Academic performance and college dropout:
Using longitudinal expectations data to estimate a learning model

Abstract: We estimate a dynamic learning model of the college dropout decision, taking advantage of unique expectations data to greatly reduce our reliance on assumptions that would otherwise be necessary for identification. We find that forty-five percent of the dropout that occurs in the first two years of college can be attributed to student learning about academic performance, but that the importance of this type of learning becomes largely irrelevant after the midway point of college. We use our model to quantify the importance of the possible avenues through which poor grade performance could influence dropout.

Section I. Introduction

The importance of understanding why many entering college students do not complete a degree has been widely recognized (Bowen et al., 2009). Dropout that arises naturally as students figure out whether their skills/interests are a good match for a career that requires a college education may not be unappealing. On the other hand, dropout due to, for example, imperfections in credit markets, difficulties in scheduling courses, avoidable social difficulties, or a lack of reasonable levels of encouragement from parents is more concerning. Unfortunately, much remains unknown about the underlying determinants of dropout.[1] In this paper we provide new evidence about this issue by using unique new expectations data to estimate a dynamic decision model which pays close attention to what a student learns about his/her academic performance after entering college.

That much remains unknown about how students make the dropout decision can be attributed, in large part, to the difficulty of obtaining ideal data. Administrative records are a natural source of information, but studies relying exclusively on data of this type may struggle with difficulties related to the proverbial black box; students can be seen entering college with certain observable characteristics and can be seen leaving college with certain outcomes, but what happens in between remains somewhat of a mystery. General longitudinal surveys present an opportunity to collect information that is not available in administrative data, but may suffer from issues related to timing and frequency. From a conceptual standpoint, the dropout outcome is best viewed as the end result of a process in which a student learns about a variety of utility-influencing factors after arriving at school (Manski, 1989; Altonji, 1993; Stange, forthcoming). If this learning tends to take place quickly, annual surveys may miss the entire period that a particular student is in school, and, more generally, will have difficulty capturing the changes in beliefs about utility-influencing factors that characterize learning.

The limitations present in existing data sources motivated our initiation of the Berea Panel Study (BPS), a unique longitudinal survey of students who entered Berea College in 2000 and 2001. Located in central Kentucky, Berea College operates with a mission of providing an education to students of "great promise but limited economic resources." Thus, given that students from low socioeconomic backgrounds are known to have much higher dropout rates than other students, the school has a demographic focus that is of particular interest to policymakers. Data collection was guided directly by

[1] Differences in college dropout by family income have been found to be at least as important as differences in college entrance by family income from the standpoint of creating differences in college degree attainment by family income (Manski and Wise, 1983; Manski, 1992; NCES, 2007).

Describing the traditional difficulties of understanding the underlying reasons for dropout, Bowen and Bok (1998) write, "One large question is the extent to which low national graduation rates are due to the inability of students and their families to meet college costs, rather than to academic difficulties or other factors." Tinto (1975) suggests that dropout is related to academic and social integration, but direct tests of this are scarce (Draper, 2005).

models in which students learn about the costs and benefits associated with the alternatives they consider when making choices. This motivated a survey design in which each student was surveyed approximately twelve times each year while in school, with, importantly, the first survey taking place immediately before the beginning of the student's freshman year and the last survey related to the schooling period taking place immediately after the student left school. As such, the dataset that results from linking the new survey data to administrative records is truly unique in the depth and detail it provides about the full college period.

In two papers that serve as background, we exhibited the benefits of our detailed case study by exploring the importance of common financial resource explanations for dropout. Stinebrickner and Stinebrickner, hereafter S&S, (2003a) documented that, similar to what is seen for students with comparable family income and college entrance scores elsewhere, between forty and fifty percent of the entering students at Berea fail to graduate (and few transfer) - even though the direct costs of schooling are zero (or perhaps negative) due to a full tuition subsidy and room and board subsidies for all students. Taking advantage of unique survey questions in the BPS, S&S (2008a) found that, while credit constraints do influence the decisions of a small number of students by making it difficult to smooth consumption between the schooling and working portions of their lives, they do not play a substantial role in determining the overall dropout rate of students at Berea. Thus, our background work shows that factors unrelated to financial resources during college per se play the prominent role in the dropout of these low income students. This motivates our current objective of examining the process through which these non-financial resource factors may influence the dropout decision.

We focus primarily on understanding the importance of the most widely recognized non-financial resource explanation - that after entering college, students learn about how well they will perform academically. In S&S (forthcoming) we began our examination into the importance of this explanation by studying dropout during one particular semester. Contributing some of the strongest direct evidence to-date to a literature recognizing the importance of learning in determining schooling outcomes (Manski, 1989; Altonji, 1993; Carneiro et al., 2005; Cunha et al., 2005; Stange, forthcoming), we found that 40% of dropout between the beginning of the second semester and the start of the third semester can be attributed to this type of learning. In this paper, we build considerably on this result in two important ways.

First, we characterize the amount of dropout that can be attributed to learning about academic performance over the entire college period. Generally speaking, it is important to widen our analysis to cover the entire period because non-trivial dropout tends to be present at all stages of college (Bowen et al., 2009). More specifically, it is important to examine whether the amount of dropout that can be attributed to this type of learning varies across stages of college. Given that higher education tends to be highly subsidized and that scarce public resources can be consumed inefficiently if persistent misperceptions lead students to remain in school longer than they otherwise would, it is desirable for students to learn quickly about how well they will perform academically.

Second, we provide some of the first evidence about *why* grade performance is consistently found to be strongly related to dropout by differentiating between three possible avenues through which grades can affect the dropout decision. The first avenue is that poorly performing students would like to stay in school but are forced out of school by grade progression cutoffs. The second avenue is that poor grade performance lowers the financial return to remaining in school. The third avenue is that poor grade performance reduces how enjoyable it is to be in school. There are many ways in which understanding why grade performance matters can be helpful to policymakers, parents, and students. As one example, a finding that dropout is primarily generated by the first two avenues might suggest that certain prevalent policy interventions which have the goal of reducing dropout at current levels of academic performance (e.g., counseling aimed at reducing stress) may not be overly effective. In the Conclusion we discuss policy implications in more detail given our specific findings.

In S&S (forthcoming), the use of a reduced form model was natural given our objective of relating what a person learned about his academic ability/performance to his schooling decision in one particular semester. Here, the estimation of an explicit dynamic, decision model is useful given our desire to understand the importance of this type of learning over the entire college period and is critical given our desire to differentiate between various possible explanations for why grade performance matters. As one example of why a model is needed, the importance of the first avenue in the previous paragraph cannot be determined simply by observing whether students have binding grade cutoffs in the semester they leave school because current period decisions can also be influenced by how likely it is that grade cutoffs will bind in the future.[2] As a second example of why a model is needed, understanding the importance of the second avenue in the previous paragraph requires characterizing, for each semester, the financial return to staying in school for an additional semester. Much of this financial return may arise because staying for an additional semester allows a student to retain the option of continuing towards higher levels of education, including graduation, in the future. The financial value of the option, which involves a weighted average over all outcomes that could occur in the future, cannot be observed directly in the data but is constructed naturally by a model in which students are forward looking.

---

[2]This may be quite relevant because schools typically have grade cutoffs that become substantially higher over time. For example, at Berea College, the minimum grade requirement is 0.0 for the second semester but eventually reaches 2.0.

The need for a model highlights the fundamental tension present in empirical micro-economics that motivated our initiation of the BPS; while structural models formed directly from economic theory represent a potentially powerful tool for understanding the mechanisms that underlie individual decision-making and for providing pre-implementation evidence about the effects of possible policy changes, their practical usefulness will be undermined if concerns about the validity of central assumptions lead to concerns about the identification of model parameters. Given the learning model described here, of central importance are beliefs about: academic performance, the financial returns to schooling, and how much a student will enjoy school. Thus, empirical work that is closely tied to theory relies heavily on the characterization of individual-specific beliefs about these factors throughout the time a student is in school. Traditionally, researchers working with models which require beliefs have relied on assumptions that allow them to characterize beliefs indirectly. For example, a common assumption, often referred to as Rational Expectations (RE), is that an individual's beliefs about a particular factor (e.g., grade performance) coincide with the actual distribution from which that factor is drawn (Das and van Soest, 2000). However, recent research such as Manski (2004) has stressed that these types of assumptions are arbitrary and untestable, in which case their use in the estimation of models of behavior may raise concerns about identification.

A natural alternative to indirect approaches for characterizing beliefs is to use carefully worded survey questions in order to elicit beliefs directly (Dominitz, 1998; Dominitz and Manski, 1996, 1997). Then, our work here is made possible because the BPS was perhaps the first sustained longitudinal survey to have a central focus on these types of questions. In S&S (forthcoming) we found evidence of the value of these questions; using reduced form models we found that certain implications of simple models of dropout were satisfied when we used our directly elicited beliefs but were not satisfied when beliefs were constructed in ways that are consistent with standard RE assumptions. This paper takes the next step of using expectations data fully in explicit models of behavior in order to reduce reliance on otherwise necessary assumptions. As such, our project provides evidence about the value of collecting survey data with very detailed issues/models in mind.[3]

Section II. The Berea Panel Study, the sample, and motivating descriptive statistics

Designed and administered by Todd Stinebrickner and Ralph Stinebrickner, the BPS is a longitudinal survey that takes place at Berea College and elicits information of relevance for

[3] For other work that uses expectations data to examine educational decisions see, for example, Zafar (2011), Arcidiacono et al. (forthcoming), and Attanasio and Kaufmann (2009). For early work that uses expectations data somewhat differently in structural models see van der Klaauw (2000) and also van der Klaauw and Wolpin (2008).

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 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. In addition to collecting detailed background information, the baseline surveys were designed to take advantage of recent advances in survey methodology (see, e.g., Barsky et al., 1997; Dominitz, 1998; and Dominitz and Manski, 1996, 1997) in order to directly elicit individual-specific expectations towards uncertain outcomes and the factors that might influence these outcomes. Substantial follow-up surveys that were administered at the beginning and end of each subsequent semester document how expectations change over time.

Because some survey questions of interest are not available for the 2000 cohort, we focus on the 2001 cohort. Approximately 88% of all students who entered Berea in the Fall of 2001 participated in the BPS survey. S&S (forthcoming) found that few students who leave Berea transfer to other four year schools. We exclude students who transfer, but stress that results change very little under different treatments of these students (e.g., treating schooling spells as being censored at the time that transfer students leave Berea). Our sample contains 341 students.

In order to obtain standard observable characteristics, X_i, the BPS survey data are linked to administrative data from Berea College. We focus primarily on a student's sex and his/her high school grade point average. The proportion of students that are male is 44.57% and the average (std. deviation) high school grade point average is 3.37 (.46). The academic credentials of students at Berea, including college entrance exam scores (average 23.35, std. deviation 3.60), are similar to those at the University of Kentucky and the University of Tennessee (S&S, 2008a). Our sample can be generally thought of as a group of students from low income families (average family income $26,000, std. deviation family income $17,000), and for most of our analysis we do not differentiate by family income within the sample.

This paper is motivated most generally by the reality that, consistent with what is seen for students from low income families elsewhere (S&S, 2008a; Manski, 1992), the dropout rate at Berea is substantial. The outcome variable we examine here is whether a student leaves school for at least a semester at any point during the first 3.5 years of school. Nine percent, 18%, 26%, 34%, 39%, and 46%, respectively, of the students in our sample have left school as of the start of the second, third, fourth, fifth, sixth, and seventh semesters, respectively. The use of the term dropout would be a misnomer to the extent that students who leave Berea return and complete a degree in the future. However, this is quite rare. For example, only ten percent of the students who left school at any time before the start of the seventh semester subsequently returned to school and were still enrolled at the start of the eighth semester. We also find that leaving school is very rare for those who have not left as of the seventh semester. For example, only two percent of the individuals who were in school for the seventh semester were not in school for the eighth semester.

From a theoretical standpoint, the presence of a substantial amount of dropout does not necessarily imply that students have updated, for example, the mean of the distribution describing their beliefs about grade performance (or other factors of relevance). For example, when the labor market returns to improved academic performance are non-linear, one might, in theory, decide to enter school knowing that he will drop out unless his grade performance is substantially better than expected. However, if a person believes at entrance that dropout is very unlikely, then an observed dropout outcome would seemingly suggest that some type of learning about expected grade performance (or other factors of relevance) has taken place. Then, the survey question below, administered at the time of entrance, provides some motivation for the learning nature of our model by showing that dropout is viewed as quite unlikely. While less than 60% of students in the sample will graduate, students, on average, believed at entrance that there was an 86% chance of graduation.

Question: What is the percent chance that you will eventually graduate from Berea College?____

Section III. A model of dropout

III.A. Basic setup and choices

We consider a simple dynamic model of sequential decision-making under uncertainty, simplifying the discussion by assuming that schooling decisions take place between semesters.[4] A student enters school with (potentially interrelated) beliefs about his future grade performance, how much he will enjoy school, and his future earnings under a variety of schooling scenarios. During the first semester, the student receives utility from being in school. At the end of the first semester, the student checks to see if he is being forced to leave college and enter the labor force (N) due to poor academic performance. If not forced to leave, he uses new information that was received during the first semester to update the beliefs he held at entrance. He then decides whether to return to school (S) for semester t or whether to enter the labor force (N) by comparing the discounted expected utilities associated with these options. Given our earlier findings that few people return to school after leaving, we assume that N is a terminal state. For students who choose to return to school, the process above is

[4] In our empirical work, a student is classified as leaving at t if he began semester t-1 and did not return for semester t. The choice of how to group students is not overly important given that the large majority of departures take place between semesters.

repeated until the student is forced out of school due to poor academic performance, chooses to leave school, or graduates.

Among the set of potential academic-related objects of interest, our focus on what a student learns about his grade performance seems natural because, among other reasons, grade performance determines whether a student fails out of school and is often of direct interest to potential employers. One might also be interested in what a student learns about his academic ability, perhaps defined to be his grade point average holding study effort and course difficulty constant. However, S&S (forthcoming) find that making a distinction between academic performance and ability is not particularly important in these data since the large majority of what a person learns about his grade performance is due to what he learns about his academic ability (rather than what he learns about his effort or course difficulty).

We have specified an extremely parsimonious choice set $\{S,N\}$. This parsimony does imply that our model cannot be used to examine how students make other important decisions such as: how much to study, what major to choose, and whether to attend graduate school after college. However, the relevant question here is whether the parsimony is, on net, beneficial given our objectives of: 1) understanding the effect that learning about academic performance has on dropout and 2) differentiating between several broad reasons for why this type of learning matters. The benefit of parsimony is model simplicity/transparency, including reductions in the number of required assumptions and the number of model parameters. The cost of parsimony would arise if modelling additional endogenous choices would allow a more accurate characterization of the current or future utility associated with the alternatives of interest, S and N. Given this tradeoff, our choice set was pushed in a parsimonious direction by the fact that survey questions in the BPS were designed to allow many of the utility effects of additional choices to be captured, even when these choices are not modelled endogenously.

As a concrete illustration, consider a student's decision about how much to study. Central to our model is what the student learns about his grade performance, his future earnings, and how much he will enjoy school. Then, if our model has the flexibility to capture the impact that decisions about how much to study have on what a person learns about each of these factors, then we can avoid modelling the study decision explicitly because it is not necessary for our purposes to characterize exactly how much of what a person learns about each of these factors is due to the study decisions per se. As will be discussed in more detail, the impact of studying on what a student learns about his grade performance is captured by frequent survey questions eliciting beliefs about grade performance. The impact of studying on what a student learns about his future income is captured by survey questions eliciting beliefs about the relationship between grade performance and income. The impact of studying on what a student learns about the enjoyability of school (e.g., the cost of foregone leisure that accompanies additional studying) is captured by frequent questions about how enjoyable it is to be in school relative to being out of school.

We also do not model the decision of how many courses to take in a particular semester. As such, we are assuming that students who do not fail out of school make steady progress towards graduation, with this progress characterized by the number of semesters attended. This does not seem overly restrictive given that Berea requires full-time attendance and given that grade cutoffs for failing out of school are meant to identify those who are not progressing in a timely fashion. Given that roughly seventy-five percent of students who graduate do so in four years, we assume that students who choose to return for the eighth semester will graduate at the end of that semester. Thus, referring to the start of semester t as "time t," with t=1 being the time of entrance and t=9 being the time of graduation, a student makes a choice from the set {S, N} at any of the times t=2, t=3,..., t=8 for which he is still in school.

III.B. Value functions

Our emphasis on understanding the importance of learning suggests the desirability of a dynamic, forward-looking model. The fundamental object needed for estimation is the discounted expected utility, or value, associated with the two options {S,N} that a person considers at each time t that he is still in school and has the option of continuing. In this subsection we describe the value functions in general terms. In subsequent subsections we describe the components of the value functions in more detail.

Let $U^N_t(\Omega_t)$ be the current period utility for a person who is in the workforce at time t with a state of Ω_t. Then, for a person who is still in school at the end of semester t-1, the value of entering the workforce (N) at time t is:

$$(1)\ V^N_t(\Omega_t) = E\sum_{\tau=t}^{T*} \beta^{\tau-t} U^N_\tau(\Omega_\tau),$$

where T* is the end of a person's utility horizon, β is the discount factor, and, to be consistent with the reality that there are two semesters in each year, each period in the workforce represents six months.

Let $U^S_t(\Omega_t)$ be the current period utility for a person who is in school at t with a state of Ω_t. For a person who is still in school at the end of semester t-1 and is not forced to leave due to poor academic performance or graduation, the value of returning to school (S) for semester t is given by the Bellman equation:

$$(2)\ V^S_t(\Omega_t) = EU^S_t(\Omega_t) + (PrFail_t)\ \beta V^N_{t+1}(\Omega_{t+1}) + (1-PrFail_t)\ \beta E\max[V^S_{t+1}(\Omega_{t+1}), V^N_{t+1}(\Omega_{t+1})].$$

The first term is the expected current reward of being in college at time t. The second term indicates that with probability $PrFail_t$ the student will fail out of school at the end of semester t, in which case he will be forced to enter the workforce permanently. The third term indicates that with probability 1-$PrFail_t$ the student will not fail out of school at the end of semester t, in which case he will have the option of returning to school for semester t+1 or entering the workforce.[5] The expected value in the third term is over all elements of Ω_{t+1} whose values are not known at time t given Ω_t and the choice of S at t.

We complete our description of the model by specifying the functions U^N_t and U^S_t in III.C and by describing the elements of Ω_t and how these elements evolve between t and t+1 in III.D.

III.C Current period utility

We assume that $U^N_t(\bullet)$ is linear in C_t, a person's consumption at time t. Letting $\varepsilon_{N,t}$ represent a period-specific, idiosyncratic shock to the utility derived from option N that is known to the individual but not the econometrician,

(3) $U^N_t(\bullet)=C_t+\varepsilon_{N,t}$.

The assumption that the utility in Eq. (3) is linear in consumption facilitates an easy interpretation of model parameters (Section III.F), is convenient for characterizing expected future utility (Section III.D), and allows us to avoid estimating potentially hard-to-identify parameters associated with the curvature in the utility function. As in other recent work in this area (Stange, forthcoming) this assumption is made primarily for convenience. However, the possible concerns about this assumption are somewhat mitigated since, as discussed in the next subsection, we do not explicitly model consumption during school, the period when consumption would most likely be at low levels where differences between a linear and non-linear assumption for the utility function would be most important. In addition, survey questions eliciting beliefs about the minimum income a person might receive in the future reveal little concern that post-college earnings might turn out to be close to zero in a particular year.

One could assume that the function U^S_t is identical in form to the function U^N_t, in which case the (average) utility difference between a schooling period (S) and a non-schooling period (N) is simply the difference in a person's consumption between the two periods. However, such an approach is worrisome because: 1) even if the amount of his own money that a student spends on consumption while in school is observed, it may be difficult to measure actual consumption while in school because there are types of consumption that are provided free of charge on a college campus (e.g., computing resources, television, etc.) and 2) the potential for certain types of leisure activities on a college campus that may not be available outside of school suggests that the mapping from consumption goods to utility may be quite different in S and N. S&S (2008a) found evidence that students believe that they are smoothing the marginal current period utility from consumption between the schooling and working portions of their lives even when they have little of their own money to spend on consumption during school.

[5]This equation generally represents the case where the student does not graduate at the end of semester t (i.e., t+1<9). However, setting $PrFail_t=1$ provides the case of a person who will graduate at end of t.

The general difficulty of understanding how much utility a person receives while in school motivated us at the beginning of each semester to use Survey Question A.1 (Appendix A) to directly measure the object of interest - how much a student enjoys being in school relative to the alternative of being in the workforce. Central to our construction of the current period utility function is the binary variable EN_t which has a value of one if a person reports at the beginning of time t that he believes that being in school is more enjoyable than being out of school (i.e., a person circles 1 or 2 on A.1).

While, in theory, EN_t might capture all academic aspects of relevance for characterizing current period utility, in practice, there are reasons that this might not be the case. First, while the effect of academic measures on enjoyability might be taken into account in answers to Question A.1, in practice, it is difficult to know exactly what students condition on when answering the question. For example, perhaps students think largely about the social part of schooling when answering A.1 or tend to consider the effect of grades in a best-case type scenario. Then, even after taking into account EN_t, $U^S_t(\bullet)$ may depend on a person's cumulative grade point average G_t at the beginning of t and his grade performance g_t in semester t. g_t may influence utility because school may be unenjoyable if a person has difficulty understanding course material and G_t may influence utility conditional on g_t because school may be particularly stressful if a person believes that he is close to failing out. Second, while our interest in understanding the full impact of learning about grade performance implies that $U^S_t(\bullet)$ should capture all non-earnings avenues through which poor academic performance may influence dropout, it is not clear whether answers to A.1 would take into account, for example, that families may provide less encouragement to stay in school when grade performance is bad. Finally, a concern in certain policy circles is that students may have a knee-jerk reaction to bad grade outcomes. In this case, if A.1 is collected somewhat after a student leaves school, EN_t may not capture the entire effect that grades had on the exit decision. Motivated by this discussion, we specify $U^S_t(\bullet)$ as

(4) $U^S_t(\bullet)=\gamma_0+\gamma_1 EN_t + \gamma_2 G_t + \gamma_3 g_t + \varepsilon_{S,t}$,

where $\varepsilon_{S,t}$ is the analog to $\varepsilon_{N,t}$. We define $\varepsilon_t=\{\varepsilon_{N,t}, \varepsilon_{S,t}\}$. Our particular interest in academic issues motivated our inclusion of the grade variables in (4). However, some of the arguments in the previous paragraph might also suggest that other factors, such as student health or whether a parent lost a job, might also not be fully captured by EN_t. Then, even with an objective of quantifying the importance of learning about academic performance, one might wish to also include these factors explicitly in (4) if it is possible that they might be correlated with what a person learns about his grade performance. We do this as a robustness check in Section VI.

An examination of Eqs. (3) and (4) reveals how, as discussed in Section III.A, our model is flexible enough to capture many of the costs and benefits that accompany certain decisions that are not modeled explicitly. Our model is one where students learn about how much they will enjoy school and about future earnings. For illustration, consider a student who decides to increase his study effort. Given our objectives, what is needed is for this change in effort to be reflected in our characterization of how enjoyable it is to be in school and our characterization of what students believe about earnings. With respect to the former, the term EN_t in Eq. (4) would account for decreases in current period utility associated with the reduction in current period leisure, while the terms G_t and g_t in Eq. (4) would allow for the possibility that studying may lead to additional current period utility benefits not captured by EN_t through improved academic performance. With respect to the latter, an improvement in grades (that might accompany increased study effort) would influence a student's beliefs about future earnings both by increasing the probability that the student will not fail out and, as discussed in more detail below, by influencing the future consumption a person receives conditional on graduation.[6] Then, while our model cannot provide direct information about issues related to studying, it does take into account the implications of studying that are important for our objectives.

III.D State variables

The set of state variables at time t, $\Omega(t)$, includes all variables whose time t values provide information about $U^S_\tau(\bullet)$ and $U^N_\tau(\bullet)$, τ=t, t+1, t+2,...

State variables providing information about $U^S_\tau(\bullet)$, τ=t, t+1, t+2,...

We first consider the state variables whose time t values provide information about U^S for the current period t. Examining Eq. (4), G_t, EN_t, and $\varepsilon_{S,t}$ are known to person i at the beginning of time t. A student's beliefs about g_t on a 0.0-4.0 scale are constructed by censoring an underlying belief (random) variable g_t*. Specifically, assuming that g_t* is normally distributed with an individual-specific mean μ_t and an individual-specific variance σ^2_t, a student's beliefs are given by:

(5) $g_t=4.0$ if $g_t^*>4.0$, $g_t=0$ if $g_t^*<0$, $g_t=g_t^*$ else, with $g_t^* \sim N(\mu_t,\sigma^2_t)$.

Then, G_t, EN_t, $\varepsilon_{S,t}$, μ_t, and σ_t are elements of $\Omega(t)$.

We next think about what time t information influences U^S in the future periods t+1, t+2,... . As described in the previous paragraph, when a student arrives at t+1, the variables that will provide information about U^S_{t+1} are G_{t+1}, EN_{t+1}, $\varepsilon_{S,t+1}$, μ_{t+1}, and σ_{t+1}. Then, given the recursive nature of the Bellman Equation in (2), what is necessary is to specify the process by which G, EN, ε_S, μ, and σ evolve

[6]One can think of how the model would adjust for other "background decisions" as well. For example, a change to a new major may influence current period utility through both changes in grade performance (as captured by G_t and g_t) and changes in how much a person enjoys studying the new subject area (as captured by EN_t). Naturally, there are some limits to the ability of our model to adjust. For example, one concern would be that a change in major might influence a person's perceptions about the relationship between grades and earnings which is discussed later.

between t and t+1.

$\varepsilon_{S,t+1}$ is not known by person i at time t. Largely for computational reasons described below, we assume that $\varepsilon_{S,t+1}$ is drawn at t+1 from an extreme value distribution, with $\varepsilon_{S,t+1}$ independent of $\varepsilon_{S,t}$. Looking ahead one subsection, we make the same distributional and intertemporal independence assumptions for $\varepsilon_{N,t}$. We also assume that $\varepsilon_{S,j}$ is independent of $\varepsilon_{N,k}$ for all periods j and k.

G_{t+1} is determined by the technical relationship between a person's cumulative grade point average (GPA) at the start of a semester and his current GPA in that semester. For example, under the implicit assumption in III.A that a person takes an equal number of courses each semester,

(6) $G_{t+1} = \frac{t}{t+1} G_t + \frac{1}{t+1} g_t$.

We assume that the binary variable EN_{t+1} depends on EN_t, g_t, and other unobserved factors $v_{EN,t+1}$:

(7) $EN_{t+1}=1$ iff $EN_{t+1}{*} = \alpha_{EN,0} + \alpha_{EN,1} EN_t + \alpha_{EN,2} g_t + v_{EN,t+1} > 0$,

so that EN_{t+1} is determined at the end of semester t after g_t is observed and $v_{EN,t+1} \sim N(0,1)$ is drawn.

Finally, the process by which μ_t and σ^2_t evolve represents learning about academic performance in the model. As discussed in detail in Section IV, because we observe μ_t and μ_{t+1} we are not forced to assume that individuals update beliefs in any specific manner. Instead we estimate the parameters of a parsimonious updating equation:

(8) $\mu_{t+1} = \alpha_{\mu,0} + \alpha_{\mu,1} \mu_t + \alpha_{\mu,2} g_t + v_{\mu,t+1}$,

with $v_{\mu,t+1} \sim N(0,\sigma^2_{\mu})$ drawn at the end of semester t.

Models of Bayesian learning are relevant for considering issues related to Eq. (8). Suppose grades are determined by $g_t=\mu+v_t$ with μ being a constant representing a student's long-run average GPA and v_t representing transitory variation in grades across semesters. Bayesian learning about μ would have the "posterior mean" as a weighted average of the "prior mean" and the "noisy signal" with the weights depending on both the amount of uncertainty at t about μ and the amount of variation in v_t (i.e., the signal-to-noise ratio). Using survey questions which ascertain individual-specific beliefs related to the signal-to-noise ratio, S&S (forthcoming) found evidence of individual-specific heterogeneity in weights, but that the very large majority of explainable heterogeneity in μ_{t+1} arises because of heterogeneity in the observed values of μ_t and heterogeneity in the observed beliefs about g_t. Thus, we simplify matters here by assuming that the coefficients in Eq. (8) are constant across students. It is also natural to believe that the coefficients in Eq. (8) might change over time. For example, in the simple Bayesian model above, the signal-to-noise ratio would be expected to change over time as individuals resolve uncertainty about μ. Thus, in our empirical work we estimate different coefficients in Eq. (8) for different stages of college.

The update σ_{t+1} is given by

(9) $\sigma_{t+1} = \alpha_{\sigma,0} + \alpha_{\sigma,1}\sigma_t + v_{\sigma,t+1}$,

with $v_{\sigma,t+1} \sim N(0,\sigma^2_{\sigma})$ drawn at the end of t and the parameters again allowed to vary across stages of college.

Eqs. (6)-(9) show that, from the perspective of a student at time t, G_{t+1}, EN_{t+1}, μ_{t+1} and σ_{t+1} are random variables whose means depend on the previously identified state variables G_t, EN_t, μ_t and σ_t that are known by the person at time t. Randomness in G_{t+1}, EN_{t+1}, μ_{t+1}, and σ_{t+1} is present due to uncertainty about g_t as characterized by μ_t, and σ_t, as well as uncertainty about the unobservables $v_{EN,t+1}$, $v_{\mu,t+1}$, and $v_{\sigma,t+1}$.

State variables influencing $U^N_t(\bullet)$

At time t, a person who is choosing between S and N must implicitly think about V_N from equation (1) for each possible time $t' \geq t$ at which he might choose to leave school. Under the linear assumption in equation (3), equation (1) becomes

(10) $$V^N_{t'}(\Omega_{t'}) = \sum_{\tau=t'}^{T*} \beta^{\tau-t'}E(C_\tau) + \beta^{\tau-t'}E(\varepsilon_{N,\tau}).$$

With respect to the first term in the sum, for each possible exit time t', a person must think about the average consumption that he would receive in each period τ after leaving. We assume that a student's beliefs about his average consumption at time τ will vary with: 1) t'-1, the number of semesters he completes before leaving, 2) $G_{t'}$, his cumulative GPA at the time he leaves, and 3) his age at τ. We write beliefs about average consumption at time τ for a student who leaves school at t'as the function $\bar{C}_\tau(t', G_{t'}, AGE(\tau))$. As discussed in Section IV, by directly eliciting information about the function $\bar{C}_\tau$ we are able to take into account that, for a variety of reasons, the function $\bar{C}_\tau$ may vary substantially across students. Given student i's individual-specific function $\bar{C}_\tau$, with t' a choice variable and a person's age at τ known, the state variables at t that influence i's beliefs about the average consumption associated with N at a future time τ are those that are related to beliefs about $G_{t'}$: μ_t, σ^2_t, and G_t.

<u>III.E. More detail about value functions</u>

Given the discussion in III.C and III.D, we can rewrite the value functions in Eqs. (10) and (2).

(11) $$V^N_t(G_t,\varepsilon_{N,t}) = \sum_{\tau=t}^{T*} \beta^{\tau-t}\bar{C}_\tau(t,G_t,AGE(\tau)) + \beta^{\tau-t}E(\varepsilon_{N,\tau})$$

(12) $$V^S_t(G_t,EN_t,\mu_t,\sigma_t,\varepsilon_{S,t}) = EU^S_t(G_t,EN_t,g_t,\varepsilon_{S,t}) + Pr(G_{t+1}<F_{t+1})\ \beta V^N_{t+1}(G_{t+1},\varepsilon_{N,t+1})$$
$$+ Pr(G_{t+1} \geq F_{t+1})\ \beta E\max[V^S_{t+1}(G_{t+1},EN_{t+1},\mu_{t+1},\sigma_{t+1},\varepsilon_{S,t+1}), V^N_{t+1}(G_{t+1},\varepsilon_{N,t+1})],$$

where we have rewritten $PrFail_t$ to make explicit that a person fails out of school if G_{t+1} is less than an institutional cumulative grade cut-off F_{t+1} at t+1.[7] With G_t, EN_t, and $\varepsilon_{s,t}$ known, the first expectation in Eq. (12) involves a one-dimensional integral over a person's beliefs at time t about g_t as characterized by μ_t and σ_t. With ε_{t+1} not observed as of time t and randomness in G_{t+1}, EN_{t+1}, μ_{t+1}, and σ_{t+1} present due to uncertainty about g_t, $v_{EN,t+1}$, $v_{\mu,t+1}$, and $v_{\sigma,t+1}$, the second expectation involves a multi-dimensional integral over a person's beliefs about g_t (as characterized by μ_t and σ_t) and over the distributions of the random variables ε_{t+1} and $v_{t+1}=\{v_{EN,t+1}, v_{\mu,t+1}, v_{\sigma,t+1}\}$.

III.F. Identification

The current period utility parameters in Eq. (4) are identified by the observed choices of S or N. Eq. (3) shows that the deterministic portion of utility from being in the workforce is a person's consumption. Because the coefficient on C in Eq. (3) is normalized to unity, the coefficients in Eq. (4) can be interpreted as utility effects measured in consumption dollars. This normalization also fixes the scale of the discrete choice problem so that it is possible to estimate the variance of ε_t. Assuming that ε has an Extreme Value distribution, we estimate the parameter τ where $Var(\varepsilon_{N,t})=Var(\varepsilon_{S,t})=\tau^2\pi^2/6$. The parameters of Eqs. (7-9) are identified because our unique data collection efforts imply that both dependent and independent variables in these equations are observed in each semester.

IV. Data

Section III indicates that solving the necessary value functions for person i (up to the value of ε_t) requires observing G_t, EN_t, μ_t, and σ^2_t for each semester that i is in school, and also requires knowledge of the person-specific function $\bar{C}_\tau$.[8] In addition, while it is *beliefs* about g_t (as given by μ_t and σ^2_t) that are used to compute value functions, it is *actual* values of g_t that will be used to estimate the parameters of the transition process in Eq. (8).

G_t and g_t are obtained for each t from administrative data. The first four rows of Table 1 show the sample mean (standard error of the sample mean) of g_t at t=1,2,...,7 for the full sample of students who were still enrolled as of t (Row 1) and for three subsamples created by stratifying the full sample on the basis of how long students remained in school (Rows 2-4). Looking across columns in Row 1 reveals that, for the full sample of students who were still enrolled as of t, there is a statistically significant increase in the mean of g_t across semesters. However, the sample mean in Row 2 for the

---

[7]The stated cutoffs at Berea were F_2 =0.0, F_3 =1.5, F_4 =1.67, F_5 =1.85, F_6 =2.0, F_7 =2.0, F_8 =2.0. However, the cutoffs were, in practice, somewhat lower because students were often able to successfully appeal suspension decisions. Under an assumption that students may know that appeals are possible, we choose to use empirical cutoffs constructed as the minimum value of G_t at which a person was observed remaining in school in our sample: F_2 =0.0, F_3 =1.37, F_4 =1.41, F_5 =1.82, F_6 =1.83, F_7 =1.89, and F_8 =2.0. Regardless, results depend little on which set is used.

[8]Also needed is the discount factor β. We assume a yearly discount factor of .95.

composition-constant subset of students who were in school for all of the seven semesters changes very little across time, suggesting that the increase over time in the first row is due largely to the change in composition that arises as poorly performing students leave school over time. Row 3 provides some support for this interpretation by showing that students who left school after completing four, five, or six semesters have sample average values of g_t that are somewhat lower than the sample average values in Row 2. Stronger evidence of changes in composition appear in Row 4 which shows that students who left school after completing one, two, or three semesters have sample average values of g_t that are much lower than the sample average values in Row 2. Thus, the results suggest that grades are indeed likely to be important determinants of dropout, especially among those that leave school early.

Moving away from administrative data, the unique feature of this project is that the BPS was designed to minimize the assumptions needed to characterize the individual-specific values of EN_t, μ_t, and σ^2_t for each t and the individual-specific function $\overline{C}_\tau(\bullet)$ for each time τ. To elicit EN_t, as discussed earlier we use Question A.1 to elicit a direct measure of how much a person enjoys school relative to being out of school. Row (14) of Table 1 shows that students tend to enter school with a very positive outlook about the utility of being in college; 89% of students in the sample believe that school will be somewhat more enjoyable or much more enjoyable than not being in college (i.e., EN_1=1). The next three rows show EN_t for those who were in school for all seven of the semesters, those who left school after completing between four and six semesters, and those who left school after completing between one and three semesters. The three groups entered school similarly optimistic. However, by the beginning of the third semester, the sample percentage of students with EN_t=1 decreased by only six percentage points for those who remained in school for all seven of the semesters (Row 15), by nine percentage points for those who left school after completing between four and six semesters (Row 16), but by twenty-five percentage points for those who left school after completing between one and three semesters (Row 17).

To elicit μ_t, and σ^2_t, we administered Question A.2 (Appendix A) at the beginning of each semester t to elicit directly each student's subjective beliefs about the distribution of g_t. The question asks each student to report the "percent chance" that g_t will fall in each of a set of mutually exclusive and collectively exhaustive categories. Importantly, students who left school were sent exit surveys immediately after leaving school. This allows us to observe beliefs about g_t at the beginning of semester t both for those who decided to stay in school for semester t and for those who were in school for t-1 but did not return for semester t.

For descriptive purposes, we compute the approximate mean of the distribution describing beliefs about g_t from a person's answers to Question A.2 by assuming that a student's beliefs are uniformly distributed within each of the grade categories. Rows 6-9 of Table 1 show the sample averages of these approximate means at times t=1,2,...,7 for the full sample of those who were still enrolled in school at the beginning of t (Row 6) and over subsamples generated by stratifying on the basis of how long students remained in school (Rows 7-9). Comparing the t=1 entry of Row 6 to the t=1 entry of Row 1, shows that, in the sample as a whole, students are, on average, substantially overoptimistic about their average grade performance at entrance. Examining the first four entries of Rows 7-9 we see that students who left school in the first three semesters were by far the most overoptimistic about future grade performance at entrance, and, subsequently, had the largest (downard) revisions to beliefs. Further motivating our interest in learning, what a student in this dropout group learned about his future grade performance substantially understates what he learned about his final cumulative GPA since, after several semesters, poor actual grade performance begins to weigh heavily on a student's final GPA.

Rows 10-13 show approximate standard deviations of the distribution describing beliefs about g_t averaged over the full sample and averaged over subsamples generated by stratifying on the basis of how long students remained in school. The results indicate that uncertainty decreases significantly over time, even under the composition-constant sample of students who remain in school for all semesters (Row 11).

Eq. (5) described our assumption, needed for estimation, that i's beliefs about g_t can be represented by censoring an underlying latent random variable $g_t^* \sim N(\mu_t, \sigma^2_t)$. At each time t, we obtain our person-specific measures of μ_t and σ_t by fitting the censored random variable to the person's self-reported probabilities from Question A.2. Specifically, for each person we choose μ_t and σ_t to minimize

$$(13) \quad \sum_{j=1}^{6} |PR_{observed}(g_t \in CAT_j) - PR_{model}(g_t \in CAT_j)| \ ,$$

where $CAT_1,...,CAT_6$ represent the grade categories [4.0,3.5), [3.5,3.0), [3.0,2.5), [2.5,2.0), [2.0,1.0), and [1.0,0], respectively, the first term in the difference is the self-reported perceived probability of category CAT_j from Question A.2 and the second term in the difference is the probability that the censored random variable produces a realization in category CAT_j. We find that a censored normal is able to fit the self-reported probabilities quite well. For example, for t=1 we find that the average value of $|PR_{observed}(g_t \in CAT_j) - PR_{model}(g_t \in CAT_j)|$ across all categories j and all sample members is .018.

Similarly, at the time of college entrance, for some combinations of possible exit times t', possible exiting grade point averages $G_{t'}$, and possible future years $\tau \geq t'$, we also utilized survey questions to directly elicit the beliefs about the expected future yearly earnings that determine the individual-specific function $\bar{C}_\tau(t', G_{t'}, AGE(\tau))$.[9] With respect to t′, we collected information about leaving college immediately, after one full year of school, after three full years of school, and at the time of graduation. With respect to $G_{t'}$, we collected information about leaving school with a GPA of 3.75, with a GPA of 3.0, and with a GPA of 2.0. With respect to τ, we collected information about earnings in the first year out of school, at the age of 28, and at the age of 38. We further reduced the number of combinations by assuming that $G_{t'}$ does not influence future earnings if a person leaves school without graduating. Thus, we collected beliefs about the expected earnings that would be received at three future points in time (first year out of school, age 28, and age 38) for each of six schooling scenarios (leave school immediately, leave school after one year, leave school after three years, graduate with a 2.0 GPA, graduate with a 3.0 GPA, and graduate with a 3.75 GPA).

Figure 1, which shows the sample mean of $\bar{C}_\tau$ at each of the three points in time for each of six schooling scenarios, indicates that students believe that earnings are strongly related to both years of completion and grade performance (if the person graduates). We discuss beliefs about the size of these differences in the V.A.

V. Solving value functions and Estimation

<u>V.A. Solving value functions</u>

Computing $V^N_{t'}$ Given our assumption that a student's GPA does not influence his future earnings if he does not graduate, it is necessary to compute $V^N_{t'}$ for 1) for every possible time t′ that a person might leave school under the scenario that he does not graduate (t′=2,3,...,8) and 2) for each possible value of $G_{t'}$ that a person could have under the scenario in which he graduates (t′=9). Then, Eq. (11) implies: 1) for each t′=2,3,...,8, $\bar{C}_\tau$ is needed for all $\tau \geq t'$ and 2) for t′=9, $\bar{C}_\tau$ is needed for all $\tau \geq t'$ for each possible value of G_9. Section IV discussed the combinations of t′, G_t and τ at which we elicited $\bar{C}_\tau$ directly. Our approach is to use these directly elicited combinations to interpolate $\bar{C}_\tau$ for all other necessary combinations.[10]

---

[9] The survey question (full question not shown) informed respondents that "when reporting incomes take into account the possibility that you will work full-time, the possibility that you will work part-time, 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."

[10] We use a straightforward interpolation approach under the following assumptions: 1) to deal with the fact that values of $\bar{C}$ are only observed directly for the time a person leaves school, at age 28, and at age 38, we assume that $\bar{C}$ is linear between the time a person leaves school and the age of 28, is linear between the age of 28 and 38, and is constant after the age of 38; 2) to deal with the fact that values of $\bar{C}$ are only observed at the dropout times t′=1, t′=3, and t′=7, we assume that $\bar{C}$ is linear between t′=1 and t′=3, is linear between t′=3 and t′=7, and is the same at t′=8 as it is at t′=7; 3) to deal with the fact that, for t′=9 (graduation), values of $\bar{C}$ are only observed for the values of G_9=2.0, G_9=3.0, and G_9=3.75, we assume that V^N_9 is linear between G_9=2.0 and G_9=3.0 and is linear between G_9=3.0 and G_9=4.0 (with the slope being identified by the values of V^N_9 at G_9=3.0 and G_9=3.75 and this

Figure 2 shows the sample mean of $V^N_{t'}()$ for the six schooling scenarios from Section IV. The perceived discounted expected lifetime return to completing the first year of college is \$527,000-\$475,000=\$52,000 (11%). The perceived return to completing an additional two years is \$123,000, or \$61,500 per year. Then, assuming that the completion of a fourth year without graduating would also be valued at \$61,500 implies a perceived lifetime sheepskin premium (graduating with a 2.0 GPA versus completing four years and not graduating) of approximately \$90,000 (roughly the return to 1.5 non-graduation years of completion). The perceived premium to graduating with a GPA of 3.0 instead of 2.0 and the perceived premium to graduating with a GPA of 3.75 instead of 3.0 are each approximately \$130,000 (roughly the return to 2 non-graduation years of completion).

Solving $V^S_{t'}$ The expected value in the last term of Eq. (12) is present because uncertainty exists at t about ε_{t+1}, EN_{t+1}, G_{t+1}, μ_{t+1}, and σ_{t+1}. The assumption that $\varepsilon_{N,t+1}$ and $\varepsilon_{S,t+1}$ have extreme value distributions implies that the Emax has a well-known closed form solution conditional on the realizations of EN_{t+1}, G_{t+1}, μ_{t+1}, and σ_{t+1}. Then, evaluating the last expected value involves summing the closed form solution over the probability function of the binary random variable EN_{t+1} and integrating over the densities of the continuous random variables G_{t+1}, μ_{t+1}, and σ_{t+1}. Appendix B describes the simulation approach that we take to evaluate this integral. This simulation approach takes into account that uncertainty about G_{t+1}, μ_{t+1}, σ_{t+1}, and EN_{t+1} is driven primarily by uncertainty about g_t.

The recursive formulation of value functions in Eq. (12) motivates a backwards recursion solution process of the general type that is standard in finite horizon, dynamic, discrete choice models. The most basic property of the algorithm is that, in order to solve all necessary value functions at time t, it is necessary to know value functions at time t+1 for each combination of the state variables in $\Omega(t+1)$ that could arise at time t+1. In Appendix B we discuss computational issues that arise when implementing the backwards recursion solution process in our particular application, including the modification that is needed to deal with the fact that we have multiple continuous, serially correlated state variables, G_{t+1}, μ_{t+1}, and σ_{t+1}.

<u>V.B Estimation</u>

We estimate the parameters of the model by Maximum Likelihood. The likelihood contribution for person i is the joint probability of observing his schooling decisions and all values of EN_t, μ_t, and σ_t that are reported after t=1. The likelihood terms associated with the reported values of μ_t and σ_t involve density evaluations with the densities determined by Eqs. (8) and (9). The likelihood term associated with the reported values of EN_t involve probability calculations as described by Eq. (7).

---

slope being used to extrapolate values of V^N_9 between G_9=3.75 and G_9=4.0).

With respect to schooling choices, we examine decisions from whether to return for the second semester (t=2) through whether to return for the fourth year (t=7). For a person who chooses to return to school in each semester through the seventh semester, the likelihood contribution associated with his observed choices is the probability that he chooses S in t=2, t=3, ..., and t=7. For a person who chooses to leave school at some time $t' \leq 7$, the likelihood contribution associated with his observed choices is the probability that he chooses S in t=2, t=3, ..., $t=t'-1$ and chooses N in $t=t'$. For a person who is forced out of school due to bad academic performance at time t', the likelihood contribution associated with his observed choices is the probability that he chooses S in t=2, t=3, ..., $t=t'-1$. At each time t, the probability of choosing S is given by $PR(V_S>V_N)$. With value functions solved up to ε_t and the components of ε_t having Extreme Value distributions, this probability has the standard closed form logit solution. The Maximum Likelihood approach is also conducive to dealing with missing data. For example, if a person does not answer a survey at time t, then EN_t, μ_t, and σ_t will be missing. We construct the joint distribution of the missing data from Eqs. (7)-(9) and compute the choice probability at t by using simulation methods to integrate the choice probability conditional on EN_t, μ_t, and σ_t over the constructed distributions.

VI. Results

We first examine estimates of the model parameters in Eq. (4) and Eqs. (7-9) which are shown in Table 2. We then use simulations to quantify the over importance of learning and the relative importance of the three avenues discussed in the Introduction.

VI.A. Estimates of parameters related to the evolution of EN_t, μ_t, and σ_t

Estimates of the parameters of Eq. (7) appear in the second panel of Column 1 of Table 2. Of particular note given our interest in grade performance, a student's grades in semester t, g_t, play an important role in determining whether he believes he will enjoy school in t+1, $EN_{t+1}=1$ (t-statistic 3.82). Further, how much a person likes school in period t+1 is also likely to be influenced by his grade performance in the past because whether someone likes school in t+1, $EN_{t+1}=1$, is found to be strongly related to whether he liked school in t, $EN_t=1$ (t-statistic 13.16).

Estimates of the parameters of the updating Eqs. (8) and (9) are shown in the third and fourth panels of Column 1. The strength of our approach in estimating Eqs. (8) and (9) is that we directly observe μ_{t+1}, μ_t, σ_{t+1}, σ_t, and g_t. Given the discussion in Section III.D that the coefficients in Eqs. (8) and (9) may vary over time, we estimate Eqs. (8) and (9) separately for updates that take place after the first and second semesters, updates that take place after the third and fourth semesters, and updates that take place during the remaining time in school. Focusing on Eq. (8), Rows 9-20 show that, for all updates, both μ_t and g_t play an important role in determining the update μ_{t+1}. Consistent with a Bayesian model in which a student resolves uncertainty about his average GPA over time, the relative influence of μ_t in determining μ_{t+1} increases over time. Rows 9-20 show that the ratio of the estimated effect of μ_t to the estimated effect of g_t (i.e., $\alpha_{\mu,1}/\alpha_{\mu,2}$) is .320/.255=1.25 for the first two updates, is .463/.212=2.18 for the next two updates, and is .558/.172=3.24 for the remainder of the updates.

VI.B. Estimates of utility parameters

Consistent with a vast amount of previous research, we find a strong reduced-form correlation between grade performance and dropout. Estimating the Logit model that results from setting β=0, τ=1, and allowing only G_t to enter current period utility (Eq. 4), we find in Column 2 that the coefficient on G_t has a t-statistic of approximately 8.5. The importance of estimating the model in this paper is that it allows the first opportunity to differentiate between three avenues through which this correlation may arise in the reduced form: 1) poor performance causes students to fail out of school immediately or causes the value of continuing in school to decrease because it increases the probability of failing out in the future, 2) poor performance reduces how enjoyable it is to be in school, and 3) poor performance reduces the value of staying in school by reducing the earnings that a person will receive in the future if he does graduate.

Estimates of the current period utility parameters associated with being in school (Eq. 4) for the full model are shown in the first panel of Column 1. The results indicate that the second avenue above will be relevant. In Section VI.A we found that grade performance influences the measure EN. Here we see that EN has a significant effect on utility (t-statistic=2.68). In addition, both G_t and g_t have a significant effect on the current period utility of being in school with the estimated effects having t-statistics of 3.73 and 1.78, respectively. With income/consumption measured in hundreds of thousands of dollars, the coefficients imply that a person with G_t=3.02, g_t=3.02, and EN_t=1 would receive, on average, about the same amount of utility in semester t (−2.186 + .479 + .292*3.02 + .304*3.02) as a person who is out of school with an annual income of roughly the average expected annual income of someone who leaves school at the beginning of college ($19,000). To get a sense of the importance of poor performance, holding g_t and EN_t constant, a .50 reduction in cumulative grade point average reduces current period utility of school by the consumption equivalent of .50*.292*$100,000=$14,600. The policy relevance of the sizeable effect that grades have on current period utility is discussed in the Conclusion.

The current period utility specification for Eq. (4) in Column 1 is parsimonious. It is worth examining whether it is useful to add to Eq. (4) student characteristics that have consistently been found to be related to dropout in previous work. Repeating the exercise in Column 2 after replacing G_t with an indicator of whether a student is male, we found that males are significantly more likely to drop out (t-statistic≈-2.00). Repeating the exercise in Column 2 after replacing G_t with a student's high school GPA (HSGPA) we found that students with strong HSGPA's are significantly less likely to drop out (t-statistic 3.31). However, when we added these two variables to the full, dynamic model in Column 1, neither was found to be statistically significant (t-statistics of .43 and .82, respectively, full results not shown). Thus, the evidence suggests that the effect of these two variables in the reduced-form arises primarily because the variables are related to actual grade performance or to beliefs about future grade performance. We also found that, as in the reduced form in S&S (forthcoming), other parameters changed very little when measures of a student's health, family income, and whether his parent lost a job were added to the current period utility Eq. (4).

From the estimates in Column 1 alone, it is not possible to quantify the importance of the second avenue in determining the dropout decision or to get any sense of whether avenues (1) and (3) above are also relevant. As a result, we now turn to our simulations.

VI.C. Simulations

For several counterfactual scenarios, which imply various changes to G_t, g_t, EN_t, μ_t, and σ_t, we use the estimates from Column 1 of Table 2 to compute the proportion of students that would drop out by the beginning of the second year (T=3), by the beginning of the third year (T=5), and by the beginning of the fourth year (T=7). For each student, the probability of dropping out at or before the start of the Tth semester is given by $1-\prod_{t=2}^{T}\Pr(V^S_t > V^N_t)$. Section V described the techniques that we use during estimation to compute the probabilities that appear in this expression. Here we require additional simulations to incorporate the changes to G_t, EN_t, μ_t, and σ_t. However, these simulations are straightforward extensions of our methods for dealing with missing data as described in Section V.

We begin with a baseline simulation. For this simulation, we wish to use actual values of G_t, EN_t, μ_t, and σ_t, so that the additional use of simulation is only necessary because, for someone who leaves school at the start of semester t', actual values of G_t, g_t, EN_t, μ_t, and σ_t are not observed in the data for $t>t'$.[11] Our baseline calculation finds that .193 of students drop out before the start of the second year (T=3, compared to .18 actual), that .355 of students drop out before the start of the third year (T=5, compared to .34 actual), and that .483 of students drop out before the start of the fourth year (T=7, compared to .46 actual).

To quantify the overall importance of learning, we next compare these baseline proportions to dropout proportions simulated under a no-learning counterfactual scenario in which a person's beliefs about grade performance do not change after the time of entrance and actual grades g_t are drawn from this perceived grade distribution. Specifically, for all t: 1) $\mu_t=\mu_1$ and $\sigma_t=\sigma_1$, 2) the distribution of actual grades g_t is determined by Eq. (5) given parameters μ_1 and σ_1 and 3) $PR(EN_t=1)$ is determined from Eq. (7) based on EN_1 and by the draws g_1, g_2,..., g_{t-1} from the distribution described in 2).

Under this no-learning scenario, we find that .106 of students would drop out by the start of the second year, .194 of students would drop out before the start of the third year, and that .309 of students would drop out before the start of the fourth year. Thus, .45=(.193-.106)/.193 of the dropout in the first year, .45=(.355-.194)/(.355) of the dropout in the first two years, and .36=(.483-.309)/(.483) of the dropout in the first three years can be attributed to what students learn about their academic performance. These cumulative numbers imply that .45 of dropout in the first year, .45 of dropout in the second year, but only .11 of dropout in the third year is caused by learning about academic performance. That dropout due to learning about academic performance tends to happen relatively early in college is consistent with the descriptive statistics discussed in the second paragraph of Section IV and is important for policy reasons discussed in the introduction.

Finally, we perform three additional simulations to provide evidence about the quantitative importance of the three broad avenues, detailed in Section VI.B, through which learning about grade performance could cause dropout. To examine the first avenue (that poor academic performance operates through grade progression cutoffs), we repeat the baseline simulation, but remove the institutional rule that students are forced to leave school due to poor academic performance. We find that the percentage of students who would drop out would decrease only trivially, from .483 to .463. Thus, the results suggest that students who perform poorly tend to learn that staying in school is not beneficial, not that they leave simply because they have lost the option to stay or believe they are

[11] Recall that, for someone who leaves school at the beginning of semester t', we collect information about beliefs at t' using an exit survey.

more likely to lose the option in the future.[12]

Differentiating between the remaining two avenues above is a matter of understanding why students find that it is not beneficial to be in school if they have performed poorly. Maintaining the assumption that students cannot fail out, we first examine the importance of avenue 3 (that poor performance reduces the earnings received upon graduation) by simulating the model under the counterfactual assumption that a person's beliefs about his earnings upon graduation are determined by his beliefs about grade performance at the start of college rather than by what he learns about his actual grade performance during college.[13] We find that .386 of students would drop out under this counterfactual scenario, so that approximately .50=(.463-.386)/(.463-.309) of the dropout that can be attributed to learning about academic performance (and is not due to the possibility of failing out) would disappear under this scenario.

Finally, continuing to maintain the assumption that students cannot fail out, we examine the importance of avenue 2 (that poor performance makes it less enjoyable to be in school) by simulating the model under the counterfactual assumption that a person's utility during school corresponds to the utility that would have been received if the student's perceptions about grade performance were correct at the time of entrance.[14] We find that .344 of students would drop out under this scenario, so that approximately .77 =(.463-.344)/(.463-.309) of the dropout that can be attributed to learning about academic performance (and is not due to grade requirements) would disappear under this counterfactual. The fact that the proportions associated with avenues 2 and 3 sum to a number greater than 1.0 (1.27) is evidence that the two avenues interact to generate the overall effect. That is, when poor performance arises, avenue 2 puts a person closer to the margin of indifference so that he is more likely to be pushed across by avenue 3 (and vice versa).

VII. Conclusion

We find that forty-five percent of the dropout that occurs in the first two years of college can

[12] We do not observe beliefs about earnings for final GPA's of less than 2.0. For this simulation, we assume that a student's beliefs about the earnings associated with GPA's less than 2.0 is the same as his beliefs about the earnings associated with a GPA of 2.0. Thus, if anything, the true effect of removing the possibility of failing out would be even smaller.

[13] Specifically, we set a person's beliefs about earnings upon graduation equal to what he would expect if he were to graduate with G_9 equal to the mean of the distribution describing his beliefs about grades at the time of entrance (i.e., the approximate mean from the t=1 response to Question A.2).

[14] When computing the current period utility in Eq. (4), we characterize G_t, g_t, and EN_t under the assumption that a person's grades in each period are equal to the mean of the distribution describing his beliefs about grades at the time of entrance (i.e., the approximate mean from the t=1 response to Question A.2).

be attributed to student learning about academic performance. However, important for policymakers concerned that scarce public resources may be consumed inefficiently if persistent misperceptions lead students to remain in school longer than they otherwise would, the importance of this type of learning becomes largely irrelevant by the midway point of college.

Our simulations show that students who perform poorly tend to learn that staying in school is not worthwhile, not that they fail out or learn that they are more likely (than they previously believed) to fail out in the future. As to why learning about academic performance makes staying in college less worthwhile, we find that poor performance both substantially decreases the enjoyability of school and substantially influences beliefs about post-college earnings. Then, given that S&S (forthcoming) found that students who perform poorly learn primarily that they are not prepared academically, the cautionary message for students is that poor performance may cause multiple future stages of life to be considerably less enjoyable. In terms of improving pre-college preparation, while the quality of elementary and secondary schools is clearly relevant, ensuring that pre-college students have correct perceptions about the level of preparation necessary to succeed in college may be important for increasing student effort at earlier stages of schooling.

Improving academic preparation seems particularly valuable because, by influencing both the enjoyability of school and the financial returns to school, students are likely to be moved a substantial distance away from the dropout margin. Given that dropout is not inherently bad, it is not easy to know the circumstances under which it would be optimal for colleges to try to retain students who are academically marginal given their current level of preparation. Regardless, because reducing dropout is often an objective of colleges and policymakers, a question of interest is whether dropout can be reduced at current levels of academic preparation by providing information or counseling during school. One potential approach would involve providing information about the financial returns to completing school. However, our unique expectations data indicate that students already believe that there is a substantial financial return to graduating - even when grade performance is not particularly strong. A second potential approach would involve providing counseling aimed at reducing the sizeable loss of enjoyability that we find accompanies poor grade performance. The fact that students are often willing to give up large amounts of future income to move away from school when grade performance is poor raises the possibility that decisions may sometimes be made in a rash manner. Unfortunately, while students making decisions in this manner might benefit the most from counseling, in practice it may be difficult to administer this counseling if students tend to leave school without warning or tend to make decisions when they are away from school (e.g., between semesters).[15] Indeed, the reality that the observed dropout in our study takes place in an environment where much thought and effort has already been put into designing counseling strategies is suggestive of natural limits to counseling. Similarly, attempts to improve retention by improving grade performance during school (at given levels of academic preparation) through programs that encourage increased study effort may be difficult both because increased study effort comes at the cost of leisure and because students at this school are already being encouraged to be conscientious in their study habits.

[15]Of course, many of the potential avenues through which grade performance may influence enjoyability (e.g., that studying may be unrewarding if a student does not comprehend course material), would seemingly be impervious to counseling interventions.

References

Altonji, Joseph. “The Demand for and Return to Education When Education Outcomes are Uncertain,” Journal of Labor Economics, 1993, vol. 11, no. 1, 48-83.

Arcidiacono, Peter. “Ability Sorting and the Returns to College Major,” Journal of Econometrics, Vol. 121, Nos. 1-2 (August, 2004), 343-375.

Arcidiacono, Peter, Hotz, Joseph, and Kang, Songman, “Modeling College Major Choices Using Elicited Measures of Expectations and Counterfactuals, Journal of Econometrics, forthcoming.

Attanasio, Orazio, and Kaufmann, Katja. “Educational Choices, Subjective Expectations, and Credit Constraints,” Working paper 15087, National Bureau of Research, 2009.

Barsky, Robert, Kimball, Miles, Juster, F. Thomas, and Shapiro, Matthew. “Preference Parameters and Behavioral Heterogeneity: An Experimental Approach in the Health and Retirement Survey,” The Quarterly Journal of Economics, May 1997, 537-579.

Black, Dan and Smith, Jeffrey. “Evaluating the Returns to College Quality with Multiple Proxies for Quality,” Journal of Labor Economics, 2006, 24(3), 701-728.

Bound, John, Todd Stinebrickner and Tim Waidmann. “Health, Economic Resources and the Work Decisions of Older Men,” Journal of Econometrics, 156(1), May 2010, 106-129.

Bound, John, Michael Lovenheim, and Sarah Turner. “Why Have College Completion Rates Declined: Marginal Students or Marginal College?” forthcoming, American Economic Journal: Applied Economics.

Bound, John and Sarah Turner. 2007. “Cohort Crowding: How Resources Affect Collegiate Attainment.” Journal of Public Economics, 91(5-6): 877-99.

Bowen, William and Bok, Derek. The Shape of the River. Princeton, New Jersey; Princeton University Press, 1998.

Bowen, William, Chingos, Matthew, and McPherson, Michael. “Crossing the Finish Line: Completing College at America’s Public Universities,” Princeton New Jersey, Princeton University Press, 2009.

Carneiro, Pedro, Hansen, Karsten, and Heckman, James, “Estimating Distributions of Counterfactuals with an Application to the Returns to Schooling and Measurement of the Effect of Uncertainty on Schooling Choice,” International Economic Review, 2005.

Cunha, Flavio, Heckman, James, and Navarro, Salvador, “Separating Uncertainty from Heterogeneity in Life Cycle Earnings,” Oxford Economic Papers, 2005, 57(2), 191-261.

Das, Marcel and van Soest, Arthur, “Expected Versus Realized Income Expectations: A Test of the Rational Expectations Hypothesis,” unpublished manuscript, 2000.

Dominitz, Jeff. “Earnings Expectations, Revisions, and Realizations,” The Review of Economics and Statistics, August 1998, 374-388.

Dominitz, Jeff and Manski, Charles. “Eliciting Student Expectations of the Returns to

Schooling," Winter 1996, 1-26.

Dominitz, Jeff and Manski, Charles. "Using Expectations Data to Study Subjective Income Expectations," Journal of American Statistical Association, September 1997, 855-867.

Dominitz, Jeff and Manski, Charles. "How Should We Measure Consumer Confidence (sentiment)? Evidence form the Michigan Survey of Consumers," Working paper 9926, National Bureau of Research, 2003.

Draper, Stephen, "Tinto's Model of Student Retention," Notes: http://www.psy.gla.ac.uk/~steve/localed/tinto.html (2005).

Heckman, James, Lochner, L, and Todd, Petra. "Earnings Functions, Rates of Return and Treatment Effects: The Mincer Equation and Beyond." In Handbook of the Economics of Education, ed. By E. Hanushek and F. Welch. Amsterdam: North-Holland, 307-458, 2006.

Heckman, James and S. Navarro, 2007." Dynamic Discrete Choice and Dynamic Treatment Effects," Journal of Econometrics, 136 (2007): 341-396

Juster, T., "Consumer Buying Intentions and Purchase Probability: An Experiment in Survey Design," Journal of the American Statistical Association 61 (1966), 658-96.

Keane, M.P., and K.I. Wolpin (1994): "The Solution and Estimation of Discrete Choice Dynamic Programming Models by Simulation and Interpolation: Monte Carlo Evidence," *Review of Economics and Statistics*, 76, 648-672.

Manski, Charles, "Schooling as Experimentation: a reappraisal of the post-secondary drop-out phenomenon," Economics of Education Review, Volume 8 number 4, 1989, 305-312.

Manski, Charles. "Parental Income and College Opportunity." Democratic Study Center Report, Washington, DC. August, 1992.

Manski, C., "The Use of Intentions Data to Predict Behavior: A Best Case Analysis," Journal of the American Statistical Association 85 (1990), 934-940.

Manski, C., "Measuring Expectations," Econometrica, 72 (5), (2004), 1329-1376.

Manski, Charles and Wise, David. College Choice in America. Cambridge, MA and London England; Harvard University Press, 1983.

National Center for Education Statistics, "Placing College Graduation Rates in Context: How Four Year College Graduation Rates Vary with Selectivity and the Size of the Low Income Enrollment." U.S. Department of Education, NCES 2007-161, 2007.

Rust, J. (1997): "Using Randomization to Break the Curse of Dimensionality," *Econometrica*, 65, 781-832.

Stange, Kevin, "An Empirical Examination of the Option Value of Schooling," American Economic Journal: Applied Economics, forthcoming.

Stinebrickner, Todd R. "Serially Correlated Variables in Dynamic, Discrete Choice Models," Journal of

Applied Econometrics, 15(6) November-December 2000, 595-624.

Stinebrickner, Todd R. and Ralph Stinebrickner. "Understanding Educational Outcomes of Students from Low Income Families: Evidence from a Liberal Arts College with a Full Tuition Subsidy Program," *Journal of Human Resources*, Summer 2003a, 591-617.

Stinebrickner, Todd R. and Ralph Stinebrickner. "Working During School and Academic Performance," *Journal of Labor Economics*, 21(2), April 2003b, 473-491.

Stinebrickner, Todd and Stinebrickner, Ralph, "Time-Use and College Outcomes," Journal of Econometrics, 121 (1-2) July-August (2004), 243-269.

Stinebrickner, Todd and Stinebrickner, Ralph, "The Effect of Credit Constraints on the College Drop-Out Decision: A Direct Approach Using a New Panel Study," American Economic Review. December (2008a)

Stinebrickner, Todd and Stinebrickner, Ralph, "The Causal Effect of Studying on Academic Performance," Frontiers in Economic Policy and Analysis (*Frontiers*), Berkeley Electronic Press (2008b).

Stinebrickner, Todd and Stinebrickner, Ralph, "Learning about Academic Ability and the College Drop-Out Decision," Journal of Labor Economics, forthcoming.

Stinebrickner, Todd and Stinebrickner, Ralph, "Math or Science? Using Longitudinal Expectations Data to Examine the Process of Choosing a College Major." NBER working paper w16869, 2011.

Tinto, Vincent. "Dropout from Higher Education: A Theoretical Synthesis of Recent Research, Review of Educational Research, 1975, 45 (1), 89-125.

van der Klaauw, Wilbert and Wolpin Kenneth, "Social Security and the Retirement and Savings Behavior of Low Income Households," Journal of Econometrics, 2008, 145 (1-2), 21-42.

van der Klaauw, Wilbert, "On the Use of Expectations Data in Estimating Structural Dynamic Models: An Analysis of Career Choices," unpublished manuscript, 2000.

Zafar, Basit, "College Major Choice and the Gender Gap," working paper (2008)

Zafar, Basit, "How Do College Students Form Expectations about Major-Specific Outcomes?" Journal of Labor Economics, 2011, 29(2), 301-348..

Table 1 sample mean (std. error of sample mean)

	t=1	t=2	t=3	t=4	t=5	t=6	t=7
(1) g_t all observations	2.81 (.04)	2.83 (.04)	2.85 (.04)	2.97 (.04)	2.96 (.04)	3.02 (.04)	3.07 (.04)
(2) g_t completed $\geq$7 semesters	3.05 (.04)	3.05 (.04)	2.96 (.04)	3.04 (.04)	3.01 (.04)	3.06 (.04)	3.07 (.04)
(3) g_t completed 4,5,6 semesters	2.95 (.08)	2.88 (.08)	2.76 (.10)	2.77 (.12)	2.72 (.16)	2.71 (.219)	
(4) g_t completed 1,2,3 semesters	2.11 (.118)	1.85 (.15)	1.95 (.29)				
(5) G_{t+1} all observations	2.81 (.04)	2.82 (.04)	2.90 (.03)	2.97 (.03)	3.00 (.03)	3.01 (.03)	3.03 (.03)
(6) belief $E(g_t)$ all observations	3.21 (.01)	3.10 (.02)	3.14 (.02)	3.13 (.02)	3.16 (.02)	3.21 (.02)	3.12 (.02)
(7) belief $E(g_t)$ completed >7 semesters	3.21 (.02)	3.14 (.02)	3.19 (.02)	3.16 (.02)	3.17 (.02)	3.19 (.02)	3.14 (.02)
(8) belief $E(g_t)$completed 4,5,6 semesters	3.21 (.03)	3.14 (.04)	3.22 (.04)	3.19 (.05)	3.14 (.05)	3.21 (.05)	
(9) belief $E(g_t)$ completed 1,2,3 semesters	3.20 (.03)	2.96 (.04)	2.84 (.04)	2.53(.15)			
(10) belief std. dev. (g_t) all observations	.53 (.01)	.48 (.01)	.45 (.01)	.43 (.01)	.42 (.01)	.38 (.01)	.39 (.01)
(11) belief std. dev (g_t) completed >7 semesters	.52 (.01)	.47 (.01)	.47 (.01)	.44 (.01)	.44 (.01)	.43 (.01)	.40 (.01)
(12) belief std dev. (g_t) completed 4,5,6 semesters	.53 (.02)	.48 (.02)	.44 (.02)	.41 (.02)	.42 (.02)	.34 (.02)	
(13) belief std dev.(g_t) completed 1,2,3 semesters	.53 (.02)	.51 (.02)	.51 (.03)	.51 (.09)			
(14) EN_t all observations	.89 (.01)	.77 (.02)	.79 (.02)	.78 (.02)	.81 (.02)	.81 (.02)	.81 (.02)
(15)EN_t completed >7 semesters	.90 (.02)	.83 (.02)	.84 (.02)	.82 (.03)	.82 (.03)	.81 (.03)	.81 (.03)
(16) EN_t completed 4,5,6 semesters	.82 (.04)	.73 (.05)	.73 (.05)	.75 (.05)	.79 (.05)	.78 (.07)	
(17) EN_t completed 1,2,3 semesters	.89 (.03)	.67 (.05)	.64 (.07)	.46 (.13)			

Table 2 Estimates of structural model: Estimate (std. error)

		1	2
Utility Parameters (Eq. 4)			
γ_0 -Constant	1	-2.186 (.434)**	-.819 (.375)**
γ_1 -Coefficient on EN	2	.479 (.179)**	
γ_2 -Coefficient on G	3	.292 (.078)**	1.14 (.135)**
γ_3 -Coefficient on g	4	.304 (.171)*	
τ - variance $\varepsilon_{N,t}$, $\varepsilon_{S,t}$ is $\tau^2\pi^2/6$	5	.986 (.049)**	1.0 (normalized)
Evolution of EN_t (Eq. 7)			
$\alpha_{EN,0}$ -Constant	6	-.288 (.207)**	
$\alpha_{EN,1}$ -Coefficient on EN	7	1.130 (.085)**	
$\alpha_{EN,2}$ -Coefficient on g	8	.212 (.055) **	
Determinants of of μ_{t+1} (Eq. 8)			
<u>t=1 and t=2</u>			
$\alpha_{\mu,0}$ -Constant	9	1.423 (.125)**	
$\alpha_{\mu,1}$ -Coefficient on μ	10	.320 (.038)**	
$\alpha_{\mu,2}$ -Coefficient on g	11	.255 (.014)**	
$Var(v_{\mu,t+1})$	12	.095 (.005)**	
<u>t=3 and t=4</u>			
$\alpha_{\mu,0}$ -Constant	13	1.055 (.119)**	
$\alpha_{\mu,1}$ -Coefficient on μ	14	.463 (.033)**	
$\alpha_{\mu,2}$ -Coefficient on g	15	.212 (.021)**	
$Var(v_{\mu,t+1})$	16	.069 (.004)**	
<u>t>4</u>			
$\alpha_{\mu,0}$ -Constant	17	.902 (.005)**	
$\alpha_{\mu,1}$ -Coefficient on μ	18	.558 (.031)**	
$\alpha_{\mu,2}$ -Coefficient on g	19	.172 (.015)**	
$Var(v_{\mu,t+1})$	20	.071 (.004)**	
Determinants of σ_{t+1} (Eq. 9)			
<u>t=1 and t=2</u>			
$\alpha_{\sigma,0}$ -Constant	21	.245 (.027)**	
$\alpha_{\sigma,1}$ -Coefficient on σ	22	.265 (.042)**	
$Var(v_{\sigma,t+1})$	23	.056 (.003)**	
<u>t=2 and t=3</u>			
$\alpha_{\sigma,0}$ -Constant	24	.133 (.037)**	
$\alpha_{\sigma,1}$ -Coefficient on σ	25	.494 (.067)**	
$Var(v_{\sigma,t+1})$	26	.052 (.003)**	
<u>t>3</u>			
$\alpha_{\sigma,0}$ -Constant	27	.066 (.031)**	
$\alpha_{\sigma,1}$ -Coefficient on σ	28	.614 (.050)**	
$Var(v_{\sigma,t+1})$	29	.039 (.003)**	
Log Likelihood		-759.648	

Figure 1 Earnings expectations

Different scenarios and different ages

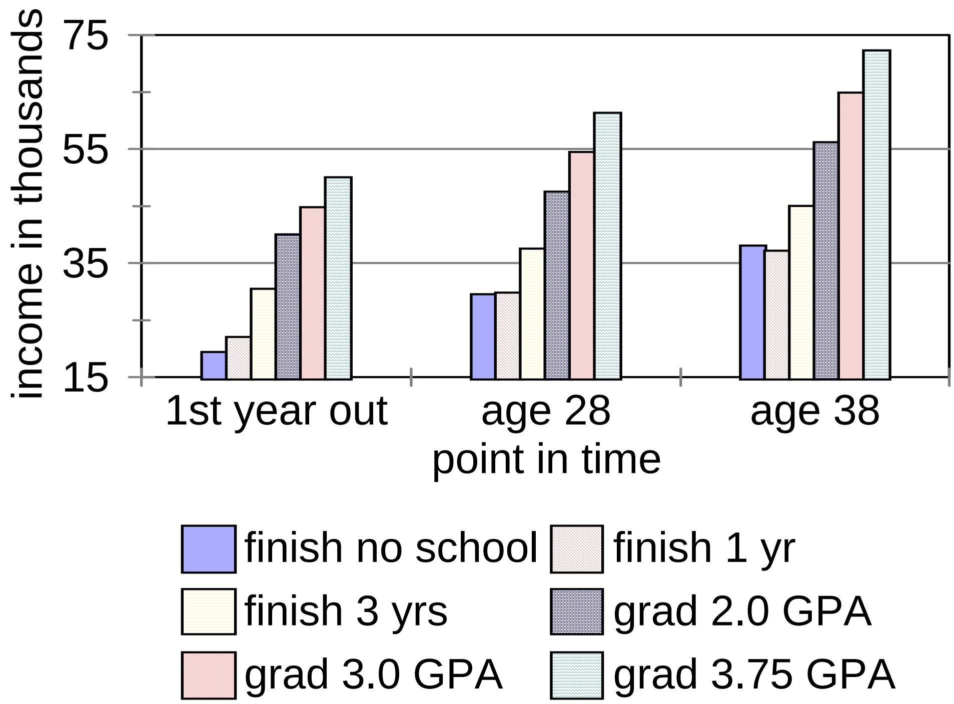
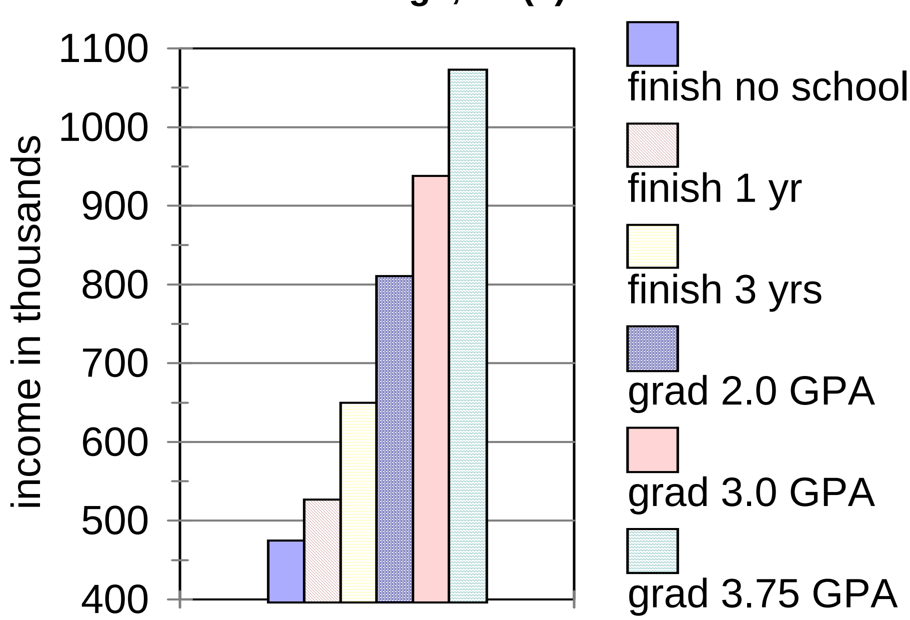


Figure 2 Discounted Expected Lifetime Earnings, VN(t')



Appendix A.

Question A.1 Circle the one answer that describes your beliefs at this time: (**Beginning of first year)**

1. I believe that being in college at Berea will be much more enjoyable than not being in college.
2. I believe that being in college at Berea will be somewhat more enjoyable than not being in college.
3. I believe that I will enjoy being in college at Berea about the same amount as I would enjoy not being in college.
4. I believe that being in college at Berea will be somewhat less enjoyable than not being in college.
5. I believe that being in college at Berea will be much less enjoyable than not being in college.

Question A.1 Circle the one answer that describes your beliefs at this time: (**Beginning of other semesters)**

1. I believe that being in college at Berea is much more enjoyable than not being in college.
2. I believe that being in college at Berea is somewhat more enjoyable than not being in college.
3. I have enjoyed being in college at Berea about the same amount as I would have enjoyed not being in college.
4. I believe that being in college at Berea is somewhat less enjoyable than not being in college.
5. I believe that being in college at Berea is much less enjoyable than not being in college.

Question A.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 this semester.

Given the amount of study-time you indicated above (question now shown here), 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.

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, .99]	_____________

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

Appendix B

The primary burden of computing value functions involves the computation of the expected future utility (Emax) of the option (S) in equation (2).

We assume that students believe that they will update μ_{t+1} and σ_{t+1} according to the predicted values from equations (8) and (9):

(A.1) $\mu_{t+1} = \alpha_{\mu,0} + \alpha_{\mu,1}\mu_t + \alpha_{\mu,2}g_t$

(A.2) $\sigma_{t+1} = \alpha_{\sigma,0} + \alpha_{\sigma,1}\sigma_t$.

Then, uncertainty about G_{t+1} and μ_{t+1} comes from uncertainty about g_t and uncertainty about EN_{t+1} comes from uncertainty about g_t and $v_{EN,,t+1}$. Letting $EMAX^*(EN_{t+1}, G_{t+1}, \mu_{t+1}, \sigma_{t+1})$ represent the well-known closed form that exists for the expected value of the maximum (conditional on EN_{t+1}, G_{t+1}, μ_{t+1}, and σ_{t+1}) when $\varepsilon_{S,t+1}$ and $\varepsilon_{N,t+1}$ have Extreme Value distributions,

(A.3) $Emax[V^{S}_{t+1}(\bullet), V^{N}_{t+1}(\bullet)] =$
$$\int\int EMAX^*(EN_{t+1}(g_t, v_{EN,t+1}), G_{t+1}(g_t)\ \mu_{t+1}(g_t), \sigma_{t+1})\ f(g_t)\ h(v_{EN,t+1})\ dg_t\, dv_{EN,t+1}$$

where f is the censored normal distribution in Eq. (5) which describes beliefs about g, and, as seen in Eq. (7), h is a standard normal random variable. A.3 can be rewritten as:

(A.4)

$$\begin{aligned}
&PR(g_t=0)^* \\
&\quad [Pr(EN_{t+1}=1|g_t=0)^*\ EMAX^*(EN_{t+1}=1, G_{t+1}(0), \mu_{t+1}(0), \sigma_{t+1}) \\
&\quad +Pr(EN_{t+1}=0|g_t=0)^*\ EMAX^*(EN_{t+1}=0, G_{t+1}(0), \mu_{t+1}(0), \sigma_{t+1})] \\
&+PR(g_t=4.0)^* \\
&\quad [Pr(EN_{t+1}=1|g_t=4)^*\ EMAX^*(EN_{t+1}=1, G_{t+1}(4), \mu_{t+1}(4), \sigma_{t+1}) \\
&\quad +Pr(EN_{t+1}=0|g_t=4)^*\ EMAX^*(EN_{t+1}=0, G_{t+1}(4), \mu_{t+1}(4), \sigma_{t+1})] \\
&+PR(0<g_{t+1}<4.0)^* \\
&\quad \int [Pr(EN_{t+1}=1|g_t)^*\ EMAX^*(EN_{t+1}=1, G_{t+1}(g_t), \mu_{t+1}(g_t), \sigma_{t+1}) \\
&\quad +Pr(EN_{t+1}=0|g_t)^*\ EMAX^*(EN_{t+1}=0, G_{t+1}(g_t), \mu_{t+1}(g_t), \sigma_{t+1})]\ f(g_t|(0<g_t<4.0)dg_t.
\end{aligned}$$

The integral in the last term of A.4 is simulated as the average value of the integrand over N draws from the conditional distribution $f(g_t|(0<g_t<4.0)$.

The most basic property of the standard solution algorithm for value functions is that, in order to solve all necessary value functions at time t, it is necessary to know value functions at time t+1 for each combination of the state variables in $\Omega(t+1)$ that could arise at time t+1.

Observable characteristics X are not burdensome because they are assumed to be exogenous and predetermined. This implies that value functions at time t+1 need to be solved only for the observed value of these variables. Similarly, ε_{t+1} is not computationally burdensome because it is assumed to be serially independent. In this case, ε_{t+1} influences $V^S_{t+1}()$ and $V^N_{t+1}()$ only through its effect on current period (t+1) utility. In general contexts, this would imply that, given $V^S_{t+1}()$ and $V^N_{t+1}()$ for some value ε_{t+1}, $V^S_{t+1}()$ and $V^N_{t+1}()$ could be obtained in a trivial manner for any other value ε_{t+1} by simply recalculating U^S_{t+1} and U^N_{t+1}. In the specific case here, where $\varepsilon_{S,t+1}$ and $\varepsilon_{N,t+1}$ have Extreme Value distributions, $V^S_{t+1}()$ and V^N_{t+1} do not have to be computed explicitly for different values of ε_{t+1} since the integration over $V^S_{t+1}()$ and V^N_{t+1} in the Emax leads to the well-known closed form solution represented by Emax* above.

The burden of solving value functions comes primarily from the variables EN_{t+1}, G_{t+1}, μ_{t+1}, and $\sigma_{t+1.}$ For each of these variables, the computational burden arises because: 1) there are multiple values for which value functions are needed at time t+1 and 2) the current period value of the variable provides information about both current and future utility. The latter characteristics implies that, in order to compute $V^S_{t+1}()$ for any particular combination of these variables, it is necessary to recompute the computationally demanding Emax in time t+2.

EN_{t+1} is a discrete (binary) variable so it can take on only two particular values at time t+1. However, the remaining variables are serially correlated continuous variables, and this causes well-known difficulties for the backwards recursion solution methods. As discussed in detail in Bound et al. (2010), Keane and Wolpin (1994), Rust (1997), and Stinebrickner (2000), quadrature or simulation methods are a useful tool for addressing the difficulties of serially correlated, continuous variables because, in effect, they served to discretize the state space - an obvious necessity given that the backwards recursion process requires that value functions be solved for all combinations of state variables. Unfortunately, while finite, the number of possible combinations of G_{t+1}, μ_{t+1}, and σ_{t+1} is in practice very large so that it is infeasible to solve value functions using standard methods for all possible combinations of EN_{t+1}, G_{t+1}, μ_{t+1}, and σ_{t+1} that could arise.

We address this issue by implementing a modified version of the backwards solution process. The first step is to determine the range of possible values that each of the variables G_{t+1}, μ_{t+1}, and σ_{t+1} could have in each time period for which the individual is making decisions. The modified backwards recursion process can then take place. At each time t in the backwards recursion process, rather than solving value functions for all possible values of G_t, μ_t, and σ_t, value functions, V^S is solved for the largest possible values of each of these variables, the smallest possible values for each of these variables, and some subset of the possible values in between the largest and smallest possible values for each of these variables. We refer to a combination of values of G_t, μ_t, and σ_t for which V^S_t is solved as a grid point. The simulation of A.4 implies that solving the value functions associated with the grid points at time t requires knowledge of value functions V^S at time t+1 for various combinations of G_{t+1}, μ_{t+1}, and σ_{t+1}. The reality that these needed combinations will not correspond to the time t+1 grid point (for which value functions were actually solved at t+1) necessitates a value function approximation. Specifically, we interpolate the t+1 value function associated with a particular combination G_{t+1}', μ_{t+1}', and σ_{t+1}' as the weighted average of the value functions associated with the eight "surrounding" grid points, where the weight associated with a particular grid point is determined by the euclidian distance between the grid point and G_{t+1}', μ_{t+1}',and σ_{t+1}'.[16] This nonparametric interpolation approach using surrounding grid points has the virtue that the interpolated value function for (G_{t+1}', μ_{t+1}', σ_{t+1}') converges to the true value function as the number of grid points increases (i.e., as the grid points used in the weighted average become close to (G_{t+1}', μ_{t+1}', σ_{t+1}') .

[16]The "surrounding" grid points are defined to be the eight possible combinations of {G_{t+1}^H, G_{t+1}^L}, { μ_{t+1}^H, μ_{t+1}^L}, and {σ_{t+1}^H, σ_{t+1}^L},where G_{t+1}^H is smallest value greater than G_{t+1}' for which value functions were solved at time t+1, G_{t+1}^L is largest value less than G_{t+1}' for which value functions were solved at time t+1, μ_{t+1}^H is smallest value greater than μ_{t+1}' for which value functions were solved at time t+1, μ_{t+1}^L is largest value less than μ_{t+1}' for which value functions were solved at time t+1,σ_{t+1}^H is smallest value greater than σ_{t+1}' for which value functions were solved at time t+1, σ_{t+1}^L is largest value less than σ_{t+1}' for which value functions were solved at time t+1. Then, the surrounding grid points form a cube around the point (G_{t+1}', μ_{t+1}', σ_{t+1}').