

2018

# 2018-5 Uncertainty about Future Income: Initial Beliefs and Resolution During College

Yifan Gong

Todd Stinebrickner

Ralph Stinebrickner

Follow this and additional works at: <https://ir.lib.uwo.ca/economicscibc>



Part of the [Economics Commons](#)

---

## Citation of this paper:

Gong, Yifan, Todd Stinebrickner, Ralph Stinebrickner. "2018-5 Uncertainty about Future Income: Initial Beliefs and Resolution During College." Centre for Human Capital and Productivity. CHCP Working Papers, 2018-5. London, ON: Department of Economics, University of Western Ontario (2018).

**Uncertainty about Future Income: Initial Beliefs  
and Resolution During College**

by

**Yifan Gong, Todd Stinebrickner and Ralph Stinebrickner**

**Working Paper# 2018-5**

**March 2018**



***Centre for Human Capital and Productivity (CHCP)***

***Working Paper Series***

Department of Economics  
Social Science Centre  
Western University  
London, Ontario, N6A 5C2  
Canada

# Uncertainty about Future Income: Initial Beliefs and Resolution During College

Yifan Gong

University of Western Ontario

Todd Stinebrickner

University of Western Ontario

Ralph Stinebrickner

Berea College

March 24, 2018

## **Abstract**

Uncertainty about future income plays a conceptually important role in college decisions. Unfortunately, characterizing how much earnings uncertainty is present for students at college entrance and how quickly this uncertainty is resolved has proven to be difficult. This paper takes advantage of unique expectations data from the Berea Panel Study to provide new evidence about this issue. We characterize initial uncertainty using survey questions that elicit the entire distribution describing one's beliefs about future earnings at an ideal time - immediately before students began their first year courses. We characterize the amount of uncertainty that is resolved during college by taking advantage of the longitudinal nature of the expectations data. Taking advantage of a variety of additional survey questions, we provide evidence about how the resolution of income uncertainty is influenced by factors such as college GPA and college major, and also examine why much income uncertainty remains unresolved at the end of college.

# 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.<sup>1</sup> 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.<sup>2</sup> Unfortunately, taking this first step has proven to be difficult (Cunha, Heckman, Navarro, 2005). This paper takes advantage of unique expectations data from the Berea Panel Study (BPS), which is described in Section 2, to provide new evidence.<sup>3</sup> From the standpoint of characterizing uncertainty, the general benefit of the expectations approach is that survey questions can be designed 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 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

---

<sup>1</sup>More generally, Friedman(1953) suggests the importance of understanding the relative role of labor market uncertainty in determining distributions of wealth.

<sup>2</sup>Throughout the paper our focus is on labor market income, and we use the terms earnings and income interchangeably.

<sup>3</sup>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).

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.<sup>4</sup> 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.<sup>5</sup> 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 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 \$13,700 a year, 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 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.<sup>6</sup> 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

---

<sup>4</sup>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.

<sup>5</sup>See also Blundell and Preston (1998) for early work using similar methods in a somewhat different substantive context.

<sup>6</sup>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.

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.<sup>7</sup> 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 15% and 17% 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.

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 end of the previous paragraph, 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.

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. 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. We do not find compelling evidence that the second explanation is of central importance when we examine hours worked and when we take advantage of unique data related to preferences about types of work. As for the first explanation, we find that uncertainty about the state of the economy at age 28 is not likely to be the whole story. Among various types of frictions that might be present, we find direct evidence suggesting that search frictions may be important.

To some extent, the expectations literature is motivated by the possibility that subjective beliefs do not necessarily correspond to what one might anticipate given distributions

---

<sup>7</sup>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.

of realized outcomes. In the last subsection of Section 4, we examine whether, consistent with what is observed in the data, students believe, at the time of college entrance, that much uncertainty will remain unresolved at the end of school. We take advantage of a survey question eliciting a student’s beliefs about the probability of dropping out of school before graduation. Intuitively, the fact that the average subjective dropout probability is found to be quite low is potentially informative because, under seemingly natural assumptions, this probability will be increasing in the fraction of earnings uncertainty that the student believes will be resolved during college.

Contributing to a recent literature that has recognized the benefits of allowing individuals to express uncertainty about outcomes that will be realized in the future, we develop a simple model in which a student’s subjective dropout probability depends explicitly on the fraction of the earnings uncertainty that will be resolved during college for the types of jobs she would receive with a college degree.<sup>8</sup> For completeness, we also allow the student to resolve earnings uncertainty for the types of jobs she would receive without a college degree. Arcidiacono, Aucejo, Maurel and Ransom (2016) suggest that, in this type of specification, it may be important to allow what a student learns about her earnings under the graduation scenario to be correlated with what a student learns about her earnings under the dropout scenario. Our expectations data provide direct evidence that this type of correlated learning is present. We note that our parsimonious model, in which students resolve uncertainty about only future income, is not particularly well-suited for providing deep insight into how the dropout decision is made. Nonetheless, it can still be useful for characterizing the fraction of uncertainty about post-graduation income that individuals expect to resolve because it seems reasonable to view the estimate of this fraction from our simple model as an upper bound for what one would obtain from a model in which more types of uncertainty were resolved. Estimating this model, we find that, at the time of college entrance, students believe that only approximately 20% of the uncertainty about post-graduation income will be resolved during college.

## 2 The Berea Panel Study

Designed and administered by Todd Stinebrickner and Ralph Stinebrickner, the BPS is a multipurpose longitudinal survey project, which 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. The BPS took place at Berea College. 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 ob-

---

<sup>8</sup>See, for example, Juster, 1966, Manski, 1990, 1999, Blass, Lach, and Manski, 2010, Stinebrickner and Stinebrickner, 2014a and Zafar and Wiswall, 2014.

jective, 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 our work 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 NLSY by Arcidiacono (2004).

The BPS consists of two cohorts. 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. Our primary sample consists of the 650 students who answered this survey.<sup>9</sup> While observable characteristics are not the primary focus of this paper, 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.<sup>10</sup>

Our primary survey questions eliciting beliefs about future earnings are of the form of baseline Survey Question 1A, which is shown in Appendix A.<sup>11</sup> Specifically, Survey Question 1A elicited the minimum, the maximum, and the three quartiles of the subjective income distribution 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). Ques-

---

<sup>9</sup>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.

<sup>10</sup>The BPS is unique in its frequency of contact; each student was surveyed approximately 12 times each year while in school.

<sup>11</sup>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 1B in Appendix A shows the portion of this question related to graduating with a 2.00 GPA.

Table 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 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 paper takes advantage of survey questions eliciting each student's subjective probability of completing a degree in different possible major groups (Question 5, Appendix A), each student's subjective probability of graduating from college (Question 4, Appendix A), and each student's belief about how much noise exists in the grade process (Question 3, Appendix A).

### 3 Uncertainty about Future Income at College Entrance

This section examines uncertainty about future income at the time of college entrance. In Section 3.1, we characterize the amount of uncertainty that exists at college entrance. In Section 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.

#### 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.<sup>12</sup> 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, detailed in Section 3.1.1, 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, detailed in Section 3.1.2, takes advantage of Survey Question 1B (Appendix A), which elicited the minimum, maximum, and 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

---

<sup>12</sup>We focus on the graduation scenario because, as we show in Section 4, this is the outcome that students overwhelmingly believe is most likely. However, when estimating our model of dropout in Section 4, we do take into account uncertainty about earnings under the scenario in which a student does not graduate.

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.

### 3.1.1 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  $C_{it}^1, C_{it}^2, C_{it}^3, C_{it}^4$  and  $C_{it}^5$ , 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  $\sigma = \log(\frac{C_{it}^4}{C_{it}^2})/2\Phi^{-1}(0.75)$  and  $\Phi$  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, 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  $E(W_{it}) = C_{it}^3$  and  $std(W_{it}) = (C_{it}^4 - C_{it}^2)/2\Phi^{-1}(0.75)$ .

c. Stepwise Uniform. The log-normal and normal distributions do not utilize information about the minimum,  $C_{it}^1$ , or the maximum,  $C_{it}^5$ , 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:

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

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 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.<sup>13</sup> Thus, the results are generally quite similar across the three distributional assumptions. The numbers in parentheses in the standard deviation column of Table 2 indicate that there is substantial heterogeneity in uncertainty across students.

Table 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	52.7181 (25.3998)	13.7561 (14.3449)	0.2531 (0.1729)
Approach 2, Normal	50.8079 (24.2972)	12.0751 (9.5888)	0.2402 (0.1432)
Approach 2, Stepwise Uniform	51.2435 (24.2952)	10.4876 (8.0323)	0.2038 (0.1141)

*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  $std(W_{i0})$ .*

### 3.1.2 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

---

<sup>13</sup>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.

is given by:

$$f_{W_{it}}(w_{it}) = \int f_{W_{it}|G_{it}=g_{it}}(w_{it})dF_{G_{it}}(g_{it}), \quad (2)$$

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 paper 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  $E(W_{it}|G_{it})$  with respect to  $G_{it}$ . In cases like this, where an expression 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,

$$E(W_{it}) = E_{G_{it}}(E(W_{it}|G_{it})).^{14} \quad (3)$$

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, 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 distributions in Section 3.1.1. As described in detail in Appendix C.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 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 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

---

<sup>14</sup> $E_{G_{it}}(E(W_{it}|G_{it})) = \int E(W_{it}|G_{it} = g_{it})dF_{G_{it}}(g_{it})$ , with  $E(W_{it}|G_{it} = g_{it}) = \int w_{it}f_{W_{it}|G_{it}=g_{it}}(w_{it})dw_{it}$ .

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.<sup>15</sup> 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 C.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:

$$std(W_{it}) = \sqrt{var_{G_{it}}(E(W_{it}|G_{it})) + E_{G_{it}}(var(W_{it}|G_{it}))}.^{16} \quad (4)$$

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 (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 C.1.

Using Approach 2, we examine the magnitude of earnings uncertainty for the same sample of 650 students as in Section 3.1.1. Results are summarized in the last three rows of Table 2. Depending on which distributional assumption is made, the average standard deviation of  $W_{i0}$  for the sample varies between \$10,487 and \$13,756 per year and the average standard deviation to mean ratio in the sample varies between 20.38% and 25.31% 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.

### 3.1.3 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-

---

<sup>15</sup>This randomness might be due to, for example, bad matches with instructors, sicknesses at inopportune times, or temporary personal problems.

<sup>16</sup> $var_{G_{it}}(E(W_{it}|G_{it})) = (E(W_{it}|G_{it} = g_{it}) - E_{G_{it}}(E(W_{it}|G_{it} = g_{it})))^2 dF_{G_{it}}(g_{it})$  and  $E_{G_{it}}(var(W_{it}|G_{it})) = \int var(W_{it}|G_{it} = g_{it}) dF_{G_{it}}(g_{it})$ , with  $var(W_{it}|G_{it} = g_{it}) = \int (w_{it} - E(W_{it}|G_{it} = g_{it}))^2 f_{W_{it}|G_{it}=g_{it}}(w_{it}) dw_{it}$ .

approach combinations in Table 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 \$1564 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 has a t-statistic greater than 1.5 for three of the six combinations, but that the average coefficient for Male over the six distribution-approach combinations is only approximately 60% of the average coefficient for Black.

We stress that understanding the exact interpretation of these results is beyond the scope of this paper. 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.

## 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 3.2.1 we compute an expectations analog to this percentage. In Section 3.2.2, we examine the robustness of our results to a measurement error correction. In Section 3.2.3, we 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.

### 3.2.1 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  $X_i$ .<sup>17</sup> 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:

$$W_i \equiv \mathcal{W}(X_i^{t-}, X_i^{t+}). \quad (5)$$

---

<sup>17</sup>Note that these random variables represent both factors related to the worker and factors related to the labor market.

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:

$$var(W_i) = E_{X_i^{t-}}(var(W_i|X_i^{t-})) + var_{X_i^{t-}}(E(W_i|X_i^{t-})). \quad (6)$$

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  $X_i^{t-}$ ).<sup>18</sup> The first term on the right hand side of equation (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 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 the two approaches in Section 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 3 shows the first (uncertainty) term from equation (6), the second column shows the second (heterogeneity) term from equation (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).

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 3.1.1 and 3.1.2), 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.

---

<sup>18</sup>In Section 3.2.3, we discuss scenarios under which the independence assumption would tend to be violated and the implications of these scenarios.



Table 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	174.4	590.3	764.7	77.19%
Approach 2, Log-normal	394.7	645.1	1039.8	62.04%
Approach 2, Normal	237.6	590.4	828.0	71.30%

*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).*

### 3.2.2 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 (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 Sections 3.1.1 and 3.1.2 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;

$$\widetilde{E}^1(W_{it}) = E(W_{it}) + \varsigma_i, \quad (7)$$

where  $\varsigma_i$  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,  $\varsigma_i$ . This can be seen by taking the variance of both sides of equation (7):

$$\text{var}(\widetilde{E}^1(W_{it})) = \text{var}(E(W_{it})) + \text{var}(\varsigma_i). \quad (8)$$

Equation (8) reveals that the true contribution of heterogeneity,  $\text{var}(E(W_{it}))$ , can be obtained by subtracting the variance of the measurement error,  $\varsigma_i$ , from the measured contribution of heterogeneity,  $\text{var}(\widetilde{E}^1(W_{it}))$ . Thus, the remainder of this section focuses on estimating the variance of  $\varsigma_i$ .

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 (7), we assume that measurement error influences the responses to questions such as 1B in a classical manner, that is,

$$\widetilde{E}(W_{it}|G_{it} = g_{it}) = E(W_{it}|G_{it} = g_{it}) + \varsigma_i^{g_{it}} \quad g_{it} = 2.00, 3.00 \text{ or } 3.75, \quad (9)$$

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 3.1.2, 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 3.1.2,  $\widetilde{E}^2(W_{it})$  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  $\widetilde{E}(W_{it}|G_{it} = 3.75)$ :

$$\widetilde{E}^2(W_{it}) = \sum_{g_{it}} \lambda_i^{g_{it}} \widetilde{E}(W_{it}|G_{it} = g_{it}) \quad g_{it} = 2.00, 3.00 \text{ or } 3.75, \quad (10)$$

where, as shown in Appendix D, 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 (9) and equation (10), we obtain the following equation:

$$\begin{aligned} \widetilde{E}^2(W_{it}) &= \sum_{g_{it}} \lambda_i^{g_{it}} E(W_{it}|G_{it} = g_{it}) + \sum_{g_{it}} \lambda_i^{g_{it}} \varsigma_i^{g_{it}} \\ &= E(W_{it}) + \sum_{g_{it}} \lambda_i^{g_{it}} \varsigma_i^{g_{it}}. \end{aligned} \quad (11)$$

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

$$\widetilde{E}^1(W_{it}) - \widetilde{E}^2(W_{it}) = \varsigma_i - \lambda_i^{git} \varsigma_i^{git}. \quad (12)$$

Using equation (12) to estimate  $var(\varsigma_i)$  requires assumptions about the joint distribution of  $\varsigma_i$ ,  $\varsigma_i^{2.0}$ ,  $\varsigma_i^{3.0}$  and  $\varsigma_i^{3.75}$ . The prior assumption that  $\varsigma_i$  and  $\varsigma_i^{git}$ 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 E,

$$var(\varsigma_i) = \frac{var(\widetilde{E}^1(W_{it}) - \widetilde{E}^2(W_{it}))}{1 + E((\lambda_i^{git})^2)}. \quad (13)$$

Note that we can compute the sample analogs of  $var(\widetilde{E}^1(W_{it}) - \widetilde{E}^2(W_{it}))$  and  $E(\lambda_i^{git})$  from data available to us.<sup>19</sup> Hence,  $var(\varsigma_i)$  can be estimated. The first column of Table 4 reports the estimates of  $var(\varsigma_i)$ . Subtracting the measurement error component from measured heterogeneity (column 2 in Table 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 4) to the sum of true heterogeneity (column 2 in Table 4) and uncertainty (column 1 in Table 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 4: Heterogeneity and Uncertainty (Measurement Error Adjusted)

# of Observations: 650	Measurement Error $var(\varsigma_i)$	Adjusted Heterogeneity	Adjusted Heterogeneity Ratio
Stepwise Uniform	84.8	407.1	72.08%
Log-normal	111.4	427.1	50.88%
Normal	94.3	389.1	64.12%

*Note: The second column (Adjusted Heterogeneity) is found by subtracting column 1, Table 4 from column 2, Table 3. The third column (Adjusted Heterogeneity Ratio) is the ratio of column 2, Table 4, to the sum of column 2, Table 4 and column 1, Table 3.*

<sup>19</sup>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.

### 3.2.3 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 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.<sup>20</sup> 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.

## 4 Uncertainty Resolution

In this section, we turn to examining when and why initial uncertainty about income is resolved. In Section 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 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 4.1 and Section 4.2 raise the possibility that much uncertainty about earnings may remain unresolved at the end of college. Section 4.3 takes advantage of the longitudinal expectations data in the BPS to show that this is the case, and Section 4.4 explores the factors that could contribute to this finding. Finally, Section 4.5 takes advantage of information about each student's subjective probability of dropping out to develop and estimate a model that allows an examination of whether students' expectations about how much

---

<sup>20</sup>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  $E(W_{it})$ .

uncertainty will be resolved during college are broadly consistent with the reality that much uncertainty remains unresolved.

## 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 3.1.2) provides a natural way to quantify the importance of uncertainty about final GPA in determining overall uncertainty about future income. Equation 4 yields a natural decomposition of income uncertainty. The first term in the 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.<sup>21</sup>

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:

$$R_{it}^G = \frac{\text{var}_{G_{it}}(E(W_{it}|G_{it}))}{\text{var}(W_{it})} \quad (14)$$

$$= \frac{\text{var}_{G_{it}}(E(W_{it}|G_{it}))}{\text{var}_{G_{it}}(E(W_{it}|G_{it})) + E_{G_{it}}(\text{var}(W_{it}|G_{it}))}.$$

Table 5: Contribution of  $R_{i0}^G$ : Mean and Quartiles

# of Observations: 650	Mean	25%	50%	75%
Stepwise Uniform	0.1729	0.0086	0.0701	0.2530
Log-normal	0.1473	0.0067	0.0533	0.2022
Normal	0.1514	0.0072	0.0571	0.2142

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

Table 5 summarizes the results for the time of entrance. The first column shows that, on average, 17% of income uncertainty is due to uncertainty about final GPA when we use the stepwise uniform assumption and that, on average, 15% of income uncertainty is

---

<sup>21</sup>Of course, it is desirable to directly investigate the importance of each of the “other” factors as thoroughly as possible. In Section 4.2 we do examine the contribution of major choice to overall earnings uncertainty, and in Section 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.

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 20% of overall income uncertainty is due to uncertainty about final GPA. For the stepwise uniform case, only 25% of students believe that more than 25% 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.

We can also provide evidence about the determinants of the heterogeneity in the Table 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 5, the first term in the denominator of equation (14) is 12 times larger for students above the median and the second term in the denominator is 35% 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. Descriptive evidence reveals that this latter source of heterogeneity is important.<sup>22</sup>

## 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 (4):

$$var(W_{it}) = var_{M_{it}}(E(W_{it}|M_{it})) + E_{M_{it}}(var(W_{it}|M_{it})), \quad (15)$$

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  $P_{ijt}$ .<sup>23</sup> The first term on the right side of equation (15) shows how the mean of  $W_{it}$  varies across

---

<sup>22</sup>The Coefficient of Variation (CV), defined as the ratio of the standard deviation to the mean of a distribution, is often used as standardized measure of dispersion. We find that the CV of the sample distribution of difference between the median income if GPA is equal to 3.75 and the median income if GPA is equal to 3.0, which represents the beliefs about how GPA translates to income, is larger than the CV of the sample distribution of standard deviation of the subjective grade distribution, which represents uncertainty about GPA. (1.8263 vs. 0.4651)

<sup>23</sup>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.

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 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:

$$R_{it}^M = \frac{\text{var}_{M_{it}}(E(W_{it}|M_{it}))}{\text{var}_{M_{it}}(E(W_{it}|M_{it})) + E_{M_{it}}(\text{var}(W_{it}|M_{it}))}. \quad (16)$$

Unfortunately, unlike what was the case for our GPA analysis in Section 4.1, the data do not include all of the information that would allow us to directly compute the two terms,  $\text{var}_{M_{it}}(E(W_{it}|M_{it}))$  and  $E_{M_{it}}(\text{var}(W_{it}|M_{it}))$ , that enter this fraction. Specifically, while our analysis in Section 4.1 took advantage of the fact that  $\text{var}(W_{it}|G_{it})$  is available in the data,  $\text{var}(W_{it}|M_{it})$  is not available. However, given information that is observed about  $E(W_{it})$ ,  $\text{var}(W_{it})$  and the probabilities  $P_{ijt}$ ,  $j = 1, \dots, 7$ , 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.

#### 4.2.1 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  $j = 1, \dots, 7$ , equation (15) shows that estimating the two terms requires knowledge of  $E(W_{i0}|M_{i0})$  and  $\text{var}(W_{i0}|M_{i0})$ . 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  $t = 0$ ,

$$\begin{aligned} E(W_{i0}|M_{i0} = j) &= \alpha_w + X_i\beta + \delta_j \\ \text{var}(W_{i0}|M_{i0} = j) &= \alpha_v + X_i\gamma + \theta_j, \end{aligned} \quad (17)$$

where  $\delta_j, j = 1, \dots, 7$  and  $\theta_j, j = 1, \dots, 7$  represent differences in the conditional means and the conditional variances, respectively, across majors.<sup>24</sup>

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  $\text{var}(W_{i0})$  can be written as  $\text{var}_{M_{i0}}(E(W_{i0}|M_{i0})) + E_{M_{i0}}(\text{var}(W_{i0}|M_{i0}))$ ,

---

<sup>24</sup>While the linear specification does not restrict the conditional means and variances in equation (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.

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 assumption as in Section 3.2.2, 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,  $\varsigma_i$ , 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

$$\begin{aligned}\widetilde{E}^1(W_{i0}) &= E_{M_{i0}}(E(W_{i0}|M_{i0})) + \varsigma_i = \sum_{j=1}^7 P_{ij0}E(W_{i0}|M_{i0} = j) + \varsigma_i \\ &= \alpha_w + X_i\beta + \sum_{j=1}^7 P_{ij0}\delta_j + \varsigma_i\end{aligned}\quad (18)$$

$$\begin{aligned}\widetilde{var}(W_{i0}) &= var_{M_{i0}}(E(W_{i0}|M_{i0})) + E_{M_{i0}}(var(W_{i0}|M_{i0})) + u_i \\ &= var_{M_{i0}}(\delta_j) + \alpha_v + X_i\gamma + \sum_{j=1}^7 P_{ij0}\theta_j + u_i\end{aligned}\quad (19)$$

Normalizing the Social Science coefficients  $\delta_7$  and  $\theta_7$  to zero, we estimate the remaining parameters,  $\alpha_w$ ,  $\beta$ ,  $\delta_j$ ,  $j = 1, \dots, 6$ ,  $\alpha_v$ ,  $\gamma$ , and  $\theta_j$ ,  $j = 1, \dots, 6$ , which are needed to estimate  $E(W_{i0}|M_{i0} = j)$ ,  $j = 1, \dots, 7$  and  $var(W_{i0}|M_{i0} = j)$ ,  $j = 1, \dots, 7$  (equation 17), and, therefore, the two terms that appear in the fraction  $R_{i0}^M$  (equation 16). We obtain estimates by:

1. Regressing  $\widetilde{E}^1(W_{i0})$  on  $X_i$  and  $P_{ij0}$ ,  $j = 1, \dots, 7$  to obtain estimates of  $\alpha_w$ ,  $\beta$  and  $\delta_j$ ,  $j = 1, \dots, 6$ .
2. Using the estimates  $\widehat{\delta}_j$ ,  $j = 1, \dots, 6$  and the normalized value  $\delta_7 = 0$  to compute an estimate of  $var_{M_{i0}}(\delta_j)$ ,  $j = 1, \dots, 7$  for each person  $i$ .
3. Regressing  $\widetilde{var}(W_{i0}) - \widehat{var}_{M_{i0}}(\delta_j)$  on  $X_i$  and  $P_{ij0}$ ,  $j = 1, \dots, 7$  to obtain estimates of  $\alpha_v$ ,  $\gamma$  and  $\theta_j$ ,  $j = 1, \dots, 6$ .

#### 4.2.2 Results

Including Black, Male, and ACT score in  $X_i$ , Table 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 6: Contribution of  $R_{i0}^M$ : 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  $R_{i0}^M$ . The final three columns show the three quartiles of the sample distribution of  $R_{i0}^M$ .

Table 7: Estimates for  $\delta_j$  and  $\theta_j$

		$j = 1$	$j = 2$	$j = 3$	$j = 4$	$j = 5$	$j = 6$	$j = 7$
$\delta_j$	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]
$\theta_j$	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 7 reports the estimates for  $\delta_j$  and  $\theta_j$ . 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.

### 4.3 Total Uncertainty Resolution

The findings in Section 4.1 and Section 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 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 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.<sup>25</sup> 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.<sup>26</sup>

In order to facilitate a comparison between total uncertainty resolution and the findings in Section 4.1 and 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 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.

<sup>25</sup>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.

<sup>26</sup>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 3 when using log-normal distribution. This increase, however, is quite small and can be reasonably attributed to measurement error.

The last three rows of Table 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, with this being consistent with an environment where learning tends to be largely about grade performance (ability) and major.

In order to keep the sample constant across columns in Table 8, the sample used includes only students who graduated. A natural question is how the results in Table 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 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 sur-

prising 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.<sup>27</sup> 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 8.

### Demographic Variables

In Section 3.1.3 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 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 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.

## **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

---

<sup>27</sup>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.

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. The question stratifies the set of possible jobs into 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). This type of uncertainty 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 that 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.<sup>28</sup> Thus, overall, these informal results do not suggest that uncertainty about preferences towards types of work are likely to be a driving force in creating income uncertainty.

The findings in the previous two paragraphs suggest that the first explanation - that individuals may be unsure about what kinds of offers they would receive at age 28 - might play an important role. We look for evidence about the importance of this explanation by considering several possible reasons for this 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

---

<sup>28</sup>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.

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 9, 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.

Table 9: 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.*

Roughly speaking, we could group the remaining reasons for the second 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 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.<sup>29</sup> While we stress that it is not possible to determine the relative importance of the different reasons/explanations described above, the results suggest that search frictions are likely to be relevant.

## 4.5 Expected Uncertainty Resolution

In this section, we examine whether students expect that much uncertainty will remain unresolved during school. We develop a simple model in which beliefs about graduating depend on the fraction of income uncertainty that a student expects to resolve during school. Our estimation of this model is made possible by Survey Question 4 (Appendix A), which, at the time of college entrance ( $t = 0$ ), elicits the probability of graduating.

### 4.5.1 A simple model

The model we estimate in this section links a student's subjective dropout probability to her expected resolution of income uncertainty. We first specify the process by which uncertainty about income is resolved, then describe how the student arrives at the subjective dropout probability that she reports at  $t = 0$ , and finally describe how the model is estimated.

#### The process by which uncertainty about income is resolved

A student enters college ( $t = 0$ ) with beliefs about the yearly earnings she will receive at age  $a$  if she graduates ( $s = 1$ ) and if she does not graduate ( $s = 0$ ). These beliefs are given by  $W_{i0}^{a,s} \sim N(\mu_{i0}^{a,s}, (\sigma_{i0}^{a,s})^2)$ ,  $s = 0, 1$ .<sup>30</sup> The student knows that, at a single time in the future  $t^*$ , she will make the decision of whether to graduate or to drop out by comparing the expected utility associated with each option, with these expected utilities depending, in large part, on beliefs about future earnings. However, in thinking about whether she will ultimately choose  $s = 1$  or  $s = 0$  at  $t^*$ , a student must take into account that her beliefs about earnings at age  $a$  will change before  $t^*$  due to the realization of income-influencing factors  $\epsilon_i^{a,s}$ ,  $s = 0, 1$ . We assume that  $\epsilon_i^{a,s}$  enters the earnings equation for choice  $s$  linearly and is normally distributed, with a mean that is normalized to zero and a standard deviation that is denoted by  $\sigma_{i0}^{\epsilon,a,s}$ . The student's beliefs about future earnings at  $t^*$  is then given by  $W_{it^*}^{a,s} \sim N(\mu_{i0}^{a,s} + \epsilon_i^{a,s}, (\sigma_{i0}^{a,s})^2 - (\sigma_{i0}^{\epsilon,a,s})^2)$ ,  $s = 0, 1$ . We further assume that  $\epsilon_i^{a,s} = \rho_s \sigma_{i0}^{a,s} v_i^s$ , where  $v_i^s \sim N(0, 1)$ , implying that  $W_{it^*}^{a,s} \sim N(\mu_{i0}^{a,s} + \rho_s \sigma_{i0}^{a,s} v_i^s, (1 - \rho_s^2)(\sigma_{i0}^{a,s})^2)$ . This indicates that the revision of the initial income beliefs associated with  $s$  depends on the realization of a factor that is relevant for  $s$ ,  $v_i^s$ , with the amount of the revision depending on the fraction of total initial uncertainty

<sup>29</sup>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.

<sup>30</sup>It is, in general, difficult to decompose a step-wise uniformly distributed random variable into multiple factors. We choose the normal distribution since it fits the expectations data slightly better than log-normal distribution does.

that is resolved before  $t^*$ ,  $\rho_s^2$ . Motivated by recent work suggesting the importance of correlated learning (Arcidiacono et al., 2016), we allow  $v_i^s$  and  $v_i^s$  to have a correlation of  $\kappa$ .

### The reported subjective dropout probability

At  $t^*$ , a student observes the realizations of  $v_i^s$ ,  $s = 0, 1$  and chooses between the two schooling options,  $s = 0$  and  $s = 1$  by comparing the expected utility of these two options. If the student chooses  $s = 1$ , she receives a constant utility  $\gamma_i$  for the remaining time in school and receives utility equal to the realization of her earnings,  $w_i^{a,1}$ , in each year (age) out of school. We assume that  $\gamma_i$  is known at  $t = 0$ , so that the only new information obtained between  $t = 0$  and  $t^*$  that is relevant for  $s = 1$  utility is the realization of  $v_i^1$ . Conditioning on the realization of  $v_i^1$ , defining  $\bar{A}$  to be the age of retirement,  $\bar{t}$  to be the age of graduation and  $\beta$  to be the discount factor, the expected utility, or value, at  $t^*$  associated with schooling scenario  $s = 1$  is given by:

$$V_{it^*}^{s=1}(v_i^1) = \gamma_i + \int_{a=\bar{t}}^{\bar{A}} \beta^{a-t^*} E(W_{it^*}^{a,1}) = \gamma_i + \int_{a=\bar{t}}^{\bar{A}} \beta^{a-t^*} \mu_{i0}^{a,1} + \rho_1 v_i^1 \int_{a=\bar{t}}^{\bar{A}} \beta^{a-t^*} \sigma_{i0}^{a,1}$$

If a student drops out, she enters the labor market immediately and receives utility equal to her earnings in each year, so the expected utility, or value, at  $t^*$  of dropping out ( $s = 0$ ) is given by:

$$V_{it^*}^{s=0}(v_i^0) = \int_{a=t^*}^{\bar{A}} \beta^{a-t^*} E(W_{it^*}^{a,0}) = \int_{a=t^*}^{\bar{A}} \beta^{a-t^*} \mu_{i0}^{a,0} + \rho_0 v_i^0 \int_{a=t^*}^{\bar{A}} \beta^{a-t^*} \sigma_{i0}^{a,0}. \quad (20)$$

At  $t = 0$ , the student reports the probability of dropping out by computing the fraction of time that her realizations of  $v_i^s$ ,  $s = 0, 1$  will lead to  $V_{it^*}^{s=0}(v_i^0) > V_{it^*}^{s=1}(v_i^1)$ . Denoting  $\bar{\mu}_{1,i} = \int_{a=\bar{t}}^{\bar{A}} \beta^{a-t^*} \mu_{i0}^{a,1}$ ,  $\bar{\mu}_{0,i} = \int_{a=t^*}^{\bar{A}} \beta^{a-t^*} \mu_{i0}^{a,0}$ ,  $\bar{\sigma}_{1,i} = \int_{a=\bar{t}}^{\bar{A}} \beta^{a-t^*} \sigma_{i0}^{a,1}$ , and  $\bar{\sigma}_{0,i} = \int_{a=t^*}^{\bar{A}} \beta^{a-t^*} \sigma_{i0}^{a,0}$ , the dropout probability can be written as:

$$P_i^D = Prob(\bar{\mu}_{0,i} + \rho_0 v_i^0 \bar{\sigma}_{0,i} > \gamma_i + \bar{\mu}_{1,i} + \rho_1 v_i^1 \bar{\sigma}_{1,i}) = \Phi\left(\frac{\bar{\mu}_{0,i} - \bar{\mu}_{1,i} - \gamma_i}{\rho_1^2 \bar{\sigma}_{1,i}^2 + \rho_0^2 \bar{\sigma}_{0,i}^2 - 2\kappa \rho_1 \rho_0 \bar{\sigma}_{1,i} \bar{\sigma}_{0,i}}\right),$$

where the last expression follows from the fact that  $v_i^1$  and  $v_i^0$  are standard normal random variables with correlation  $\kappa$ .

The general focus of our paper on income uncertainty under the graduation scenario motivates a particular interest in  $\rho_1$ . The effect of  $\rho_1$  on the subjective dropout probability depends on the value of other parameters. However, some intuition about how subjective dropout probabilities are related to  $\rho_1$  can be obtained by considering a seemingly reasonable scenario in which students resolve relatively little uncertainty about income under the dropout scenario (i.e.,  $\rho_0$  is small). 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, it is typically negative with its absolute value being related to the distance that a student is from the margin of dropping out at the time of entrance.<sup>31</sup> An increase in  $\rho_1$  increases the amount that a student learns 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 situation where it is optimal to leave school. Thus, all else equal, in the seemingly most likely scenario in which the numerator is negative, the dropout probability will tend to be increasing in  $\rho_1$ .

### Estimation

$\bar{\mu}_{0,i} - \bar{\mu}_{1,i}$ ,  $\bar{\sigma}_{0,i}$  and  $\bar{\sigma}_{1,i}$  can be constructed from data.<sup>32</sup> Thus, assuming that  $\gamma_i$  is normally distributed, equation (21) reveals that the three parameters  $\rho_0$ ,  $\rho_1$ , and  $\kappa$  (as well as the distribution of  $\gamma_i$ ) can be estimated using Maximum Likelihood.<sup>33</sup> However, from a practical standpoint, given the our relatively small sample size and the potentially high correlation between  $\bar{\sigma}_{0,i}$  and  $\bar{\sigma}_{1,i}$ , it may be difficult in practice to estimate all three parameters. Hence, as discussed in Appendix G, we choose to assume that  $\kappa$  is equal to its realization counterpart and estimate  $\kappa$  outside the model in a manner that takes advantage of the longitudinal aspects of our data.<sup>34</sup>

### 4.5.2 Results and Discussion

Students in the 2000 cohort were not asked to answer the dropout probability question, Question 4, in their freshman year. As a result, we restrict our sample to 349 respondents from cohort 2001. For this sample, we find that the average subjective probability of dropping out reported on Question 4 is only 0.14. At  $t = 0$ , the sample average of expected lifetime income associated with the graduation scenario and the dropout scenario are approximately \$1,040,000 and \$729,000, respectively. On average, there is more uncertainty about earnings under the graduation scenario than there is about

<sup>31</sup>Of course, from a theoretical standpoint, when experimentation plays a role in the decision to enter school, a student might enter college even if she has a positive numerator.

<sup>32</sup> $\bar{\mu}_{1,i}$  and  $\bar{\sigma}_{1,i}$  are weighted sums (across ages  $a$ ) of the means and standard deviations of subjective earnings distributions for the graduation scenario, elicited at  $t = 0$  using Question 1A.  $\bar{\mu}_{0,i}$  and  $\bar{\sigma}_{0,i}$  are the weighted sum (across ages  $a$ ) of means and standard deviations of subjective earnings distributions for the dropout scenario, elicited at  $t = 0$  using questions analogous to Question 1A.

<sup>33</sup>We assume that  $\beta = 0.95$ . Following Stinebrickner and Stinebrickner (2014b), to deal with the fact that values of  $\mu_{i0}^{a,s}$  and  $\sigma_{i0}^{a,s}$  are only observed directly for the first year a student leaves school, at age 28, and at age 38, we assume that both  $\mu_{i0}^{a,s}$  and  $\sigma_{i0}^{a,s}$  are linear between the first year out of college and the age of 28, are linear between the ages of 28 and 38, and are constant after the age of 38.

Our model implies that the reported dropout probability should be strictly between 0 and 1. In practice, following Stinebrickner and Stinebrickner (2014a), we set  $P_i^D = 0.01$  if  $i$  reports a dropout probability that is smaller than 0.01, and  $P_i^D = 0.99$  if  $i$  reports a dropout probability that is greater than 0.99. Our results are robust to slight changes in these assumptions.

<sup>34</sup>Consistent with our belief that estimating all of the parameters that are technically identified might be challenging in practice, we had some difficulty obtaining convergence when we tried to estimate the full model. Our decision to estimate  $\kappa$  outside the model was motivated by the symmetry present between  $\rho_0$  and the parameter  $\rho_1$  which has been the focus of previous sections.

earnings under the dropout scenario: The sample average of  $\bar{\sigma}_{1,i}$  and  $\bar{\sigma}_{0,i}$  are \$256,000 and \$175,000, respectively.

The estimation results are shown in Table 10. Consistent with the notion that learning is highly correlated, the estimate for  $\kappa$  is 0.5605. However, the estimate of  $\rho_0$ , 0.2170, indicates that the amount of uncertainty that is resolved about earnings under the dropout scenario is not particularly large.

The estimate of  $\rho_1$ , our parameter of primary interest, is 0.4695, implying that, during college, students expect to resolve 22.04% ( $\rho_1^2$ ) of their uncertainty about earnings under the scenario in which they graduate from college. Hence, consistent with evidence on actual uncertainty resolution, our results suggest that the majority of income uncertainty is expected to remain at the end of college.<sup>35</sup>

It is notable that income represents the only source of uncertainty in our model. As a result, we do not believe that our model is particularly well-suited for providing general evidence about how uncertainty influences the dropout decision. Nonetheless, our model is still valuable for our quite narrow objective of providing a rough characterization of the fraction of uncertainty that students expect to resolve in college. This is the case because, from a conceptual standpoint, the dropout probability will tend to be increasing in the total amount of uncertainty that is resolved about all factors that a student is uncertain about at college entrance. Therefore, introducing another potential source of uncertainty resolution that could help explain the observed subjective dropout probabilities would likely lead to a lower estimate of  $\rho_1$ .<sup>36</sup>

Table 10: Estimation Results

# of Observations: 349	$\rho_1$	$\bar{\gamma}$	$\rho_2$	$\kappa$
Correlated Learning	0.4695 (13.0265)	-158.7077 (7.8757)	0.2170 (2.6665)	0.5605

Note: t-statistics are in the parenthesis.

## 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.

<sup>35</sup>The fact that the estimated disutility associated with being in school is quite large is consistent with other research, e.g., Stinebrickner and Stinebrickner (2014a/b) and Stange (2012).

<sup>36</sup>Similarly, at a given value of  $\rho_1$ , the subjective dropout probabilities from our conceptual framework would tend to increase if students were given more points in time at which they could choose to drop out. This implies that the value of  $\rho_1$  that would be needed to explain the observed subjective dropout probabilities would tend to be lower if we relaxed the assumption that students make their dropout decision at a single point in time,  $t^*$ .

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 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.

# Appendices

## 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

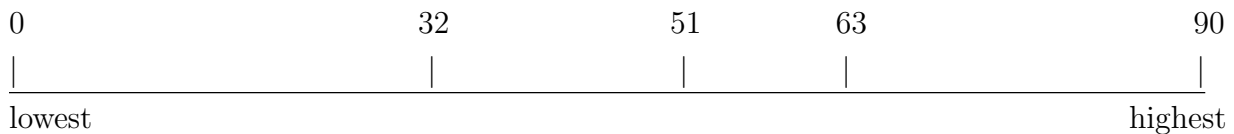
2) The person would then divide the values between 14 and 100 into four intervals so that she thinks that there is a 25% chance that her grade will be in each interval. For example, suppose that the person marked three points between 14 and 100 and labeled them 52, 80 and 92.



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 do 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



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 and 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.**

<u>Interval</u>	<u>Percent Chance(number of chances out of 100)</u>
[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 up to 100).

**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
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	_____

## B 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:

$$AE(D) = \frac{1}{N} \sum_{i=1}^N [(F(C_{it}^3; D) - F(C_{it}^2; D) - 0.25)^2 + (F(C_{it}^4; D) - F(C_{it}^3; D) - 0.25)^2], \quad (21)$$

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

Using the same sample as in Section 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.

## C Approach 2: Computation Details

### C.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  $g_{it} = 2.00, 3.00, \text{ or } 3.75$ . 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  $g_{it} = 4.00$ ).

## C.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,  $k=1, \dots, 8$ , subject to the constraint that the student obtains the 2.0 average that is needed to graduate. Thus,  $G_{i0}$  is given by:

$$G_{i0} = \frac{\sum_{k=1}^8 G_{i0}^k}{8}, \text{ if } \frac{\sum_{k=1}^8 G_{i0}^k}{8} \geq 2, \quad (22)$$

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  $G_{i0}^k$ ,  $k = 1, \dots, 8$ . 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:

$$\tilde{G}_{i0}^k = a_{i0} + \xi_{i0}^k, \text{ where } G_{i0}^k = 0 \text{ if } \tilde{G}_{i0}^k < 0, G_{i0}^k = 4 \text{ if } \tilde{G}_{i0}^k > 4, \text{ and } G_{i0}^k = \tilde{G}_{i0}^k, \text{ otherwise.} \quad (23)$$

$a_{i0}$  represents student  $i$ 's ( $t = 0$ ) beliefs about permanent (academic) ability and  $\xi_{i0}^k$  describes  $i$ 's ( $t = 0$ ) beliefs about the mean-zero transitory shock component of grades which is independent across semesters  $k$ . Thus, the  $G_{i0}^k$ '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  $\xi_{i0}^k$  are normally distributed:  $a_{i0} \sim N(\mu_{i0}^a, \sigma_{i0}^a)$  and  $\xi_{i0}^k \sim N(0, \sigma_{i0}^\xi)$ . For each student, we numerically search for the set of parameters  $\{\mu_{i0}^a, \sigma_{i0}^a, \sigma_{i0}^\xi\}$  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.<sup>37</sup> Formally, we have:

$$\begin{aligned} \{\widehat{\mu}_{i0}^a, \widehat{\sigma}_{i0}^a, \widehat{\sigma}_{i0}^\xi\} = \underset{cat_j^g \in CAT^g}{argmin} & Pr_{model}(G_{i0}^k \in cat_j^g)(Pr_{obs}(G_{i0}^k \in cat_j^g) - Pr_{model}(G_{i0}^k \in cat_j^g))^2 \\ & + \underset{cat_j^\xi \in CAT^\xi}{Pr_{model}(\xi_{i0}^k \in cat_j^\xi)(Pr_{obs}(\xi_{i0}^k \in cat_j^\xi) - Pr_{model}(\xi_{i0}^k \in cat_j^\xi))^2}, \quad (24) \end{aligned}$$

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  $\{\mu_{i0}^a, \sigma_{i0}^a, \sigma_{i0}^\xi\}$  are estimated, we can approximate the distribution of  $G_{i0}$  by simulation using equation (22) and (23).

## D 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,

$$\begin{aligned} E(W_{it}) &= E_{G_{it}}(E(W_{it}|G_{it})) = \int_2^4 E(W_{it}|g_{it})dF_{G_{it}}(g_{it}) \\ &= \int_2^3 [E(W_{it}|2.00) + \frac{E(W_{it}|3.00) - E(W_{it}|2.00)}{3.00 - 2.00}(g_{it} - 2)]dF_{G_{it}}(g_{it}) \\ &+ \int_3^4 [E(W_{it}|3.00) + \frac{E(W_{it}|3.75) - E(W_{it}|3.00)}{3.75 - 3.00}(g_{it} - 3)]dF_{G_{it}}(g_{it}) \\ &= \int_2^3 [E(W_{it}|2.00)(1 - \frac{g_{it} - 2}{3.00 - 2.00}) + E(W_{it}|3.00)\frac{g_{it} - 2}{3.00 - 2.00}]dF_{G_{it}}(g_{it}) \\ &+ \int_3^4 [E(W_{it}|3.00)(1 - \frac{g_{it} - 3}{3.75 - 3.00}) + E(W_{it}|3.75)\frac{g_{it} - 3}{3.75 - 3.00}]dF_{G_{it}}(g_{it}) \\ &= \lambda_{it}^G E(W_{it}|G) \quad G = 2.00, 3.00 \text{ or } 3.75, \quad (25) \end{aligned}$$

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  $\lambda_i^{3.75} = \int_3^4 \frac{g_{it}-3}{0.75}dF_{G_{it}}(g_{it})$ .

<sup>37</sup>We have also estimated a non-weighted version. The results are similar.

## E Magnitude of the Measurement Error

In this section, we show that equation (12), along with additional assumptions, implies equation (13). Recall that equation (12) states:

$$\widetilde{E}^1(W_{it}) - \widetilde{E}^2(W_{it}) = \varsigma_i - \lambda_i^{g_{it}} \varsigma_i^{g_{it}}. \quad (12 \text{ revisited})$$

Taking the expectation of the square of both sides, we have:

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

Therefore,

$$\text{var}(\varsigma_i) = \frac{\text{var}(\widetilde{E}^1(W_{it}) - \widetilde{E}^2(W_{it}))}{1 + E((\lambda_i^{g_{it}})^2)}. \quad (13 \text{ revisited})$$

## F Taking into Account Interpolation Errors

In Section 3.2.2, 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  $\text{var}(\varsigma_i)$ , 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  $F_{G_{it}}^\Delta(g_{it}) = \widetilde{F}_{G_{it}}(g_{it}) - F_{G_{it}}(g_{it})$ .

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  $\mathbf{E}_{G_{it}}^W$ , and a vector that includes  $(\widetilde{E}(W_{it}|G_{it} = 2.00), \widetilde{E}(W_{it}|G_{it} =$



3.00),  $\tilde{E}(W_{it}|G_{it} = 3.75)$ ) as  $\tilde{\mathbf{E}}_{\mathbf{G}_{it}}^{\mathbf{W}}$ . 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  $\tilde{\mathbf{E}}_{\mathbf{G}_{it}}^{\mathbf{W}}$  to  $\tilde{E}(W_{it}|G_{it} = g_{it})$ ,  $g_{it} \neq 2.00, 3.00, 3.75$ . We denote this mapping as  $\tilde{E}^W(g_{it}; \tilde{\mathbf{E}}_{\mathbf{G}_{it}}^{\mathbf{W}})$ . Note that the difference between the computed value of the conditional mean,  $\tilde{E}^W(g_{it}; \tilde{\mathbf{E}}_{\mathbf{G}_{it}}^{\mathbf{W}})$ , and the true value of conditional mean,  $E(W_{it}|G_{it} = g_{it})$ , is a result of both the measurement error,  $\tilde{\mathbf{E}}_{\mathbf{G}_{it}}^{\mathbf{W}} - \mathbf{E}_{\mathbf{G}_{it}}^{\mathbf{W}} = (\varsigma_i^{2.00}, \varsigma_i^{3.00}, \varsigma_i^{3.75})$ , and the interpolation error,  $\tilde{E}^W(g_{it}; \tilde{\mathbf{E}}_{\mathbf{G}_{it}}^{\mathbf{W}}) - E(W_{it}|G_{it} = g_{it})$ .

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

$$\begin{aligned}
\tilde{E}^2(W_{it}) &= \int_2^4 \tilde{E}(W_{it}|G_{it} = g_{it}) d\tilde{F}_{G_{it}}(g_{it}) = \int_2^4 \tilde{E}^W(g_{it}; \tilde{\mathbf{E}}_{\mathbf{G}_{it}}^{\mathbf{W}}) d\tilde{F}_{G_{it}}(g_{it}) \\
&= \int_2^4 E(W_{it}|G_{it} = g_{it}) d\tilde{F}_{G_{it}}(g_{it}) + \int_2^4 (\tilde{E}^W(g_{it}; \tilde{\mathbf{E}}_{\mathbf{G}_{it}}^{\mathbf{W}}) - E(W_{it}|G_{it} = g_{it})) d\tilde{F}_{G_{it}}(g_{it}) \\
&= \int_2^4 E(W_{it}|G_{it} = g_{it}) dF_{G_{it}}(g_{it}) + \int_2^4 E(W_{it}|G_{it} = g_{it}) dF_{G_{it}}^\Delta(g_{it}) \\
&\quad + \int_2^4 (\tilde{E}^W(g_{it}; \tilde{\mathbf{E}}_{\mathbf{G}_{it}}^{\mathbf{W}}) - E(W_{it}|G_{it} = g_{it})) d\tilde{F}_{G_{it}}(g_{it}) \\
&= E(W_{it}) + \int_2^4 E(W_{it}|G_{it} = g_{it}) dF_{G_{it}}^\Delta(g_{it}) + \int_2^4 (\tilde{E}^W(g_{it}; \tilde{\mathbf{E}}_{\mathbf{G}_{it}}^{\mathbf{W}}) - \tilde{E}^W(g_{it}; \mathbf{E}_{\mathbf{G}_{it}}^{\mathbf{W}})) d\tilde{F}_{G_{it}}(g_{it}) \\
&\quad + \int_2^4 (\tilde{E}^W(g_{it}; \mathbf{E}_{\mathbf{G}_{it}}^{\mathbf{W}}) - E(W_{it}|G_{it} = g_{it})) d\tilde{F}_{G_{it}}(g_{it}) \tag{26}
\end{aligned}$$

Following steps similar to those in Section D, we can show that:

$$\int_2^4 (\tilde{E}^W(g_{it}; \tilde{\mathbf{E}}_{\mathbf{G}_{it}}^{\mathbf{W}}) - \tilde{E}^W(g_{it}; \mathbf{E}_{\mathbf{G}_{it}}^{\mathbf{W}})) d\tilde{F}_{G_{it}}(g_{it}) = \int_{g_{it}} \tilde{\lambda}_i^{g_{it}} \varsigma_i^{g_{it}}, \quad g_{it} = 2.00, 3.00 \text{ or } 3.75, \tag{27}$$

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

Denoting  $\Delta_{it} \equiv \int_2^4 E(W_{it}|G_{it} = g_{it}) dF_{G_{it}}^\Delta(g_{it}) + \int_2^4 (\tilde{E}^W(g_{it}; \mathbf{E}_{\mathbf{G}_{it}}^{\mathbf{W}}) - E(W_{it}|G_{it} = g_{it})) d\tilde{F}_{G_{it}}(g_{it})$ , equation (26) can be written as:

$$\tilde{E}^2(W_{it}) = E(W_{it}) + \int_{g_{it}} \tilde{\lambda}_i^{g_{it}} \varsigma_i^{g_{it}} + \Delta_{it} \quad g_{it} = 2.00, 3.00 \text{ or } 3.75. \tag{28}$$

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

$$\tilde{E}^1(W_{it}) - \tilde{E}^2(W_{it}) = \varsigma_i - \int_{g_{it}} \tilde{\lambda}_i^{g_{it}} \varsigma_i^{g_{it}} - \Delta_{it} \quad g_{it} = 2.00, 3.00 \text{ or } 3.75. \tag{29}$$

Recall that  $\varsigma_i$  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  $\Delta_{it}$  since none of them show up in the expression of  $\Delta_{it}$ . Taking the variance of both sides of equation (29), we find:

$$\begin{aligned}
\text{var}(\widetilde{E}^1(W_{it}) - \widetilde{E}^2(W_{it})) &= \text{var}(\varsigma_i - \widetilde{\lambda}_i^{g_{it}} \varsigma_i^{g_{it}}) + \text{var}(\Delta_{it}) \\
&= \text{var}(\varsigma_i) [1 + \underset{g_{it}}{E((\widetilde{\lambda}_i^{g_{it}})^2)}] + \text{var}(\Delta_{it}) \\
&\geq \text{var}(\varsigma_i) [1 + \underset{g_{it}}{E((\widetilde{\lambda}_i^{g_{it}})^2)}].
\end{aligned} \tag{30}$$

Therefore,

$$\text{var}(\varsigma_i) \leq \frac{\text{var}(\widetilde{E}^1(W_{it}) - \widetilde{E}^2(W_{it}))}{1 + \underset{g_{it}}{E((\widetilde{\lambda}_i^{g_{it}})^2)}}. \tag{31}$$

Since both  $\widetilde{\lambda}_i^{g_{it}}$  in this section and  $\lambda_i^{g_{it}}$  in Section 3.2.2 are computed using the same distribution of  $G_{it}$  (we assume that there is no error in the distribution of  $G_{it}$  in Section 3.2.2), they are numerically identical. Thus, the right side of equation (31) is numerically identical to the right side of equation (13). As a result, equation (31) shows that our estimates of  $\text{var}(\varsigma_i)$  reported in Table 4 should be considered as upper bounds for the true value of  $\text{var}(\varsigma_i)$ .

## G Estimation of $\kappa$

We show that  $\kappa$  can be estimated from the evolution of individual income beliefs. Recall that student  $i$ 's expectation about  $w_i^{a,s}$  at the beginning of college and at the end of the third year are denoted as  $EW_{i0}^{a,s}$  and  $EW_{i3}^{a,s}$ , respectively. As explained in the text:

$$\begin{aligned}
EW_{i0}^{a,s} &= \mu_{i0}^{a,s} \\
EW_{i3}^{a,s} &= \mu_{i0}^{a,s} + \rho_s v_i^s \sigma_{i0}^{a,s}.
\end{aligned} \tag{32}$$

Taking the difference of the two equations, we have:

$$EW_{i3}^{a,s} - EW_{i0}^{a,s} = \rho_s v_i^s \sigma_{i0}^{a,s}. \tag{33}$$

Denote the covariance matrix of  $(\frac{EW_{i3}^{a,0} - EW_{i0}^{a,0}}{\sigma_{i0}^{a,0}}, \frac{EW_{i3}^{a,1} - EW_{i0}^{a,1}}{\sigma_{i0}^{a,1}})$  as  $\Pi$ , with the  $(p, q)$ th entry denoted by  $\Pi_{pq}$ . Equation 33 implies that:

$$\begin{aligned} \Pi &= \begin{bmatrix} \rho_0^2 \text{var}(v_i^0) & \rho_0 \rho_1 \text{corr}(v_i^0, v_i^1) \\ \rho_0 \rho_1 \text{corr}(v_i^0, v_i^1) & \rho_1^2 \text{var}(v_i^1) \end{bmatrix} \\ &= \begin{bmatrix} \rho_0^2 \text{var}(v_i^0) & \rho_0 \rho_1 \kappa \sqrt{\text{var}(v_i^0) \text{var}(v_i^1)} \\ \rho_0 \rho_1 \kappa \sqrt{\text{var}(v_i^0) \text{var}(v_i^1)} & \rho_1^2 \text{var}(v_i^1) \end{bmatrix}. \end{aligned} \quad (34)$$

Hence,  $\kappa = \frac{\overline{\Pi_{12}\Pi_{21}}}{\overline{\Pi_{11}\Pi_{22}}}$ .

In the BPS dataset, students were asked to report their expectations about future income at the time of college entrance and at the end of each academic year. Therefore, we are able to compute both  $\frac{EW_{i3}^{a,0} - EW_{i0}^{a,0}}{\sigma_{i0}^{a,0}}$  and  $\frac{EW_{i3}^{a,1} - EW_{i0}^{a,1}}{\sigma_{i0}^{a,1}}$  for students who remain in the sample at the end of the third year. The sample analog of  $\Pi$  can be computed accordingly. However, due to the potentially non-random attrition of our sample, this sample analog might not consistently estimate  $\Pi$ . Therefore, we also consider the following alternative.

We further decompose  $\rho_s v_i^s$  into independently distributed factors that are realized in Year 1, Year 2 and Year 3, respectively;

$$\rho_s v_i^s = \prod_{j=1}^3 \rho_{s,j} v_i^{s,j}, \quad (35)$$

where  $v_i^{s,j}$ s are standard normal and  $\prod_{j=1}^3 \rho_{s,j}^2 = \rho_s^2$ . It follows that:

$$EW_{i1}^{a,s} - EW_{i0}^{a,s} = \rho_{s,1} v_i^{s,1} \sigma_{i0}^{a,s}. \quad (36)$$

Denote  $\delta_j = \text{corr}(v_i^{0,j}, v_i^{1,j})$ . Data on  $EW_{i1}^{a,s}$  and  $EW_{i0}^{a,s}$  are collected at the beginning and end of the first year, respectively. Since the majority of dropout takes place after the end of the first year, sample attrition is arguably random. Hence,  $\kappa_1$  can be consistently estimated from the sample covariance matrix of  $\frac{EW_{i1}^{a,0} - EW_{i0}^{a,0}}{\sigma_{i0}^{a,0}}$  and  $\frac{EW_{i1}^{a,1} - EW_{i0}^{a,1}}{\sigma_{i0}^{a,1}}$ . Under the assumption that  $\kappa_j$  is constant over  $j$ , it can be shown that  $\kappa = \kappa_1$ .

Table 11: Estimates of  $\kappa$

	$a = 1$ Year Out	$a = 28$	Average
$EW_{i3}^{a,s} - EW_{i0}^{a,s}$	0.4069	0.4999	0.4534
$EW_{i1}^{a,s} - EW_{i0}^{a,s}$	0.5393	0.5818	0.5605

Note that we can compute the sample covariance matrix of  $(\frac{EW_{i3}^{a,0} - EW_{i0}^{a,0}}{\sigma_{i0}^{a,0}}, \frac{EW_{i3}^{a,1} - EW_{i0}^{a,1}}{\sigma_{i0}^{a,1}})$  and  $(\frac{EW_{i1}^{a,0} - EW_{i0}^{a,0}}{\sigma_{i0}^{a,0}}, \frac{EW_{i1}^{a,1} - EW_{i0}^{a,1}}{\sigma_{i0}^{a,1}})$  and estimate  $\kappa$  for both  $a = 4$  (first year out of college) and  $a = 10$  (age 28 or 10 years after college entrance). Hence, in total, we can obtain 4 estimates of  $\kappa$ .

Estimation results are summarized in Table 11. Depending on which sample covariance matrix is used,  $\kappa$  is estimated to be between 0.4069 and 0.5818. We find that, the estimate of  $\kappa$  is reasonably robust to the choice of  $a$ . In the main text, we choose to

set  $\kappa$  to 0.5605, which is the average of the two estimates computed using the sample covariance matrix of  $(\frac{EW_{i1}^{a,0}-EW_{i0}^{a,0}}{\sigma_{i0}^{a,0}}, \frac{EW_{i1}^{a,1}-EW_{i0}^{a,1}}{\sigma_{i0}^{a,1}})$ .

## References

- [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] Peter Arcidiacono, Esteban Aucejo, Arnaud Maurel, and Tyler Ransom. College attrition and the dynamics of information revelation. Technical report, National Bureau of Economic Research, 2016.
- [7] 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.
- [8] 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.
- [9] Julian R Betts. What do students know about wages? evidence from a survey of undergraduates. *Journal of human resources*, pages 27–56, 1996.
- [10] Asher A Blass, Saul Lach, and Charles F Manski. Using elicited choice probabilities to estimate random utility models: Preferences for electricity reliability. *International Economic Review*, 51(2):421–440, 2010.
- [11] Richard Blundell and Ian Preston. Consumption inequality and income uncertainty. *The Quarterly Journal of Economics*, 113(2):603–640, 1998.
- [12] Martin Browning and Jesus Carro. Heterogeneity and microeconometrics modeling. *Econometric Society Monograph*, 43:47, 2007.

- [13] 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.
- [14] 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.
- [15] 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.
- [16] Flavio Cunha, James J Heckman, and Salvador Navarro. Separating uncertainty from heterogeneity in life cycle earnings. *Oxford Economic Papers*, 57(2):191–261, 2005.
- [17] Jeff Dominitz and Charles F Manski. Perceptions of economic insecurity: Evidence from the survey of economic expectations. *Public Opinion Quarterly*, pages 261–287, 1997a.
- [18] 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.
- [19] Milton Friedman. Choice, chance, and the personal distribution of income. *Journal of Political Economy*, 61(4):277–290, 1953.
- [20] Robert Gibbons and Lawrence Katz. Does unmeasured ability explain inter-industry wage differentials? *The Review of Economic Studies*, 59(3):515–535, 1992.
- [21] 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.
- [22] F Thomas Juster. Consumer buying intentions and purchase probability: An experiment in survey design. *Journal of the American Statistical Association*, 61(315):658–696, 1966.
- [23] Charles F Manski. The use of intentions data to predict behavior: A best-case analysis. *Journal of the American Statistical Association*, 85(412):934–940, 1990.
- [24] 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.

- [25] Charles F Manski. Analysis of choice expectations in incomplete scenarios. *Journal of Risk and Uncertainty*, 19(1-3):49–66, 1999.
- [26] Charles F Manski. Measuring expectations. *Econometrica*, 72(5):1329–1376, 2004.
- [27] Costas Meghir and Luigi Pistaferri. Income variance dynamics and heterogeneity. *Econometrica*, 72(1):1–32, 2004.
- [28] 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.
- [29] Kevin M Stange. An empirical investigation of the option value of college enrollment. *American Economic Journal: Applied Economics*, 4(1):49–84, 2012.
- [30] 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.
- [31] Ralph Stinebrickner and Todd Stinebrickner. Learning about academic ability and the college dropout decision. *Journal of Labor Economics*, 30(4):707–748, 2012.
- [32] 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.
- [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, 2014b.
- [34] Ralph Stinebrickner, Todd Stinebrickner, and Paul Sullivan. Dynamics of the gender wage gap among the college educated. *Unpublished Manuscript*, 2018.
- [35] Matthew Wiswall and Basit Zafar. Determinants of college major choice: Identification using an information experiment. *The Review of Economic Studies*, 82(2):791–824, 2014.
- [36] Basit Zafar. How do college students form expectations? *Journal of Labor Economics*, 29(2):301–348, 2011.