Quantitative Economics 10 (2019), 607–641 1759-7331/20190607

Uncertainty about future income: Initial beliefs and resolution during college

Yifan Gong
Department of Economics, The University of Western Ontario

Todd Stinebrickner
Department of Economics, The University of Western Ontario

Ralph Stinebrickner
Department of Mathematics, Berea College

We use unique data from the Berea Panel Study to characterize how much earnings uncertainty is present for students at college entrance and how quickly this uncertainty is resolved. We characterize uncertainty using survey questions that elicit the entire distribution describing one's beliefs about future earnings. Taking advantage of the longitudinal nature of the expectations data, we find that roughly two-thirds of the income uncertainty present at the time of entrance remains at the end of college. 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. This paper also contributes to a literature interested in understanding the relative importance of uncertainty and heterogeneity in determining observed earnings distributions.

Keywords. Income uncertainty, uncertainty resolution, heterogeneity, expectations data.

JEL classification. J30.

1. Introduction

From a conceptual standpoint, it is clear that the decision to enter or not enter college, as well as other college decisions, will depend on the amount of uncertainty about future income that is present at the time of college entrance.[1] However, college decisions

Yifan Gong: ygong48@uwo.ca
Todd Stinebrickner: trstineb@uwo.ca
Ralph Stinebrickner: stinebricknerr@berea.edu
The project was made possible by generous support from the Mellon Foundation, the Spencer Foundation, the National Science Foundation, and the Social Sciences and Humanities Research Council. We would like to thank Charles Leung, Yuxi Yao, Martin Luccioni, and Phuong Vu and conference participants at the 2018 summer meetings of The Econometric Society and the Canadian Economic Association for comments.

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

 https://doi.org/10.3982/QE954

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 toward understanding how income uncertainty influences college decisions involves characterizing how much income uncertainty is present for students at the time of college entrance and how quickly (and why) this uncertainty is resolved.[2] Unfortunately, taking this first step has proven to be difficult (Cunha, Heckman, and 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.[3] 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 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) developed methods for separating uncertainty from het-

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

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

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

Our results in Section 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 $9600 a year to an average of $14,100 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.[6] This issue is directly linked to the question of why uncertainty exists. For example, one particularly prominent potential source of uncertainty is college grade point average (GPA), which is widely viewed as the best available proxy for human capital at the time of college graduation. By definition, all uncertainty about final college GPA will be resolved by the end of college. Thus, if uncertainty about GPA is an important contributor to the initial uncertainty about earnings, then students will expect much of the uncertainty about earnings to be resolved at some point during college and that this resolution will take place early in college if learning about academic ability tends to happen quickly.[7] We are able to

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

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

[6]An exception is Navarro and Zhou (2017) who develop a model that identifies the path of uncertainty resolution over multiple periods. With each period having a length of 6 years, their first period (age 18–24) corresponds to the time that our sample spends in college and the first 2 years in the workforce.

[7]See Stinebrickner and Stinebrickner (2012, 2014b) and Zafar (2011) for research that used 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.

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 18% of the variance representing (age 28) earnings uncertainty at the time of college entrance can be attributed to uncertainty about grade performance at the time of college entrance. A related analysis finds that between 11% and 17% of the earnings uncertainty at the time of college entrance can be attributed to uncertainty about college major at the time of college entrance. Moreover, when combined, uncertainty about these two factors together account for about 18% to 24% of overall initial uncertainty about future income.

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

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

2. 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 objective, provides a full tuition subsidy to all students. Thus, as always, it is necessary to be appropriately cautious about the exact extent to which results from one school would generalize to other institutions. However, important for the notion that the basic lessons from 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.[8] 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.

Crucial to our analysis, the BPS was the first comprehensive longitudinal survey with a central focus on the collection of beliefs (expectations) data. Initial beliefs were elicited on the baseline survey. An important aspect of the BPS in our context is that substantial follow-up surveys, which were administered at the beginning and end of each subsequent semester, documented how beliefs change over time.[9]

Our primary survey questions eliciting beliefs about future earnings are of the form of baseline survey Question 1A, which is shown in Appendix A.[10] Specifically, survey Question 1A elicited the minimum, the maximum, and the three quartiles of the subjective income 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. An almost identical set of questions (not shown) was used to elicit beliefs under the scenario in which the student does not graduate from college. A baseline survey question also elicited beliefs about earnings conditional on graduating with three particular levels of GPA $(2.00, 3.00, 3.75)$. Question 1B in Appendix A shows the portion of this question related to graduating with a 2.00 GPA.

Table 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

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

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

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

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	49.1923	60.5161	41.8088	48.1623	54.7238
	(18.3900)	(21.9922)	(36.7525)	(21.7551)	(23.7830)	(26.3292)
Interquartile range	12.6773	15.3221	19.2754	12.3756	13.8969	15.9806
	(10.3599)	(12.8526)	(30.7070)	(10.6551)	(12.1135)	(13.6923)

Note: The unit of measurement for all entries is $1000. 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 $39,548.00 and a sample standard deviation of $18,390.00 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 $41,808.80 and a sample standard deviation of $21,755.10 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.

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.

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, 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), 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.[11] The general object of interest is the distribution describing a student's subjective beliefs about her future income, which, as noted earlier, we often refer to as the student's subjective income distribution. While this entire section focuses on beliefs at the time of entrance, which we often refer to as "initial" beliefs,

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

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 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 $\text{std}(W_{it}) = E(W_{it})\sqrt{e^{\sigma^2} - 1}$, where $\sigma = \log(\frac{C_{it}^4}{C_{it}^2})/2\Phi^{-1}(0.75)$ and Φ is the standard normal cumulative distribution function.

b. Normal. The log-normal distribution imposes an asymmetry that may or may not be present in the data. While the log-normal does have the appealing feature of ruling out negative income, the probability of negative income will tend to be small for the normal distribution when, as we find in our data, the mean is relatively large compared to the standard deviation. As described in Appendix C (available in the Online Supplementary Material (Gong, Stinebrickner, and Stinebrickner (2019))), we find that the fit of the two distributions is quite similar with, if anything, the normal having a slightly

TABLE 2. Earnings beliefs at entrance.

# of Observations: 650	$E(W_{i0})$	std(W_{i0})	$\frac{\text{std}(W_{i0})}{E(W_{i0})}$
Approach 1, Log-normal	51.1742	13.0641	0.2417
	(23.2062)	(15.5580)	(0.2055)
Approach 1, Normal	49.1524	11.3152	0.2295
	(21.9879)	(9.4768)	(0.1617)
Approach 1, Stepwise uniform	49.7633	9.6529	0.1895
	(22.1799)	(8.0391)	(0.1165)
Approach 2, Log-normal	53.3440	14.0952	0.2540
	(26.3336)	(15.6768)	(0.1754)
Approach 2, Normal	51.3609	12.3235	0.2414
	(24.9312)	(10.1175)	(0.1464)
Approach 2, Stepwise uniform	51.8350	10.6937	0.2043
	(25.0832)	(8.5515)	(0.1141)

Note: The unit of measurement for W_{i0} is \$1000. 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}).

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 $\text{std}(W_{it}) = (C_{it}^4 - C_{it}^2)/2\Phi^{-1}(0.75)$.

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

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

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 $\text{std}(W_{it}) = \sqrt{\sum_{n=1}^4 \frac{(C_{it}^{n+1})^2 + C_{it}^{n+1} C_{it}^n + (C_{it}^n)^2}{12} - (E(W_{it}))^2}$.

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 \$9653 and \$13,064 per year and the average standard deviation to mean ratio in the sample varies between 18.95% and 24.17% per year.[12] Thus, the re-

[12]Using log-normal distributions leads to the largest mean and standard deviation approximations and using stepwise uniform distributions leads to the smallest. Note that the distributions constructed using

sults 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.

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 is given by

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

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, $\text{std}(W_{it})$, 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,

$$E(W_{it}) = E_{G_{it}}\big(E(W_{it}|G_{it})\big).^{13} \tag{3}$$

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 B.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 average GPA (i.e., the cumulative GPA) she will receive over all remaining (future) semesters

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.

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

in school. The primary source of information used to construct these beliefs is survey Question 2 (Appendix A), which elicits beliefs about semester GPA. However, even making the natural assumption that Question 2 represents a student's beliefs about semester GPA in each future semester, Question 2 alone is not enough to determine how uncertain a student is about the average GPA she will receive over all remaining semesters. This is the case because one's uncertainty about average GPA over multiple semesters will depend on beliefs about the correlation in semester GPA across semesters. For example, if uncertainty about semester GPA arises because of uncertainty about a factor such as ability that is permanent in nature, and, therefore, will tend to influence grades in each semester, then the uncertainty about semester GPA expressed in Question 2 will tend to be a good indicator of the student's uncertainty about average GPA over multiple semesters. On the other hand, if uncertainty about semester GPA arises because of semester-specific randomness in grades which is transitory in nature and, therefore, will tend to average out to some extent over multiple semesters, then the uncertainty about semester GPA expressed in Question 2 might substantially overstate the student's uncertainty about average GPA over multiple semesters.[14] Our approach for characterizing a student's subjective beliefs about the cumulative GPA she will receive over all remaining semesters differentiates between these two types of possibilities by taking advantage of a novel survey question (Question 3 in Appendix A), which elicited beliefs about the importance of the semester-specific randomness. Appendix B.2 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, $\text{std}(W_{it})$, which is given by

$$\text{std}(W_{it}) = \sqrt{\text{var}_{G_{it}}\big(E(W_{it}|G_{it})\big) + E_{G_{it}}\big(\text{var}(W_{it}|G_{it})\big)}.^{15} \tag{4}$$

The value of $\text{std}(W_{it})$ 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 $\text{var}(W_{it}|G_{it})$ at realizations of G_{it} other than 2.00, 3.00 or 3.75. The details of our interpolation approach are described in Appendix B.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 \$10,694 and \$14,095 per year and the average standard deviation to mean ratio in the sample varies between 20.43% and 25.40% 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.

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

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

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-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 \$1536 higher than nonblacks. Further, the Black coefficient has a t-statistic greater than 1.5 in four of the six distribution-approach combinations, with the maximum t-statistic having a value of 2.6. Comparing these findings to those for our other binary variable, Male, we find that the coefficient for Male also has a t-statistic greater than 1.5 for four of the six combinations, but that the average coefficient for Male over the six distribution-approach combinations is only approximately 62% of the average coefficient for Black.

We stress that understanding the exact interpretation of these results is beyond the scope of this 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 versus 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.[16] Further decompose X_i into factors that are observed by the students at t, X_i^{t-}, and those that are not, X_i^{t+}, and define $X_i \equiv (X_i^{t-}, X_i^{t+})$. Then we can write the future income of student i, W_i, as

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

[16]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 $\text{var}(W_i)$ can be written as a function of the conditional distributions that we observe:

$$\text{var}(W_i) = E_{X_i^{t-}}\big(\text{var}\big(W_i|X_i^{t-}\big)\big) + \text{var}_{X_i^{t-}}\big(E\big(W_i|X_i^{t-}\big)\big). \tag{6}$$

Under the assumption that X_i is independently distributed across students, taking an expectation with respect to X_i^{t-} is, in essence, averaging across individuals (whose beliefs about income at time t differ only through X_i^{t-}).[17] The first term on the right-hand side of equation (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 $\text{var}(W_{it})$. 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, that is, 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.

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

[17] 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	187.4	629.2	816.5	77.05%
Approach 2, Log-normal	444.1	693.5	1137.5	60.96%
Approach 2, Normal	254.1	621.6	875.6	70.98%

Note: The unit of measurement for W_{i0} is \$1000. 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).

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;

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

where ς_i is the classical measurement error attached to the true value $E(W_{it})$. Dispersion in the computed value, $\widetilde{E^1}(W_{it})$, across students originates from both dispersion in the true value, $E(W_{it})$, across students and randomness caused by measurement error, ς_i. This can be seen by taking the variance of both sides of equation (7):

$$\operatorname{var}\big(\widetilde{E^1}(W_{it})\big) = \operatorname{var}\big(E(W_{it})\big) + \operatorname{var}(\varsigma_i). \tag{8}$$

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

Turning to Approach 2, the value $\widetilde{E^2}(W_{it})$ is computed from the responses to questions eliciting beliefs about income conditional on the three particular realizations of final GPA (questions such as 1B) as well as questions eliciting beliefs about grade performance (Questions 2 and 3). Similar to the assumption made in equation (7), we assume that measurement error influences the responses to questions such as 1B in a classical manner, that is,

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

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 $\widetilde{E}(W_{it}|G_{it}=3.75)$:

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

where, as shown in Appendix D (available in the Online Supplementary Material), 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 (available in the Online Supplementary Material) 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:

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

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

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

Using equation (12) to estimate $\text{var}(\varsigma_i)$ requires assumptions about the joint distribution of ς_i, $\varsigma^{2.0}$, $\varsigma^{3.0}$, and $\varsigma^{3.75}$. The prior assumption that ς_i and $\varsigma_i^{g_{it}}$s represent classical measurement error implies that they have mean zero and are independent of other factors. In addition, we assume that the four measurement error terms are independent and identically distributed.

TABLE 4. Heterogeneity and uncertainty (measurement error adjusted).

# of Observations: 650	Measurement Error $\text{var}(\varsigma_i)$	Adjusted Heterogeneity	Adjusted Heterogeneity Ratio
Stepwise uniform	79.0	413.0	72.37%
Log-normal	110.9	427.6	50.91%
Normal	90.6	392.8	64.34%

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.

Under these assumptions, as shown in Appendix E (available in the Online Supplementary Material),

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

Note that we can compute the sample analogs of $\text{var}(\widetilde{E^1}(W_{it}) - \widetilde{E^2}(W_{it}))$ and $E((\lambda_i^{g_{it}})^2)$ from data available to us.[18] Hence, $\text{var}(\varsigma_i)$ can be estimated. The first column of Table 4 reports the estimates of $\text{var}(\varsigma_i)$. 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 $\text{var}(E(W_{it}))$, 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.

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/1B) 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

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

to uncertainty.[19] Thus, it is reasonable to conclude that our findings reinforce the strong message in Cunha and Heckman (2007) that taking into account heterogeneity is essential for characterizing the amount of uncertainty that exists about future earnings at the time of college entrance.

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, finally, Section 4.4 explores the factors that could contribute to this finding.

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.[20]

[19]The former is true if, for example, 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 $E(W_{it})$.

[20]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.

TABLE 5. Contribution of R_{i0}^G: mean and quartiles.

# of Observations: 650	Mean	25%	50%	75%
Stepwise uniform	0.1819	0.0120	0.0737	0.2709
Log-normal	0.1535	0.0090	0.0573	0.2054
Normal	0.1580	0.0102	0.0613	0.2258

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.

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:

$$R_{it}^G = \frac{\text{var}_{G_{it}}\big(E(W_{it}|G_{it})\big)}{\text{var}(W_{it})} = \frac{\text{var}_{G_{it}}\big(E(W_{it}|G_{it})\big)}{\text{var}_{G_{it}}\big(E(W_{it}|G_{it})\big) + E_{G_{it}}\big(\text{var}(W_{it}|G_{it})\big)}. \tag{14}$$

Table 5 summarizes the results for the time of entrance. The first column shows that, on average, 18% of income uncertainty is due to uncertainty about final GPA when we use the stepwise uniform assumption and that, on average, 15% to 16% of income uncertainty is due to uncertainty about final GPA when we use the log-normal or normal distributions. The final three columns show the three quartiles for the three distributional assumptions. For the log-normal and normal distributions, only roughly 25% of students believe that more than roughly 21% to 23% of overall income uncertainty is due to uncertainty about final GPA. For the stepwise uniform case, only 25% of students believe that more than 27% of income uncertainty is due to uncertainty about final GPA. Hence, we conclude that, while uncertainty about grade performance has a nontrivial 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 third (Normal) row of Table 5, the first term in the denominator of equation (14) is 11 times larger for students above the median and the second term in the denominator is 51% smaller for students above the median. Differences in the amount of income uncertainty that is due to GPA could arise, not only because of differences in uncertainty about GPA, but also because of differences in beliefs about how GPA translates to income. We find evidence that, in practice, both of these sources of heterogeneity matter.[21]

[21]Evidence about the importance of the first source of heterogeneity can be seen by computing the sample interquartile range of $\text{var}_{G_{i0}}(E(W_{i0}|G_{i0}))$ assuming that, conditional on a given realization of GPA, all students have identical beliefs about the mean of the subjective conditional income distribution. In practice, we set these means equal to their sample averages. Evidence about the importance of the second

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):

$$\operatorname{var}(W_{it}) = \operatorname{var}_{M_{it}}\big(E(W_{it}|M_{it})\big) + E_{M_{it}}\big(\operatorname{var}(W_{it}|M_{it})\big), \tag{15}$$

where M_{it} is a discrete random variable describing student i's beliefs about final major at time t, which takes on one of seven possible majors j with probability P_{ijt}.[22] The first term on the right-hand side of equation (15) shows how the mean of W_{it} varies across different majors. Thus, it measures the contribution of uncertainty about major to income uncertainty. The second term is an average (across major realizations) of how much uncertainty is present conditional on a particular realization of final major. Thus, it measures the contribution of other factors to 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:

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

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, $\operatorname{var}_{M_{it}}(E(W_{it}|M_{it}))$ and $E_{M_{it}}(\operatorname{var}(W_{it}|M_{it}))$, that enter this fraction. Specifically, while our analysis in Section 4.1 took advantage of the fact that $\operatorname{var}(W_{it}|G_{it})$ is available in the data, $\operatorname{var}(W_{it}|M_{it})$ is not available. However, given information that is observed about $E(W_{it})$, $\operatorname{var}(W_{it})$ and the probabilities P_{ijt}, $j = 1, \ldots, 7$, 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 $j = 1, \ldots, 7$, equation (15) shows that estimating the two terms requires knowledge of $E(W_{i0}|M_{i0})$ and $\operatorname{var}(W_{i0}|M_{i0})$. We estimate these conditional means and conditional variances under the assumption that

source of heterogeneity can be seen by computing the sample interquartile range of $\operatorname{var}_{G_{i0}}(E(W_{i0}|G_{i0}))$ assuming that all students have identical beliefs about final GPA. In practice, we set the parameters of the subjective GPA distribution equal to their sample averages. We find that, depending on which of the three distributional assumptions is used, the interquartile range for the first source of heterogeneity is roughly 35% to 40% as large as the interquartile range for the second source of heterogeneity.

[22]The numbers $1, \ldots, 7$ 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.

they are homogeneous across students conditional on observable characteristics, X_i, that are known to the student at time $t = 0$,

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

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

The unconditional mean $E(W_{i0})$ can be written as $E_{M_{i0}}(E(W_{i0}|M_{i0}))$ and, therefore, is a function of $E(W_{i0}|M_{i0})$ and the random variable M_{i0}. Similarly, the unconditional variance $\operatorname{var}(W_{i0})$ can be written as $\operatorname{var}_{M_{i0}}(E(W_{i0}|M_{i0})) + E_{M_{i0}}(\operatorname{var}(W_{i0}|M_{i0}))$ and, therefore is a function of $E(W_{i0}|M_{i0})$, $\operatorname{var}(W_{i0}|M_{i0})$, 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, ς_i, to the true unconditional mean, $E(W_{i0})$. Similarly, the unconditional variance, $\widetilde{\operatorname{Var}}(W_{i0})$, that is computed from survey Question 1A using Approach 1 is determined by adding classical measurement error, u_i, to the true unconditional variance, $\operatorname{var}(W_{i0})$. This implies that

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

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

Normalizing the Social Science coefficients δ_7 and θ_7 to zero, we estimate the remaining parameters, α_w, β, δ_j, $j = 1, \ldots, 6$, α_v, γ, and θ_j, $j = 1, \ldots, 6$, which are needed to estimate $E(W_{i0}|M_{i0} = j)$, $j = 1, \ldots, 7$ and $\operatorname{var}(W_{i0}|M_{i0} = j)$, $j = 1, \ldots, 7$ (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}, $j = 1, \ldots, 7$ to obtain estimates of α_w, β, and δ_j, $j = 1, \ldots, 6$.

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

[23]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.

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}.

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

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

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.

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

TABLE 7. Estimates for δ_j and θ_j.

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

Note: Note: Equal-tail bootstrap P-values are in the parenthesis. Ranks are in the brackets. 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.

TABLE 8. Uncertainty resolution.

# of Observations: 246		Beginning	Year 1	Year 2	Year 3	End
Sample	Stepwise uniform	10.1310	9.1084	8.3859	8.2887	8.2874
average	Log-normal	13.4582	11.8160	11.0484	11.0632	10.7536
of std(W_{it})	Normal	11.7686	10.7123	10.0112	9.6912	9.6384
Percentage of	Stepwise uniform	N.A.	0.1917	0.3148	0.3306	0.3308
uncertainty	Log-normal	N.A.	0.2291	0.3261	0.3242	0.3615
resolved	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.

of the first year, the end of the second year, the end of the third year, and the time of graduation (End)—for each of our three distributional assumptions, using Approach 1.[24] We restrict our sample to students who answered income expectations questions at all five

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

points. Looking across columns, as would be expected, students become increasingly certain about their future income as they progress through college.[25]

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

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 2 years of college, with little uncertainty resolved after the end of the third year. This finding suggests that learning about future income happens relatively quickly in college. Given evidence that uncertainty about grade performance and major is resolved relatively quickly, the finding is consistent with an environment where learning about grade performance (ability) and major contribute heavily to the total resolution of income uncertainty.

Further, comparing the sum of the contribution of GPA uncertainty (Table 5) and major uncertainty (Table 6) to the results in the last three rows of Table 8 provides some direct evidence about whether this is the case. However, this sum would give a biased view of the joint contribution of GPA and major if these two factors tend to be correlated. The joint contribution of GPA and major is determined by an equation analogous to equation (4) and equation (15):

$$\text{var}(W_{it}) = \text{var}_{G_{it},M_{it}}\big(E(W_{it}|G_{it},M_{it})\big) + E_{G_{it},M_{it}}\big(\text{var}(W_{it}|G_{it},M_{it})\big). \tag{20}$$

The first term on the right-hand side of equation (20), which is the variance of $E(W_{it}|G_{it},M_{it})$ over the joint distribution of G_{it} and M_{it}, represents the joint contribution of uncertainty about final GPA and major to total income uncertainty. The second term on the right-hand side of equation (20), which is the mean of $\text{var}(W_{it}|G_{it},M_{it})$ over the joint distribution of G_{it} and M_{it}, represents the contribution of other factors to total initial income uncertainty. Analogous to equation (14) and equation (16), we define the contribution of final GPA and major to total income uncertainty, R_{it}^{GM}, as the ratio of the first term to the sum of the two terms.

We compute R_{it}^{GM} for the time of entrance ($t = 0$) using a method described in Appendix G (available in the Online Supplementary Material). Table 9 summarizes the results. The first column shows that, on average, 24% of initial income uncertainty is due to uncertainty about final GPA and major when we use the stepwise uniform assumption,

[25]The only exception is a slight increase of sample average of $\text{std}(W_{it})$ 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.

TABLE 9. Contribution of R_{i0}^{GM}: mean and quartiles.

# of Observations: 588	Mean	25%	50%	75%
Stepwise uniform	0.2379	0.0926	0.1993	0.3156
Log-normal	0.1781	0.0589	0.1307	0.2279
Normal	0.2156	0.0864	0.1734	0.2839

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

on average, 18% of initial income uncertainty is due to uncertainty about final GPA and major when we use the log-normal assumption, and, on average, 22% of initial income uncertainty is due to uncertainty about final GPA and major when we use the normal assumption. Thus, the results in Table 9 along with the results in the last three rows of Table 8 do indicate a very substantial role for final GPA and major in the resolution of uncertainty.

4.3.1 *Selection* 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 surprising given that a broad literature finds that transitory components play an important role in the earnings process (Blundell and Preston (1998), Meghir and Pistaferri (2004)). Consistent with these findings, using our post-college data to estimate a random effects model of earnings, we find that the transitory component has a standard deviation of approximately \$9000.[26] While a variety of concerns could arise from comparing this standard deviation from the realized earnings data to standard deviations elicited using expectations questions, it does seem generally relevant that \$9000 is nontrivial when viewed next to the standard deviations in Table 8.

4.3.2 *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 2 years, we regress $\text{std}(W_{i2})$ 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 $\text{std}(W_{i2}) - \text{std}(W_{i0})$, 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 0.1 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 nonblacks.

4.4 *What factors account for end-of-college income uncertainty?*

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

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

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) found that this reason is unlikely to be of particular importance for our recent cohort of college graduates; the large majority of both men and women work full-time throughout their first decade in the labor force, with even departures for children tending to be short.

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

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

We consider several possible reasons for the first explanation. The first reason we consider is that uncertainty may exist about the state of the economy at age 28. To examine this reason, we take advantage of the fact that, as students approached the end of college, the BPS elicited beliefs about not only earnings at the age of 28, but also about earnings in the first year out of college. As shown in the first column of Table 10, at the

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

TABLE 10. Earnings beliefs at the end of college.

	$\text{std}(W_{i4}^{a,1})$		
# of Observations: 359	$a = 1$ Year Out	$a = 28$	$a = 38$
Stepwise uniform	6.3281	8.6233	10.9518
	(4.9854)	(6.9139)	(9.3353)
Log-normal	9.0638	11.4029	14.1175
	(13.0912)	(11.7530)	(14.5838)
Normal	7.2301	9.9188	12.5319
	(5.6186)	(7.9017)	(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 \$1000. A particular entry in the table shows the sample mean and standard deviation of $\text{std}(W_{i4}^{a,1})$ for a particular age a. For example, row 1, column 1 shows a sample mean of \$6328.10 and a sample standard deviation of \$4985.40 for $\text{std}(W_{i4}^{a,1})$ for the age a corresponding to the first post-college year.

end of college ($t = 4$), the average standard deviation of the subjective distribution of earnings in the first post-college year is between 6000 and 9000 dollars, depending on the distributional assumption that is employed. This standard deviation tends to be approximately 75% of the standard deviation associated with age 28 (second column) and approximately 60% of the standard deviation associated with age 38 (third column). The fact that much uncertainty exists for the first year out of school suggest that, at the very least, factors other than the state of the economy are influencing income uncertainty.

Roughly speaking, we could group the remaining reasons for the first explanation under the heading of frictions. One possibility is that information frictions are present. For example, students may begin school with uncertainty about the type of job opportunities that tend to be available for college graduates, and this uncertainty may not be entirely resolved even by the end of college (Betts (1996)). It is somewhat difficult to provide direct evidence about the importance of this type of friction. However, we are able to provide some evidence about a second potential type of frictions—labor market/search frictions. The first piece of evidence comes from survey Question 6. Although we found that more than 80% of students prefer a Degree-My-Area job, Question 6 indicates that, on average, students believe there is only a 50% chance of ending up in such a job in the first year. Further, while almost no students prefer a No-Degree-Needed job, on average, students believe there is almost a 20% chance of being forced to accept this type of job. The second piece of evidence comes from survey Question 8. On average, students believe that there is a 22% probability that it will take five or more months of search to find a job. Further, on average, students believe that there is only a 20% chance of obtaining a job with less than one month of search.[28] While we stress that it is not possible to determine the relative importance of the different reasons for the first explanation, the results suggest that search frictions are likely to be relevant.

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

5. Conclusion

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

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

Appendix A: Survey questions

Question 1. The following questions will ask you about the income you might earn in the future at different ages under several hypothetical scenarios. We 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.

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.

NOTE TO READER. The remainder of Question 1B (not shown) was identical to parts I, II, and III of Question 1A. In additional parts Question 1 (also not shown), students were asked to consider scenarios in which they graduate with other grade point averages (GPA 3.00 and 3.75).

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

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

Interval	**Percent Chance (number of chances out of 100)**
[3.5, 4.00]	_____
[3.0, 3.49]	_____
[2.5, 2.99]	_____
[2.0, 2.49]	_____
[1.0, 1.99]	_____
[0.0, 0.99]	_____

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

QUESTION 3. Your grades are influenced by your academic ability/preparation and how much you decide to study. However, your grades may also be influenced to some extent by good or bad luck which may vary from term to term and may be out of your control. Examples of "luck" may include (note: examples omitted due to space considerations).

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? _____

NOTE. Two subsequent (identical) questions (not shown) asked about the percentage of semesters that bad luck would lower GPA by (1) between 0.26 and 0.50 points and (2) by more than 0.51 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.

NOTE. The three questions in this section (not shown) followed closely the questions in the BAD LUCK section, eliciting the percentage of semesters that good luck would raise GPA by: (1) between 0.00 and 0.25 points, (2) between 0.26 and 0.50 points, and (3) greater than 0.51 points or more.

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). In the third column, please write down the grade point average (GPA) you would expect to receive in a typical semester in the future if you had each of these areas of study.

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

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

QUESTION 6. After graduating, there are different types of jobs that you may hold. For Questions 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

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

NOTE. The wording of 7.2 (not shown) is identical to the wording of 7.1 with DEGREE-ANYAREA replacing NO-DEGREE-NEEDED, and NO-DEGREE-NEEDED replacing DEGREE-ANYAREA.

7.3 IF you circled DEGREE-MYAREA

NOTE. The wording of 7.3 (not shown) is identical to the wording of 7.1 with DEGREE-MYAREA replacing NO-DEGREE-NEEDED, and NO-DEGREE-NEEDED replacing DEGREE-MYAREA.

QUESTION 8. Suppose during this school year that you searched seriously for a job. You may not know exactly how long it would take to find a job. What is the percent chance that it would take the following amounts of time to receive a job offer from the time you start searching seriously?

Note: A serious job search is one that involves actively looking for a job by participating in activities such as on-campus interviewing, reading and responding to want ads, or contacting potential employees even if they have not posted want ads.

Amount of time to find a job-interval	Percent chance
[0, 1) months	_____
[1, 2) months	_____
[2, 3) months	_____
[3, 5) months	_____
[5, 6) months	_____
6 months or more	_____

APPENDIX B: APPROACH 2: COMPUTATION DETAILS

B.1 *Construction of* $E(W_{it}|G_{it} = g_{it})$ *and* $\text{std}(W_{it}|G_{it} = g_{it})$ *(or, equivalently,* $\text{var}(W_{it}|G_{it} = g_{it})$*) 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 $\text{std}(W_{it}|G_{it} = g_{it})$ (henceforth, $\text{std}(W_{it}|g_{it})$) for $g_{it} = 2.00, 3.00$, or 3.75. However, we need to approximate $E(W_{it}|g_{it})$ and $\text{std}(W_{it}|g_{it})$ for all other possible values of g_{it}. Following a straightforward

interpolation approach adopted in Stinebrickner and Stinebrickner (2014b), we assume that both $E(W_{it}|g_{it})$ and std$(W_{it}|g_{it})$ are linear between $g_{it} = 2.00$ and $g_{it} = 3.00$. We also assume that $E(W_{it}|g_{it})$ and std$(W_{it}|g_{it})$ are linear between $g_{it} = 3.00$ and $g_{it} = 4.00$, with the slope being identified by the observed values at $g_{it} = 3.00$ and $g_{it} = 3.75$ (i.e., we extrapolate values of $E(W_{it}|g_{it})$ and std$(W_{it}|g_{it})$ between $g_{it} = 3.75$ and $g_{it} = 4.00$).

B.2 *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, $k = 1, \ldots, 8$, subject to the constraint that the student obtains the 2.0 average that is needed to graduate. Thus, G_{i0} is given by

$$G_{i0} = \sum_{k=1}^{8} G_{i0}^k/8, \quad \text{if } \sum_{k=1}^{8} G_{i0}^k/8 \geq 2, \tag{21}$$

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, $k = 1, \ldots, 8$. 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:

$$\begin{aligned} &\tilde{G}_{i0}^k = a_{i0} + \xi_{i0}^k, \quad \text{where} \\ &G_{i0}^k = 0 \quad \text{if } \tilde{G}_{i0}^k < 0, \qquad G_{i0}^k = 4 \quad \text{if } \tilde{G}_{i0}^k > 4, \quad \text{and} \quad G_{i0}^k = \tilde{G}_{i0}^k, \quad \text{otherwise.} \end{aligned} \tag{22}$$

a_{i0} represents student i's ($t = 0$) beliefs about permanent (academic) ability and ξ_{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 ξ_{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.[29] Formally, we have

$$\begin{aligned}\{\widehat{\mu_{i0}^a}, \widehat{\sigma_{i0}^a}, \widehat{\sigma_{i0}^{\xi}}\} = \operatorname{argmin} & \sum_{\text{cat}_j^g \in \text{CAT}^g} \Pr_{\text{model}}\big(G_{i0}^k \in \text{cat}_j^g\big) \\ & \times \big(\Pr_{\text{obs}}\big(G_{i0}^k \in \text{cat}_j^g\big) - \Pr_{\text{model}}\big(G_{i0}^k \in \text{cat}_j^g\big)\big)^2 \\ & + \sum_{\text{cat}_j^{\xi} \in \text{CAT}^{\xi}} \Pr_{\text{model}}\big(\xi_{i0}^k \in \text{cat}_j^{\xi}\big) \\ & \times \big(\Pr_{\text{obs}}\big(\xi_{i0}^k \in \text{cat}_j^{\xi}\big) - \Pr_{\text{model}}\big(\xi_{i0}^k \in \text{cat}_j^{\xi}\big)\big)^2, \end{aligned} \tag{23}$$

where $\text{CAT}^g = \{[3.5, 4.00], [3.0, 3.49], [2.5, 2.99], [2.0, 2.49], [1.0, 1.99], [0.0, 0.99]\}$, and $\text{CAT}^{\xi} = \{(-\infty, -0.5], (-0.5, -0.25], (-0.25, 0], (0, 0.25], (0.25, 0.5], (0.5, \infty)\}$.

Once parameters $\{\mu_{i0}^a, \sigma_{i0}^a, \sigma_{i0}^{\xi}\}$ are estimated, we can approximate the distribution of G_{i0} by simulation using equations (21) and (22).

References

Acemoglu, D. and D. Autor (2011), "Skills, tasks and technologies: Implications for employment and earnings." In *Handbook of Labor Economics*, Vol. 4, 1043–1171, Elsevier. [631]

Altonji, J. G. (1993), "The demand for and return to education when education outcomes are uncertain." *Journal of Labor Economics*, 11 (1, Part 1), 48–83. [609]

Altonji, J. G., P. Arcidiacono, and A. Maurel (2016), "The analysis of field choice in college and graduate school." *Handbook of the Economics of Education*, 5, 305–396. [624]

Altonji, J. G., E. Blom, and C. Meghir (2012), "Heterogeneity in human capital investments: High school curriculum, college major, and careers." *Annual Review of Economics*, 4 (1), 185–223. [624]

Arcidiacono, P. (2004), "Ability sorting and the returns to college major." *Journal of Econometrics*, 121 (1), 343–375. [611]

Attanasio, O. P. and K. M. Kaufmann (2014), "Education choices and returns to schooling: Mothers' and youths' subjective expectations and their role by gender." *Journal of Development Economics*, 109, 203–216. [611]

[29] We have also estimated a nonweighted version. The results are similar.

Autor, D. H. and M. J. Handel (2013), "Putting tasks to the test: Human capital, job tasks, and wages." *Journal of Labor Economics*, 31 (S1), S59–S96. [631]

Betts, J. R. (1996), "What do students know about wages? Evidence from a survey of undergraduates." *Journal of Human Resources*, 27–56. [632]

Blundell, R. and I. Preston (1998), "Consumption inequality and income uncertainty." *The Quarterly Journal of Economics*, 113 (2), 603–640. [609, 630]

Browning, M. and J. Carro (2007), "Heterogeneity and microeconometrics modeling." *Econometric Society Monograph*, 43, 47. [609]

Carneiro, P., K. T. Hansen, and J. J. Heckman (2003), "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. [608]

Cunha, F., J. J. Heckman, and S. Navarro (2004), "Separating heterogeneity from uncertainty an Aiyagari–Laitner economy." In *Goldwater Conference on Labor Markets in Arizona*. [608]

Cunha, F., J. J. Heckman, and S. Navarro (2005), "Separating uncertainty from heterogeneity in life cycle earnings." *Oxford Economic Papers*, 57 (2), 191–261. [608, 609, 633]

Cunha, F. and J. J. Heckman (2007), "Identifying and estimating the distributions of ex post and ex ante returns to schooling." *Labour Economics*, 14 (6), 870–893. [609, 617, 621, 622]

Dominitz, J. and C. F. Manski (1997a), "Perceptions of economic insecurity: Evidence from the survey of economic expectations." *Public Opinion Quarterly*, 261–287. [608]

Dominitz, J. and C. F. Manski (1997b), "Using expectations data to study subjective income expectations." *Journal of the American Statistical Association*, 92 (439), 855–867. [608]

Friedman, M. (1953), "Choice, chance, and the personal distribution of income." *Journal of Political Economy*, 61 (4), 277–290. [607]

Gibbons, R. and L. Katz (1992), "Does unmeasured ability explain inter-industry wage differentials?" *The Review of Economic Studies*, 59 (3), 515–535. [631]

Gong, Y., Stinebrickner, T. Stinebrickner, R. (2019), "Supplement to 'Uncertainty about future income: Initial beliefs and resolution during college'." *Quantitative Economics Supplemental Material*, 10, https://doi.org/10.3982/QE954. [613]

Heckman, J. J. and G. Sedlacek (1985), "Heterogeneity, aggregation, and market wage functions: An empirical model of self-selection in the labor market." *Journal of Political Economy*, 93 (6), 1077–1125. [631]

Manski, C. F. (1993), "Adolescent econometricians: How do youth infer the returns to schooling?" In *Studies of Supply and Demand in Higher Education*, 43–60, University of Chicago Press. [608]

Manski, C. F. (2004), "Measuring expectations." *Econometrica*, 72 (5), 1329–1376. [608, 613]

Meghir, C. and L. Pistaferri (2004), "Income variance dynamics and heterogeneity." *Econometrica*, 72 (1), 1–32. [630]

Navarro, S. and J. Zhou (2017), "Identifying agent's information sets: An application to a lifecycle model of schooling, consumption and labor supply." *Review of Economic Dynamics*, 25, 58–92. [609]

Stinebrickner, R. and T. Stinebrickner (2008), "The effect of credit constraints on the college drop-out decision: A direct approach using a new panel study." *American Economic Review*, 98 (5), 2163–2184. [610]

Stinebrickner, R. and T. Stinebrickner (2012), "Learning about academic ability and the college dropout decision." *Journal of Labor Economics*, 30 (4), 707–748. [609, 622]

Stinebrickner, R. and T. Stinebrickner (2014a), "A major in science? Initial beliefs and final outcomes for college major and dropout." *Review of Economic Studies*, 81 (1), 426–472. [610, 624]

Stinebrickner, R. and T. Stinebrickner (2014b), "Academic performance and college dropout: Using longitudinal expectations data to estimate a learning model." *Journal of Labor Economics*, 32 (3), 601–644. [609, 615, 622, 638]

Stinebrickner, R., T. Stinebrickner, and P. Sullivan (2018), "Job tasks and the gender wage gap among college graduates." NBER Working Papers 24790. [631]

Zafar, B. (2011), "How do college students form expectations?" *Journal of Labor Economics*, 29 (2), 301–348. [609]

Co-editor Peter Arcidiacono handled this manuscript.

Manuscript received 18 August, 2017; final version accepted 25 August, 2018; available online 2 October, 2018.