 - 3.00}(g_{it} - 3)]dF_{G_{it}}(g_{it}) \\
&= \int_2^3 [E(W_{it}|2.00)(1 - \frac{g_{it} - 2}{3.00 - 2.00}) + E(W_{it}|3.00)\frac{g_{it} - 2}{3.00 - 2.00}]dF_{G_{it}}(g_{it}) \\
&+ \int_3^4 [E(W_{it}|3.00)(1 - \frac{g_{it} - 3}{3.75 - 3.00}) + E(W_{it}|3.75)\frac{g_{it} - 3}{3.75 - 3.00}]dF_{G_{it}}(g_{it}) \\
&= \sum_G \lambda_{it}^G E(W_{it}|G) \qquad G = 2.00, 3.00 \text{ or } 3.75,
\end{aligned}
\tag{B.5}
$$

where $\lambda_i^{2.00} = \int_2^3 (3 - g_{it})dF_{G_{it}}(g_{it})$, $\lambda_i^{3.00} = \int_2^3 (g_{it} - 2)dF_{G_{it}}(g_{it}) + \int_3^4 (1 - \frac{g_{it}-3}{0.75})dF_{G_{it}}(g_{it})$ and $\lambda_i^{3.75} = \int_3^4 \frac{g_{it}-3}{0.75}dF_{G_{it}}(g_{it})$.

B.4 Magnitude of the Measurement Error

In this section, we show that equation (2.12), along with additional assumptions, implies equation (2.13). Recall that equation (2.12) states:

$$
\widetilde{E^1}(W_{it}) - \widetilde{E^2}(W_{it}) = \varsigma_i - \sum_{g_{it}} \lambda_i^{g_{it}} \varsigma_i^{g_{it}}. \tag{2.12 revisited}
$$

Taking the variance of both sides, we have:

$$
\begin{aligned}
var(\widetilde{E^1}(W_{it}) - \widetilde{E^2}(W_{it})) &= var(\varsigma_i - \sum_{g_{it}} \lambda_i^{g_{it}} \varsigma_i^{g_{it}}) \\
&= var(\varsigma_i) + \sum_{g_{it}} var(\lambda_i^{g_{it}} \varsigma_i^{g_{it}}) && \text{(independence of MEs)} \\
&= var(\varsigma_i) + \sum_{g_{it}} E((\lambda_i^{g_{it}})^2)E((\varsigma_i^{g_{it}})^2) - (E(\lambda_i^{g_{it}})E(\varsigma_i^{g_{it}}))^2 && (\lambda_i^{g_{it}} \perp\!\!\!\perp \varsigma_i^{g_{it}}) \\
&= var(\varsigma_i) + \sum_{g_{it}} E((\lambda_i^{g_{it}})^2) var(\varsigma_i^{g_{it}}) && (E(\varsigma_i) = 0 \text{ and } E(\varsigma_i^{g_{it}}) = 0) \\
&= var(\varsigma_i)[1 + \sum_{g_{it}} E((\lambda_i^{g_{it}})^2)]. && (var(\varsigma_i) = var(\varsigma_i^{g_{it}}))
\end{aligned}
$$

Therefore,

$$
var(\varsigma_i) = \frac{var(\widetilde{E^1}(W_{it}) - \widetilde{E^2}(W_{it}))}{1 + \sum_{g_{it}} E((\lambda_i^{g_{it}})^2)}. \tag{2.13 revisited}
$$

B.5 Taking into Account Interpolation Errors

In Section 2.3.2, we note that interpolation error could be introduced into our computations because it is necessary to interpolate the means of subjective income distributions conditional on values of GPA other than 2.00, 3.00 or 3.75. In addition, errors can be introduced because it is necessary to compute distributions of final GPA from data. In this appendix, we show that taking into account these errors would lead to a smaller value of $var(\varsigma_i)$, implying a larger estimate of our measure of true heterogeneity.

We start by describing how we incorporate both types of errors into our analysis. With respect to the potential error introduced during the computation of the distribution of final GPA, we denote $F_{G_{it}}(g_{it})$ and $\widetilde{F}_{G_{it}}(g_{it})$ as the true CDF and the computed CDF of G_{it}, respectively. We allow the CDFs to potentially differ from each other and denote the difference as $F^{\Delta}_{G_{it}}(g_{it}) = \widetilde{F}_{G_{it}}(g_{it}) - F_{G_{it}}(g_{it})$.

For ease of notation, we denote a vector that includes $(E(W_{it}|G_{it} = 2.00), E(W_{it}|G_{it} = 3.00), E(W_{it}|G_{it} = 3.75))$ as $\mathbf{E^W_{G_{it}}}$, and a vector that includes $(\widetilde{E}(W_{it}|G_{it} = 2.00), \widetilde{E}(W_{it}|G_{it} = 3.00), \widetilde{E}(W_{it}|G_{it} = 3.75))$ as $\widetilde{\mathbf{E}}^{\mathbf{W}}_{\mathbf{G_{it}}}$. The interpolation approach that we use to compute the mean of subjective income distributions conditional on values of GPA other than 2.00, 3.00, or 3.75 is essentially a mapping from $\widetilde{\mathbf{E}}^{\mathbf{W}}_{\mathbf{G_{it}}}$ to $\widetilde{E}(W_{it}|G_{it} = g_{it})$, $g_{it} \neq 2.00, 3.00, 3.75$. We denote this mapping as $\widetilde{E}^W(g_{it}; \widetilde{\mathbf{E}}^{\mathbf{W}}_{\mathbf{G_{it}}})$. Note that the difference between the computed value of the conditional mean, $\widetilde{E}^W(g_{it}; \widetilde{\mathbf{E}}^{\mathbf{W}}_{\mathbf{G_{it}}})$, and the true value of conditional mean, $E(W_{it}|G_{it} = g_{it})$, is a result of both the measurement error, $\widetilde{\mathbf{E}}^{\mathbf{W}}_{\mathbf{G_{it}}} - \mathbf{E}^{\mathbf{W}}_{\mathbf{G_{it}}} = (\varsigma_i^{2.00}, \varsigma_i^{3.00}, \varsigma_i^{3.75})$, and the interpolation error, $\widetilde{E}^W(g_{it}; \mathbf{E}^{\mathbf{W}}_{\mathbf{G_{it}}}) - E(W_{it}|G_{it} = g_{it})$.

The mean of subjective income distribution computed using Approach 2, $\widetilde{E^2}(W_{it})$, is then given by,

$$\begin{aligned}
\widetilde{E^2}(W_{it}) &= \int_2^4 \widetilde{E}(W_{it}|G_{it} = g_{it}) d\widetilde{F}_{G_{it}}(g_{it}) = \int_2^4 \widetilde{E}^W(g_{it}; \widetilde{\mathbf{E}}^{\mathbf{W}}_{\mathbf{G_{it}}}) d\widetilde{F}_{G_{it}}(g_{it}) \\
&= \int_2^4 E(W_{it}|G_{it} = g_{it}) d\widetilde{F}_{G_{it}}(g_{it}) + \int_2^4 (\widetilde{E}^W(g_{it}; \widetilde{\mathbf{E}}^{\mathbf{W}}_{\mathbf{G_{it}}}) - E(W_{it}|G_{it} = g_{it})) d\widetilde{F}_{G_{it}}(g_{it}) \\
&= \int_2^4 E(W_{it}|G_{it} = g_{it}) dF_{G_{it}}(g_{it}) + \int_2^4 E(W_{it}|G_{it} = g_{it}) dF^{\Delta}_{G_{it}}(g_{it}) \\
&\quad + \int_2^4 (\widetilde{E}^W(g_{it}; \widetilde{\mathbf{E}}^{\mathbf{W}}_{\mathbf{G_{it}}}) - E(W_{it}|G_{it} = g_{it})) d\widetilde{F}_{G_{it}}(g_{it}) \\
&= E(W_{it}) + \int_2^4 E(W_{it}|G_{it} = g_{it}) dF^{\Delta}_{G_{it}}(g_{it}) + \int_2^4 (\widetilde{E}^W(g_{it}; \widetilde{\mathbf{E}}^{\mathbf{W}}_{\mathbf{G_{it}}}) - \widetilde{E}^W(g_{it}; \mathbf{E}^{\mathbf{W}}_{\mathbf{G_{it}}})) d\widetilde{F}_{G_{it}}(g_{it}) \\
&\quad + \int_2^4 (\widetilde{E}^W(g_{it}; \mathbf{E}^{\mathbf{W}}_{\mathbf{G_{it}}}) - E(W_{it}|G_{it} = g_{it})) d\widetilde{F}_{G_{it}}(g_{it}) \qquad \text{(B.6)}
\end{aligned}$$

Following steps similar to those in Section B.3, we can show that:

$$\int_2^4 (\widetilde{E}^W(g_{it}; \widetilde{\mathbf{E}}^{\mathbf{W}}_{\mathbf{G_{it}}}) - \widetilde{E}^W(g_{it}; \mathbf{E}^{\mathbf{W}}_{\mathbf{G_{it}}})) d\widetilde{F}_{G_{it}}(g_{it}) = \sum_{g_{it}} \widetilde{\lambda}_i^{g_{it}} \varsigma_i^{g_{it}}, \qquad g_{it} = 2.00, 3.00 \text{ or } 3.75, \quad \text{(B.7)}$$

where $\widetilde{\lambda}_i^{2.00} = \int_2^3 (3 - g_{it}) d\widetilde{F}_{G_{it}}(g_{it})$, $\widetilde{\lambda}_i^{3.00} = \int_2^3 (g_{it} - 2) d\widetilde{F}_{G_{it}}(g_{it}) + \int_3^4 (1 - \frac{g_{it}-3}{0.75}) d\widetilde{F}_{G_{it}}(g_{it})$ and $\widetilde{\lambda}_i^{3.75} = \int_3^4 \frac{g_{it}-3}{0.75} d\widetilde{F}_{G_{it}}(g_{it})$.

Denoting $\Delta_{it} \equiv \int_2^4 E(W_{it}|G_{it} = g_{it}) dF^{\Delta}_{G_{it}}(g_{it}) + \int_2^4 (\widetilde{E}^W(g_{it}; \mathbf{E}^{\mathbf{W}}_{\mathbf{G_{it}}}) - E(W_{it}|G_{it} = g_{it})) d\widetilde{F}_{G_{it}}(g_{it})$, equation (B.6) can be written as:

$$\widetilde{E^2}(W_{it}) = E(W_{it}) + \sum_{g_{it}} \widetilde{\lambda}_i^{g_{it}} \varsigma_i^{g_{it}} + \Delta_{it} \qquad g_{it} = 2.00, 3.00 \text{ or } 3.75. \qquad \text{(B.8)}$$

Taking the difference between the mean computed using Approach 1 and the mean computed using Approach 2, we obtain:

$$\widetilde{E^1}(W_{it}) - \widetilde{E^2}(W_{it}) = \varsigma_i - \sum_{g_{it}} \widetilde{\lambda}_i^{g_{it}} \varsigma_i^{g_{it}} - \Delta_{it} \qquad g_{it} = 2.00, 3.00 \text{ or } 3.75. \qquad \text{(B.9)}$$

Recall that ς_i and $\varsigma_i^{g_{it}}$, $g_{it} = 2.00, 3.00$ or 3.75, are, by assumption, independent of other factors. Hence, they are independent of Δ_{it} since none of them show up in the expression of Δ_{it}. Taking the variance of both sides of equation (B.9), we find:

$$
\begin{aligned}
var(\widetilde{E^1}(W_{it}) - \widetilde{E^2}(W_{it})) &= var(\varsigma_i - \sum_{g_{it}} \widetilde{\lambda}_i^{g_{it}} \varsigma_i^{g_{it}}) + var(\Delta_{it}) \\
&= var(\varsigma_i)[1 + \sum_{g_{it}} E((\widetilde{\lambda}_i^{g_{it}})^2)] + var(\Delta_{it}) \\
&\geq var(\varsigma_i)[1 + \sum_{g_{it}} E((\widetilde{\lambda}_i^{g_{it}})^2)]. \qquad \text{(B.10)}
\end{aligned}
$$

Therefore,

$$
var(\varsigma_i) \leq \frac{var(\widetilde{E^1}(W_{it}) - \widetilde{E^2}(W_{it}))}{1 + \sum_{g_{it}} E((\widetilde{\lambda}_i^{g_{it}})^2)}. \qquad \text{(B.11)}
$$

Since both $\widetilde{\lambda}_i^{g_{it}}$ in this section and $\lambda_i^{g_{it}}$ in Section 2.3.2 are computed using the same distribution of G_{it} (we assume that there is no error in the distribution of G_{it} in Section 2.3.2), they are numerically identical. Thus, the right side of equation (B.11) is numerically identical to the right side of equation (2.13). As a result, equation (B.11) shows that our estimates of $var(\varsigma_i)$ reported in Table 2.4 should be considered as upper bounds for the true value of $var(\varsigma_i)$.

B.6 Joint Decomposition

In Section 2.4.1 and Section 2.4.2, we estimated the fraction of total initial uncertainty that is explained by uncertainty about GPA and major, respectively. In this appendix we explain how to examine how much of total initial income uncertainty is due to uncertainty about both of the two factors combined.

We start by decomposing total income uncertainty into the contribution of uncertainty about both final GPA and major and the contribution of uncertainty about other factors, following an equation similar to Equation (2.4) and Equation (2.15):

$$
\begin{aligned}
var(W_{it}) &= var_{G_{it},M_{it}}(E(W_{it}|G_{it}, M_{it})) + E_{G_{it},M_{it}}(var(W_{it}|G_{it}, M_{it})) \\
&= \{var_{G_{it}}[E_{M_{it}|G_{it}}(E(W_{it}|G_{it}, M_{it}))] + E_{G_{it}}[var_{M_{it}|G_{it}}(E(W_{it}|G_{it}, M_{it}))]\} + E_{G_{it},M_{it}}(var(W_{it}|G_{it}, M_{it})) \\
&= \{var_{G_{it}}(E(W_{it}|G_{it})) + E_{G_{it}}[var_{M_{it}|G_{it}}(E(W_{it}|G_{it}, M_{it}))]\} + E_{G_{it}}[E_{M_{it}|G_{it}}(var(W_{it}|G_{it}, M_{it}))]. \qquad \text{(B.12)}
\end{aligned}
$$

The sum of the two terms in the fancy bracket corresponds to the contribution of uncertainty about both final GPA and major to total initial income uncertainty, while the last term corresponds to the contribution of uncertainty about all other factors. Analogous to Equation (2.14) and Equation (2.16), we define the contribution of final GPA and major to total income uncertainty as follows:

$$R_{it}^{GM} = \frac{var_{G_{it}}(E(W_{it}|G_{it})) + E_{G_{it}}[var_{M_{it}|G_{it}}(E(W_{it}|G_{it}, M_{it}))]}{\{var_{G_{it}}(E(W_{it}|G_{it})) + E_{G_{it}}[var_{M_{it}|G_{it}}(E(W_{it}|G_{it}, M_{it}))]\} + E_{G_{it}}[E_{M_{it}|G_{it}}(var(W_{it}|G_{it}, M_{it}))]}. \quad \text{(B.13)}$$

B.6.1 Estimation

We focus on the time of entrance ($t = 0$). In order to compute the joint contribution of final GPA and major, we need to compute all three terms on the RHS of Equation (B.12). The first term can be computed using exactly the same method as in Section 2.3.1. We now explain how to estimate the second and third term on the RHS.

Note that we can compute $E(W_{i0}|G_{i0})$ and $var(W_{i0}|G_{i0})$ for $G_{i0} = 2.00, 3.00, 3.75$. Hence, if we have data on the distribution of $M_{i0}|G_{i0}$, we can apply the method detailed in Section 2.4.2 to estimate $E(W_{i0}|G_{i0}, M_{i0})$ and $var(W_{i0}|G_{i0}, M_{i0})$ for all M_{i0} and $G_{i0} = 2.00, 3.00, 3.75$ and compute $var_{M_{i0}|G_{i0}}(E(W_{i0}|G_{i0}, M_{i0}))$ and $E_{M_{i0}|G_{i0}}(var(W_{i0}|G_{i0}, M_{i0}))$ for $G_{i0} = 2.00, 3.00, 3.75$. Then, we can interpolate their values at other realizations of G_{i0} ($G_{i0} \neq 2.00, 3.00, 3.75$) and compute $E_{G_{i0}}[var_{M_{i0}|G_{i0}}(E(W_{i0}|G_{i0}, M_{i0}))]$ and $E_{G_{i0}}[E_{M_{i0}|G_{i0}}(var(W_{i0}|G_{i0}, M_{i0}))]$ using a simulation-based method.

Unfortunately, the distribution of $M_{i0}|G_{i0}$ is not directly available in the data. To deal with this issue, we propose a method to estimate it using data on the unconditional distribution of M_{i0}, P_{ij0}, the distribution of G_{i0}, $F_{G_{i0}}(g_{i0})$ and the expectation of $G_{i0}|M_{i0}$, $E(G_{i0}|M_{i0})$.[2]

Denote the conditional probability of major, Prob($M_{i0} = j|G_{i0} = g_{i0}$), as $P^C_{ij0}(g_{i0})$. Furthermore, we assume that $P^C_{ij0}(g_{i0})$ has the following form:

$$P^C_{ij0}(g_{i0}; \rho^0_{i10}, \rho^1_{i10}, ...) = \frac{exp(\rho^0_{ij0} + \rho^1_{ij0} g_{i0})}{\sum_{j'} exp(\rho^0_{ij'0} + \rho^1_{ij'0} g_{i0})}, \quad \text{(B.14)}$$

where ρ^0_{i70} and ρ^1_{i70} are normalized to 0. This leaves us $2 \times (7 - 1) = 12$ parameters to estimate. Note that this specification actually corresponds to the case where final major is determined by a multinomial logistic model with final GPA as the regressor.

We start by writing $E(G_{i0}|M_{i0})$ as a function of P_{ij0}, $F_{G_{i0}}(g_{i0})$ and $P^C_{ij0}(g_{i0})$.

[2] More precisely, what we observe in the data (Question 5 in Appendix A) is the conditional expectation of semester GPA, $E(G^k_{i0}|M_{i0})$, instead of the conditional expectation of final GPA, $E(G_{i0}|M_{i0})$. The two would be identical if there does not exist a GPA minimum requirement for graduation. In practice, because most students believe that receiving grades less than the minimum is highly unlikely (and do not think they will drop out), in this section we simply approximate $E(G_{i0}|M_{i0})$ by $E(G^k_{i0}|M_{i0})$.

$$\begin{aligned} E(G_{i0}|M_{i0}) &= \int g_{i0} dF_{G_{i0}|M_{i0}}(g_{i0}) \\ &= \int \frac{P^C_{ij0}(g_{i0})}{P_{ij0}} g_{i0} dF_{G_{i0}}(g_{i0}), \end{aligned} \tag{B.15}$$

where the second line follows from the Bayes rule.

We can rearrange the terms in Equation (B.15) to derive an expression for P_{ij0}:

$$\begin{aligned} P_{ij0} &= \frac{1}{E(G_{i0}|M_{i0})} \int P^C_{ij0}(g_{i0}) g_{i0} dF_{G_{i0}}(g_{i0}) \\ &= \frac{1}{E(G_{i0}|M_{i0})} \int \frac{exp(\rho^0_{ij0} + \rho^1_{ij0} g_{i0})}{\sum_{j'} exp(\rho^0_{ij'0} + \rho^1_{ij'0} g_{i0})} g_{i0} dF_{G_{i0}}(g_{i0}). \end{aligned} \tag{B.16}$$

Note that, by definition, P_{ij0} also satisfies the following equation:

$$\begin{aligned} P_{ij0} &= \int P^C_{ij0}(g_{i0}) dF_{G_{i0}}(g_{i0}) \\ &= \int \frac{exp(\rho^0_{ij0} + \rho^1_{ij0} g_{i0})}{\sum_{j'} exp(\rho^0_{ij'0} + \rho^1_{ij'0} g_{i0})} dF_{G_{i0}}(g_{i0}). \end{aligned} \tag{B.17}$$

Equation (B.16) and (B.17) allow us to express P_{ij0} as two different functions of $(\rho^0_{ij0}, \rho^1_{ij0})$, $j = 1, 2, 3, ..., 7$. We label them as $\widetilde{P}^1_{ij0}(\cdot)$ and $\widetilde{P}^2_{ij0}(\cdot)$, respectively. We then define the estimator of $(\rho^0_{ij0}, \rho^1_{ij0})$, $j = 1, 2, 3, ..., 7$ to be the minimizer of the sum of squared differences between P_{ij0} and $\widetilde{P}^1_{ij0}(\rho^0_{i10}, \rho^1_{i10}, ...)$ and between P_{ij0} and $\widetilde{P}^2_{ij0}(\rho^0_{i10}, \rho^1_{i10}, ...)$. Formally, we have:

$$\{\widehat{\rho}^0_{i10}, \widehat{\rho}^1_{i10}, ...\} \equiv argmin \sum_{q=1}^{2} \sum_{j=1}^{7} [\widetilde{P}^q_{ij0}(\rho^0_{i10}, \rho^1_{i10}, ...) - P_{ij0}]^2. \tag{B.18}$$

Once $\{\widehat{\rho}^0_{i10}, \widehat{\rho}^1_{i10}, ...\}$ are estimated, we can use Equation (B.14) to compute the distribution of $M_{i0}|G_{i0}$ for any realization of G_{i0} and compute the three terms in Equation (B.12) in the way described in the second paragraph of this subsection.

Appendix C

Appendices for Chapter 3

C.1 Computation of $\tilde{\mu}^{Y,s}_{it_0}$

In this appendix we show how $\tilde{\mu}^{Y,s}_{it_0}$ can be computed using students' responses to Question 1. Recall that $Y^{s=1}_i = \sum_{a=\bar{t}}^{\bar{A}} \beta^{a-t^*} w^{a,s=1}_i$ and $Y^{s=0}_i = \sum_{a=t^*}^{\bar{A}} \beta^{a-t^*} w^{a,s=0}_i$. Denoting the mean of $\tilde{w}^{a,s}_{it_0}$, the random variable describing student i's beliefs about $w^{a,s}_i$ at t_0, as $\tilde{\mu}^{a,s}_{it_0}$, we have:

$$\tilde{\mu}^{Y,s=1}_{it_0} = \sum_{a=\bar{t}}^{\bar{A}} \beta^{a-t^*} \tilde{\mu}^{a,s=1}_{it_0}, \qquad \tilde{\mu}^{Y,s=0}_{it_0} = \sum_{a=t^*}^{\bar{A}} \beta^{a-t^*} \tilde{\mu}^{a,s=0}_{it_0}. \tag{C.1}$$

Similar to $\tilde{\sigma}^{a,s}_{it_0}$, we can obtain $\tilde{\mu}^{a,s}_{it_0}$ using students' reported quartiles of the distribution of $\tilde{w}^{a,s}_{it_0}$. Specifically, the normality assumption that we imposed on $\tilde{w}^{a,s}_{it_0}$ implies that $\tilde{\mu}^{a,s}_{it_0}$ is equal to $Q^{2,a,s}_{it_0}$, the second quartile (median) of $\tilde{w}^{a,s}_{it_0}$. Hence, adopting the same interpolation and timing assumptions as in Section 3.4.2, Equation (C.1) allows us to compute $\tilde{\mu}^{Y,s}_{it_0}$ for $s = 0, 1$.

C.2 Robustness: Allowing for Learning about $Y^{s=0}_i$

Our analysis in Section 3.4.2-3.4.4 assumed that students learn only about the future earnings associated with the graduation alternative, $Y^{s=1}_i$. The simplifying assumption that students do not learn about the future earnings associated with the dropout alternative, $Y^{s=0}_i$, has the virtue of allowing for a more transparent discussion of identification and the virtue of allowing results to be discussed in a straightforward manner. It is also consistent with the intuitively appealing notion that college is best suited for providing information about one's ability to perform high skilled jobs. Nonetheless, this section recognizes the benefit of providing some evidence that this is a reasonable assumption. We find that this is the case. Both the actual and perceived amounts of uncertainty resolved about $Y^{s=0}_i$ are much smaller than the corresponding amounts resolved about $Y_i^{s=1}$. Further, in part because of this result and in part because what a student learns about earnings under the graduation scenario is informative about earnings under the dropout alternative, allowing students to also resolve uncertainty about $Y_i^{s=0}$ does not change our substantive conclusion in Section 3.4.3 and Section 3.4.4 - that students underestimate the option value and overestimate the net continuation value.

C.2.1 Defining σ_i in a Correlated Learning Environment

Allowing students to learn about the future earnings associated with the dropout alternative leads to a modification of Equation (3.8). The relevant new information, Δ_i, is now given by:

$$\begin{aligned}\Delta_i &= \sum_{a=\bar{t}}^{\bar{A}} \beta^{a-t^*}[E_{t=t^*}(w_i^{a,s=1}) - E_{t=t_0}(w_i^{a,s=1})] - \sum_{a=t^*}^{\bar{A}} \beta^{a-t^*}[E_{t=t^*}(w_i^{a,s=0}) - E_{t=t_0}(w_i^{a,s=0})] \\ &= \sum_{a=\bar{t}}^{\bar{A}} \beta^{a-t^*}(\epsilon_{i\tau_2}^{a,s=1}) - \sum_{a=t^*}^{\bar{A}} \beta^{a-t^*}(\epsilon_{i\tau_2}^{a,s=0}),\end{aligned} \tag{C.2}$$

where, analogous to $\epsilon_{i\tau_2}^{a,s=1}$, $\epsilon_{i\tau_2}^{a,s=0}$ is the component of $w_i^{a,s=0}$ that is observed by student i between t_0 and t^*. Similarly, we assume that the $\epsilon_{i\tau_2}^{a,s=0}$ are normally distributed with standard deviation $\sigma_{i\tau_2}^{a,s=0}$ and are perfectly correlated across all a.

Motivated by recent work suggesting the importance of correlated learning (Arcidiacono et al., 2016), we allow $\epsilon_{i\tau_2}^{a,s=1}$ and $\epsilon_{i\tau_2}^{a',s=0}$ to have correlation κ for all a, a' pairs. Under these assumptions, Equation (C.2) implies that the standard deviation of Δ_i, σ_i, is given by:

$$\sigma_i = \sqrt{\left[\sum_{a=\bar{t}}^{\bar{A}} \beta^{a-t^*}(\sigma_{i\tau_2}^{a,s=1})\right]^2 + \left[\sum_{a=t^*}^{\bar{A}} \beta^{a-t^*}(\sigma_{i\tau_2}^{a,s=0})\right]^2 - 2\kappa\left[\sum_{a=\bar{t}}^{\bar{A}} \beta^{a-t^*}(\sigma_{i\tau_2}^{a,s=1})\right]\left[\sum_{a=t^*}^{\bar{A}} \beta^{a-t^*}(\sigma_{i\tau_2}^{a,s=0})\right]}. \tag{C.3}$$

As shown in Section 3.4.2, $\sum_{a=\bar{t}}^{\bar{A}} \beta^{a-t^*}(\sigma_{i\tau_2}^{a,s=1})$ can be written as a fraction ρ of $\tilde{\sigma}_{it_0}^{Y,s=1}$, student i's initial uncertainty about lifetime earnings associated with alternative $s = 1$. Similarly, we can write $\sum_{a=\bar{t}}^{\bar{A}} \beta^{a-t^*}(\sigma_{i\tau_2}^{a,s=1})$ as a fraction ρ^0 of $\tilde{\sigma}_{it_0}^{Y,s=0} \equiv \sum_{a=t^*}^{\bar{A}} \beta^{a-t^*}(\tilde{\sigma}_{it_0}^{a,s=0})$, student i's initial uncertainty about lifetime earnings associated with alternative $s = 0$. Equation (C.3) becomes:

$$\sigma_i = \sqrt{(\rho\tilde{\sigma}_{it_1}^{Y,s=0})^2 + (\rho^0\tilde{\sigma}_{it_0}^{Y,s=0})^2 - 2\kappa\rho\rho^0\tilde{\sigma}_{it_0}^{Y,s=1}\tilde{\sigma}_{it_0}^{Y,s=0}}. \tag{C.4}$$

In Section 3.4.2, we showed how to obtain $\tilde{\sigma}_{it_0}^{Y,s=1}$ from students' responses to earnings expectations questions in the BPS. Since students report their beliefs about future earnings under both alternatives ($s = 0$ and $s = 1$), $\tilde{\sigma}_{it_0}^{Y,s=0}$ can be obtained using the same method. The second column of Table 3.1 shows that the sample average of $\tilde{\sigma}_{it_0}^{Y,s=0}$ is \$163,000, roughly 30% smaller than the sample average of $\tilde{\sigma}_{it_1}^{Y,s=1}$, implying that, at t_0, on average there is more uncertainty about earnings under the graduation scenario than there is about earnings under the dropout scenario.

With data on $\tilde{\sigma}_{it_0}^{Y,s}$ for $s = 0, 1$, computation of σ_i, and therefore option values, requires information on ρ, ρ^0, and κ. In the next two subsections we discuss how to estimate the actual and perceived values of these objects.

C.2.2 Actual Option Values

Allowing for learning about the value of the dropout alternative has no bearing on our estimation of ρ_A; the estimate of ρ_A remains 0.51. The value of ρ_A^0 can be estimated in the same manner. We find an estimate of 0.28 for ρ_A^0, suggesting that students resolve a smaller fraction of their initial uncertainty about $Y_i^{s=0}$ than about $Y_i^{s=1}$. Since students were less uncertain about $Y_i^{s=0}$ than about $Y_i^{s=1}$ to begin with, we conclude that the actual uncertainty resolution about $Y_i^{s=0}$ is much smaller than that about $Y_i^{s=1}$.

The correlation κ can be estimated from the evolution of individual earnings beliefs. Recall that student i's expectation about Y_i^s at the beginning of college and at the end of the third year are denoted as $\tilde{\mu}_{it_0}^{Y,s}$ and $\tilde{\mu}_{it^*}^{Y,s}$, respectively. Equation (3.7) along with our timing assumptions imply that:

$$\begin{aligned}\tilde{\mu}_{it^*}^{Y,s=1} - \tilde{\mu}_{it_0}^{Y,s=1} &= \sum_{a=\bar{t}}^{\bar{A}} \beta^{a-t^*} \epsilon_{i\tau_2}^{a,s=1}, \\ \tilde{\mu}_{it^*}^{Y,s=0} - \tilde{\mu}_{it_0}^{Y,s=0} &= \sum_{a=t^*}^{\bar{A}} \beta^{a-t^*} \epsilon_{i\tau_2}^{a,s=0}.\end{aligned} \tag{C.5}$$

Under our assumptions that 1) $\epsilon_{i\tau_2}^{a,s=1}$ and $\epsilon_{i\tau_2}^{a',s=0}$ have a correlation of κ for any pair (a, a'), and 2) $\epsilon_{i\tau_2}^{a,s}$ are perfectly correlated across a (for a given s), we can show that the correlation of $\sum_{a=\bar{t}}^{\bar{A}} \beta^{a-t^*} \epsilon_{i\tau_2}^{a,s=1}$ and $\sum_{a=t^*}^{\bar{A}} \beta^{a-t^*} \epsilon_{i\tau_2}^{a,s=0}$ is also κ. Hence, for a random sample of students, κ can be consistently estimated by the correlation of $\tilde{\mu}_{it^*}^{Y,s=1} - \tilde{\mu}_{it_0}^{Y,s=1}$ and $\tilde{\mu}_{it^*}^{Y,s=0} - \tilde{\mu}_{it_0}^{Y,s=0}$.

However, in practice, a complication exists because the sample of students who remained at the end of third year is, by construction, not random. Indeed, in the context of our model, students choose to remain in school precisely because the realization of $\sum_{a=\bar{t}}^{\bar{A}} \beta^{a-t^*} \epsilon_{i\tau_2}^{a,s=1} - \sum_{a=t^*}^{\bar{A}} \beta^{a-t^*} \epsilon_{i\tau_2}^{a,s=0}$ is sufficiently high. To deal with this selection issue, we take advantage of the fact that selection should not be problematic when estimating the correlation between

$\tilde{\mu}^{Y,s=1}_{i(t_0+1)} - \tilde{\mu}^{Y,s=1}_{it_0}$ and $\tilde{\mu}^{Y,s=0}_{i(t_0+1)} - \tilde{\mu}^{Y,s=0}_{it_0}$ since very few students drop out before the end of the first year (i.e., we have a random sample for the first year). Data on $\tilde{\mu}^{Y,s}_{it_0}$ and $\tilde{\mu}^{Y,s}_{i(t_0+1)}$ are collected at the beginning and end of the first year, respectively. We compute this correlation to be 0.63. In the end of this subsection (Section C.2.2), we show that, with additional assumptions on how uncertainty about future earnings is resolved over time between t_0 and t^*, this correlation represents a consistent estimator of κ_A.

With ρ_A, ρ_A^0, and κ_A estimated using the methods described above, we compute the actual value of σ_i for each student. The average actual value of σ_i is \$96,780. This is smaller than the value of \$115,430 obtained using the values of ρ_A and $\tilde{\sigma}^{Y,s=1}_{it_0}$ from Section 3.4.2 under the previous assumption that students only resolve uncertainty about $Y_i^{s=1}$. Hence, allowing students to also resolve uncertainty about $Y_i^{s=0}$ leads students to learn less about the gap between the value of the two alternatives. This is primarily because students learn about the two alternatives in a positively correlated fashion: a positive information shock to the graduation alternative is likely to be accompanied with a positive information shock to the dropout alternative. Consequently, the actual option values computed under this correlated learning environment are also somewhat smaller than their counterparts in the baseline scenario. The average actual OV and NCV are now \$21,020 and \$63,720, respectively (versus \$25,040 and \$76,130, respectively, in Section 3.4.3 and Section 3.4.4).

$\kappa_1 = \kappa$: Assumptions and Proof

We show that, with additional assumptions on how uncertainty about future earnings are resolved between t_0 and t^*, we can consistently estimate κ_A using the correlation between $\bar{\mu}^{s=1}_{i(t_0+1)} - \bar{\mu}^{s=1}_{it_0}$ and $\bar{\mu}^{s=0}_{i(t_0+1)} - \bar{\mu}^{s=0}_{it_0}$. We start by further decomposing $\epsilon^{a,s}_{i\tau_2}$ into independently distributed factors that are realized in Year 1, Year 2 and Year 3, respectively;

$$\epsilon^{a,s}_{i\tau_2} = \sum_{j=1}^{3} \epsilon^{a,s}_{i\tau_2^j}, \tag{C.6}$$

where $\epsilon^{a,s}_{i\tau_2^j}$ normally distributed with standard deviation $\sigma^{a,s}_{i\tau_2^j}$. It follows that:

$$\begin{aligned} \tilde{\mu}^{Y,s=1}_{i(t_0+1)} - \tilde{\mu}^{Y,s=1}_{it_0} &= \sum_{a=\bar{t}}^{\bar{A}} \beta^{a-t^*} \epsilon^{a,s=1}_{i\tau_2^1}, \\ \tilde{\mu}^{Y,s=0}_{i(t_0+1)} - \tilde{\mu}^{Y,s=0}_{it_0} &= \sum_{a=t^*}^{\bar{A}} \beta^{a-t^*} \epsilon^{a,s=0}_{i\tau_2^1}. \end{aligned} \tag{C.7}$$

We assume that the correlation between $\epsilon^{a,s=1}_{i\tau_2^j}$ and $\epsilon^{a',s=0}_{i\tau_2^j}$, given any a, a' pair, is κ_j. This implies that the correlation between $\tilde{\mu}^{Y,s=1}_{i(t_0+1)} - \tilde{\mu}^{Y,s=1}_{it_0}$ and $\tilde{\mu}^{Y,s=0}_{i(t_0+1)} - \tilde{\mu}^{Y,s=0}_{it_0}$ is also κ_j. Under the assumption that both the correlation κ_j and the ratio of signal strength $\frac{\sigma^{a,s=1}_{i\tau_2^j}}{\sigma^{a,s=0}_{i\tau_2^j}}$ are constant over j, it can be shown that $\kappa = \kappa_1$.

Proof We first show that $\frac{\sigma^{a,s=1}_{i\tau_2}}{\sigma^{a,s=0}_{i\tau_2}} = \frac{\sigma^{a,s=1}_{i\tau_2^1}}{\sigma^{a,s=0}_{i\tau_2^1}}$;

$$\begin{aligned}
\frac{\sigma^{a,s=1}_{i\tau_2}}{\sigma^{a,s=0}_{i\tau_2}} &= \frac{\sqrt{\sum_{j=1}^{3}(\sigma^{a,s=1}_{i\tau_2^j})^2}}{\sqrt{\sum_{j=1}^{3}(\sigma^{a,s=0}_{i\tau_2^j})^2}} \\
&= \sqrt{\frac{\sum_{j=1}^{3}(\sigma^{a,s=0}_{i\tau_2^j})^2(\frac{\sigma^{a,s=1}_{i\tau_2^j}}{\sigma^{a,s=0}_{i\tau_2^j}})^2}{\sum_{j=1}^{3}(\sigma^{a,s=0}_{i\tau_2^j})^2}} \\
&= \sqrt{\frac{\sum_{j=1}^{3}(\sigma^{a,s=0}_{i\tau_2^j})^2(\frac{\sigma^{a,s=1}_{i\tau_2^1}}{\sigma^{a,s=0}_{i\tau_2^1}})^2}{\sum_{j=1}^{3}(\sigma^{a,s=0}_{i\tau_2^j})^2}} \\
&= \frac{\sigma^{a,s=1}_{i\tau_2^1}}{\sigma^{a,s=0}_{i\tau_2^1}}.
\end{aligned} \tag{C.8}$$

Then, we can show that:

$$\begin{aligned}
\kappa = corr(\epsilon^{a,s=1}_{i\tau_2}, \epsilon^{a,s=0}_{i\tau_2}) &= \frac{cov(\epsilon^{a,s=1}_{i\tau_2}, \epsilon^{a,s=0}_{i\tau_2})}{\sqrt{var(\epsilon^{a,s=1}_{i\tau_2})var(\epsilon^{a,s=0}_{i\tau_2})}} \\
&= \frac{\sum_{j=1}^{3} cov(\epsilon^{a,s=1}_{i\tau_2^j}, \epsilon^{a,s=0}_{i\tau_2^j})}{\sigma^{a,s=1}_{i\tau_2}\sigma^{a,s=0}_{i\tau_2}} \\
&= \frac{\sum_{j=1}^{3} \kappa_j \sigma^{a,s=1}_{i\tau_2^j}\sigma^{a,s=0}_{i\tau_2^j}}{\sigma^{a,s=1}_{i\tau_2}\sigma^{a,s=0}_{i\tau_2}} \\
&= \kappa_1 \frac{\sum_{j=1}^{3}(\sigma^{a,s=1}_{i\tau_2^j}/\sigma^{a,s=0}_{i\tau_2^j})(\sigma^{a,s=0}_{i\tau_2^j})^2}{(\sigma^{a,s=1}_{i\tau_2}/\sigma^{a,s=0}_{i\tau_2})(\sigma^{a,s=0}_{i\tau_2})^2} \\
&= \kappa_1 \frac{(\sigma^{a,s=1}_{i\tau_2^1}/\sigma^{a,s=0}_{i\tau_2^1})\sum_{j=1}^{3}(\sigma^{a,s=0}_{i\tau_2^j})^2}{(\sigma^{a,s=1}_{i\tau_2}/\sigma^{a,s=0}_{i\tau_2})(\sigma^{a,s=0}_{i\tau_2})^2} \\
&= \kappa_1
\end{aligned} \tag{C.9}$$

C.2.3 Perceived Option Values

Analogous to Equation (3.15), substituting the expressions for $V_{it^*}^s$ (shown in Equation 3.5) and σ_i^P (shown in Equation C.4 with ρ, ρ^0, and κ replaced by ρ_P, ρ_P^0, and κ_P, respectively) into Equation (3.2), we obtain:

$$P_i^{P,s=0} = \Phi\left(\frac{\tilde{\mu}_{it_0}^{Y,s=0} - \tilde{\mu}_{it_0}^{Y,s=1} + \tilde{\gamma}_i}{\sqrt{(\rho_P \tilde{\sigma}_{it_0}^{Y,s=1})^2 + (\rho_P^0 \tilde{\sigma}_{it_0}^{Y,s=0})^2 - 2\kappa_P \rho_P \rho_P^0 \tilde{\sigma}_{it_0}^{Y,s=1} \tilde{\sigma}_{it_0}^{Y,s=0}}}\right). \tag{C.10}$$

Parallel to Section 3.4.2, here we rewrite Equation (C.10) as a linear equation and explicitly allow for measurement error in expectations variables.

$$\Phi^{-1}(P_i^{P,s=0})\tilde{\sigma}_{it_0}^Y = \frac{\bar{\gamma}}{\rho_P} + [\tilde{\mu}_{it_0}^{Y,s=0} - \tilde{\mu}_{it_0}^{Y,s=1}]\frac{1}{\rho_P} + \frac{\tilde{\gamma}_i - \bar{\gamma}}{\rho_P} - \frac{\Delta\mu_i^Y}{\rho_P} + \Delta y_i, \tag{C.11}$$

where $\tilde{\sigma}_{it_0}^Y \equiv \sqrt{(\tilde{\sigma}_{it_0}^{Y,s=1})^2 + (\theta_P \tilde{\sigma}_{it_0}^{Y,s=0})^2 - 2\kappa_P \theta_P \tilde{\sigma}_{it_0}^{Y,s=1} \tilde{\sigma}_{it_0}^{Y,s=0}}$, and $\theta_P \equiv \frac{\rho_P^0}{\rho_P}$. Similarly, we assume that the observed measure of $\tilde{\mu}_{it_0}^{Y,s=0} - \tilde{\mu}_{it_0}^{Y,s=1}$ contains classical measurement error $\Delta\mu_i^Y$ and that the computed value of $\Phi^{-1}(P_i^{P,s=0})\tilde{\sigma}_{it_0}^{Y,}$ contains classical measurement error Δy_i.

To apply the measurement-error-robust approach detailed in Section 3.4.2 and Appendix C.4, we need to compute $\tilde{\sigma}_{it_0}^Y$ for each student. Note that $\tilde{\sigma}_{it_0}^{Y,s=0}$ and $\tilde{\sigma}_{it_0}^{Y,s=1}$ can be directly computed from the data. We impose the assumption that the perceived values of the ratio of signal strength θ_P and the correlation κ are equal to their actual counterparts, which have been estimated in Section C.2.2 ($\theta_A \equiv \frac{\rho_A^0}{\rho_A} = 0.55$ and $\kappa_A = 0.63$).

With $\tilde{\mu}_{it_0}^{Y,s=0}$ and $\tilde{\mu}_{it_0}^{Y,s=1}$ directly constructed from the data and $\Phi^{-1}(P_i^{P,s=0})\tilde{\sigma}_{it_0}^Y$ computed as above, we consistently estimate $\frac{\bar{\gamma}}{\rho_P}$ and $\frac{1}{\rho_P}$ using the approach described in Section 3.4.2. The estimates of ρ_P and ρ_P^0 are 0.55 and 0.29, respectively. Comparing $\rho_P = 0.55$ and $\rho_P^0 = 0.29$ to $\rho_A = 0.51$ and $\rho_A^0 = 0.28$ (Section C.2.2), we continue to find, as in Section 3.4, that students have quite accurate perceptions about the magnitude of uncertainty resolution. Then, as expected, the perceived value of σ_i is equal to \$103,140, which is very close to its actual counterpart (\$96,780). The resulting average perceived OV and average perceived NCV are \$7,680 and \$155,590, respectively, which are almost identical to the average values computed in Section 3.4. Comparing these numbers to the actual analogs found in Section C.2.2, (\$21,020 and \$63,720), our main conclusion that students underestimate the option value and overestimate the net continuation value remains appropriate in this slightly modified learning environment.

C.3 Robustness: Allowing for Learning about $\gamma_i^{s=1}$

In this appendix, we examine the implications of allowing students to also obtain relevant information about the non-pecuniary benefits associated with the graduation scenario, $\gamma_i^{s=1}$. In particular, we show that, under assumptions that are broadly consistent with the setting in Stinebrickner and Stinebrickner (2012, 2014b), our estimates of actual option values in Section 3.4.3 tend to be downward biased while our estimates of perceived option values in Section 3.4.3 remain consistent.

Recall that Section 3.3 shows that the option value is multiplicatively separable in the dropout probability $P_i^{s=0}$ and the amount of uncertainty resolved in college σ_i. Since both actual and perceived values of $P_i^{s=0}$ are obtained from the data in somewhat direct ways, we only need to examine whether our estimates of the actual and perceived σ_i tend to be consistent when students are also learning about $\gamma_i^{s=1}$. For the purpose of clarity, here we denote the estimates of actual and perceived ρ computed in Section 3.4.2 as $\underline{\rho}_A$ and $\underline{\rho}_P$, respectively, and denote the estimates of actual and perceived σ_i computed in Section 3.4.2 as $\underline{\sigma}_i^A$ and $\underline{\sigma}_i^P$, respectively.

The relevant new information $\Delta_i \sim N(0, \sigma_i^2)$ is given by:

$$
\begin{aligned}
\Delta_i &= (V_{it^*}^{s=1} - V_{it^*}^{s=0}) - E_{t=t_0}(V_{it^*}^{s=1} - V_{it^*}^{s=0}) \\
&= [E_{t=t^*}(Y_i^{s=1}) - E_{t=t_0}(Y_i^{s=1})] + [E_{t=t^*}(\gamma_i^{s=1}) - E_{t=t_0}(\gamma_i^{s=1})] \\
&\equiv \Delta Y_i^{s=1} + \Delta\gamma_i^{s=1},
\end{aligned}
\tag{C.12}
$$

where $\text{std}(\Delta Y_i^{s=1}) = \rho\tilde{\sigma}_{it_0}^{Y,s=1}$ (Equation 3.11).

Motivated by Stinebrickner and Stinebrickner (2012, 2014b), we consider a case where, between t_0 and t^*, students resolve uncertainty about $Y_i^{s=1}$ and $\gamma_i^{s=1}$ through a common signal s_i. For example, in their setting, grade performance is a signal that is found to influence both beliefs about earnings and the non-pecuniary benefits of school. In this case, both $\Delta Y_i^{s=1}$ and $\Delta\gamma_i^{s=1}$ are functions of s_i. Under a linearity assumption for the two functions, we have that $\Delta\gamma_i^{s=1}$ is proportional to $\Delta Y_i^{s=1}$, i.e. $\Delta\gamma_i^{s=1} = \alpha(\Delta Y_i^{s=1})$.[1] It implies that:

$$
\Delta_i = (1+\alpha)\Delta Y_i^{s=1} \text{ and } \sigma_i = (1+\alpha)\rho\tilde{\sigma}_{it_0}^{Y,s=1}. \tag{C.13}
$$

We first examine the consistency of our estimates of actual option values in Section 3.4.3. Recall from Section 3.4.2 that the actual fraction ρ_A is estimated using observed data on $\tilde{\sigma}_{it_0}^{Y,s=1}$ and $\tilde{\sigma}_{it^*}^{Y,s=1}$. Therefore, our estimates of ρ_A and $\rho_A\tilde{\sigma}_{it_0}^{Y,s=1}$ are consistent regardless of whether

[1] Both $\Delta Y_i^{s=1}$ and $\Delta\gamma_i^{s=1}$ have a mean of zero, by construction.

students are also resolving uncertainty about non-pecuniary benefits $\gamma_i^{s=1}$, i.e. $\underline{\sigma}_i^A$ consistently estimates $\rho\tilde{\sigma}_{it_0}^{Y,s=1}$. In the likely case where $\alpha > 0$, the actual value of σ_i would be greater than $\rho_A\tilde{\sigma}_{it_0}^{Y,s=1}$.[2] Thus, $\underline{\sigma}_i^A$ underestimates the actual value of σ_i, which implies that, for each student, our estimate of actual option value reported in Section 3.4.3 underestimates its true value.

We then examine the consistency of our estimates of perceived option values in Section 3.4.3. Allowing students to learn about non-pecuniary benefits associated with the graduation scenario leads to a modification of Equation (3.15).

$$\begin{aligned} P_i^{P,s=0} &= \Phi(\frac{E_{t=t_0}[(Y_i^{s=0} + \gamma_i^{s=0}) - (Y_i^{s=1} + \gamma_i^{s=1})]}{(1+\alpha)\rho_P\tilde{\sigma}_{it_0}^{Y,s=1}}) \\ &= \Phi(\frac{E_{t=t_0}(Y_i^{s=0}) - E_{t=t_0}(Y_i^{s=1}) + E_{t=t_0}(\gamma_i^{s=0} - \gamma_i^{s=1})}{(1+\alpha)\rho_P\tilde{\sigma}_{it_0}^{Y,s=1}}) \\ &\equiv \Phi(\frac{\tilde{\mu}_{it_0}^{Y,s=0} - \tilde{\mu}_{it_0}^{Y,s=1} + \tilde{\gamma}_i}{(1+\alpha)\rho_P\tilde{\sigma}_{it_0}^{Y,s=1}}). \end{aligned} \tag{C.14}$$

Consequently, the main estimation equation (Equation 3.17) can be modified as follows:

$$\Phi^{-1}(P_i^{P,s=0})\tilde{\sigma}_{it_0}^{Y,s=1} + \Delta y_i = \frac{\bar{\gamma}}{(1+\alpha)\rho_P} + [\tilde{\mu}_{it_0}^{Y,s=0} - \tilde{\mu}_{it_0}^{Y,s=1} + \Delta\mu_i^Y]\frac{1}{(1+\alpha)\rho_P} + \frac{\tilde{\gamma}_i - \bar{\gamma}}{(1+\alpha)\rho_P}. \tag{C.15}$$

The only difference between Equation (3.17) and Equation (C.15) is that $(1+\alpha)\rho_P$ shows up in Equation (C.15) at places where ρ_P shows up in Equation (3.17). Therefore, $\underline{\rho}_P$ consistently estimates $(1+\alpha)\rho_P$, which implies that $\underline{\sigma}_i^P$ consistently estimates σ_i as well. Hence, for each student, the estimate of perceived option value reported in Section 3.4.3 consistently estimates its true value.

C.4 Measurement Error Correction

C.4.1 Estimating the Variance of $\Delta\mu_i^Y$

Appendix C.1 describes how to obtain measures of $\tilde{\mu}_{it_0}^{Y,s=1}$ and $\tilde{\mu}_{it_0}^{Y,s=0}$ using our measures of $\tilde{\mu}_{it_0}^{a,s=1}$ and $\tilde{\mu}_{it_0}^{a,s=0}$. Let $\Delta\mu_i^{a,s}$ denote the measurement error that is present in our measure of $\tilde{\mu}_{it_0}^{a,s}$. Equation (C.1) implies that $var(\Delta\mu_i^Y)$ is given by:

$$var(\Delta\mu_i^Y) = var(\sum_{a=\bar{t}}^{\bar{A}}\beta^{a-t^*}\Delta\mu_i^{a,s=1} - \sum_{a=t^*}^{\bar{A}}\beta^{a-t^*}\Delta\mu_i^{a,s=0}). \tag{C.16}$$

[2]This is consistent with a scenario where the common factor is grade performance; Having a high realized grade would tend to positively influence a student's perceptions about both the pecuniary and non-pecuniary benefits associated with the graduation scenario.

Recall that the unconditional earnings expectations questions in the BPS were asked for three specific ages a: the first year after graduation (age 23), age 28, and age 38, and for both schooling scenarios: graduation ($s = 1$) and dropout ($s = 0$). The linear interpolation assumption we employed to impute $\tilde{\mu}_{it_0}^{a,s}$ for other ages implies that $\Delta\mu_i^{a,s}$ is a linear combination of a subset of $\{\Delta\mu_i^{23,s}, \Delta\mu_i^{28,s}, \Delta\mu_i^{38,s}\}$ for all a.

We further assume that (1) the distribution of measurement error is the same for each of the six unconditional earnings expectations questions; (2) measurement errors are uncorrelated across schooling scenarios s, but are perfectly correlated within schooling scenarios.[3] Under these assumptions, we have:

$$
\begin{aligned}
var(\Delta\mu_i^Y) &= var(\sum_{a=\bar{t}}^{\bar{A}} \beta^{a-t^*} \Delta\mu_i^{a,s=1} - \sum_{a=t^*}^{\bar{A}} \beta^{a-t^*} \Delta\mu_i^{a,s=0}) \\
&= var(\sum_{a=\bar{t}}^{\bar{A}} \beta^{a-t^*} \Delta\mu_i^{28,s=1}) + var(\sum_{a=t^*}^{\bar{A}} \beta^{a-t^*} \Delta\mu_i^{28,s=0}) \\
&= var(\Delta\mu_i^{28,s=1})[(\sum_{a=\bar{t}}^{\bar{A}} \beta^{a-t^*})^2 + (\sum_{a=t^*}^{\bar{A}} \beta^{a-t^*})^2]. \qquad \text{(C.17)}
\end{aligned}
$$

Following the method developed in Gong, Stinebrickner and Stinebrickner (2019), we estimate the variance of the measurement error contained in students' reported value of $\tilde{\mu}_{it_0}^{28,s=1} \equiv E_{t=t_0}(w_i^{28,s=1})$ for the 2001 cohort, $\Delta\mu_i^{28,s=1}$. The approach takes advantage of the fact that the BPS includes two separate sets of expectations questions that can be used to compute $\tilde{\mu}_{it_0}^{28,s=1}$. The difference between the two computed values of $\tilde{\mu}_{it_0}^{28,s=1}$ provides evidence about the magnitude of measurement error. The estimate of $var(\Delta\mu_i^{28,s=1})$ is 109.54 (earnings measured in \$1,000 units). Using Equation (C.17), we estimate that $var(\Delta\mu_i^Y)$ is 67236 (earnings measured in \$1,000 units).

C.4.2 ME Correction Formula

Let vector $\boldsymbol{z}_i$ denote the independent variables that are accurately measured and x_i denote the independent variable that is measured with classical measurement error η_i. We allow the variance of η_i to depend on observable $\boldsymbol{g}_i$ and denote this variance $\sigma_{ME}^2(\boldsymbol{g}_i)$. Let $\tilde{x}_i = x_i + \eta_i$ denote the measured value of x_i. Then, the dependent variable y_i is given by:

[3] Assumption (2) captures the notion that factors that affect students' beliefs about earnings under the college alternative ($s = 1$) are likely different from those affecting students' beliefs about earnings under the non-college alternative ($s = 0$).

$$\begin{aligned} y_i &= \boldsymbol{z}_i'\boldsymbol{a} + bx_i + \epsilon \\ &= \boldsymbol{z}_i'\boldsymbol{a} + b\tilde{x}_i + (\epsilon - b\eta_i). \end{aligned} \tag{C.18}$$

By construction, $\tilde{x}$ and $\epsilon - b\eta_i$ are correlated. Hence, the OLS estimator is biased. To correct for this bias, we notice that:

$$E\left[(y_i - (\boldsymbol{z}_i'\boldsymbol{a} + b\tilde{x}_i))\begin{pmatrix} \boldsymbol{z}_i \\ \tilde{x}_i \end{pmatrix} + \begin{pmatrix} \boldsymbol{0} \\ b\sigma^2_{ME}(\boldsymbol{g}_i) \end{pmatrix}\right] = E\left[(\epsilon - b\eta_i)\begin{pmatrix} \boldsymbol{z}_i \\ \tilde{x}_i \end{pmatrix} + \begin{pmatrix} \boldsymbol{0} \\ b\sigma^2_{ME}(\boldsymbol{g}_i) \end{pmatrix}\right] = \boldsymbol{0}. \tag{C.19}$$

Equation system (C.19) has the same number of equations and parameters which are equal to the number of observables. Hence, it can be estimated using the Method of Moments, i.e., the estimator of $\begin{pmatrix} \boldsymbol{a} \\ b \end{pmatrix}$ is the solution to the sample analog of the moment conditions defined by Equation (C.19). It is easy to show that this estimator has an easy-to-implement matrix-form expression. Letting $\boldsymbol{c}$ denote $\begin{pmatrix} \boldsymbol{a} \\ b \end{pmatrix}$ and $\boldsymbol{q}_i$ denote $\begin{pmatrix} \boldsymbol{z}_i \\ \tilde{x}_i \end{pmatrix}$, we have:

$$\hat{\boldsymbol{c}} = \left[Q'Q - \begin{pmatrix} \boldsymbol{0} & 0 \\ 0 & \sum_i \sigma^2_{ME}(\boldsymbol{g}_i) \end{pmatrix}\right]^{-1} Q'Y, \tag{C.20}$$

where and Y and Q are the matrices of y_i and $\boldsymbol{q}_i$, respectively.

Appendix D

Appendices for Chapter 4

D.1 Proof of Lemma 4.3.2

Lemma 4.3.2 *For any $(x, z) \in supp((X_{i0}, Z_i)) \cap supp((X_{i1}, Z_i))$, if Assumption 4.3.1-4.3.7 are satisfied, then $\tilde{D}_1(x, z, s)$ is continuous and invertible in s and $\tilde{D}_1^{-1}(x, z, d)$ is continuous in d.*

Proof Fix $(x, z) \in \text{supp}((X_{i0}, Z_i)) \cap \text{supp}((X_{i1}, Z_i))$. Following Equation (4.6), the conditional mean function $\tilde{D}_1(x, z, s)$ is given by:

$$\tilde{D}_1(x, z, s) = E_{\xi_i}\bar{D}(x, B(z) + \Gamma_0(s - B(z); z, \xi_i)). \tag{D.1}$$

Assumption 4.3.5-4.3.7 imply that, for any ξ_i, $\bar{D}(x, B(z) + \Gamma_0(s - B(z); z, \xi_i))$ is a continuous function of s and is bounded on $\text{supp}(S_{i0}|(x, z))$. Then, by the theorem of Continuity under Integral Sign, $\tilde{D}_1(x, z, s) = E_{\xi_i}\bar{D}(x, B(z) + \Gamma_0(s - B(z); z, \xi_i))$ is also a continuous function of s defined on $\text{supp}(S_{i0}|(x, z))$.

As shown in the proof of Theorem 4.3.1, $\tilde{D}_1(x, z, s)$ is invertible in s, under Assumption 4.3.2 and 4.3.3. Using the well known result that the inverse of a continuous function that maps an interval to the real line is also continuous, the continuity of $\tilde{D}_1(x, z, s)$ implies the continuity of $\tilde{D}_1^{-1}(x, z, d)$, under Assumption 4.3.7.

D.2 Examples

D.2.1 Effort Choice and Beliefs about Productivity

Example Consider a worker who is working for piece rate w for each unit of output y_{it}. Her production function is given by:

$$y_{it} = s_{it}d_{it}, \tag{D.2}$$

where s_{it} and d_{it} are worker i's productivity and effort choice at period t, respectively.

Assume that productivity s_{it} is the sum of a permanent factor A_i and a mean zero, independently distributed shock v_{it}. Let μ_{it} denote worker i's subjective mean of A_i at period t. Between each two consecutive periods, worker i observes the realization of $s_{it} = \frac{y_{it}}{d_{it}}$ and update subjective mean μ_{it}.

Worker i's utility is given by the difference between her income $wd_{it}s_{it}$ and her (psychological) cost $C(d_{it}) = \alpha_1 d_{it} + \alpha_2 d_{it}^2$. At each period t, given piece rate w and subjective expectation $\mu_{it} = E_t^b A_i = E^b s_{it}$, worker i optimally chooses effort level d_{it} to maximize her expected utility. Formally, the maximization problem for worker i is given by:

$$\max_{d_{it}} wd_i\mu_{it} - \alpha_1 d_i - \alpha_2 d_i^2. \tag{D.3}$$

I assume that when solving the policy function of maximization problem (D.3), worker i makes a mean zero optimization error ϵ_{it} which is independent from any other factors. This implies that:[1]

$$d_{it} = -\frac{\alpha_1}{2\alpha_2} + \frac{w}{2\alpha_2}\mu_{it} + \epsilon_{it}. \tag{D.4}$$

It is easy to verify that the average decision $E(d_{it}) = -\frac{\alpha_1}{2\alpha_2} + \frac{w}{2\alpha_2}\mu_{it}$ is a time-invariant, strictly monotonic and continuous function in μ_{it}, and is bounded on any finite interval in $\mathbb{R}$.

D.2.2 Rural-urban Migration and Beliefs about Earnings

Example At the beginning of each period t, individual i who lives in the rural area needs to decide whether to work in the urban area for that period. Let ϵ_{ijt} and w_{ijt} denote individual i's cost and earnings from working in area $j \in \{R, U\}$. I assume that w_{iUt} is a linear function of factors x_{iUt}, shocks v_{iUt}, and individual i's skill type A_i, while w_{iRt} only depends on factors x_{iRt} and shocks v_{iRt}:

$$\begin{aligned} w_{iUt} &= x'_{iUt}\alpha_U + A_i + v_{iUt}, \\ w_{iRt} &= x'_{iRt}\alpha_R + v_{iRt}, \end{aligned} \tag{D.5}$$

where v_{iUt} and v_{iRt} have mean zero and are independently distributed.

Individual i observes ϵ_{ijt} and x_{ijt}, but is uncertain about A_i and v_{ijt}. Let μ_{it} denote customer i's subjective mean of A_i at period t. At the end of each period t, an individual who chose to

[1]For the purpose of illustration, I assume interior solutions.

work in the urban area observes her earnings w_{iUt} and uses the effective signal $s_{it} = w_{iUt} - x'_{iUt}\alpha_U$ to update her subjective mean μ_{it}. Note that although s_{it} is not directly observed by the econometrician, it can be consistently estimated using data on s_{it} and w_{iUt}.

I assume that individuals are myopic and maximize the difference between their expected current period earnings and cost when choosing where to work. The probability of working in the urban area is given by:

$$\begin{aligned}\text{Prob}(d_{it} = U) &= \text{Prob}(x'_{iUt}\alpha_U + \mu_{it} - \epsilon_{iUt} > x'_{iRt}\alpha_R - \epsilon_{iRt}) \\ &= 1 - F_{\epsilon_{iRt}-\epsilon_{iUt}}(x'_{iUt}\alpha_U - x'_{iRt}\alpha_R + \mu_{it}),\end{aligned} \tag{D.6}$$

where $F_{\epsilon_{iRt}-\epsilon_{iUt}}(x)$ is the CDF of $\epsilon_{iRt} - \epsilon_{iUt}$.

Under the assumption that $F_{\epsilon_{iRt}-\epsilon_{iUt}}(x)$ is time-invariant, strictly increasing, and continuous, $\text{Prob}(d_{it} = U)$ is a time-invariant, strictly increasing, and continuous function in μ_{it} and is bounded on any finite interval in $\mathbb{R}$.

D.3 The Case Where $D(x_{it}, \epsilon_{it}, \mu_{it})$ and $A(z_i)$ are Arbitrary Parametric Functions

In this appendix, I show that the feasible semiparametric estimator developed in Section 4.4.1 can be applied to the case where $D(x_{it}, \epsilon_{it}, \mu_{it})$ and $A(z_i)$ are arbitrary parametric functions. I also impose parametric assumptions on the distribution of ϵ_{it}.

Let $\tilde{D}(x_{i0}, (z_i, k_i))$ denote $E_{\epsilon_{i0}} D(x_{i0}, \epsilon_{i0}, \mu_{i0}) = E_{\epsilon_{i0}} D(x_{i0}, \epsilon_{i0}, A(z_i) + \tilde{\pi}_B(k_i))$. Since $D(x_{it}, \epsilon_{it}, \mu_{it})$, $A(z_i)$, and the distribution of ϵ_{i0} are fully parametric, $\tilde{D}(x_{i0}, (z_i, k_i))$ is also a parametric function of x_{i0}, z_i, and k_i, hence can be estimated parametrically. Let $\hat{D}(x_{i0}, (z_i, k_i))$ denote a consistent estimator of $\tilde{D}(x_{i0}, (z_i, k_i))$.

I maintain the assumption on s_{i0} that $s_{i0} = A(z_i) + \eta_i + v_{i0}$, where η_i and v_{i0} are independent from other factors. $A(z_i)$ can be estimated parametrically. Let $\hat{A}(z_i)$ denote a consistent estimator of $A(z_i)$.

Let $\tilde{d}_{i1} \equiv d_{i1} - \hat{D}_0(x_{i1}, (z_i, k_i))$ and $\tilde{s}_{i0} \equiv s_{i0} - \hat{A}(z_i)$. Then,

$$
\begin{aligned}
\tilde{d}_{i1} &= D(x_{i1}, \epsilon_{i1}, \mu_{i1}) - \tilde{D}_0(x_{i1}, (z_i, k_i)) + [\tilde{D}_0(x_{i1}, (z_i, k_i)) - \hat{D}_0(x_{i1}, (z_i, k_i))] \\
&= D(x_{i1}, \epsilon_{i1}, A(z_i) + \tilde{\pi}_B(k_i) + \Gamma_{i0}(s_{i0} - A(z_i) - \tilde{\pi}_B(k_i))) - \tilde{D}_0(x_{i1}, (z_i, k_i)) + o_p(1) \\
&= D(x_{i1}, \epsilon_{i1}, A(z_i) + \tilde{\pi}_B(k_i) + \Gamma_{i0}(\tilde{s}_{i0} - \tilde{\pi}_B(k_i))) - \tilde{D}_0(x_{i1}, (z_i, k_i)) + o_p(1) \\
&\quad + [D(x_{i1}, \epsilon_{i1}, A(z_i) + \tilde{\pi}_B(k_i) + \Gamma_{i0}(s_{i0} - A(z_i) - \tilde{\pi}_B(k_i))) - D(x_{i1}, \epsilon_{i1}, A(z_i) + \tilde{\pi}_B(k_i) + \Gamma_{i0}(\tilde{s}_{i0} - \tilde{\pi}_B(k_i)))] \\
&= D(x_{i1}, \epsilon_{i1}, A(z_i) + \tilde{\pi}_B(k_i) + \Gamma_{i0}(\tilde{s}_{i0} - \tilde{\pi}_B(k_i))) - \tilde{D}_0(x_{i1}, (z_i, k_i)) + o_p(1), \qquad \text{(D.7)}
\end{aligned}
$$

where the $o_p(1)$ term represents the estimation error in $\hat{D}(x_{i0}, (z_i, k_i))$ and $\hat{A}(z_i)$, which converges in probability to zero as the number of observations goes to infinity.

Since k_i is a finite type variable, it is feasible to stratify individuals by k_i. For each type k_i, $\tilde{\pi}_B(k_i)$ is a constant and $\tilde{d}_{i1}$ only depends on one "observable" $\tilde{s}_{i0}$ (and the unobserved factor ϵ_{i1}).

To establish the consistency of the feasible semiparametric estimator proposed in Section 4.4.1, it suffices to show that, for each k_i, the probability limit of $E(\tilde{d}_{i1}|\tilde{s}_{i0}, k_i)$ is a continuous and strictly monotonic function of $\tilde{s}_{i0}$ and equals zero when $\tilde{s}_{i0} = \tilde{\pi}_B(k_i)$.

Recall that $\Gamma_{i0}(s) = \Gamma_0(s; z_i, \xi_i)$. Let $\tilde{D}_1(x_{i1}, (z_i, k_i), \tilde{s}_{i0}) \equiv E_{(\epsilon_{i1}, \xi_i)|x_{i1}, z_i, s_{i0}} D(x_{i1}, \epsilon_{i1}, A(z_i) + \tilde{\pi}_B(k_i) + \Gamma_0(\tilde{s}_{i0} - \tilde{\pi}_B(k_i); z_i, \xi_i))$. Under Assumption 4.3.1-4.3.7, I can show that $\tilde{D}_1(x_{i1}, (z_i, k_i), \tilde{s}_{i0}) - \tilde{D}_0(x_{i1}, (z_i, k_i))$ is a continuous and strictly monotonic function of $\tilde{s}_{i0}$ and equals zero when $\tilde{s}_{i0} = \tilde{\pi}_B(k_i)$, following the same arguments as in Section 4.3.2.

Equation (D.7) implies that:

$$
\begin{aligned}
E(\tilde{d}_{i1}|x_{i1}, (z_i, k_i), \tilde{s}_{i0}) &= \tilde{D}_1(x_{i1}, (z_i, k_i), \tilde{s}_{i0}) - \tilde{D}_0(x_{i1}, (z_i, k_i)) + o_p(1) \\
&\equiv \Delta\tilde{D}_1(x_{i1}, (z_i, k_i), \tilde{s}_{i0}) + o_p(1). \qquad \text{(D.8)}
\end{aligned}
$$

For each k_i, $E(\tilde{d}_{i1}|\tilde{s}_{i0}, k_i)$ can be obtained through integrating the function $E(\tilde{d}_{i1}|x_{i1}, (z_i, k_i), \tilde{s}_{i0})$ over the distribution of (X_{i1}, Z_i). The probability limit of $E(\tilde{d}_{i1}|\tilde{s}_{i0}, k_i)$ is given by:

$$
\begin{aligned}
plimE(\tilde{d}_{i1}|\tilde{s}_{i0}, k_i) &= plim(E_{(X_{i1}, Z_i)} \Delta\tilde{D}_1(x_{i1}, (z_i, k_i), \tilde{s}_{i0}) + o_p(1)) \\
&= E_{(X_{i1}, Z_i)} \Delta\tilde{D}_1(x_{i1}, (z_i, k_i), \tilde{s}_{i0}). \qquad \text{(D.9)}
\end{aligned}
$$

Since $\Delta\tilde{D}_1(x_{i1}, (z_i, k_i), \tilde{\pi}_B(k_i)) = 0$ for all (x_{i1}, z_i), $E_{(X_{i1}, Z_i)} \Delta\tilde{D}_1(x_{i1}, (z_i, k_i), \tilde{\pi}_B(k_i)) = 0$ as well. The strict monotonicity of $\Delta\tilde{D}_1(x_{i1}, (z_i, k_i), \tilde{s}_{i0})$ is preserved by integration. By the theorem of Continuity under Integral Sign, the continuity of $\Delta\tilde{D}_1(x_{i1}, (z_i, k_i), \tilde{s}_{i0})$ is preserved if this function is bounded on supp(X_{i1}, Z_i) for all $\tilde{s}_{i0}$. Hence, under this boundedness assumption,

$plimE(\tilde{d}_{i1}|\tilde{s}_{i0}, k_i)$ is a continuous and strictly monotonic function of $\tilde{s}_{i0}$, and equals zero when $\tilde{s}_{i0} = \tilde{\pi}_B(k_i)$.

D.4 Theorem 4.6.1: Proof

Let $V_1(s_{i0})$ denote individual i's expectation about $\max_{j\in\{0,1\}}(u_{ij1})$ before observing the realizations of ϵ_{i11} and ϵ_{i01}. It is given by:

$$\begin{aligned} V_1(s_{i0}) &= E_{\epsilon_{i11},\epsilon_{i01}} \max_{j\in\{0,1\}}(u_{ij1}) \\ &= E_{\epsilon_{i11},\epsilon_{i01}} \max(\alpha_1 + \beta_1(\bar{\mu} + \Gamma_0(s_{i0} - \bar{\mu}) + \epsilon_{i11}, \epsilon_{i01}) \\ &= E_{\epsilon_{i11}-\epsilon_{i01}} \max(\alpha_1 + \beta_1(\bar{\mu} + \Gamma_0(s_{i0} - \bar{\mu}) + \epsilon_{i11} - \epsilon_{i01}, 0) \\ &= E_{\epsilon_{i11}-\epsilon_{i01}} \max(F^{-1}_{\Delta\epsilon}(\tilde{D}_1(s_{i0})) + \epsilon_{i11} - \epsilon_{i01}, 0). \end{aligned} \tag{D.10}$$

Note that $F^{-1}_{\Delta\epsilon}(d)$, $\tilde{D}_1(s)$, and the distribution of $\epsilon_{i11} - \epsilon_{i01}$ are all known by the econometrician. Hence, $V_1(s)$ is a known function as well. Now consider individual i's (subjective) expectation about $V_1(s_{i0})$ before receiving the signal s_{i0}, which is given by $E_{S^b_{i0}} V_1(s_{i0}) = E_{S_{i0}+\bar{\mu}-\bar{s}} V_1(s_{i0})$ (Assumption 4.6.6). Both $V_1(s)$ and the distribution of S_{i0} are known by the econometrician. Hence, $E_{S_{i0}+\bar{\mu}-\bar{s}} V_1(s_{i0})$ is a known function of the individual's prior mean $\bar{\mu}$. I denote this function as $\bar{V}_1(\bar{\mu})$. Below I show that if $\tilde{D}_1(s)$ is strictly increasing (decreasing) in s, then $\bar{V}_1(\bar{\mu})$ is strictly increasing (decreasing) in $\bar{\mu}$.

Without loss of generality, consider the case where $\tilde{D}_1(s)$ is strictly increasing in s. I first show $V_1(s)$ is strictly increasing in s. Fix $s_1 > s_2$. Since both $F^{-1}_{\Delta\epsilon}(d)$ and $\tilde{D}_1(s)$ are strictly increasing, $\max(F^{-1}_{\Delta\epsilon}(\tilde{D}_1(s_1)) + \epsilon_{i11} - \epsilon_{i01}, 0) \geq \max(F^{-1}_{\Delta\epsilon}(\tilde{D}_1(s_2)) + \epsilon_{i11} - \epsilon_{i01}, 0)$ with the inequality being strict when $\epsilon_{i11} - \epsilon_{i01} > -F^{-1}_{\Delta\epsilon}(\tilde{D}_1(s_2))$. Assumption 4.6.2 implies that $\epsilon_{i11} - \epsilon_{i01} > -F^{-1}_{\Delta\epsilon}(\tilde{D}_1(s_2))$ takes place with positive probability. Therefore, $V_1(s_1) \equiv E_{\epsilon_{i11}-\epsilon_{i01}} \max(F^{-1}_{\Delta\epsilon}(\tilde{D}_1(s_1)) + \epsilon_{i11} - \epsilon_{i01}, 0) > E_{\epsilon_{i11}-\epsilon_{i01}} \max(F^{-1}_{\Delta\epsilon}(\tilde{D}_1(s_2)) + \epsilon_{i11} - \epsilon_{i01}, 0) \equiv V_1(s_2)$.

Fix $\bar{\mu}_1 > \bar{\mu}_2$. Assumption 4.6.4 implies that S_{i0} has a full support. Hence $S_{i0} + \bar{\mu}_1 - \bar{s}$ is different from and first order stochastically dominates $S_{i0} + \bar{\mu}_2 - \bar{s}$. Since $V_1(s)$ is strictly increasing in s, using the well-known property of first order stochastic dominance, I obtain $\bar{V}_1(\bar{\mu}_1) \equiv E_{S_{i0}+\bar{\mu}_1-\bar{s}} V_1(s_{i0}) > E_{S_{i0}+\bar{\mu}_2-\bar{s}} V_1(s_{i0}) \equiv \bar{V}_1(\bar{\mu}_2)$.

D.5 Partially Dependent X_{it} and Z_i

In this appendix, I show that the method proposed in Section 4.6.2 is valid whenever X_{it} and Z_i are linearly independent. Since X_{it} and Z_i are linearly independent, there are some random variables contained in Z_i that are not linear combinations of X_{it}. Let Z_i^N denote these random variables and Z_i^Y denote the rest of the random variables contained in Z_i. The full model is given by:

$$\begin{aligned} d_{it} &= x_{i0}'\alpha + \beta\mu_{it} + \epsilon_{it}, \\ \mu_{i0} &= (z_i^Y)'\pi_B^Y + (z_i^N)'\pi_B^N + v_i^B, \\ \mu_{i1} &= \mu_{i0} + \theta(s_{i0} - \mu_{i0}), \\ s_{it} &= (z_i^Y)'\pi_R^Y + (z_i^N)'\pi_R^N + v_i^R + \eta_i + v_{it}. \end{aligned} \tag{D.11}$$

I rewrite this linear model to eliminate unobserved subjective means μ_{it}:

$$d_{i0} = x_{i0}'\alpha + (z_i^Y)'(\pi_B^Y\beta) + (z_i^N)'(\pi_B^N\beta) + \beta v_i^B + \epsilon_{i0}, \tag{D.12}$$

$$d_{i1} = x_{i1}'\alpha + (z_i^Y)'[\pi_B^Y\beta(1-\theta)] + (z_i^N)'[\pi_B^N\beta(1-\theta)] + \beta\theta s_{i0} + \beta(1-\theta)v_i^B + \epsilon_{i1}, \tag{D.13}$$

$$s_{it} = (z_i^Y)'\pi_R^Y + (z_i^N)'\pi_R^N + v_i^R + \eta_i + v_{it}. \tag{D.14}$$

The objects of primary interest are $E_{v_i^B}(\mu_{i0}) = (z_i^Y)'\pi_B^Y + (z_i^N)'\pi_B^N$ and β.

Note that (z_i^Y) is a linear combination of x_{i0}. Hence, there exist α_0 and α_1 such that $x_{i0}'\alpha_0 = x_{i0}'\alpha + (z_i^Y)'(\pi_B^Y\beta)$ and $x_{i1}'\alpha_1 = x_{i1}'\alpha + (z_i^Y)'[\pi_B^Y\beta(1-\theta)]$. Equation (D.12) and (D.13) can be rewritten as:

$$d_{i0} = x_{i0}'\alpha_0 + (z_i^N)'(\pi_B^N\beta) + \beta v_i^B + \epsilon_{i0}, \tag{D.15}$$

$$d_{i1} = x_{i1}'\alpha_1 + (z_i^N)'[\pi_B^N\beta(1-\theta)] + \beta\theta s_{i0} + \beta(1-\theta)v_i^B + \epsilon_{i1}, \tag{D.16}$$

Uncorrelated Unobserved Components

Equation (D.15) shows that α_0 and $\pi_B^N\beta$ can be estimated by regressing d_{i0} on x_{i0} and z_i^N. Similarly, if v_i^R and v_i^B are uncorrelated, OLS regression of d_{i1} on x_{i1}, z_i^N, and s_{i0} gives estimators of α_1, $\pi_B^N\beta(1-\theta)$, and $\beta\theta$. Hence, θ can be consistently estimated by $1 - \frac{\widehat{\pi_B^N\beta(1-\theta)}}{\widehat{\pi_B^N\beta}}$, β can be consistently estimated by $\frac{\widehat{\beta\theta}}{\hat{\theta}}$, and π_B^N can be consistently estimated by $\frac{\widehat{\pi_B^N\beta}}{\hat{\beta}}$.

Note that $x_{i0}'\alpha_0 - x_{i0}'\alpha_1 = x_{i0}'\alpha + (z_i^Y)'(\pi_B^Y\beta) - x_{i0}'\alpha - (z_i^Y)'[\pi_B^Y\beta(1-\theta)] = (z_i^Y)'(\pi_B^Y\beta\theta)$. Hence, $(z_i^Y)'\pi_B^Y$ can be consistently estimated by $x_{i0}'\frac{(\hat{\alpha}_0 - \hat{\alpha}_1)}{\widehat{\beta\theta}}$.

Correlated Unobserved Components

If v_i^R and v_i^B are correlated, I can use $\Delta s_{i0} = s_{i0} - s_{i1}$ as an IV to consistently estimate α_1, $\pi_B^N \beta(1-\theta)$, and $\beta\theta$ based on Equation (D.16) and consistently estimate $E_{v_i^B}(\mu_{i0}) = (z_i^Y)'\pi_B^Y + (z_i^N)'\pi_B^N$ and β following the steps above.

Curriculum Vitae

Name:	Yifan Gong
Post-Secondary Education and Degrees:	University of Science and Technology of China Hefei, China 2007 - 2011 B.A.
	City University of Hong Kong Hong Kong SAR, China 2011 - 2012 M.Sc.
	University of Western Ontario London, Ontario, Canada 2013 - 2020 Ph.D.
Honours and Awards:	Department of Economics Scholarship University of Western Ontario 2016
	Graduate Fellowship University of Western Ontario 2013 - 2018
Related Work Experience:	Teaching and Research Assistant The University of Western Ontario 2013 - 2020

Publications:

Uncertainty about Future Income: Initial Beliefs and Resolution during College (joint with Todd Stinebrickner, and Ralph Stinebrickner), Quantitative Economics, 10(2), 607-641, 2019.