An Empirical Investigation of the Option Value of College Enrollment

Kevin Stange*

University of Michigan

Ford School of Public Policy

kstange@umich.edu

November 11, 2010

Abstract

This paper quantifies the option value arising from sequential schooling decisions made in the presence of uncertainty and learning about academic ability. College attendance has option value since enrolled students have the option, but not obligation, to continue in school after learning their aptitude and tastes. I estimate that option value accounts for 14% of the total value of the opportunity to attend college for the average high school graduate and is greatest for moderate-aptitude students. Students' ability to make decisions sequentially in response to new information increases welfare and also makes educational outcomes less polarized by background.

*I am particularly thankful to David Card for many helpful discussions and invaluable advice throughout the duration of this project and to David S. Lee, Enrico Moretti, Steve Raphael, and Emmanuel Saez for their guidance. This work also benefited from discussions with Peter Arcidiacono, Raymundo Campos, Ada Chen, Raj Chetty, Avraham Ebenstein, Dan Hartley, Jane Herr, Robert Johnson, Damon Jones, Ashley Langer, Ron Lee, Nate Miller, Paulina Oliva Vallejo, Marit Rehavi, Megan Tompkins-Stange, and numerous seminar participants. Financial support for this work was provided by the National Institute for Child Health and Human Development (Interdisciplinary Training Grant No. T32-HD007275), the U.C. Berkeley Institute for Business and Economics Research, and the Center for Labor Economics at U.C. Berkeley. All errors are of course my own.

I Introduction

Since the pioneering work of Becker and Mincer, the application of investment theory to the study of individuals' education decisions has become commonplace. People are assumed to weigh short-term costs against future benefits and choose the schooling level that maximizes welfare. This static framework abstracts from uncertainty and suggests that few people should drop out if the marginal earnings gain from graduating is high, as it appears to be. In reality, schooling decisions involve much uncertainty, outcomes often deviate from expectations, and dropout is common.[1] Despite its salience and its importance to investment generally, uncertainty has historically received relatively little attention in the study of education.[2]

This paper examines the consequences of educational uncertainty using a structural model in which schooling decisions are sequential and academic ability is learned through grades. Since psychic schooling costs depend on ability, people refine their expectations of them over time. This set-up is analogous to Pindyck's (1993) model of "technical" cost uncertainty, where the cost of completing a long-term project is revealed only as investment proceeds. Option value arises in this context since students have the option, but not obligation, to continue in school after learning their aptitude and tastes. My estimates suggest option value is substantial for the average high school graduate and is greatest for moderate-ability students. Their decisions are particularly sensitive to new information, so they derive the most value from learning it. The ability to condition sequential education decisions on new information increases welfare and also makes educational outcomes less polarized by background. Option value also rationalizes dropout in the presence of the substantial degree or "sheepskin" effects.

One implication is that policies that restrict dynamic flexibility curtail welfare most for those closest to the decision margin. School tracking, for example, will have the greatest impact on the welfare of students who are most uncertain about their fit with vocational or academic tracks. The general setup can be used to examine a wide range of phenomenon - job choices, marital decisions, health investments - in which decisions are sequential, partially irreversible, and responsive to new

[1]For instance, only 51% of 1982 high school seniors who intended to earn a Bachelor's degree had done so by 1992, while 16% of those planning to earn less than a four-year degree eventually did according to the National Center for Educational Statistics, *2004 Digest of Educational Statistics* Table 307.

[2]Uncertainty is at the heart of a burgeoning body of very recent empirical work on schooling, as surveyed in Heckman, Lochner, and Todd (2006). In the investment literature, most relevant here is the work related to real options collected in Dixit and Pindyck (1994).

information.

This paper quantifies the importance of uncertainty and computes option value through simulations of a structural dynamic model, which is estimated using postsecondary transcript data on a recent cohort of U.S. men from the National Educational Longitudinal Study (NELS). The model encompasses enrollment decisions and grade outcomes over four years as well as the decision to start at a two-year (community) or four-year college. I simulate educational outcomes and welfare using the dynamic model and compare this to the counterfactual scenario wherein individuals commit to an educational outcome before enrolling in college. The welfare difference between these two scenarios is the value of the option to respond to the information received during college.[3]

I assume that enrollment reveals three pieces of information. The first is collegiate aptitude, which influences the persistent psychic costs (or benefits) from school attendance. Enrollment provides information in the form of course grades which are used to predict the future desirability of school. Non-persistent shocks to the relative cost (or benefit) of schooling are the second. These shocks combine many factors - getting ill, having a parent lose a job, having a winning football team - that are not expected to persist over time. The final source of uncertainty is about labor market opportunities associated with higher levels of education. Expected lifetime income increases with education but the specific realization is unknown ex-ante. Individuals learn of these opportunities only if they actually enroll.[4] Since decisions can be conditioned on all this information, acquiring it has value.

Estimates suggest that uncertainty about college completion is empirically important; unanticipated taste shocks are half as large as the returns to the final year of college and dwarf direct tuition fees at public colleges. There is also evidence of learning about ability - over time people put increased weight on course grades in their continuation decisions. Because of this uncertainty, the average high school graduate would be willing to pay $14,900 (in 1992 dollars) to maintain the ability to decide sequentially, with moderate-ability students (for whom educational outcomes are most uncertain) willing to pay even more (up to $25,000 in 1992 dollars). Option value accounts

[3]In order to isolate the value of new information, I adopt Dixit and Pindyck's (1994) definition of option value, which nets out the continuation value that arises even with no uncertainty if returns are nonlinear. Heckman, Lochner, and Todd (2006), Heckman and Navarro (2007), and Heckman and Urzua (2008) define option value inclusive of this continuation value. See Section IIC.

[4]Such preference and labor market shocks are common features in the dynamic structural models of Keane and Wolpin (1997) and others.

for 14% of the total value of the opportunity to attend college among all high school graduates and 32% for those closer to the enrollment margin. Approximately 60% of this value comes from the information received in the first year of college. The ability to make decisions sequentially increases both enrollment and dropout, but also closes a quarter of the welfare gap between the first-best scenario (individuals maximize welfare ex-post) and the static one (individuals commit to outcomes ex-ante).

Though most previous treatment of this subject has been theoretical, recent empirical work also underscores the importance of schooling uncertainty and option value.[5] For instance, Altonji (1993) finds large differences between mean ex-ante and ex-post returns to starting college and Cunha, Heckman, and Navarro (2005) conclude that 30% of people would change their schooling decisions if they had perfect information. Chen (2008) estimates that 80% of potential wage variation reflects uncertainty and this share varies across education levels. Uncertainty is clearly important empirically.

This paper is in the tradition of the multi-period dynamic structural schooling models exemplified by Keane and Wolpin (1997), but with two key contributions. First, I augment their basic model to include learning about ability through course grades, similar to Arcidiacono (2004). Psychic costs (which depend on ability) are very important to schooling decisions, but their nature is not well understood.[6] Heckman and Navarro (2007) discuss identification of a general model which permits learning about serially-persistent attributes (such as psychic costs), but leave estimation for future work. Learning about academic ability is one source of option value not present in previous empirical work.[7]

Second, I examine the properties and consequences of option value using a fully estimated dynamic structural model. Heckman, Lochner, and Todd (2006) caution that rates of return to schooling depend on the empirical importance of option value, yet previous work ignores this. They provide preliminary estimates of it using a calibrated model with exogenous dropout, con-

[5]Weisbrod (1962) was the first to point out that education has option value. Also see the theoretical work of Comay, Melnick, and Pollatschek (1973), Dothan and Williams (1981), and Manski (1989).

[6]See Heckman, Lochner, and Todd (2006) for a discussion of recent evidence on the importance of psychic costs.

[7]Arcidiacono (2004) estimates the returns to various majors after controlling for dynamic selection, using course grades as a signal of subject-specific unobserved ability. His model and estimates could be used to quantify the option value arising from learning about subject-specific ability, as is also examined in Altonji (1993). Since he only examines students admitted to four-year colleges, however, his estimates cannot be used to investigate the importance of learning to enrollment decisions more generally.

cluding that much more work is needed on the subject.[8] This paper uses a simple theoretical model to show how uncertainty creates option value and influences enrollment decisions, particularly for those at the margin. These properties are quantified using the estimated structural model, in order to examine the empirical importance of option value to educational attainment, welfare, and policy.[9]

The rest of the paper proceeds as follows. Section II uses a simple two-period model to analyze how option value arises in the presence of educational uncertainty. This section states the definition of option value used in this study and discusses several of its properties. Section III presents the full empirical model and discusses issues related to its estimation. Estimation results are presented in Section IV, which also includes a discussion of model fit. Section V uses the estimated model to calculate the option value created by the sequential nature of schooling decisions. Section VI concludes by identifying directions for future work as well as other applications.

II Modeling Educational Investment

A The college dropout puzzle

The static model of educational investment widely used in the literature is inconsistent with high levels of college dropout if degree wage effects are large. Consider a simple version of the traditional model first developed by Becker (1964) as discussed in Card (1999). Individuals are assumed to maximize lifetime utility, which is a function of lifetime earnings and the (monetary and psychic) cost of schooling, $U = \ln y(S) - c(S)$, where $c(S)$ is some increasing and convex function of years of schooling. If $y(S)$ and $c(S)$ are continuous and differentiable, then the optimal schooling level (S_i^*) satisfies the first order condition $\frac{dy_i(S_i^*)}{dS}\frac{1}{y_i(S_i^*)} = \frac{dc_i(S_i^*)}{dS}$. The benefit of an additional year of schooling (higher earnings) just offsets the additional costs (delayed earnings and psychic costs) at the optimum.

However, the returns to college appear to be highly non-linear with substantial degree or

[8] Work in progress by Heckman and Urzua (2008) is also quantifying option value using an estimated dynamic model of schooling.

[9] Reduced form techniques are inadequate for quantifying option value, but they can be used to explore its importance to various decisions. For instance, Eide and Waehrer (1998) examine whether students consider the likelihood of graduate school (and accompanying wage gains) when choosing a college major choice.

"sheepskin effects."[10] Figure I presents estimates of the earnings production function for male high school graduates from the National Longitudinal Survey of Youth 1979 (NLSY79). The present discounted value of lifetime earnings minus tuition jumps discretely at four years of college, but is unrelated to schooling attainment until then. If psychic schooling costs are smooth, individuals should bunch at this discontinuity and very few people should fall in the intermediate ranges. Figure I also plots the distribution of postsecondary schooling attainment for men aged 35, who have presumably all completed their schooling. Consistent with the traditional model, zero (39% of the sample) and four years (17%) of college are the most frequent schooling outcomes. Ten percent attend college for two years, which partially reflects Associate's degree attainment. Contrary to the theory, however, there are many people whose schooling level puts them on the flat part of the earnings production function. Fully 28% of high school graduates drop out before finishing their fourth year of college. From the perspective of traditional human capital theory where individuals optimally choose their schooling level to equate the known marginal costs and benefits of an additional year, these individuals seemingly present an unexplained puzzle.

However, dropout can be rationalized when schooling decisions are sequential and the feasibility and desirability of degree completion is unknown ex-ante. As pointed out by Altonji (1993), uncertainty about the difficulty of graduating can interact with nonlinearities in the ex-post returns to schooling to create option value. Students with schooling outcomes on the flat part of the earnings curve may therefore be people for whom option value made enrollment worthwhile, even though the return was negative ex-post.

B A simple dynamic model of college enrollment and completion

Now consider a simple dynamic model with two periods, which correspond to the first and second half of college.[11] Utility is in dollars, individuals are assumed to be risk-neutral, and time discounting is ignored. At period one, individuals decide whether or not to enroll in college. Entering the labor market immediately provides zero utility but enrollment provides an individual-specific net return to the first half of college ($\varepsilon_{i,1}$) which is known throughout. At period two, those who

[10] There is a substantial literature that documents the existance of nonlinearities (or "sheepskin" effects) in the returns to education. See Hungerford and Solon (1987), Jaeger and Page (1996), Park (1999), and Heckman, Lochner, and Todd (2006).

[11] A similar two-period setup was used by Manski (1989), Altonji (1993), Taber (2000), and Arcidiacono (2004).

enrolled decide whether or not to graduate. Dropping out provides no further utility but graduating provides additional utility of $\rho\varepsilon_{i,1} + \varepsilon_{i,2}$, where $\varepsilon_{i,2}$ is revealed at the start of period two and $Cov[\varepsilon_{i,1}, \varepsilon_{i,2}] = 0$. Individual-specific returns to the second half of college have a component that is known when the initial enrollment decision is being made ($\rho\varepsilon_{i,1}$) and one that is only learned after enrollment ($\varepsilon_{i,2}$). This specification allows the returns in each period to be correlated, so first period returns provide information about the desirability of attending the second period. For expositional simplicity, I normalize mean returns to zero in each period, $E[\varepsilon_{i,1}] = E[\varepsilon_{i,2}] = 0$.[12] I focus on the case where returns are non-negatively correlated, $\rho \geq 0$. Figure II illustrates the structure and payoffs of the model.

Static model. First consider the fully static case where individuals make a single schooling decision between the three schooling outcomes (no enrollment, dropout, complete) at period one. Since they have no knowledge of $\varepsilon_{i,2}$, they set it to its expected value when evaluating the payoffs. The decision rules of individual i are thus:

$$Enroll\ if : \varepsilon_{i,1} + \max\{0, \rho\varepsilon_{i,1}\} > 0$$

$$Complete\ if : \rho\varepsilon_{i,1} > 0$$

Individuals will enroll and complete if $\varepsilon_{i,1} > 0$ and not enroll otherwise. Here the static model predicts no dropouts; anyone for whom enrollment is desirable will also want to complete college. To see this, note that payoffs are 0, $\varepsilon_{i,1}$, and $\varepsilon_{i,1}(1+\rho)$ for non-enrollees, dropouts, and completers, respectively, so completing college dominates dropping out if $\rho > 0$. The presence of non-linear returns (e.g. $E[\varepsilon_{i,2}] = r > E[\varepsilon_{i,1}]$) will only magnify this result. With negatively correlated returns ($\rho < 0$), the static model predicts that some people will drop out but all who do will have positive ex-post payoffs.[13]

Dynamic model. Now consider the dynamic case, where individuals only have to make the enrollment decision at period 1. People will enroll if the expected utility from doing so is greater than zero, where expectations are taken over the distribution of the unknown second-period returns

[12] The model can easily incorporate nonlinearities in returns by setting $E[\varepsilon_{i,2}] = r > E[\varepsilon_{i,1}]$. Nonlinearities are not necessary to create option value, but simply highlight option value's importance in explaining dropout. The model I actually estimate uses the empirical returns to each year of college, which permit nonlinearities.

[13] This is a key difference between the static and dynamic models. While people who drop-out in the static model will have positive ex-post payoffs, some drop-outs in the dynamic context will have negative ex-post payoffs.

$\varepsilon_{i,2}$. The model is solved starting with the completion decision in period two, when all parameters are known. The decision rules of individual i are thus:

$$\begin{aligned} Enroll\ if &: \varepsilon_{i,1} + E[\max\{0, \rho\varepsilon_{i,1} + \varepsilon_{i,2}\}] > 0 \\ Complete\ if &: \rho\varepsilon_{i,1} + \varepsilon_{i,2} > 0 \end{aligned}$$

The enrollment decision incorporates not only the immediate payoffs ($\varepsilon_{i,1}$) but also the expectation of future ones ($E[\max\{0, \rho\varepsilon_{i,1} + \varepsilon_{i,2}\}]$). Now the enrollment and completion decisions are not completely coupled since completion can be conditioned on the realized value of $\varepsilon_{i,2}$. This property has several implications for the level of enrollment, dropout, and welfare, to which I now turn.

C The option value of college enrollment

A key feature of the dynamic model where dropout is endogenous is that the expected net utility gain from completing college is truncated at zero. If $\varepsilon_{i,2}$ is sufficiently adverse, then individuals will choose to drop out rather than assume this adverse shock. By providing information about the desirability of completion, enrollment thus has value beyond the utility provided in the first period directly. This section defines the option value created by uncertainty and discusses the implications of option value for educational outcomes and welfare.

Enrollment is valuable because it leads to outcomes people may want to commit to ex-ante and because it provides information about the desirability of completion. The value of the opportunity to enroll can be decomposed into these two parts.

$$V_{dynamic}(\varepsilon_{i,1}) = V_{static}(\varepsilon_{i,1}) + OptionValue(\varepsilon_{i,1}) \tag{1}$$

$V_{dynamic}(\varepsilon_{i,1})$ is the value of the opportunity to enroll for individual i (as a function of $\varepsilon_{i,1}$) in the dynamic setting where individuals can drop out if continuation ends up being undesirable. $V_{static}(\varepsilon_{i,1})$ is the value of the enrollment opportunity in the static case, where individuals commit to an educational outcome ex-ante. Define $\bar{\varepsilon}_{d,1}$ as the critical value above which enrollment is optimal in the dynamic setting and $\overline{\varepsilon}_{s,1}$ analogously in the static setting.[14]

From above we have $V_{dynamic}(\varepsilon_{i,1}) = \max\left(0, \varepsilon_{i,1} + E[\max\{0, \rho\varepsilon_{i,1} + \varepsilon_{i,2}\}]\right)$ and $V_{static}(\varepsilon_{i,1}) = \max\left(0, \varepsilon_{i,1} + \max\{0, E\left[\rho\varepsilon_{i,1} + \varepsilon_{i,2}\right]\}\right)$. Thus option value can be written as:

$$\begin{aligned} OptionValue(\varepsilon_{i,1}) &= \underbrace{\max\left(0, \varepsilon_{i,1} + E[\max\{0, \rho\varepsilon_{i,1} + \varepsilon_{i,2}\}]\right)}_{V_{dynamic}(\varepsilon_{i,1})} \\ &\quad - \underbrace{\max\left(0, \varepsilon_{i,1} + \max\{0, E\left[\rho\varepsilon_{i,1} + \varepsilon_{i,2}\right]\}\right)}_{V_{static}(\varepsilon_{i,1})} \end{aligned} \tag{2}$$

This definition of option value nets out the continuation value arising from nonlinear returns with no uncertainty, consistent with Dixit and Pindyck (1994). If completing the first year of college is required in order to enter the second year, then the first year has continuation value. Continuation value may cause people with negative first year returns to enroll if second year returns are sufficiently high. However, if second period returns are uncertain and future decisions can be conditioned on new information, then even individuals who expect negative returns in both periods ($\varepsilon_{i,1} < 0$) may find it optimal to enroll. In this paper, I focus on this latter effect. Heckman, Lochner, and Todd (2006), Heckman and Navarro (2007), and Heckman and Urzua (2008) define option value inclusive of the continuation value, which is appropriate given their interest in estimating total returns.[15] While my estimates of the opportunity to attend college includes both continuation and option value, this paper primarily focuses on the latter. Proposition 1 describes the properties of option value as defined in this paper.

Proposition 1 *(The properties of option value).*

a. $OptionValue(\varepsilon_{i,1})$ *is non-negative for all* $\varepsilon_{i,1}$.

b. $OptionValue(\varepsilon_{i,1})$ *is greatest for individuals at the enrollment margin in the static model.*

c. $OptionValue(\varepsilon_{i,1})$ *is increasing and the critical value* $\overline{\varepsilon}_{d,1}$ *is decreasing in the level of uncertainty (variance of* $\varepsilon_{i,2}$*).*

[14] Here $\overline{\varepsilon}_{s,1} = 0$ and $\overline{\varepsilon}_{d,1}$ solves $V_{dynamic}(\varepsilon_{i,1}) = 0$.

[15] Roughly speaking, this distinction is a matter of how to treat the extent to which the option is "in the money" when it is granted. In the above notation, Heckman, Lochner, and Todd (2006) would define option value as :
$OV(\varepsilon_{i,1}) = \max\left(0, \varepsilon_{i,1} + E[\max\{0, \rho\varepsilon_{i,1} + \varepsilon_{i,2}\}]\right) - \max\left(0, \varepsilon_{i,1}\right)$

d. $OptionValue(\varepsilon_{i,1})$ *reduces the dependence of educational outcomes on* $\varepsilon_{i,1}$.

e. *The option to drop out improves welfare.*

Proof. See Appendix 1. ■

Figure III illustrates these features of option value in this context through simulations. The left panel plots the value of the enrollment opportunity for a range of values of $\varepsilon_{i,1}$ and for different levels of uncertainty about $\varepsilon_{i,2}$. The dotted line is the value of the enrollment opportunity in the static case, $V_{static}(\varepsilon_{i,1})$. This value is zero for those who choose not to enroll ($\varepsilon_{i,1} < 0$) and then increases linearly with $\varepsilon_{i,1}$. The dashed lines plot the value of the enrollment opportunity in the dynamic situation where $\varepsilon_{i,2}$ is uncertain, $V_{dynamic}(\varepsilon_{i,1})$, for two different levels of uncertainty about $\varepsilon_{i,2}$. The vertical distance between the dashed and dotted lines represents the $OptionValue(\varepsilon_{i,1})$. For comparison, the solid line plots the average welfare in the full information counterfactual scenario where individuals can make education decisions to maximize welfare ex-post, after learning $\varepsilon_{i,2}$. Figure III confirms that $OptionValue(\varepsilon_{i,1})$ is increasing in σ. In contrast to the standard view that uncertainty reduces welfare if agents are risk averse, here uncertainty combined with the ability to respond dynamically actually increases welfare by increasing the option value. As $OptionValue(\varepsilon_{i,1})$ increases due to increased uncertainty about $\varepsilon_{i,2}$, enrollment becomes desirable to more people. This can also be seen in Figure III: $\overline{\varepsilon}_{d,1}$ is where the dashed lines intersect the horizontal axis. Even without nonlinearities, option value will make enrollment desirable to people for whom the first half of college is unproductive ($\varepsilon_{i,1} < 0$). In Figure III, the vertical distance between the solid line and the others represents the welfare loss resulting from incomplete information about $\varepsilon_{i,2}$. The ability to drop out after learning $\varepsilon_{i,2}$ (the dashed line) closes much of this welfare gap.

The sources of the welfare gains coming from the ability to drop out can be seen more clearly by looking at educational outcomes under the various scenarios. The right panel of Figure III plots the fraction enrolling in (Panel A) and completing college (Panel B) under the static, dynamic, and full-information scenarios described above. Individuals in Group A receive no schooling in either the static or dynamic settings, though some (with high $\varepsilon_{i,2}$) would enroll and graduate if they knew $\varepsilon_{i,2}$ with certainty. Individuals in Group B are compelled to enroll despite their negative first period returns because of the informational value. Though many will eventually drop out, others will graduate and the costs of experimenting are not too high. This group receives considerably more education in the dynamic setting. Interestingly, a small subset of these individuals actually continue to graduation due to the sunk-cost nature of their period 1 investment, despite this being suboptimal ex-post. Group C benefits from the dynamic setting because they have the option to drop out if continuation is undesirable. In the static model, all commit to graduating, even if it is undesirable ex-post. Option value increases the welfare of this group by reducing their educational attainment.

D Implications for empirical work

A simple dynamic model of college enrollment and completion was motivated by the failure of the static model to explain high rates of college dropout. In a dynamic setting, dropout occurs when new information reveals that continuation is not desirable. The opportunity to drop out in response to this information creates option value, which was shown to have important consequences for educational outcomes and welfare. Specifically, option value increases the incentive to enroll, particularly for those at the enrollment margin in the static model. Any model that ignores this value will necessarily understate the incentive to enroll and mischaracterize the social desirability of college dropout.

III Empirical Implementation

To characterize schooling uncertainty quantitatively, I estimate an empirical model that is a much richer version of the basic model presented above. The empirical model describes enrollment decisions and grade outcomes at four time periods and allows individuals to start at either a two-year or four-year college. The model includes several sources of uncertainty. Like many dynamic models, I include unanticipated shocks to the relative desirability of school and labor market entry at each point in time. For example, receiving an unusually favorable outside job offer or getting ill influences the relative desirability of schooling and work at a single period. These shocks are assumed to be serially uncorrelated. The second source of uncertainty is about academic aptitude, which influences taste for schooling throughout college. Students do not know for certain whether they are a "B" or "C" college student until they enroll. Grades following enrollment provide a signal of this unobserved ability and students learn about their aptitude through their grades. This section presents the key elements of the empirical model and discusses issues related to its estimation. A complete description of the full model is contained in the Appendix, which also includes a discussion of several important extensions.

A Data

The model is estimated on a panel of 2055 men participating in the National Educational Longitudinal Study (NELS). NELS participants were first interviewed in 1988, while in 8th grade, then again in 1990, 1992, 1994, and 2000. Complete college transcripts were obtained in 2000 for most participants. The NELS transcript and survey data are used to construct the main variables used in the analysis: college enrollment indicators, grade outcomes, and baseline characteristics. I supplemented the NELS dataset with institutional characteristics obtained from the Integrated Postsecondary Education Data System (IPEDS) 1992 Institutional Characteristics survey. For each NELS individual, I merged on distance to the nearest two-year and four-year college (in miles) and average tuition levels at public two-year and four-year colleges in each state.[16]

I define a time period as one academic year and classify individuals by years of continuous college enrollment following high school graduation. Students are considered enrolled during year t if they attempted at least six course units (approximately part-time status) at either a two-year or four-year school in both Fall and Spring of the academic year. Since income measures as adults do not appear in the NELS dataset, I estimate conditional expectations of lifetime income using data from an earlier cohort, male high school graduates from the National Longitudinal Survey of Youth 1979 (NLSY79). Using variables that are common in both the NLSY79 and the NELS (such as high school GPA, parental education, AFQT, ethnicity, urban and region), I estimate the parameters of a lifetime income equation using OLS and predict counterfactual lifetime income for individuals in the NELS sample. Essentially, I assume that individuals in my sample look at the experience of "similar" individuals twelve years older to form their income expectations. This approach is similar to the "reference group expectations" referred to by Manski (1991).

[16] Characteristics of the specific schools students attend (e.g. tuition) is not used in this analysis. Average tuition levels in each state are a more exogenous source of variation in the price of college than own-school tuition, which varies considerably between public and private institutions and is endogenous. Tuition levels are set at their 1992 levels thoughout, assuming students don't reoptimize in response to short-term tuition changes.

I restrict the dataset to on-time high school graduates with complete information on key baseline variables (high school GPA, AFQT, parents' education, family income, distance to nearest colleges) and complete college transcripts (unless no claim of college attendance). I also exclude residents of Alaska, Hawaii, and the District of Columbia. After these restrictions the final dataset contains 2055 men. Appendix 2 contains summary statistics and more details on how the dataset was constructed. Though these restrictions reduce the sample considerably, the final unweighted analysis sample is very similar to a nationally representative sample of the high school class of 1992. Appendix 3 describes the counterfactual lifetime income estimation procedure.

B Model description

I model schooling decisions in the four academic years after high school graduation. During the first period individuals decide whether to start at a four-year or two-year college, which I refer to as pathway choice, or to not enroll in college. The pathway chosen affects the level and timing of direct schooling costs (which may differ across individuals) and unmodeled college amenities. At each time period t an individual chooses whether to enter the labor market (receiving payoff $u^w_{i,t}$) or continue in school for another year, receiving payoff $u^s_{i,j,t}$ in period t and the option to make an analogous work-school decision in period $t+1$, where $j = 2, 4$ denotes type of school (two-year or four-year). After period two, students that started at a two-year college must attend a four-year college if they want to continue in school. After period four, there are no more decisions to make and all individuals enter the labor market. Figure IV depicts the structure of choices, information, and payoffs in the full empirical model, where the individual subscripts have been omitted.

Utility is in dollars. The indirect utility from discontinuing school and entering the labor market at period t equals the expected present discounted value of lifetime income from period t to age 62 ($Income_{i,t}$) plus a random component $\varepsilon^w_{i,t}$. Note that t subscripts a decision period so it is collinear with years of education in this model. Thus predicted lifetime income depends implicitly on years of schooling, which is determined directly by when students leave school.

$$u^w_{i,t} = Income_{i,t} + \varepsilon^w_{i,t} \tag{3}$$

The expected indirect utility derived from attending school during year t, $u^s_{i,j,t}$, depends linearly on a type-specific intercept ($\alpha_{m,j}$), expected unknown ability (A_i), direct tuition and commuting costs, and a random component $\varepsilon^s_{i,j,t}$. $Distance_{i,j,t}$ and $Tuition_{i,j,t}$ vary by the type of school, so individuals that start at a two-year school will pay community college tuition for the first two years then four-year college tuition for their third and fourth years. The random shocks ($\varepsilon^s_{i,j,t}$,$\varepsilon^w_{i,t}$) are revealed to the individual prior to making the period t decision.

$$u^s_{i,j,t} = \alpha_{0,j} + \alpha_{m,j} + \alpha_A E_t[A_i] - (\alpha_D Distance_{i,j,t} + Tuition_{i,j,t}) + \varepsilon^s_{i,j,t} \tag{4}$$

The term $\alpha_A E_t[A_i]$ captures the preference for school (in dollar terms) that covaries with its expected difficulty.[17] Individuals do not know A_i at any time, so they form expectations of it when making their period-t decisions. I assume that individuals form rational expectations of their performance in school.[18] In period one, I make the parametric assumption that the conditional expectation of A_i on baseline characteristics depends linearly on a type-specific intercept (γ_m), high school grade point average ($HSgpa_i$), percentile score on the AFQT, and whether a parent has a college degree ($ParBA_i$):

$$E_1[A_i] = E[A_i|X_i] = \gamma_0 + \gamma_m + \gamma_G HSgpa_i + \gamma_T AFQT_i + \gamma_P ParBA_i \tag{5}$$

At the end of each year, students enrolled in college learn their performance during that year, which is measured by the college grade point average (on a four-point scale) during period t. I assume that grades provide a noisy signal of A_i: $g_{i,t} = A_i + \varepsilon^g_{i,t}$. Grade shocks are assumed to be serially uncorrelated and normally distributed: $\varepsilon^g_{i,t} \sim N(0, \sigma_{Gt})$. With learning, individuals update their belief about A_i in response to new information received through grades. I make the parametric assumption that the conditional expectation of A_i is a weighted average of the unconditional expectation and students' cumulative grade point average. The weights are parameters to

[17] This specification can be motivated by a model where the difficulty of year t is distributed around a fixed and unobserved individual-specific mean, so $A_{i,t} = A_i + \varepsilon^a_{i,t}$. Individuals learn $A_{i,t}$ after each year, but cannot separate A_i from $\varepsilon^a_{i,t}$. If $\varepsilon^a_{i,t}$ is mean zero and serially uncorrelated, then $E_t[A_{i,t}] = E_t[A_i]$. Also, since I have assumed risk neutrality, the variance of $\varepsilon^a_{i,t}$ has no impact on expected utility or decisions, so can be ignored.

[18] Stinebrickner and Stinebrickner (2008) have direct evidence that students are over-confident about their likely performance in college. How students form expectations about college is ripe area for future research.

be estimated.[19]

$$E_t[A_i] = \gamma_{Xt} E[A_i|X_i] + (1-\gamma_{Xt}) \sum_{q=1}^{q=t-1} \frac{g_{i,q}}{t-1} \text{ if } t > 1 \tag{6}$$

To permit a general structure of correlation between unobservable preferences and ability, I specify that $\alpha_{m,j}$ and γ_m come from a mass point distribution which describe the ability and schooling preferences of M different types of individuals.[20] γ_m measures the unobserved academic aptitude of people of "type" m and $\alpha_{m,j}$ is their preference for school of type j. Type is known to the individual throughout, but is unknown to the econometrician. Essentially, the specification permits the intercepts of academic performance and of indirect utility to each take on three different values, corresponding to the three unobserved types. As a special case, I will also estimate models with no unobserved heterogeneity, which assumes that all correlation between preference for school and academic aptitude are captured linearly through $\alpha_A E_t[A_i]$. $\overline{u}^s_{i,j,1}(\cdot)$ represents the non-stochastic component of the indirect utility of attending school. Individuals know baseline characteristics (X_i) as well as the first period shocks ($\varepsilon^s_{i,2,1}$,$\varepsilon^s_{i,4,1}$,$\varepsilon^w_{i,1}$) when making the initial enrollment decision, but learn future shocks and grade outcomes only after enrolling. All other parameters of the model are known to the individual throughout.

At each period t , the individual maximizes the expected discounted value of lifetime utility by choosing whether to discontinue schooling and receive $u^w_{i,t}$ or continue school for at least one more year. Solving the model consists of finding the value functions for each alternative at each point in time: $V^s_{i,2,t}$, $V^s_{i,4,t}$, and $V^w_{i,t}$. These value functions take the following form:

$$\begin{aligned} V^w_{i,t} &= u^w_{i,t} \\ V^s_{i,j,t} &= u^s_{i,j,t} + \beta E\left[\max\left\{V^w_{i,t+1}, V^s_{i,j,t+1}\right\}\right] \end{aligned} \tag{7}$$

The decision problem can be solved for each individual by backwards recursion. In order to get a closed form solution for the $E\left[\max\left\{.,.\right\}\right]$ term, I assume these shocks are drawn from an Ex-

[19]This is an approximation of the normal learning model, which imposes that $\gamma_{Xt} = \left(\frac{1/\sigma_a^2}{1/\sigma_a^2+(t-1)/\sigma_g^2}\right)$, where σ_a^2 is the variance of A_i and σ_g^2 is the variance of $(g_{i,t} - A_i)$. Instead of imposing that the learning process follow this structure, I estimate γ_{Xt} and the variance of the residual $g_{i,t} - E_t[A_i]$ as parameters.

[20]The use of a mass-point distribution to approximate the distribution of preferences known to the agent but unknown to the econometrician is discussed by Heckman and Singer (1984) and is widely used in dynamic structural work such as Keane and Wolpin (1997) and Eckstein and Wolpin (1999). Here I estimate models with up to three points of support.

treme Value Type I distribution with location and scale parameters zero and τ, respectively. The derivation of these value functions is contained in the Appendix.

C Interpretation of parameters

The indirect utility functions $\left\{V_{i,j,t}^{s}, V_{i,t}^{w}\right\}_{t=1}^{t=4}$ provide expressions for the relative desirability of entering the labor market or continuing in school at time t. This relative value depends on a number of primitive parameters. The direct and opportunity costs as well as financial returns are captured in the terms $Cost_{i,t}$ and $Income_{i,t}$. Their importance to educational decisions have been the topic of much examination. Less frequently studied is the contribution of academic ability to continuation decisions. This is captured by α_A and the parameters of the grade function. I have modeled family background and ability as influencing educational decisions primarily through expected scholastic aptitude (grades). This model can be used to quantify the contribution of family background to educational outcomes that operates through college academic performance. Family background influences academic performance, which in turn influences educational decisions.

The value of enrollment is also influenced by the amount of uncertainty and the speed at which it is revealed, as parameterized by τ and $\{\gamma_{Xt}, \sigma_{Gt}\}_{t=1}^{t=4}$. If τ is high, then preference shocks have a high variance, which increases the value of college enrollment and continuation. Future decisions take these preference shocks into account, so a greater variance increases the likelihood that either the schooling or work shock will be high, thus increasing the option value.

Option value decreases with the variance of grade shocks (σ_{Gt}). Since grades provide a noisy signal of unobserved ability (which influences utility through academic performance), greater variance decreases the signal value of grade realizations and thus the option value created by the ability to learn about aptitude through grades. If grades provided no signal value (either because they were completely random or because there is no uncertainty about ability), the value of enrollment would be diminished.

The temporal nature of learning about ability is parameterized by $\{\gamma_{Xt}\}_{t=1}^{t=4}$. If academic ability is learned quickly, then γ_{Xt} should decline rapidly at first then level off. If subsequent grade shocks continue to provide new information about ability, γ_{Xt} should continue to decline throughout college. The normal learning model imposes that γ_{Xt} follow a specific decreasing pattern over time.

D Estimation and identification

The parameters of the model are estimated with full information maximum likelihood using data on the enrollment decisions, academic performance, and baseline characteristics of a panel of individuals. With no unobserved heterogeneity, individual $i's$ contribution to the likelihood function is $L_i = L_i^1 \cdot L_i^2 \cdot L_i^3,$ where:

$$\begin{aligned} \text{Period 1:} \qquad & L_i^1 = \Pr(S_{i,2,1}=1)^{S_{i,2,1}} \Pr(S_{i,4,1}=1)^{S_{i,4,1}} \Pr(S_{i,1}=0)^{1-S_{i,1}} \\ \text{Periods 2 to 4:} \qquad & L_i^2 = \prod_{t=2}^{4} \Pr(S_{i,t}=1)^{S_{i,t}} \Pr(S_{i,t}=0)^{1-S_{i,t}} \\ \text{Grades} \quad : \quad & L_i^3 = \prod_{t=1}^{4} \Pr(g_{i,t}) \end{aligned} \tag{8}$$

where $S_{i,2,1}$ and $S_{i,4,1}$ indicate pathway choice in period 1 and $S_{i,t}$ is an indicator for enrollment in either type of school during period t. With the extreme value assumption on the preference shocks (which are unobserved to the econometrician), choice probabilities take the familiar logit form and the likelihood of grade outcomes given by the normal probability density function.

When unobserved (to the econometrician) heterogeneity is included, the likelihood contribution of individual i must be integrated over the joint distribution of γ_m and $\alpha_{m,j}$. Since this distribution is assumed to have M mass points, the type-specific likelihood contribution must be summed over the M possible types, weighted by the probability of being each type.

$$L_i = \sum_{m=1}^{M} p_m L_{im}$$

where p_m is the probability of being "type" m, which is a parameter to be estimated.

With no heterogeneity, there are 16 parameters to estimate: five in the utility function ($\alpha_{0,2}$, $\alpha_{0,4}$, α_A, α_D, τ) and eleven in the grade equations (γ_0, γ_G, γ_T, γ_P, σ_{G1}, γ_{X2}, σ_{G2}, γ_{X3}, σ_{G3}, γ_{X4}, σ_{G4}). Unobserved heterogeneity adds four parameters ($\alpha_{m,2}, \alpha_{m,4}, \gamma_m, p_m$) for each additional type.

The parameters in the utility function ($\alpha_{0,2}, \alpha_{0,4}, \alpha_A, \alpha_D$) are identified from the educational choices up to the scale parameter τ. For example, the difference in enrollment rates between individuals with high expected grades and low expected grades but all else equal identifies the ratio α_A/τ. Since utility is in dollar units, τ is identified from variation in $Tuition_{i,t}$ and $Income_{i,t}$ across individuals and across periods. Holding all other variables constant, the estimate of τ is the magnitude of preference shocks that is needed to rationalize the proportions of people dropping out in each year, given the financial costs and benefits from doing so and the parametric distribution assumed on the shocks. For instance, if the financial return to completing a fourth year of college is much higher than completing the third year, then more people should drop out before the third year than the fourth. The magnitude of this enrollment difference identifies τ - if the dropout rates are similar then the variance of preference shocks must be high (τ must be large) to rationalize the data. Cross-state tuition differences contribute to the identification of τ in the same way. It should be noted that the estimate of τ will be affected by any bias in the estimate of the return to each year of schooling. If the least squares estimated return to each year of school is biased upwards by unobserved factors, then the estimate of τ will also be overstated. However, most IV and twins estimates suggest that ability bias in OLS estimates is not too severe.[21]

The parameters of the grade function are identified primarily from the grade outcomes in the typical manner, though the educational choices also help identify these parameters.

Parameters associated with unobserved heterogeneity are identified by common behavior which is contrary to the model. For instance, there may be individuals with poor academic performance but who still persist to graduation due to unmodeled parental pressure. If there are a sufficient number of similar individuals, then a model that permits for this type of behavior will fit the data better (i.e., have a higher likelihood). In practice, it is difficult to identify the discount factor β separately from τ. In the current specification, I fix β at 0.95.[22]

[21]See Card (1999) for a review.

[22]I have estimated the model with $\beta = 0.90$ and the results are qualitatively similar. The estimate of the dollar value of the option value decreases by one third reflecting a decrease in the estimated scale parameter τ, but the importance of option value relative to the value of enrollment and welfare is unchanged.

IV Estimation Results

A Parameter estimates

Table I provides estimates of the structural parameters. Columns (1) and (2) provide estimates from a base model with no learning about academic aptitude while columns (3) and (4) provide estimates from the full model described above. Both models are estimated with and without allowing for up to three points of unobserved heterogeneity. Standard errors were computed by taking the inverse of the numerical Hessian at the estimated parameter values.

In the model without learning, expectations about grade realizations are based exclusively on baseline characteristics and type, so $E_t[A_i] = E[A_i|X_i, Type]$ for all t. The parameter estimates all have the expected signs and are statistically significant. Since utility is in units of dollars, these estimates are immediately interpretable as the dollar value (in $100,000) associated with a one-unit change in the independent variable. With no unobserved heterogeneity or learning (column (1)), the estimates imply that four-year colleges have amenities valued at $32,300 over two-year colleges. Expecting to do well in school is also valuable. Each additional grade point (e.g. going from a C-student to a B-student) is equivalent to $70,700. Living 100 miles from a college is equivalent to paying an additional $12,100 in tuition. A key parameter is τ, which parameterizes the variance of the preference shocks. At the estimated parameters, the preference shocks have a standard deviation of $65,500 $\left(= \tau \frac{\Pi}{\sqrt{6}}\right)$. As expected, the grade parameter estimates show a strong positive correlation between academic performance and baseline characteristics such as academic performance in high school, AFQT test scores, and parent's education.

The estimate of α_A in column (1) could be biased if people with high academic ability also have a stronger preference for attending school, independent of the causal effect of aptitude on schooling ease. Column (2) addresses this concern by allowing for several different "unobserved types," each with an arbitrary correlation between schooling preference and academic aptitude. Permitting unobserved heterogeneity improves model fit considerably. Relative to type 1 individuals, type 2 individuals (17% of sample) are higher ability ($\gamma_{type2} > 0$), but have a stronger dislike of 4-year colleges ($\alpha^{S4}_{type2} < 0$) and are neutral to two-year colleges. These individuals can be thought of as good students from disadvantaged families. By contrast, type 3 individuals (36%) are lower ability ($\gamma_{type3} < 0$), have a stronger preference for 4-year colleges ($\alpha^{S4}_{type3} > 0$) and dislike two-year colleges ($\alpha^{S2}_{type2} < 0$), though this latter effect is not statistically significant. Incorporating unobserved heterogeneity does not qualitatively change the other parameter estimates. However, the estimated deviation of the preference shocks increases to $100,000. Consequently, the magnitude of the other parameter estimates also increases. Interestingly, the relationship between expected academic ability and enrollment probabilities (α_A/τ) changed little, increasing from 1.4 to 1.6 when unobserved heterogeneity is permitted. The estimated variance of the grade shocks decreases because a greater share of the performance variance is captured by baseline characteristics (including type).

Columns (3) and (4) present estimates from the full learning model presented in Section 3. The parameter estimates are very similar to estimates from the no-learning model, both qualitatively and quantitatively. With learning, individuals estimate future academic performance by calculating a weighted average of performance predicted with baseline characteristics (including type) and cumulative grade point average, where the weights (γ_{x2}, γ_{x3}, and γ_{x4}) are parameters to be estimated. The normal learning model predicts that the weight placed on baseline characteristics should decrease with t (γ_{x1} is normalized to one), as should the residual grade variance (σ_{gt}). The estimates in column (3), which do not control for unobserved heterogeneity, support this implication of the normal learning model. The best predictor of year-two grades weighs baseline characteristics and first-year grades approximately equally (48% vs. 52%). Fourth-year grades, however, are best predicted by placing only 19% of the weight on baseline characteristics and 81% on three-year cumulative grade point average.

Due to unobserved heterogeneity, however, these estimates can overstate the amount of learning taking place. $E[A_i|X_i]$ may not fully capture all information about future academic performance available to individuals, so the increasing weight placed on cumulative academic performance may simply capture the revelation of private information to the econometrician. Column (4) addresses this concern (and the potential bias of α_A/τ discussed earlier) by allowing for several different unobserved types, each with different levels of academic aptitude, known ex-ante, and preferences for two- and four-year school. The estimates in column (4), which allow for three different types, imply that learning about academic ability continues to occur through the end of college. Controlling for unobserved heterogeneity does not change the learning parameters much.[23]

[23]These results assume that I have specified the information set used by individuals correctly. If students possess information about future grades beyond that modeled here, these estimates overstate the extent of uncertainty and learning and understate the extent of hetergeneity. The methods presented in Cuhna, Heckman, and Navarro (2005)

The types identified in the learning model are slightly different than those revealed in the non-learning model. Relative to type 1, type 2 individuals (8% of the sample) have higher academic aptitude, greater-than-average preference for two-year colleges, and less preference for four-year colleges. Type 3 individuals (63%) reflect students with poor academic aptitude who have lower than expected preference for two- and four-year schools. Accounting for unobserved heterogeneity again increases τ and the scale of most other parameters, though the relationship between expected academic ability and enrollment probabilities (α_A/τ) changes little. The estimated deviation of the preference shocks is about $82,000 in this preferred specification. The overall model fit also improves when unobserved heterogeneity is permitted. I now discuss model fit more directly.

B Model fit

To examine model fit, I simulate the grade outcomes and educational choices of individuals in my estimation sample 100 times and compare the predicted outcomes to the actual observed outcomes. In this section I discuss simulations that use the preferred estimated parameter values, from model (4) from Table I. In the Appendix, I also examine model fit for the models that do not incorporate unobserved heterogeneity and learning simultaneously (models (1) to (3) in Table I). In general, the preferred specification provides a much better fit of the data than the simpler models. I examine model fit in two ways. First, I compare actual to predicted enrollment outcomes, including initial pathway choice, dropout, and college completion. This comparison is also done by demographic characteristics which are not explicitly incorporated in the model. I then examine the relationship between grade outcomes and subsequent enrollment decisions. It should be noted that if the model contained utility intercepts that differ over time, by school, and by academic performance, then the moments presented below would not constitute a true test of "fit." Such a fully saturated and calibrated model would fit the data perfectly. The model I employ is much more parsimonious, as I discuss below.

Figure V compares the predicted enrollment decisions to the actual decisions made by individuals in the estimation sample. Overall, the model predictions fit the distribution of actual enrollment decisions reasonably well considering how unsaturated the model is. Forty-five percent of individuals are predicted not to enroll, two percentage points below the actual share. Consequently,

could be used to distinguish between these two sources of variability.

enrollment in four-year colleges is over-predicted by three percent. The fraction of individuals enrolling in two-year colleges is identical between actual and predicted. The goodness of initial enrollment decision fit is not surprising since the model includes separate constants for two- and four-year schools in the utility function ($\alpha_{0,j}$). If the parameters were estimated using only the initial enrollment decision, these shares would fit exactly.

The fit of dropout behavior following initial enrollment decision is a better test of the ability of the model to predict behavior. Since the utility intercepts do not vary over time, predicted differential dropout between different periods is driven entirely by between-period differences in the financial returns (lifetime earnings gain minus costs) and changes in expected academic performance ($E_t[A_i]$). Figure VI depicts the fraction of two- and four-year enrollments who drop out in each year or graduate. There are two primary discrepancies between the model predictions and actual outcomes. First, the model slightly underpredicts the fraction of people beginning at community college that drop out after one or three years and consequently over-predicts completion. The second discrepancy is that the model over-predicts dropout after the first year among people that start at a four-year college and consequently underpredicts four-year college graduation.

Figure VII compares actual and predicted enrollment shares by whether students come from a high- or low-income family. Family income does not enter the model at all, so this is a pretty strong test of model fit. Any correlation between family income and enrollment outcomes must operate through the correlation between family income and the modeled background characteristics (high school performance, AFQT, and parental education). Additionally, these characteristics do not enter individuals' preferences for school directly. Higher parental education, for instance, increases academic aptitude, which in turn makes schooling more desirable. Higher parental education also increases predicted lifetime income, which reduces individuals' sensitivity to schooling costs. Nonetheless, the model still captures several important features of the data, namely the strong positive correlation between family income, college enrollment, and degree completion.

Enrollment decisions and grades are related for several reasons. First, students with adverse baseline characteristics (e.g., poor grades in high school) have low expected college aptitude, which increases the disutility of school. Consequently, students with low expected academic performance will be less likely to enroll and more likely to drop out if they do enroll. Second, if students learn about the desirability of college through their grades, then students who persist to graduation will have consistently received high grades while those who dropped out will have received low grades. Figure VIII displays the actual and simulated fraction of students that complete their fourth year by their first-year grade point average. The overall slope and curvature of the grade-graduation relationship is matched very closely. Like the actual data, predicted completion is increasing most quickly in the middle grade span, where grade signals are expected to be most influential.

C Discussion of estimates and fit

To summarize, the parameter estimates suggest that uncertainty is an important feature of postsecondary schooling outcomes. The preferred estimates (column (4) from Table I) indicate that the deviation of unanticipated shocks to the relative preference for enrollment and labor market entry is equivalent to $82,300 in lifetime earnings. These shocks have the same order of magnitude as the incremental gain from completing a college degree. Thus, unanticipated preference shocks are an important determinant of educational outcomes. It should be noted that the model assumes that individuals face no credit constraints. My specification does not permit me to distinguish between large shocks and small shocks whose effects are magnified by credit constraints. My estimates reflect the combination of these two factors.[24] The estimates also suggest that students learn about their ex-ante unknown academic aptitude through college grades. Lastly, the estimates suggest that academic aptitude does predict enrollment outcomes and that much of the relationship between family background and schooling outcomes can be captured through the effect of background on academic performance.

Predictions from simulations using the estimated model parameters do match many features of the actual data on enrollments and grade outcomes. The overall distributions of predicted and actual outcomes is roughly similar and the model captures several main features of the relationship between grade outcomes and enrollment decisions. Importantly, the model also replicates educational differences by background characteristics, despite the strong restriction that they operate entirely through expected academic performance.

[24]Incorporating credit constraints would require a different structural model. Cameron and Heckman (2001) use such a model and conclude that long-run factors associated with family background, not short-term credit constraints, explain much of the observed racial disparity in college education.

V The Importance of Option Value

In this section, I estimate the option value created by the ability of students to make educational decisions sequentially and in response to new information. To do this, I treat the estimated structural model as the actual data generating process and simulate educational choices and welfare under alternative assumptions about individuals' information set.[25] In the static model, I simulate outcomes when individuals are restricted to commit to educational choices before enrolling in college. They base their decision only on information available before college enrollment: baseline characteristics (high school GPA, AFQT, parent education, and type), predicted lifetime earnings, direct tuition and commuting costs, and first-period shocks ($\varepsilon^s_{i,2,1}, \varepsilon^s_{i,4,1}, \varepsilon^w_{i,1}$). As a basis of comparison, I also simulate the choices and welfare in the first-best scenario, where individuals make decisions with perfect knowledge of all future shocks.

A Educational outcomes

Figure IX summarizes the importance of option value to educational decisions. The top panel plots the average number of years of college by expected academic ability, separately for the first-best full information (solid), baseline dynamic (dashed), and static (dotted) models. The static model predicts that education would be much more bifurcated if students were forced to commit ex-ante with limited information. People with low expected performance would get very little education while high ability students would get much more. Compared to the first-best outcome, this bifurcation reduces welfare because some ex-ante low-ability students should go to or graduate from college, while some higher ability students should not. Sequential decision-making permits individuals to come closer to the first-best outcome.

This can be seen more clearly in the middle and bottom panels, which plot the simulated enrollment and graduation rates by expected ability. These figures are the empirical analog to the right panel of Figure III, where $E[A_i|X_i, Type]$ is analogous to $\varepsilon_{i,1}$. Option value increases the enrollment rates of all individuals, particularly those in the middle who are on the enrollment

[25]To implement the simulations, I first replicate each observation 100 times. For each of these simulated observations, I then draw preference and grade shocks from the appropriately scaled EV(1) and normal distributions and assign an unobserved "type" based on the estimated probabilities. The optimal choices for each individual are then computed by utility comparisons, incorporating these shocks.

margin in the static model. Many of these individuals would choose to enroll if they knew their shocks with certainty but would not if they were forced to commit ex-ante. For low- to moderate-ability students, option value only slightly increases college completion. The biggest effect of option value on completion is to reduce it for high ability students. Some high-ability students expect to graduate - so would commit to doing so ex-ante - but then learn that completion is undesirable and would prefer to drop out. Allowing them to do so reduces completion rates but improves their welfare.

B Quantifying option value

Figure X quantifies the option value of college enrollment. The figure plots the average value of the opportunity to enroll in college by expected academic ability for the same three scenarios and is the empirical analog of the left panel of Figure III. This value is zero for those who do not enroll. The value of the opportunity to enroll is increasing in expected ability both because enrollment increases with ability and because school is less costly for high ability people, so value conditional on enrollment is also increasing. The vertical distance between the solid and dotted lines represents individuals' total welfare loss from being forced to commit to an educational outcome ex-ante, compared to the first-best situation with full information. This loss is greatest for moderate-ability individuals. Since sequential decision making helps more individuals obtain their optimal level of education, it partially closes this welfare gap, as indicated by the dashed line. The difference between the dashed and dotted lines thus represents the value of the option to drop out whenever continuation turns out to be undesirable.

Table II summarizes the option value by expected ability category. On average, students would be willing to pay \$14,900 (in 1992 dollars) to maintain the ability to make enrollment decisions sequentially in response to information. Given the precision of the parameter estimates, total option value is fairly precisely estimated with a 90% confidence interval of \$11,400 to \$18,100.[26] Consistent with the simple theoretical model, option value varies considerably with ability. Moderate-ability students, for whom educational outcomes are most uncertain, are willing to pay up to

[26] Since the option value is a highly nonlinear and complicated function of the parameters, I rely on simulations to compute the confidence intervals. Confidence intervals were computed by performing the option value simulation for 500 different draws of the parameter vector from its estimated distribution.

$25,000, while the lowest ability students derive virtually no value from the option. The option is also worth less to higher ability students because their enrollment decisions do not depend on it.[27]

Table II also normalizes the option value in two ways. My estimates imply that option value accounts for 14% of the total value of the opportunity to enroll in college. For low to moderate ability students, this fraction is even higher. Option value also represents approximately one quarter of the welfare loss associated with moving from the full information to static scenarios.

Additional simulations are used to allocate the total option value into the years in which new information is learned. The first three years of college each provide new information about academic ability (in the form of grade signals) and the relative desirability of schooling and work ($\varepsilon^s_{i,2,t}, \varepsilon^s_{i,4,t}, \varepsilon^w_{i,t}$). To do this decomposition, I simulate educational choices and welfare when individuals are restricted to commit to educational choices before enrolling in college (the static model discussed above), after the first year, after the second year, and after the third year (the baseline dynamic model). For moderate-ability students, the most valuable information is that which is learned in the first year of college, when the wisdom of their enrollment decision is most uncertain. Higher ability students derive relatively more value from information received later, when graduation decisions are made. Approximately 60% of the total option value derives from information learned in the first year, while the other two years account for about 20% each.

To summarize, the value of the option to drop out is considerable, particularly for moderate ability students who have the most uncertainty about their net benefit from schooling. The option to drop out has value both because it encourages more people to enroll, who may not want to if forced to commit ex-ante, and it because it permits dropout if graduation is undesirable among those who would commit to graduate ex-ante. In aggregate, the former is greater than the latter. Furthermore, the majority of the aggregate option value comes from the information received in the first year of college.

[27]These estimates are not directly comparable to those presented in Table 7 of Heckman, Lochner, and Todd (2006) because their model is one of exogenous dropout and their estimates include continuation value. That said, their estimate of the option value of college attendance is of a similar order of magnitude as that reported here.

C Option value in a more general setting

There are several ways in which the model could be generalized. The current specification assumes that (1) labor market entry is costless but irreversible, or (2) individuals do not learn about the relative desirability of schooling and work while in the labor market, and (3) labor market draws persist following labor market entry. This is a special case of a more general model (discussed in the Appendix) in which attending school and working both provide information and where decisions are not completely irreversible. Relaxing these assumptions would affect my estimate of the option value. In the extreme case, with no switching costs and completely symmetric learning (i.e. people learn as much about their tastes while working as they do attending school), enrollment and labor market entry would provide equal option value, so the net total informational value from enrollment itself would be zero. Though the information learned in school is valuable, this value is offset by the cost of lost information that could be gained by working. Welfare overall would be much greater in this scenario but the net return to enrolling rather than working would be lower than if re-enrollment were not permitted.

Another asymmetry in the current specification is that individuals receive new labor market draws only if enrolled in school, so enrollment lets people delay labor market entry until receiving a favorable draw. Relaxing this restriction so that people receive new labor market draws while not in school will also reduce the estimated option value. The appropriateness of this assumption can be examined using annual data on labor market outcomes, which the current dataset does not contain.

Though the maintained assumptions seem plausible, my estimates should be considered an upper bound of the net option value associated with enrollment. Extensions that permit dynamic considerations after initial labor market entry, such as re-enrollment and repeated labor market draws, would make labor market entry more desirable and diminish the relative benefit of college enrollment.

This paper has chosen to focus on the flexibility afforded to students' binary enrollment decisions, but there are many other schooling choice dimensions over which students can re-optimize after enrollment. For instance, college students can change majors, transfer schools (adjusting college quality), or adjust course sequencing in response to new information. The value of the decision flexibility in these dimensions is included in the value of the enrollment opportunity both in the dynamic and static settings, so is netted out from the option value estimate presented here. Quantifying the option value created by major, course, and school choice flexibility is an important task for future research, as these attributes are directly controlled by schools and policy-makers.

D Policy consequences of educational uncertainty

Many education policies have a temporal dimension, making option value considerations potentially important. For instance, giving students a bonus for graduating directly alters the financial gain to the final year of college but not the first three. Both community colleges and the Federal Hope tax credit explicitly alter the tuition gradient by making the first few years of college cheaper than the last half. If students are forward-looking, the time path of incentives will enter enrollment and continuation decisions in a different way than if decisions were static. While a complete assessment of specific policies is beyond the scope of this paper, I briefly sketch how static and dynamic models of schooling may result in different policy effects.[28]

A static model will generally under-predict the effect of community colleges on enrollment. In addition to directly making college less expensive, community colleges increase the option value of enrollment because dropout is less costly so more people experiment with school. The static model does not fully incorporate this added informational benefit. In contrast, a static model will over-state the effects of across-the-board tuition subsidies on college completion. In the presence of large degree effects, a static model predicts a bimodal distribution of education outcomes, with many non-enrollees and many graduates, but few dropouts. Consequently, more enrollees are predicted to continue through to graduation in response to an across-the-board subsidy than would be the case in a dynamic setting. A static model which ignores endogenous dropout will also make similar predictions for front- and back-loaded tuition subsidies, but these policies can have quite different effects if choices are dynamic. A static model also over-predicts the graduation consequences of increasing academic preparedness in high school. Expected performance in college - which depends heavily on high school GPA - has much more influence on educational outcomes in the static model. If decision-making is dynamic, however, less weight is placed on

[28] A companion paper is using the estimated model to examine the option value aspects of several policy interventions specifically.

baseline anticipated performance as new information is acquired during each year. The bottom line is that uncertainty is an important feature of educational decisions and failing to account for it may provide misleading estimates of policy effects. This is particularly true when comparing policies that have different temporal characteristics, such as community colleges (which alter the tuition gradient) or across-the-board tuition reductions (which do not).

VI Summary and Conclusions

This paper examines the empirical importance of uncertainty and option value to college enrollment. It is the first to quantify the magnitude of the option value that arises when individuals make decisions to invest in a college education sequentially and when the desirability of doing so is uncertain. Estimates suggest that this value is substantial. In contrast to a scenario where individuals must commit to an educational outcome ex-ante, the current flexible system increases welfare by $14,900 on average. This represents 14% of the overall value of the opportunity to enroll in college. Moderate-ability students, who have the most uncertainty about the desirability of schooling, derive even more value from this flexibility. The traditional human capital model ignores this value.

The finding that enrollment choice flexibility substantially improves welfare has direct implications for the potential costs of student "tracking." This paper suggests that, at least in the U.S. postsecondary context, students learn quite a bit about their ability and preferences in the first few years of college. Forcing students to commit ex-ante makes educational outcomes more polarized by background and reduces welfare, particularly for students at the margin. This welfare loss must be weighed against any efficiency gains resulting from greater specialization through earlier tracking, such as that identified by Malamud (2008). The temporal dimension of many other education policies - for instance, whether to subsidize tuition at the beginning or end of college - have received very little attention despite their importance if schooling decisions are dynamic.

The general framework developed herein could also be used in a number of different contexts in which decisions are partially irreversible and made in the presence of uncertainty.[29] One po-

[29]Retirement decisions are one topic in labor economics to which this framework has been applied. See Stock and Wise (1990) and Coile and Gruber (2007) for an application of option value to the study of retirement decisions.

tential application is the use of "take-it-or-leave-it" job offers. Firms hiring many law or business school graduates force students to commit to a job early in the fall, possibly before their industry or locational preferences are finalized. The model implies that firms would have to compensate individuals for this loss of flexibility, through a signing bonus or higher salary. Marriage and fertility decisions are also partially irreversible and made in the presence of uncertainty. The ability to wait and acquire more information before committing to a decision thus creates option value. The effects of policies that alter the ability to reverse a decision (e.g. divorce costs) operate through this channel. Investments in health can also be understood as motivated by option value considerations. Since many health conditions (e.g. diabetes onset, lung cancer) are partially irreversible, forward-looking individuals should make costly health investments when young in order to preserve the option of being healthy when old. Subsidies for preventative care, a healthy diet, and exercise among the young can be rationalized by this option value if individuals are not completely forward-looking.

References

[1] Altonji, Joel. 1993. "The Demand for and Return to Education When Education Outcomes Are Uncertain." *Journal of Labor Economics*, 11: 48-83.

[2] Arcidiacono, Peter. 2004. "Ability Sorting and the Returns to College Major." *Journal of Econometrics,* 121: 343-375.

[3] Becker, Gary, 1964. *Human Capital: A Theoretical and Empirical Analysis with Special Reference to Education*. New York: Columbia University Press.

[4] Cameron, Stephen and James J. Heckman. 2001. “The Dynamics of Educational Attainment for Black, Hispanic, and White Males.” *Journal of Political Economy*, 109 (3): 455-499.

[5] Card, David, 1999. “The Causal Effect of Education on Earnings,” in *Handbook of Labor Economics*, Volume 3A, ed. by O. Ashenfelter and D. Card. New York: Elsevier, 1801-1864.

[6] Chen, Stacey, 2008. “Estimating the Variance of Wages in the Presence of Selection and Unobservable Heterogeneity.” *Review of Economics and Statistics*, 90(2).

[7] Coile, Courtney and Jonathan Gruber. 2007. "Future Social Security Entitlements and the Retirement Decision." *Review of Economics and Statistics*, 89(2):234-246.

[8] Comay, Y., A. Melnik, M.A. Pollatschek. 1973. “The Option Value of Education and the Optimal Path for Investment in Human Capital.” *International Economic Review*, 14 (2): 421-35.

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

[10] Dixit, A.K., R.S. Pindyck. 1994. *Investment under Uncertainty*. Princeton, N.J.: Princeton University Press.

[11] Domencich, Thomas A. and Daniel McFadden,.1975. *Urban Travel Demand: A Behavioral Analysis*, Amsterdam: North-Holland Publishing Company.

[12] Dothan, Uri and Joseph Williams. 1981. "Education as an Option," *Journal of Business*, 51(1).

[13] Eckstein, Z. and K. Wolpin. 1999. “Why Youths Drop out of High School: The Impact of Preferences, Opportunities, and Abilities.” *Econometrica*, 67(6): 1295-1339.

[14] Eide, Eric and Geetha Waehrer. 1998. "The Role of the Option Value of College Attendance in College Major Choice." *Economics of Education Review*, 17 (1): 72-83.

[15] Heckman, J.J., L. Lochner, and P. Todd. 2006. "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.

[16] Heckman, J.J. and S. Navarro. 2007. "Dynamic Discrete Choice and Dynamic Treatment Effects." *Journal of Econometrics*, 136: 341-396.

[17] Heckman, J.J. and B. Singer. 1984. "A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data," *Econometrica*, 52 (2): 271-320.

[18] Heckman, J.J. and S. Urzua. 2008. "The Option Value of Education and the Rate of Return to Educational Choices," unpublished.

[19] Hungerford, T. and Gary Solon. 1987. “Sheepskin Effects in the Returns to Education,” *Review of Economics and Statistics*, 69: 75-177.

[20] Jaeger, D. and M. Page. 1996. “Degrees Matter: New Evidence on Sheepskin Effects in Returns to Education," *Review of Economics and Statistics*,78.

[21] Kane, Thomas J. and Cecilia Elena Rouse. 1995. “Labor Market Returns to Two- and Four-Year College,” *The American Economic Review*, 85 (3): 600-614.

[22] Keane, Michael P. and Kenneth I. Wolpin. 1997. “The Career Decisions of Young Men,” *Journal of Political Economy*, 105 (3): 473-522.

[23] Kilburn, M. Rebecca, Lawrence M. Hanser, and Jacob A. Klerman. 1998. "Estimating AFQT Scores for NELS Respondents," RAND, MR-818-OSD/A, 1998.

[24] Levhari, David, and Yoram Weiss. 1974. "The Effect of Risk on the Investment in Human Capital," *American Economic Review,* 64 (Dec): 950-963.

[25] Malamud, Ofer, 2008. "Breadth vs. Depth: The Timing of Specialization in Higher Education." *Harris School Working Paper Series 08.08.*

[26] Manski, Charles. 1989. “Schooling as Experimentation: A Re-Appraisal of the Postsecondary Dropout Phenomenon." *Economics of Education Review,* 8 (4): 305-312.

[27] Manski, Charles, 1991. “Nonparametric Estimation of Expectations in the Analysis of Discrete Choice Under Uncertainty” in W. Barnett, J. Powell, and G. Tauchen (editors), *Nonparametric and Semiparametric Methods in Econometrics and Statistics*, New York: Cambridge University Press: 259-275.

[28] Navarro, Salvador, 2007. "Using Observed Choices to Infer Agent’s Information: Reconsidering the Importance of Borrowing Constraints, Uncertainty and Preferences in College Attendance.” Unpublished.

[29] Olson, Lawrence; Halbert White, H.M. and Shefrin.1979. “Optimal Investment in Schooling When Income Are Risky.” *Journal of Political Economy*, 87(3): 522-39.

[30] Oreopoulos, Philip, Till von Wachter and Andrew Heisz. 2006. "The Short- and Long-Term Career Effects of Graduating in a Recession: Hysteresis and Heterogeneity in the Market for College Graduates." *NBER Working Paper No. 12159*, April 2006.

[31] Park, Jin Heum. 1999. "Estimation of Sheepskin Effects Using the Old and the New Measures of Educational Attainment in the Current Population Survey." *Economics Letters*, 62: 237-240.

[32] Pindyck, Robert S..1993. "Investments of Uncertain Cost." *Journal of Financial Economics*, 34:53-76.

[33] Rust, John. 1987. "Optimal Replacement of GMC Bus Engine: An Empirical Model of Harold Zurcher." *Econometrica*, 55(5): 999-1033.

[34] Stinebrickner, Todd and Raph Stinebrickner. 2008. "Learning about Academic Ability and the College Drop-out Decision." Unpublished.

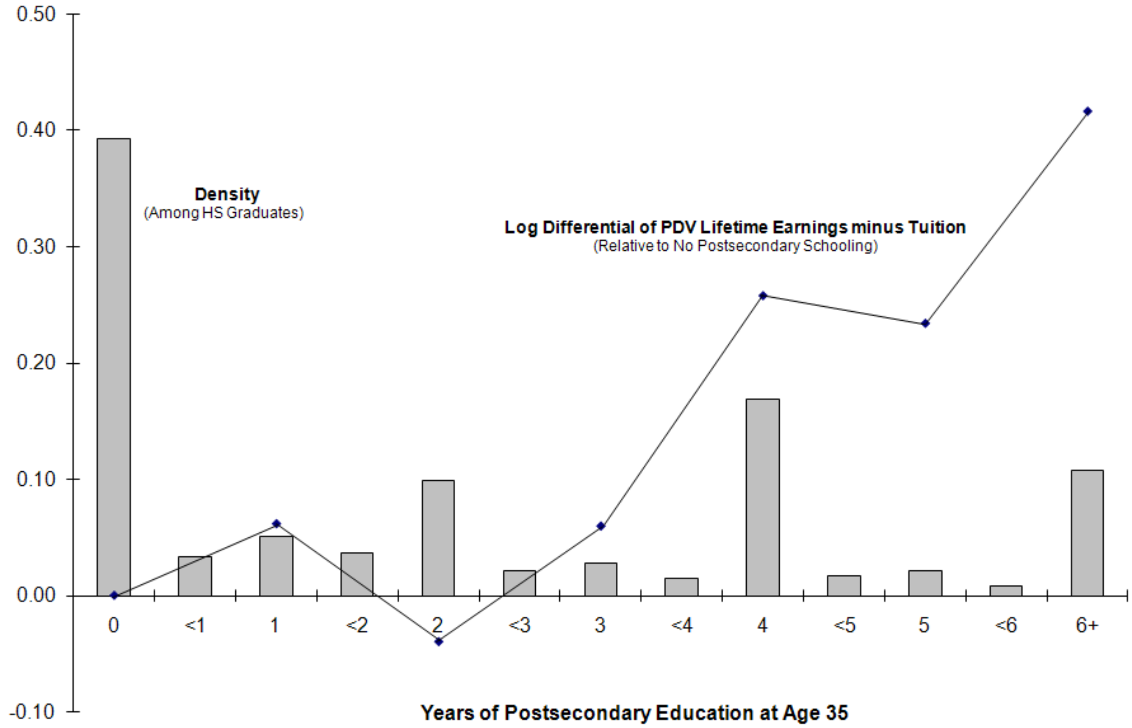
[35] Stock, James H. and David A. Wise. 1990. "Pensions, the Option Value of Work, and Retirement." *Econometrica,* 58(5): 1151-1180.

[36] Taber, Christopher. 2000. “Semiparametric Identification and Heterogeneity in Dynamic Programming Discrete Choice Models,” *Journal of Econometrics*, 117.

[37] Weisbrod, B.. 1962. “Education and Investment in Human Capital.” *The Journal of Political Economy*, 70(5).

[38] Williams. Joseph T. 1979. "Uncertainty, and the Accumulation of Human Capital over the Life Cycle." *Journal of Busrness,* 52(4): 521-548.

Figure I: Returns to and Distribution of Postsecondary Education, Men



Notes: Density is from IPUMS-CPS years 1985 to 1990 restricted to 35 year old male high school graduates. The solid line plots the coefficients from a linear regression of log lifetime earnings (minus average tuition) on a set of schooling level dummies and control variables using data from the NLSY. PDV of lifetime earnings are computed from age 18 to 62 assuming real income is constant from age 38 to 62 and a discount rate of 5%. Linear controls include dummies for ethnicity, four regions, urban, parents' education, high school GPA, AFQT, and the pairwise interactions between these last three variables. These OLS estimates only partially address the endogeneity and selection problems which complicate earnings comparisons by schooling level.

Figure II: Simple Dynamic Model of College Enrollment and Completion

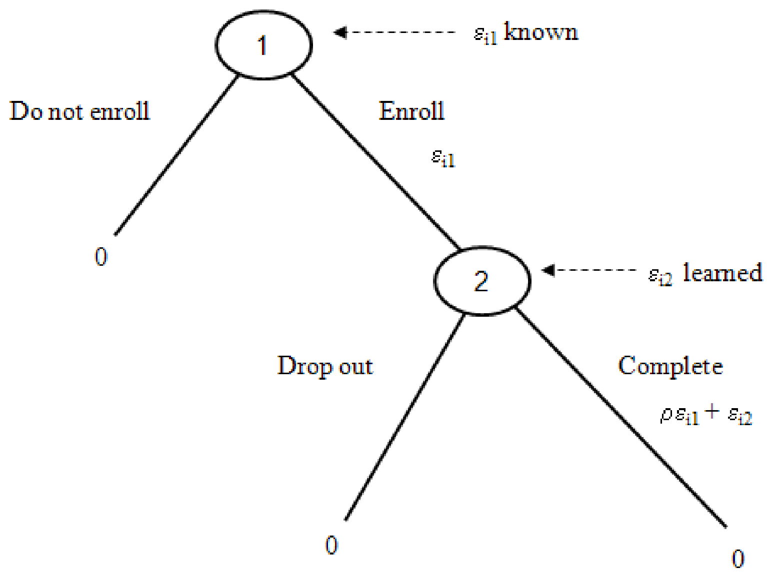
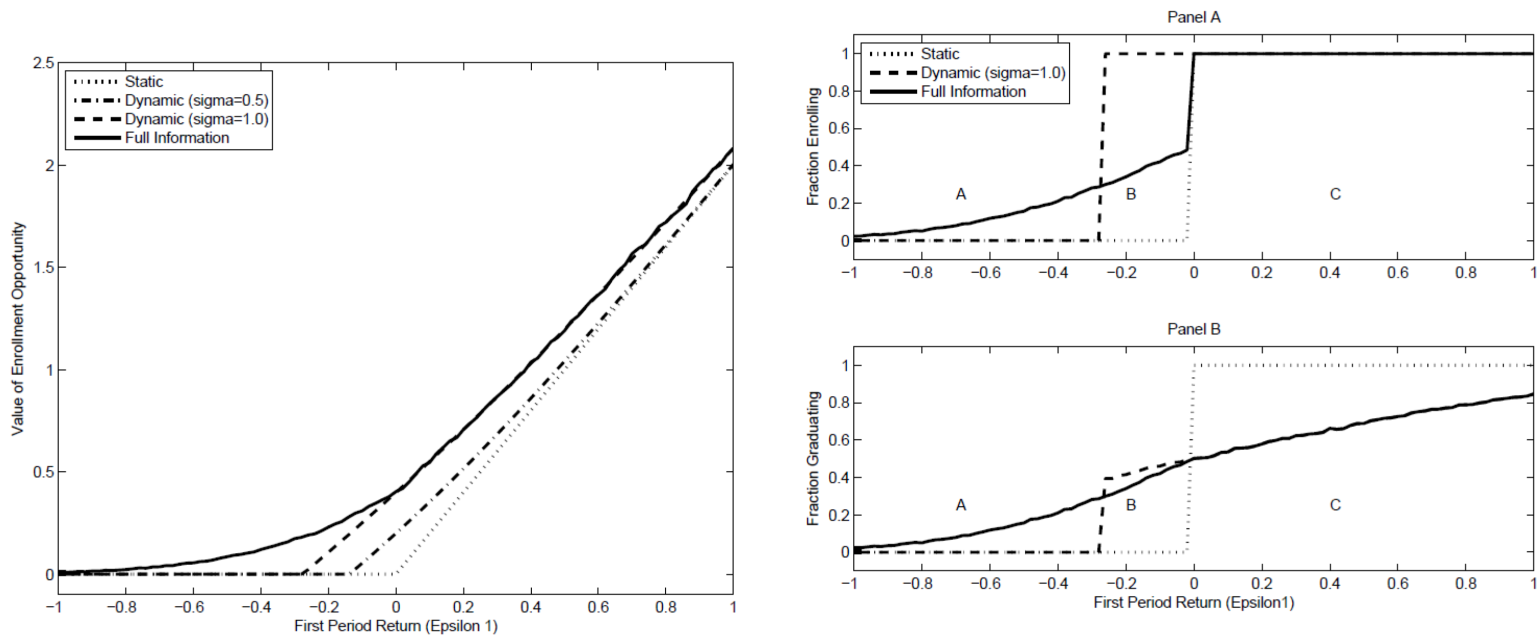


Figure III: Welfare and Educational Consequences of Uncertainty



Notes: Figures plot the average welfare and schooling outcomes across 10,000 random draws of $\varepsilon_{i,2}$ from a normal distribution with mean zero and variance σ^2 for each value of $\varepsilon_{i,1}$, with $\rho = 1$. Agents are assumed to follow the choice model described in Section IIB. Static scenario assumes agents make enrollment and graduation decisions based only on $\varepsilon_{i,1}$. Dynamic scenario assumes they make enrollment decisions based on $\varepsilon_{i,1}$ but learn $\varepsilon_{i,2}$ before making graduation decision. Full information scenario assumes agents know both $\varepsilon_{i,1}$ and $\varepsilon_{i,2}$ when making enrollment and graduation decisions.

Figure IV: Full Empirical Dynamic Model

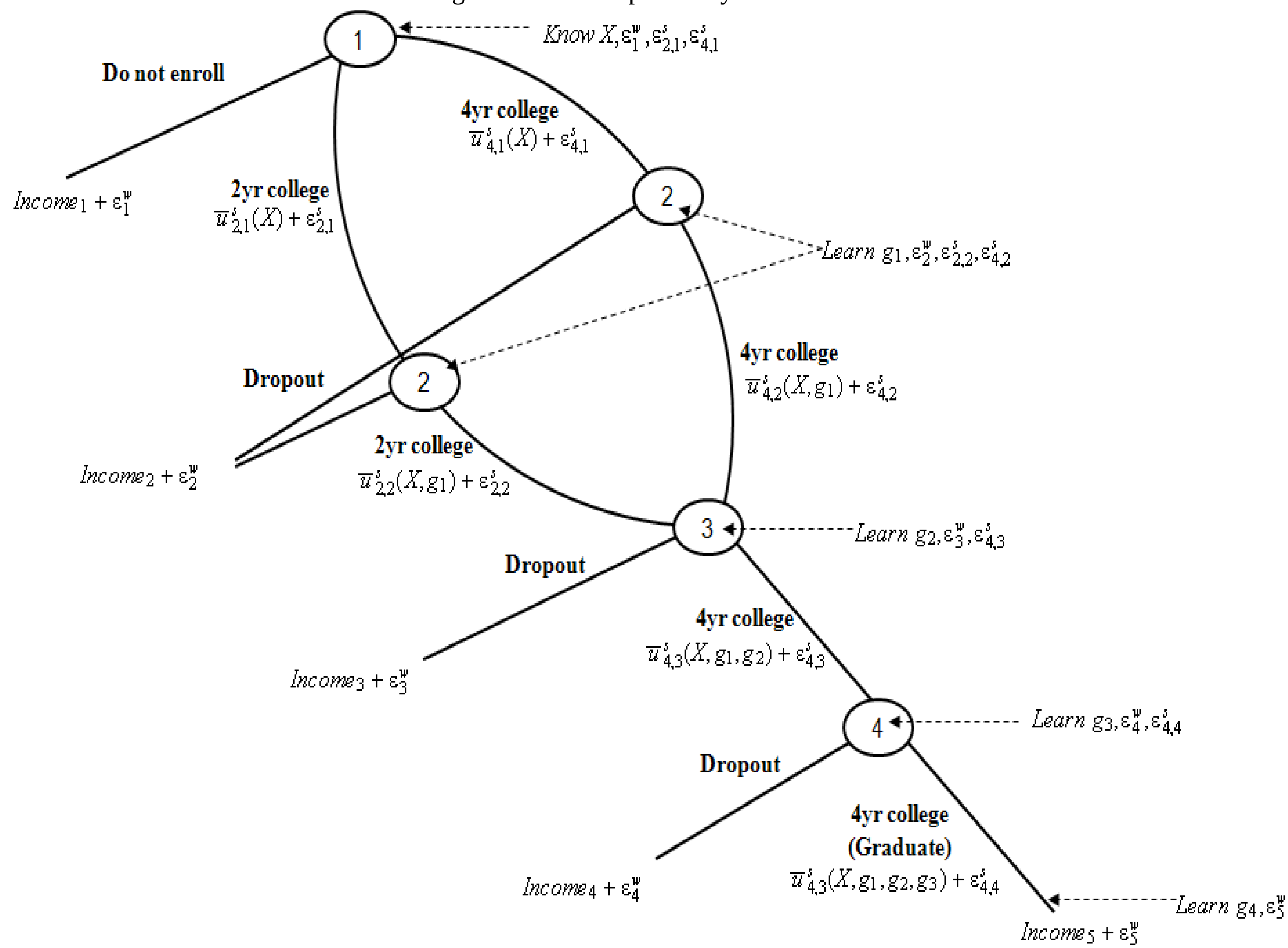
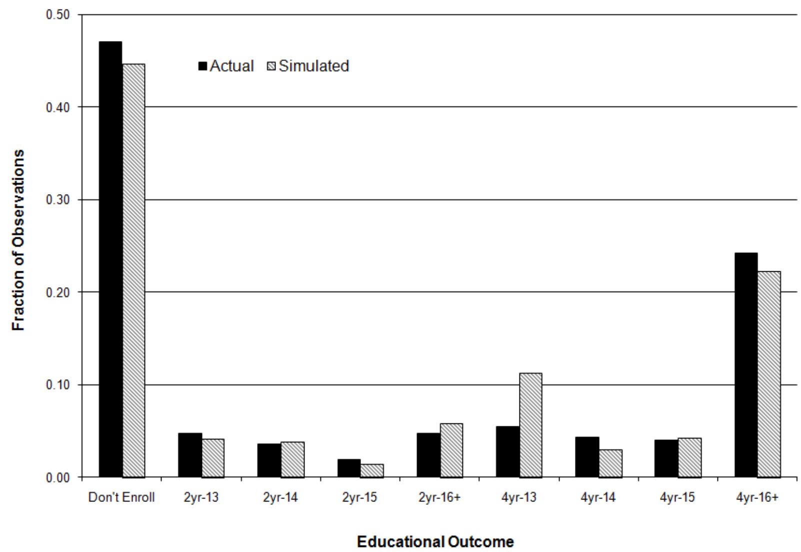


Table I: Estimates of Structural Parameters

	No Learning		Learning	
	One type	Three types	One type	Three types
	(1)	(2)	(3)	(4)
Utility parameters				
constant (2yr)	-2.911	-4.346	-2.569	-3.187
	(0.150)	(0.442)	(0.121)	(0.378)
constant (4yr)	-2.588	-3.765	-2.220	-2.845
	(0.137)	(0.391)	(0.105)	(0.332)
E[Ai]	0.707	1.242	0.591	1.009
	(0.049)	(0.161)	(0.039)	(0.154)
distance (100)	0.121	0.277	0.139	0.220
	(0.034)	(0.074)	(0.034)	(0.065)
tau	0.511	0.780	0.513	0.642
	(0.022)	(0.074)	(0.023)	(0.070)
Grade parameters				
constant (gpa)	1.192	0.835	0.802	0.659
	(0.056)	(0.102)	(0.072)	(0.087)
HS GPA	0.383	0.394	0.436	0.523
	(0.019)	(0.026)	(0.025)	(0.029)
AFQT	0.411	0.702	0.581	0.695
	(0.039)	(0.082)	(0.057)	(0.072)
ParBA	0.206	0.297	0.281	0.336
	(0.017)	(0.033)	(0.026)	(0.033)
E[A|X] period 1				
(fixed)				
E[A|X] period 2			0.482	0.528
			(0.030)	(0.034)
E[A|X] period 3			0.319	0.343
			(0.038)	(0.046)
E[A|X] period 4			0.188	0.206
			(0.046)	(0.057)
sd_gpa	0.645	0.478		
	(0.008)	(0.007)		
sd_gpa1			0.657	0.617
			(0.014)	(0.016)
sd_gpa2			0.534	0.521
			(0.013)	(0.013)
sd_gpa3			0.526	0.520
			(0.014)	(0.014)
sd_gpa4			0.547	0.545
			(0.016)	(0.016)
Type-specific parameters				
constant (gpa) - T2		0.634		0.256
		(0.024)		(0.088)
constant (2yr) - T2		0.124		0.603
		(0.290)		(0.196)
constant (4yr) - T2		-0.244		-2.387
		(0.185)		(0.519)
probability T2		0.174		0.075
		(0.022)		(0.011)
constant (gpa) - T3		-0.889		-0.536
		(0.042)		(0.067)
constant (2yr) - T3		-0.201		-1.646
		(0.245)		(0.496)
constant (4yr) - T3		0.271		-0.441
		(0.088)		(0.120)
probability T3		0.359		0.625
		(0.030)		(0.040)
Observations	2055	2055	2055	2055
lnL (total)	6328	5844	5888	5719

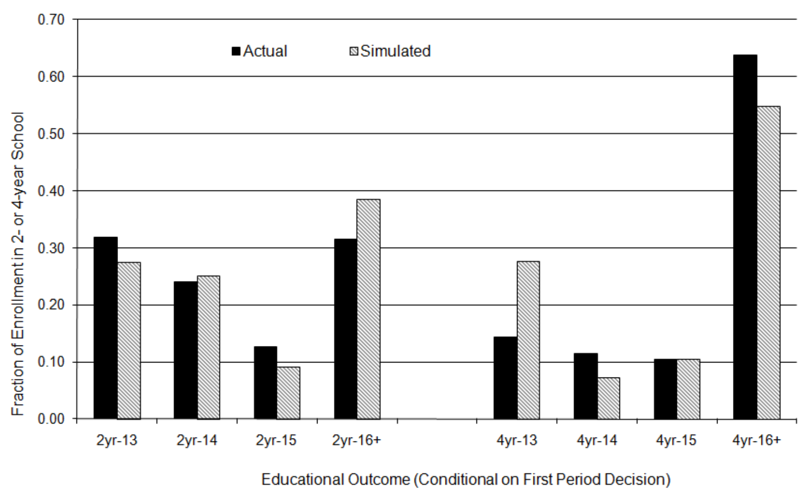
Notes: Utility is in units of \$100,000. Income specification (1) from Table A3 was used to generate counterfactual income estimates. Standard errors (in parentheses) were caclulated from the inverse of the numerical Hessian. Specifications (3) and (4) uses seventeen GPA categories for Emax approximation (0.0, 0.25, 0.50, . . . , 4.0)

Figure V: Actual vs. Simulated Educational Outcomes



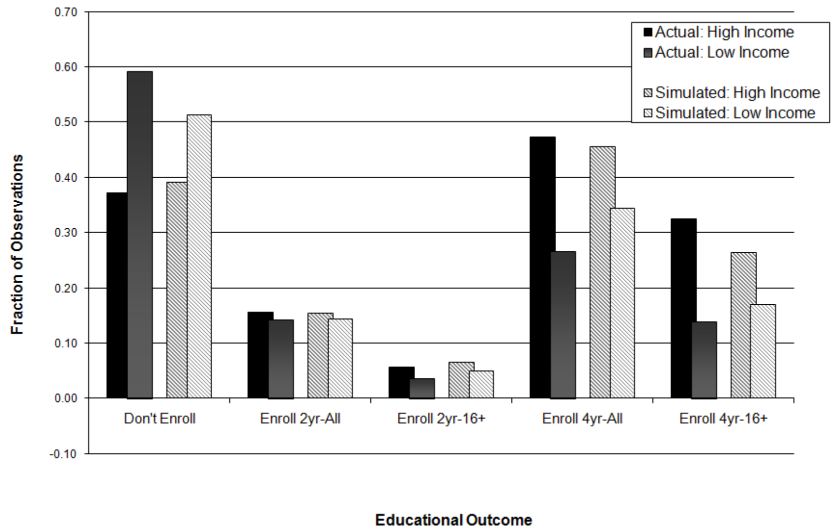
Notes: To generate simulated outcomes, the unobserved type, grade and preference shocks, and choices of each observation is simulated 100 times, assuming that individuals follow the full model described in the text with parameter values equal to those in specification (4) of Table I.

Figure VI: Actual vs. Simulated Outcomes Conditional on Enrollment



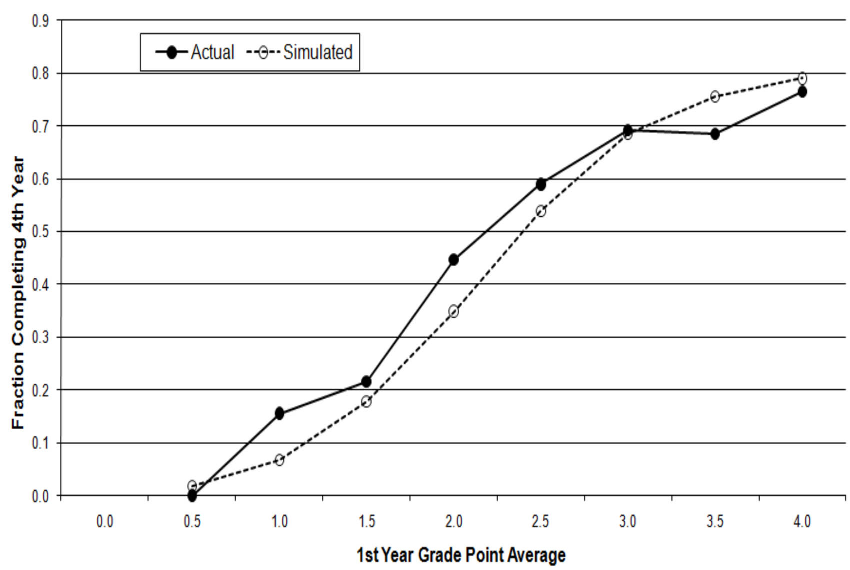
Notes: To generate simulated outcomes, the unobserved type, grade and preference shocks, and choices of each observation is simulated 100 times, assuming that individuals follow the full model described in the text with parameter values equal to those in specification (4) of Table I.

Figure VII: Model Fit of Educational Outcome Differentials by Familiy Income



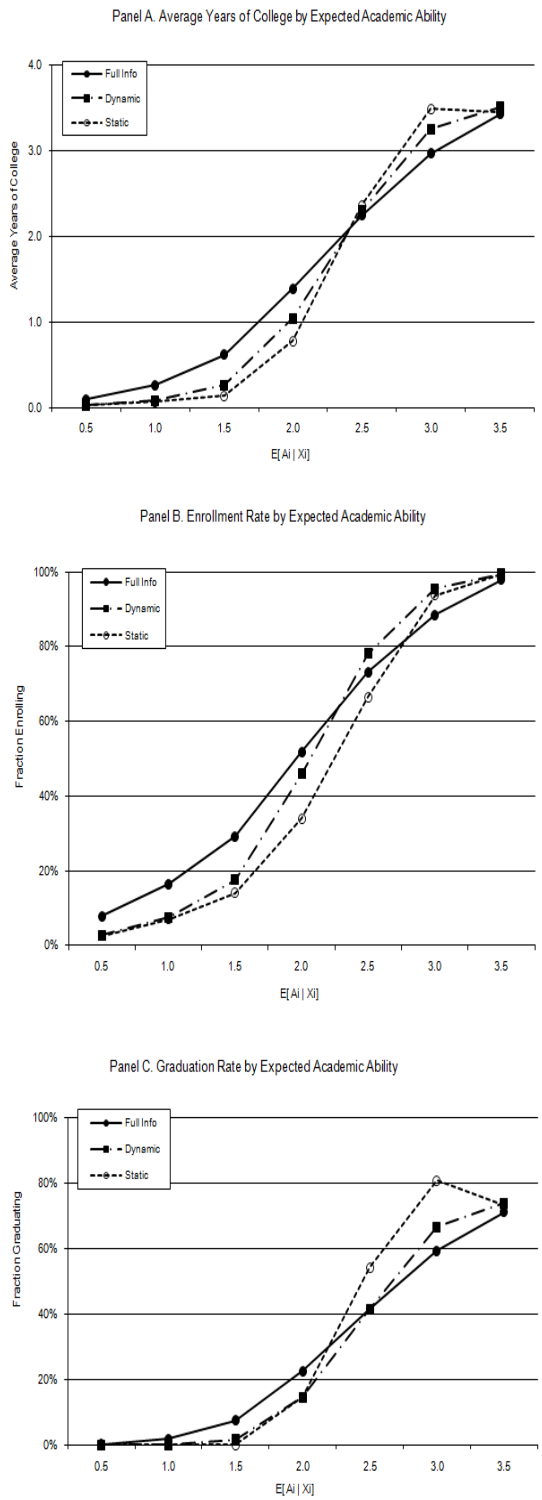
Notes: To generate simulated outcomes, the unobserved type, grade and preference shocks, and choices of each observation is simulated 100 times, assuming that individuals follow the full model described in the text with parameter values equal to those in specification (4) of Table I.

Figure VIII: Actual vs. Simulated Graduation Rates by 1st Year GPA



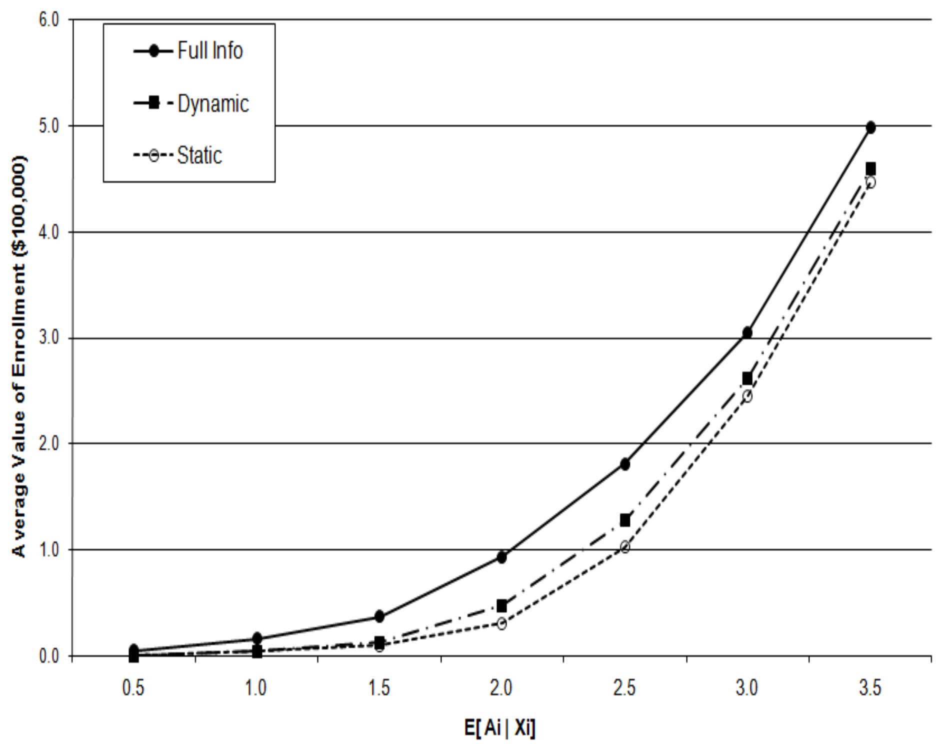
Notes: To generate simulated outcomes, the unobserved type, grade and preference shocks, and choices of each observation is simulated 100 times, assuming that individuals follow the full model described in the text with parameter values equal to those in specification (4) of Table I.

Figure IX: Effect of Uncertainty on Educational Outcomes



Notes: To generate simulated outcomes, the unobserved type, grade and preference shocks, and choices of each observation is simulated 100 times, assuming that individuals follow the static, full dynamic, or perfect information scenarios described in the text. Parameter values are assumed equal to those in specification (4) of Table I.

Figure X: Average Value of College Enrollment Opportunity by Expected Academic Ability



Notes: To generate simulated outcomes, the unobserved type, grade and preference shocks, and choices of each observation is simulated 100 times, assuming that individuals follow the static, full dynamic, or perfect information scenarios described in the text. Parameter values are assumed equal to those in specification (4) of Table I.

Table II: Estimated Option Value, by Expected Academic Ability

	Option value ($1,000)			O.V. as % total value of enrollment in dynamic scenario			O.V. as % welfare loss between full information and static scenarios		
$E[A_i \mid X_i]$	Estimate	5%	95%	Estimate	5%	95%	Estimate	5%	95%
1.0	0.3	0.2	1.7	7%	4%	19%	3%	2%	12%
1.5	3.2	2.8	5.1	25%	22%	29%	12%	11%	17%
2.0	16.8	13.4	20.5	35%	31%	38%	27%	25%	29%
2.5	25.0	18.1	31.2	19%	17%	23%	32%	29%	34%
3.0	16.6	12.7	21.3	6%	5%	8%	28%	25%	31%
3.5	12.2	4.7	19.2	3%	1%	5%	24%	9%	32%
All	14.9	11.4	18.1	14%	12%	16%	27%	25%	29%

Notes: For a given parameter vector, option value is calculated as the average welfare difference between the static and dynamic scenarios when the type, shocks, and choices of each observation is simulated 100 times. Confidence intervals are computed by performing this option value calculation for 500 different draws of the parameter vector from its estimated distribution.

Appendix Materials (Not for Publication)

Appendix I. Proof of Proposition 1

Proposition 1 (The properties of option value).

a. $OptionValue(\varepsilon_{i,1})$ *is non-negative for all* $\varepsilon_{i,1}$.

b. $OptionValue(\varepsilon_{i,1})$ *is greatest for individuals at the enrollment margin in the static model.*

c. $OptionValue(\varepsilon_{i,1})$ *is increasing and the critical value* $\overline{\varepsilon}_{d,1}$ *is decreasing in the level of uncertainty (variance of* $\varepsilon_{i,2}$*).*

d. $OptionValue(\varepsilon_{i,1})$ *reduces the dependence of educational outcomes on* $\varepsilon_{i,1}$.

e. *The option to drop out improves welfare.*

Proof. Consider three groups of individuals which together span the space of $\varepsilon_{i,1}$. *Group A* ($\varepsilon_{i,1} < \overline{\varepsilon}_{d,1}$) does not enroll under either the static or dynamic settings. Since they do not enroll, they get no value from the option to drop out. *Group C* ($\overline{\varepsilon}_{s,1} < \varepsilon_{i,1}$) enrolls in both the dynamic and static settings. Their option value equals $E[\max\{-\rho\varepsilon_{i,1}, \varepsilon_{i,2}\}]$. This expression is decreasing in $\varepsilon_{i,1}$ and positive since $E[\varepsilon_{i,2}|\varepsilon_{i,2} > Z] \geq E[\varepsilon_{i,2}] = 0$ for any value Z. *Group B* ($\overline{\varepsilon}_{d,1} < \varepsilon_{i,1} < \overline{\varepsilon}_{s,1}$) enrolls in the dynamic setting but would not if they were forced to commit to their educational decision ex-ante. For these individuals, the option value is pivotal to enrollment. This option value is equal to $\varepsilon_{i,1} + E[\max\{0, \rho\varepsilon_{i,1} + \varepsilon_{i,2}\}]$. In this region, this expression is positive (by definition of $\overline{\varepsilon}_{d,1}$) and monotonically increasing in $\varepsilon_{i,1}$. Option value of individuals in this group is maximized at the boundary $\varepsilon_{i,1} = \overline{\varepsilon}_{s,1} = 0$ where the option value equals $E[\max\{0, \varepsilon_{i,2}\}]$. This is greater than the option value of any individuals in the other two groups. Properties *a* and *b* follow. For a given level of variance of $\varepsilon_{i,2}$, the truncation point is fixed (at $-\rho\varepsilon_{i,1}$ for *Group A* and 0 for *Group B*). Since increased variance increases the truncated conditional expectation of a random variable, property *c* follows. Like a financial option, the value of the dropout option increases in the variance of the value of the underlying asset ($\varepsilon_{i,2}$). Also, as $OptionValue(\varepsilon_{i,1})$ increases due to increased uncertainty about $\varepsilon_{i,2}$, enrollment becomes desirable to more people, reducing the enrollment threshold. Property *d* can be seen from the decision rules in the previous section. In the fully static case, educational outcomes are fully determined by information available in the first period. This in not true when schooling decisions are sequential. Property *e* is a corollary of *a*: since option value is non-negative, it improves welfare. ■

Appendix II. Dataset Construction

The dataset used in estimation and simulation was constructed from several sources. Table A-I provides an overview of the main variables used in the analysis. The sample of individuals comes from the National Educational Longitudinal Study (NELS). The NELS is a longitudinal survey of a representative sample of U.S. 8th graders in 1988. Interviews were conducted in 1988, 1990, 1992, 1994, and 2000 and complete college transcripts were obtained for most individuals in 2000. The core schooling outcome variables, including yearly grade point average and indicators for enrollment were constructed directly from the college transcripts. The transcripts consist of course-specific records, including student ID, college IPEDS ID number, subject, month and year, credits, letter grade, and standardized numeric grade on a four-point scale. Course-level records were aggregated up to the student x college x term level to identify the primary school enrolled in, and then to the student x year level. The final transcript data contains student x year records of credits attempted, credits earned, grade point average, and several other variables. Individuals were considered enrolled during academic year t if they attempted at least six course credits (the traditional definition of part-time enrollment) at a two- or four-year college during both the Fall and Spring semesters of year t.[30] The model describes college dropout, so I categorize people according to their number of years of continuous enrollment. Students who "stop-out," but eventually return and possibly graduate are grouped with students who dropout permanently in the same year. From the 1992 NELS surveys I utilize high school grade point average, standardized test scores, parents' highest education level, and family income during high school. I convert NELS senior year test scores into AFQT percentile scores using the cross-walk developed by RAND researchers in Kilburn, Hanser, and Klerman (1998).

[30] Enrollment at private (profit or non-profit) two-year colleges, for-profit four-year colleges, or less than two-year schools were counted as non-enrollment.

I supplemented the NELS dataset with institutional characteristics obtained from the Department of Education's 1992 Integrated Postsecondary Education Data System (IPEDS) Institutional Characteristics survey. IPEDS surveys the universe of public and private two- and four-year colleges in the United States. From the IPEDS, I calculated average tuition levels at public two-year and four-year colleges in each state and merged this data onto the NELS. Latitude/longitude coordinates were then assigned to each college in IPEDS and high school in the NELS by zip code from the US Census 1990 Gazetteer Files (http://www.census.gov/geo/www/gazetteer/gazette.html). From this, I calculated distance from each NELS high school to the nearest public two-year and four-year college (in miles). Table A-II displays summary statistics.

One limitation of the NELS dataset is that respondents are relatively young (approximately 26 years old) at the time of the final survey year. Income at this age is a poor indicator of ultimate lifetime income due to job instability, graduate school attendance, and the steep return to initial labor market experience. I instead estimate individuals' expectation of lifetime income using data from an earlier cohort. This procedure is described in the next section.

I restrict the dataset to on-time high school graduates with complete information on key baseline variables and complete college transcripts (unless no claim of college attendance). I also exclude residents of Alaska, Hawaii, and the District of Columbia. From the initial 5,782 men in the NELS, these restrictions eliminate the following number of observations: not 1992 high school graduate (1,068), incomplete transcripts (314), high school missing or in AK/DC/HI (62), missing high school GPA (1,078), missing AFQT (408), missing parent education (193), missing family income (170), missing distance to nearest colleges (412, mostly private high schools for which address is not available), missing college GPA if enrolled (13). After these restrictions the final dataset contains 2,055 men. Though these restrictions reduce the sample considerably, the final unweighted analysis sample is very similar to a nationally representitive sample of U.S. high school graduates. Panel A in Table A-III compares the analysis sample to the full NELS sample of 1992 high school graduates and 12th graders (weighted and unweighted). The unweighted analysis sample is generally very similar to the full representative sample, thus my results can be generalized to all U.S. high school graduates from 1992.

Appendix III. Estimating Conditional Income Expectations

Expectations of lifetime income under different schooling outcomes are a key factor in educational choices. One limitation of the NELS dataset is that respondents are relatively young (approximately 26 years old) at the time of the final survey year. Since income at this age may be a poor indicator of ultimate lifetime income, I do not estimate expectations using individual's actual labor market outcomes. Instead I estimate individuals' expectation of lifetime income using data from a cohort about 12 years earlier, the National Longitudinal Survey of Youth 1979 (NLSY79). This approach assumes students form "reference group expectations" referred to by Manski (1991).

The NLSY79 is a Department of Labor longitudinal survey of 12,686 men and women who were 14-22 years old in 1979. They have been surveyed annually or biennially since. Using variables that are common in both the NLSY79 and NELS (such as high school GPA, parental education, AFQT, ethnicity, urban and region), I first estimate the parameters of a lifetime income equation on the NLSY79 data. My NLSY79 analysis sample consists of all male high school graduates with non-missing covariates, including oversamples of minority and poor individuals. Panel B of Table A-III compares this analysis sample to the NLSY cross-section sample (which doesn't include these oversamples) and to my NELS analysis sample. The NLSY analysis sample is more disadvantaged than NLSY high school graduates generally and than members of my NELS analysis sample. A lack of comparability between the NLSY and the NELS could affect my option value estimates if the returns estimated using the NLSY are not reflective of all high school graduates in the NELS. I examine these comparability issues both by letting returns differ with student background and by restricting analysis to the NLSY cross-section sample (excluding the poor and minority oversamples).

Equation A1 is estimated on the NLSY high school graduate sample using OLS and is used to predict counterfactual lifetime income for individuals in the NELS sample.

$$\begin{aligned} Income_i &= \omega_0 + \omega_{13}1(S_i(t_i) = 13) + \omega_{14}1(S_i(t_i) = 14) + \omega_{15}1(S_i(t_i) = 15) + \omega_{16}1(S_i(t_i) \geq 16) \\ &\quad + \omega_b Black_i + \omega_l Latino_i + \omega_c Central_i + \omega_s South_i + \omega_w West_i + \omega_u Urban_i \\ &\quad + \omega_g HSgpa_i + \omega_a AFQT_i + \omega_p ParentEd_i \\ &\quad + \omega_{ga} HSgpa_i * AFQT_i + \omega_{gp} HSgpa_i * ParentEd_i + \omega_{ap} AFQT_i * ParentEd_i + \epsilon_i^{\omega} \end{aligned} \tag{A1}$$

The dependent variable $Income_i$ is the present discounted value of lifetime income from the period of first labor market entry (t_i) to age 62. The key dependent variable is years of continuous enrollment in school, $S_i(t_i)$, which is entered as a set of four dummy variables and is determined mechanically by period of first labor market entry (t_i) since re-enrollment is ignored. Since NLSY79 individuals are ages 39 to 47 in 2004, the most recent year for which data is available, so I assume that earnings are constant from age 39 to 62. The base specification permits the intercept of lifetime income to vary with observable background and ability variables, but restricts the lifetime income returns to each year of college to be constant across individuals. An alternative specification allows the return to some college (S = 13, 14, or 15) and a BA (S $\geq$ 16) to vary with high school gpa, AFQT, and parent's education. If returns to education differ with student background, then permitting heterogeneous returns will partially mitigate concerns about the comparability of the NLSY and NELS analysis samples. In practice, these interactions are insignificant, so my main analysis uses the constant-returns estimates. Table A-IV provides estimates of the parameters of the lifetime income equation for both the base and heterogeneous-returns model for different assumed values of the discount rate. The last two columns exclude the poor and minority oversamples from the analysis. Again, the estimated returns to each year of college are very similar using the full and smaller samples, so I use the former in my main analysis.

For each individual in the NELS analysis sample, the model estimated in A1 is used to predict counterfactual lifetime income for the five possible schooling levels: $Income_{i1}$ (corresponding to $S_i = 12$) through $Income_{i5}$ (corresponding to $S_i \geq 16$). Table A-V presents the predicted lifetime income counterfactuals for the NELS sample.

Appendix IV. Full Model and Solution

Structure of Choices and Preferences

I model the college enrollment and continuation decisions at four periods in time, corresponding to the four academic years after high school graduation. During the first period individuals decide whether to start at a four-year or two-year college, which I refer to as pathway choice, or to not enroll in college. The pathway chosen affects the level and timing of direct schooling costs (which may differ across individuals) and unmodeled college amenities. At each time period t an individual chooses whether to enter the labor market (receiving payoff $u^w_{i,t}$) or continue in school for another year, receiving payoff $u^s_{i,j,t}$ in period t and the option to make an analogous work-school decision in period $t+1$, where $j = 2, 4$ denotes the type of school currently attending. After period two, students that started at a two-year college must attend a four-year college if they want to continue in school.[31] After period four, there are no more decisions to make and all individuals enter the labor market.[32]

Utility is in dollars. The indirect utility from discontinuing school and entering the labor market at period t equals the expected present discounted value of lifetime income from period t to age 62 ($Income_{i,t}$) plus a random component $\varepsilon^w_{i,t}$.

$$u^w_{i,t} = Income_{i,t} + \varepsilon^w_{i,t} \tag{A2}$$

The expected indirect utility derived from attending school during period t, $u^s_{i,j,t}$, depends linearly on a heterogeneous intercept ($\alpha_{i,j}$, specified later), expected unknown ability (A_i), direct tuition and commuting costs, and a random component $\varepsilon^s_{i,j,t}$. $Distance_{i,j,t}$ and $Tuition_{i,j,t}$ vary by the type of school currently attending (2-year or 4-year), so individuals that start at a two-year school will pay community college tuition for the first two years then four-year college tuition for their third and fourth years.

$$u^s_{i,j,t} = \alpha_{i,j} + \alpha_A E_t[A_i] - (\alpha_D Distance_{i,j,t} + Tuition_{i,j,t}) + \varepsilon^s_{i,j,t} \tag{A3}$$

The random shocks ($\varepsilon^s_{i,j,t}$,$\varepsilon^w_{i,t}$) are learned by the individual prior to making the period t decision. The term $\alpha_A E_t[A_i]$ captures the preference for school (in dollar terms) that covaries with its expected difficulty.[33] Individuals do not know A_i at any time, so they form expectations of it

[31] In the estimation, I do not actually distinguish between people attending two- and four-year schools in their third year. I simplify by assuming that anyone who started at a two-year school that is enrolled in their third year faces the four-year school cost structure, even if they are actually enrolled in a two-year school.

[32] The model does not currently permit two-year and four-year colleges to affect earnings differently or allow for heterogeneity among four-year colleges. Kane and Rouse (1995) find that the return to education received at two- and four-year institutions is comparable. They estimate that the average college student earned about 5% more than similar high school graduates for every year of credits completed, regardless of where those credits were earned.

[33] This specification can be motivated by a model where the difficulty of year t is distributed around a fixed and unobserved individual-specific mean, so $A_{i,t} = A_i + \varepsilon^a_{i,t}$. Individuals learn $A_{i,t}$ after each year, but cannot separate A_i from $\varepsilon^a_{i,t}$. If $\varepsilon^a_{i,t}$ is mean zero and serially uncorrelated, then $E_t[A_{i,t}] = E_t[A_i]$. Also, since I have assumed risk

when making their period-t decisions. Utility is cumulative so individuals who attend a two-year school for two years then enter the labor market, for instance, will receive total lifetime utility of $u^s_{i,2,1} + \beta u^s_{i,2,2} + \beta^2 u^w_{i,3}$, where β is a discount factor.

$\overline{u}^s_{i,j,1}(\cdot)$ represents the non-stochastic component of the indirect utility of attending school. Individuals know baseline characteristics (X_i) as well as the first period shocks ($\varepsilon^s_{i,2,1}$,$\varepsilon^s_{i,4,1}$,$\varepsilon^w_{i,1}$) when making the initial enrollment decision, but learn future shocks and grade outcomes only after enrolling. All other parameters of the model are known to the individual throughout.

Academic Performance

At the end of each year, students enrolled in college learn their performance during that year. Academic performance is measured by the college grade point average (on a four-point scale) during period t. I assume that grades provide a noisy signal of A_i:

$$g_{i,t} = A_i + \varepsilon^g_{i,t} \tag{A4}$$

The $\varepsilon^g_{i,t}$ is the component of grade outcomes that is not serially correlated. This represents idiosyncratic determinants of academic performance that do not persist across time. The conditional expectation of A_i on baseline characteristics (X_i) is given by the heterogeneous term γ_i, which is specified in the next subsection.

$$E[A_i|X_i] = \gamma_i \tag{A5}$$

Heterogeneity

The variables $\alpha_{i,j}$ and γ_i represent persistent preferences for school and persistent determinants of academic aptitude, respectively, which may be correlated in the population. $\alpha_{i,j}$ varies with school type (j) so that individuals may have different tastes for attending a two- or four-year school. To permit a general structure of correlation between unobservable preferences and ability, I specify that $\alpha_{i,j}$ and γ_i come from a mass point distribution which describe the ability and schooling

neutrality, the variance of $\varepsilon^a_{i,t}$ has no impact on expected utility or decisions, so can be ignored.

preferences of M different types of individuals.[34] Type is known to the individual throughout, but is unknown to the econometrician. I also make the parametric assumption that the conditional expectation of A_i on baseline characteristics is linear in high school grade point average ($HSgpa_i$), percentile score on the AFQT, and whether a parent has a college degree ($ParBA_i$).

$$\alpha_{i,j} = \alpha_{0,j} + \alpha_{m,j} \qquad \text{for } m = 1, 2, ..., M \tag{A6}$$

$$\gamma_i = \gamma_0 + \gamma_m + \gamma_G HSgpa_i + \gamma_T AFQT_i + \gamma_P ParBA_i \tag{A7}$$

where γ_m measures the unobserved academic aptitude of people of "type" m and $\alpha_{m,j}$ is their preference for school of type j. I estimate models permitting up to three types ($M = 3$). For Type I individuals, γ_m and $\alpha_{m,j}$ are normalized to zero. Essentially, the specification permits the intercepts of academic performance and of indirect utility to each take on three different values, corresponding to the three unobserved types. As a special case, I will also estimate models with no unobserved heterogeneity, which assumes that all correlation between preference for school and academic aptitude are captured linearly through $\alpha_A E_t[A_i]$.

Solution

At each time t , the individual maximizes the expected discounted value of lifetime utility by choosing whether to discontinue schooling and receive $u^w_{i,t}$ or continue school for at least one more year. The decision problem can be solved for each individual by backwards recursion and by assuming a distribution for the preference and grade shocks ($\varepsilon^s_{i,j,t}$, $\varepsilon^w_{i,t}$, $\varepsilon^g_{i,t}$). Throughout I assume that $\varepsilon^s_{i,2,t}$,$\varepsilon^s_{i,4,t}$, and $\varepsilon^w_{i,t}$ are drawn from an Extreme Value Type I distribution with location and scale parameters zero and τ, respectively. Grade shocks are assumed to be normally distributed with $\varepsilon^g_{i,t} \sim N(0, \sigma_{Gt})$.

With learning, individuals update their belief about A_i in response to new information received through grades. I make the parametric assumption that the conditional expectation of A_i is a weighted average of the unconditional expectation and students' cumulative grade point average.

[34]The use of a mass-point distribution to approximate the distribution of preferences known to the agent but unknown to the econometrician is discussed by Heckman and Singer (1984) and is widely used in dynamic structural work such as Keane and Wolpin (1997) and Eckstein and Wolpin (1999).

The weights are parameters to be estimated.

$$\begin{aligned} E_t[A_i] &= E[A_i|X_i] \text{ if } t=1 \\ &= \gamma_{Xt}E[A_i|X_i] + (1-\gamma_{Xt})\sum_{q=1}^{q=t-1}\frac{g_{i,q}}{t-1} \text{ if } t>1 \end{aligned} \tag{A8}$$

This specification is an approximation of the normal learning model. The normal learning model imposes that $\gamma_{Xt} = \left(\frac{1/\sigma_a^2}{1/\sigma_a^2+(t-1)/\sigma_g^2}\right)$, where σ_a^2 is the variance of A_i and σ_g^2 is the variance of $(g_{i,t} - A_i)$. I have not imposed that the timing of learning follow the behavior implied by the normal learning model. Instead, I estimate γ_{Xt} and the variance of the residual $g_{i,t} - E_t[A_i]$ as parameters.

At period 4 the final enrollment decision is made by comparing the lifetime utility of entering the labor market without graduating to that of continuing for one more year. In periods 2 through 4, I omit the j subscripts.

$$\begin{aligned} V_{i,4}^w &= Income_{i,4} + \varepsilon_{i,4}^w \\ V_{i,4}^s &= \alpha_0 + \alpha_m + \alpha_A E_4[A_i] - Cost_{i,4} + \beta E_4[V_{i,5}] + \varepsilon_{i,4}^s \end{aligned} \tag{A9}$$

where $Cost_{i,4} = \alpha_D Distance_{i,4} + Tuition_{i,4}$. At period 4, expectations are taken over the distribution of labor market shocks in period 5 ($\varepsilon_{i,5}^w$) and grade shocks in period 4 ($g_{i,4}$). Since all individuals enter the workforce upon reaching period 5, $V_{i,5} = V_{i,5}^w = Income_{i,5} + \varepsilon_{i,5}^w$ and $E_4[V_{i,5}] = Income_{i,5} + \tau\lambda$ from the extreme value assumption [$\lambda = 0.577$ is Euler's constant]. Future utility is discounted at the rate β. If individuals learn about unobserved ability through grades, then $E_4[A_i]$ is a weighted average of the unconditional expectation and previous grade realizations:

$$V_{i,4}^s = \alpha_0 + \alpha_m + \alpha_A\left[\gamma_{X4}E[A_i|X_i] + (1-\gamma_{X4})\sum_{q=1}^{q=3}\frac{g_{i,q}}{3}\right] - Cost_{i,4} + \beta[Income_{i,5} + \tau\lambda] + \varepsilon_{i,4}^s \tag{A10}$$

Individuals will continue to graduation if $V_{i,4}^s > V_{i,4}^w$.

At periods 2 and 3, the enrollment and continuation decisions are made by comparing the lifetime utility of entering the labor market immediately to that of continuing school for one more year.

$$
\begin{aligned}
V^w_{i,t} &= Income_{i,t} + \varepsilon^w_{i,t} \\
V^s_{i,t} &= \alpha_0 + \alpha_m + \alpha_A E_t[A_i] - Cost_{i,t} + \beta E_t[V_{i,t+1}] + \varepsilon^s_{i,t}
\end{aligned}
$$

where $V_{i,t+1} = \max(V^w_{i,t+1}, V^s_{i,t+1})$. Expectations are again taken over the distribution of all future preference shocks ($\varepsilon^w_{i,q}, \varepsilon^s_{i,q}$ for $q > t$) and grade shocks ($g_{i,q}$ for $q \geq t$), but now both of these influence future educational decisions. Integrating out the grade shocks (due to conditional independence between grades and shocks, see Rust (1987)), the $E\max$ term can be written as:

$$
E_t\left[\max(V^w_{i,t+1}, V^s_{i,t+1})\right] = \int E_t\left[\max(V^w_{i,t+1}, V^s_{i,t+1}) | g_{i,t}\right] \cdot \Pi(dg_{i,t}|X_i, \{g_{i,1}...g_{i,t-1}\})
$$

where $\Pi(dg_{i,t}|X_i, \{g_{i,1}...g_{i,t-1}\})$ is the pdf of the t-period grade outcome conditional on information available at time t. The conditional expectation is taken only over the future preference shocks ($\varepsilon^w_{i,q}, \varepsilon^s_{i,q}$ for $q > t$). Again with the assumption that the preference shocks are not serially correlated and are drawn from an extreme value distribution, this expectation has a closed-form representation[35]:

$$
\begin{aligned}
& E_t\left[\max(V^w_{i,t+1}, V^s_{i,t+1})\right] \\
= & \int \left[\tau\lambda + \tau \log\left\{\exp\left(\frac{1}{\tau}\overline{V}^s_{i,t+1}(g_{i,t})\right) + \exp\left(\frac{1}{\tau}\overline{V}^w_{i,t+1}\right)\right\}\right] \cdot \Pi(dg_{i,t}|X_i, \{g_{i,1}...g_{i,t-1}\})
\end{aligned}
$$

In order to actually solve and estimate the model, I discretize $g_{i,t}$ into K values and approximate $\Pi(dg_{i,t}|X_i, \{g_{i,1}...g_{i,t-1}\})$ with a discretized version $p(g^k_{i,t}|X_i, \{g_{i,1}...g_{i,t-1}\})$.[36] The $E\max$ term

[35]Domencich and McFadden (1975, Chapter 4) show that the expected value of the maximum of an EV(1) random variable has this closed form representation.

[36]See Rust (1987). Since grades are distributed normally, the transition probabilities can be computed directly using the standard normal cumulative distribution function. $p(g^k_{i,t}|X_i, \{g_{i,1}...g_{i,t-1}\}) = \Phi\left(\frac{g^k_{i,t} + (0.5)*kstep - E_t[g_{i,t}]}{\sigma_{t,g}}\right) - \Phi\left(\frac{g^k_{i,t} - (0.5)*kstep - E_t[g_{i,t}]}{\sigma_{t,g}}\right)$ where kstep is the distance between the points in the discretized grade space.

can then be written as

$$
\begin{aligned}
& E_t\left[\max(V^w_{i,t+1}, V^s_{i,t+1})\right] \\
= & \sum_{k=1}^{K}\left[\tau\lambda + \tau\log\left\{\exp\left(\frac{1}{\tau}\overline{V}^s_{i,t+1}(g^k_{i,t})\right) + \exp\left(\frac{1}{\tau}\overline{V}^w_{i,t+1}\right)\right\}\right]\cdot p(g^k_{i,t}|X_i,\{g_{i,1}...g_{i,t-1}\})
\end{aligned}
$$

And the indirect utility function becomes:

$$
\begin{aligned}
V^s_{i,t} = {} & \alpha_0 + \alpha_m + \alpha_A\left[\gamma_{Xt}E[A_i|X_i] + (1-\gamma_{Xt})\sum_{q=1}^{q=t-1}\frac{g_{i,q}}{t-1}\right] - Cost_{i,t} \\
& +\beta\left[\sum_{k=1}^{K}\left[\tau\lambda + \tau\log\left\{\exp\left(\frac{1}{\tau}\overline{V}^s_{i,t+1}(g^k_{i,t})\right) + \exp\left(\frac{1}{\tau}\overline{V}^w_{i,t+1}\right)\right\}\right]\cdot p(g^k_{i,t}|X_i,\{g_{i,1}...g_{i,t-1}\})\right] + \varepsilon^s_{i,t}
\end{aligned}
\tag{A11}
$$

Individuals will continue their education if $V^s_{i,t} > V^w_{i,t}$.

At period 1, the value of the two enrollment options takes a similar form:

$$
\begin{aligned}
V^s_{i,j,1} = {} & \alpha_{0,j} + \alpha_{m,j} + \alpha_A E[A_i|X_i] - Cost_{i,j,t} \\
& +\beta\left[\sum_{k=1}^{K}\left[\tau\lambda + \tau\log\left\{\exp\left(\frac{1}{\tau}\overline{V}^s_{i,2}(g^k_{i,1})\right) + \exp\left(\frac{1}{\tau}\overline{V}^w_{i,2}\right)\right\}\right]\cdot p(g^k_{i,1}|X_i,)\right] + \varepsilon^s_{i,j,1}
\end{aligned}
\tag{A12}
$$

At period 1, individuals maximize expected lifetime utility by choosing between $V^s_{i,2,1}, V^s_{i,4,1}$, and $V^w_{i,1}$.

Appendix V. Model Alternatives and Extensions

The empirical model places several important restrictions on individuals' choice and information sets. I assume that (1) labor market entry is costless but irreversible. People cannot return to school after entering the workforce. I also assume that (2) individuals do not learn about the relative desirability of schooling and work while in the labor market. In combination, these two restrictions mean that individuals "exercise their option" by leaving school. Lastly, I assume that (3) labor market draws persist following labor market entry. Individuals do not receive another labor market draw while in the workforce.

To see these assumptions more clearly, consider a general model in which attending school and working both provide information - people learn about their enjoyment of each only through doing them (individual subscripts have been omitted):

$$
\begin{aligned}
u_t^w &= \alpha^w + Income_t^w(Exp_t, S_t) + \alpha_A^w E[A^w|\Im_o, \Im_t^w] + SwitchCost^w \cdot 1(Attend_{t-1} = 1) + \varepsilon_t^w \\
u_t^s &= \alpha^s + Income_t^s(Exp_t, S_t) + \alpha_A^s E[A^s|\Im_o, \Im_t^s] + SwitchCost^s \cdot 1(Attend_{t-1} = 0) + \varepsilon_t^s
\end{aligned}
$$

Individuals form expectations of the state-specific unknown component of indirect utility (A^w and A^s) based on information available at baseline ($\Im_o$), and that learned while working ($\Im_t^w$) and attending school ($\Im_t^s$) up to that point. Moving between the labor market and school is costly. Income while working or attending school depends on labor market experience (Exp_t) and years of completed schooling (S_t) up to that point. I assume $Income_t^s(Exp_t, S_t) = -Cost_t = -(\alpha_D Distance_t + Tuition_t)$. Any income earned during school will be absorbed in the estimate of α^s.

Assumption (1) corresponds to the restriction that $SwitchCost^s = \infty$ and $SwitchCost^w = 0$. While it is possible to re-enter college after leaving or taking time out, few people do so in practice. Table A-VI presents the fraction of students enrolled during each year, by the number of years of continuous schooling. In my sample, the fraction of students who return in the year after their first year of non-enrollment is 17%, 19%, 28%, and 27% for those whose first year of non-enrollment is year 1 to 4, respectively. Relatively few of these eventually earn a B.A. degree. This restriction can be relaxed and $SwitchCost^s$ and $SwitchCost^w$ can be estimated directly from the data.[37] Also, the value of being able to return to school is partially embedded in my estimates of $Income_{i,t}^w$. I combine the earnings of people who enter the labor market at period t and never return to school with those who eventually do return to school. Therefore my estimate of $Income_{i,t}^w$ is inclusive of the expected financial gains of being able to return to school after entering the labor market at time t.

Assumption (2) corresponds to the restriction that $\Im_t^w = \Im_o$ for all t. High school graduates' expectation of the enjoyment of future work does not depend on their past experience. This assumption is innocuous if people are not able to return to school upon discovering that they don't like

[37] Keane and Wolpin (1997) estimate the cost of returning to school after dropping out to be \$23,000 during high school and \$10,000 during college.

working. I also assume that $\Im_o = \{HSgpa,\ AFQT,\ ParBA,\ Type\}$ and $\Im_t^s = \{g_1, ..., \ g_{t-1}\}$. The fact that returning to school is rare could be due to high switching costs or to limited learning while working, so my specification requires that only one of these assumptions holds. Allowing for learning about tastes for work is an important extension, but one that may need to be pursued with a different dataset. I use course grades to measure academic aptitude and to serve as a proxy for taste for school. The current dataset does not contain an obvious analog proxy for individuals' enjoyment of work.

Assumption (3) corresponds to replacing the labor market shock ε_t^w with $\varepsilon_t^w \cdot 1(Attend_{t-1} = 1)$. Individuals only receive a new labor market draw if they are currently attending school. I assume that each year of college provides access to a new set of labor market opportunities previously unavailable, which increases mean earnings and generates a new draw. Consistent with this assumption, Oreopoulos, von Wachter, and Heisz (2006) find that temporary labor market shocks (e.g. graduating college during a recession) have permanent effects on lifetime earnings. Significant initial earnings losses fade only after 8 to 10 years, generating large losses in the total present value of lifetime earnings.

These generalizations are beyond the scope of this current paper, but their implications for my empirical results are discussed in the body of the paper.

Appendix VI. Model Fit

Figures A-I to A-V extend Figures V to VIII to include the model fit for all four models estimated in Table I. Generally, the preferred specification (Column (4) in Table I) provides the best fit of the data.

Table A- I: Variable Descriptions and Sources

Variable	Description	Source
high school gpa	Cumulative grade point average in high school on 4.0 scale	NELS.
afqt score	Armed Forces Qualifying Test percentile score	Constructed from NELS test score variables using method developed by RAND (see text).
parent education	Years of school attended by most educated parent	NELS.
parent has ba	Indicator for whether at least one parent earned a BA degree	NELS. Constructed from pareduc variable.
low income family	Indicator for whether family income during high school was below $35,000 (approximately the median)	NELS. Constructed from faminc variable.
urban	Attended urban high school	NELS. Constructed from phsurban variable.
region northeast	High school in Northeast	NELS. NLSY categorization.
region northcentral	High school in Northcentral	NELS. NLSY categorization.
region south	High school in South	NELS. NLSY categorization.
region west	High school in West	NELS. NLSY categorization.
white	Ethnicity white	NELS
black	Ethnicity black	NELS
latino	Ethnicity latino	NELS
distance to 2year	Distance from high school to nearest public two-year college.	Computed from lat/long coordinates of high school (NELS) and each public 2-year college in state (IPEDS)
distance to 4year	Distance from high school to nearest public four-year college.	Computed from lat/long coordinates of high school (NELS) and all public 4-year college in state (IPEDS)
tuition at public 2year	Average tuition ($1992) of public two-year colleges in high school state	IPEDS
tuition at public 4year	Average tuition ($1992) of public four-year colleges in high school state	IPEDS
income1	Expected present discounted value of lifetime income if do not enter college in first year after high school. (thousands of $1992)	Estimated using out-of-sample prediction from NLSY (see text).
income2	Expected present discounted value of lifetime income if exit college after first year (thousands of $1992)	Estimated using out-of-sample prediction from NLSY (see text).
income3	Expected present discounted value of lifetime income if exit college after second year (thousands of $1992)	Estimated using out-of-sample prediction from NLSY (see text).
income4	Expected present discounted value of lifetime income if exit college after third year (thousands of $1992)	Estimated using out-of-sample prediction from NLSY (see text).
income5	Expected present discounted value of lifetime income if complete four years of college (thousands of $1992)	Estimated using out-of-sample prediction from NLSY (see text).
gpa(t)	Grade point average during year (t) of college	Computed from NELS college transcripts for all courses taken for credit (including failures).
enroll(t)	Indicator for enrollment in college during year (t)	Computed from NELS college transcripts. Individual must have attempted at least six units of college credit (approx part-time) in each semester during year (t).
contenroll	Years of continuous enrollment in college after high school graduation.	Constructed from enroll(t).
fouryear(t)	Indicator for enrollment in four-year college during year (t)	Constructed from enroll(t) and college type from IPEDS. Equals one if enroll(t) = 1 and enrolled in a four-year school in either semester
twoyear(t)	Indicator for enrollment in two-year college during year (t)	Constructed from enroll(t) and fouryear(t)

Table A- II: Summary Statistics

Variable	Mean	Standard Deviation	Min	Max
Baseline variables				
high school gpa	2.70	0.68	0.14	4.00
afqt score	46.7	26.9	1	99
parent education (years)	14.2	2.2	10	19
parent has ba	0.28	0.45	0	1
low income family	0.55	0.50	0	1
urban	0.62	0.49	0	1
region northeast	0.16	0.37	0	1
region northcentral	0.31	0.46	0	1
region south	0.32	0.47	0	1
region west	0.20	0.40	0	1
white	0.73	0.45	0	1
black	0.08	0.28	0	1
latino	0.11	0.31	0	1
distance to 2year	15.5	20.5	0	162
distance to 4year	24.0	26.9	0	234
tuition at public 2year	1482	874	280	3476
tuition at public 4year	2298	770	1251	4265
Educational outcomes				
enroll year 1	0.53	0.50	0	1
year 2	0.50	0.50	0	1
year 3	0.43	0.50	0	1
year 4	0.40	0.49	0	1
start at 2year	0.15	0.36	0	1
start at 4year	0.38	0.49	0	1
gpa year 1	2.42	0.86	0.00	4.00
year 2	2.47	0.90	0.00	4.00
year 3	2.63	0.88	0.00	4.00
year 4	2.75	0.85	0.00	4.00
yrs of continuous enrollment	13.81	2.11	12	19
don't enroll	0.47	0.50	0	1
enroll year 1 only	0.10	0.30	0	1
enroll years 1-2 only	0.08	0.27	0	1
enroll years 1-3 only	0.06	0.24	0	1
enroll at least 4 years	0.29	0.45	0	1

Notes: All variables have 2,055 observations, with the exception of GPA variables which are restricted to those enrolled in each year

Table A- III: Representativeness and Comparability of NELS and NLSY Samples

Panel A. NELS sample using full sample and sample weights

	Base		All (unweighted)		All (w=f4qwt92g)		All (w=f4f2pnwt)	
	obs	mean	obs	mean	obs	mean	obs	mean
high school gpa	2055	2.70	3520	2.69	3520	2.64	3474	2.65
afqt score	2055	46.66	3855	48.93	3855	46.44	3819	46.95
parent education (years)	2055	14.18	4315	14.47	4315	14.46	4276	14.45
parent has ba	2055	0.28	4315	0.35	4315	0.34	4276	0.35
low income family	2055	0.55	4007	0.60	4007	0.60	3972	0.62
urban	2055	0.62	4707	0.69	4707	0.69	4632	0.69
region northeast	2055	0.16	4672	0.20	4672	0.20	4624	0.19
region northcentral	2055	0.31	4672	0.28	4672	0.27	4624	0.26
region south	2055	0.32	4672	0.32	4672	0.34	4624	0.35
region west	2055	0.20	4672	0.20	4672	0.20	4624	0.20
white	2055	0.73	4711	0.72	4711	0.74	4636	0.74
black	2055	0.08	4711	0.08	4711	0.11	4636	0.10
latino	2055	0.11	4711	0.11	4711	0.09	4636	0.09
distance to 2year	2055	15.54	3672	14.90	3672	15.04	3614	15.02
distance to 4year	2055	24.02	3672	23.30	3672	22.90	3614	22.13
tuition at public 2year	2055	1482.12	4668	1473.77	4668	1477.21	4620	1458.47
tuition at public 4year	2055	2298.27	4672	2298.17	4672	2287.92	4624	2277.31
Total observations	2055		4714		4714		4638	

Panel B. NLSY sample vs. NELS sample

	Base		Cross-section only		Cross-section (weighted)		NELS sample	
	obs	mean	obs	mean	obs	mean	obs	mean
predicted pdv lifetime income	1982	529.70	1352	580.59	1352	601.14	2055	594.08
black	1982	0.26	1352	0.08	1352	0.03	2055	0.08
latino	1982	0.15	1352	0.05	1352	0.01	2055	0.11
regionnc	1982	0.29	1352	0.36	1352	0.38	2055	0.31
regionso	1982	0.35	1352	0.28	1352	0.25	2055	0.32
regionwe	1982	0.19	1352	0.17	1352	0.17	2055	0.20
urban14	1982	0.78	1352	0.75	1352	0.77	2055	0.62
gpahs	1982	2.39	1352	2.50	1352	2.54	2055	2.70
afqt89	1982	49.87	1352	57.25	1352	60.56	2055	46.66
parented	1982	12.34	1352	13.03	1352	13.33	2055	14.18
don't enroll	1982	0.59	1352	0.56	1352	0.54	2055	0.47
enroll year 1 only	1982	0.08	1352	0.07	1352	0.07	2055	0.10
enroll years 1-2 only	1982	0.10	1352	0.10	1352	0.10	2055	0.08
enroll years 1-3 only	1982	0.06	1352	0.06	1352	0.06	2055	0.06
enroll at least 4 years	1982	0.17	1352	0.21	1352	0.23	2055	0.29

Notes: In Panel A, "All" refers to all male 1992 high school graduates in the NELS. Weight f4qwt92g corresponds to 1992 high school graduates and weight f4f2pnwt corresponds to 1992 12th graders. Number of observations varies by column due to missing values. In Panel B, samples include male high school graduates with non-missing covariates.

Table A- IV: Parameter Estimates from Lifetime Income Equation

	Dependent variable: PDV of lifetime income post-school					
	All Men in NLSY				NLSY Cross-section	
	d = 5%	d =10%	d = 5%	d =10%	d = 5%	d = 5%
	No weights	No weights	No weights	No weights	No weights	Weighted
	(1)	(2)	(3)	(4)	(5)	(6)
contenroll = 13	34.64	19.36	95.61	56.34	20.83	23.86
	(22.09)	(10.63)	(75.61)	(36.57)	-(29.86)	-(34.23)
contenroll = 14	55.54	32.10	126.74	76.04	62.53	53.26
	(24.77)	(11.35)	(82.67)	(39.35)	-(32.06)	-(35.60)
contenroll = 15	165.89	88.99	238.77	133.56	189.16	202.90
	(37.88)	(17.43)	(85.66)	(40.95)	-(48.70)	-(54.29)
contenroll > 15	328.13	183.50	-82.26	-6.75	327.79	333.55
	(30.08)	(14.17)	(154.00)	(72.66)	-(35.90)	-(38.25)
ParentEd	7.65	4.63	13.15	6.96	13.02	11.88
	(8.58)	(4.02)	(8.22)	(3.85)	-(13.52)	-(14.90)
Black	-81.17	-44.76	-80.82	-44.71	-56.53	-66.94
	(19.14)	(8.97)	(19.06)	(8.95)	-(31.02)	-(32.43)
Latino	5.93	2.64	1.23	0.71	3.29	-22.28
	(23.21)	(11.01)	(22.81)	(10.84)	-(39.90)	-(39.34)
NorthCentral	-45.96	-24.49	-42.48	-22.52	-40.72	-41.11
	(24.85)	(11.74)	(24.89)	(11.75)	-(29.02)	-(32.07)
South	-56.99	-29.31	-54.99	-28.10	-56.79	-50.44
	(24.54)	(11.61)	(24.52)	(11.60)	-(31.56)	-(35.96)
West	-56.01	-29.65	-52.13	-27.85	-63.45	-77.30
	(25.08)	(11.78)	(25.12)	(11.76)	-(31.05)	-(35.36)
Urban	32.60	13.06	31.70	12.65	39.88	42.69
	(15.71)	(7.51)	(15.57)	(7.43)	-(19.36)	-(21.76)
HSgpa	42.78	24.61	74.22	39.36	43.67	59.40
	(41.40)	(19.49)	(40.77)	(19.42)	-(64.64)	-(74.50)
AFQT	1.52	0.99	3.03	1.64	2.32	0.98
	(1.35)	(0.64)	(1.34)	(0.64)	-(1.95)	-(2.29)
HSgpa*AFQT	0.25	0.04	-0.19	-0.14	0.28	0.35
	(0.43)	(0.20)	(0.42)	(0.19)	-(0.56)	-(0.62)
HSgpa*ParentEd	-0.55	-0.36	-2.03	-1.14	-1.29	-2.51
	(3.95)	(1.84)	(3.81)	(1.80)	-(5.86)	-(6.50)
AFQT*ParentEd	-0.03	-0.03	-0.08	-0.05	-0.11	-0.05
	(0.08)	(0.04)	(0.09)	(0.04)	-(0.12)	-(0.15)
(s13-s15)*AFQT			-0.16	-0.27		
			(0.67)	(0.32)		
s16*AFQT			1.91	0.65		
			(1.33)	(0.64)		
(s13-s15)*HSgpa			0.47	-3.39		
			(28.99)	(13.57)		
s16*HSgpa			46.76	29.88		
			(47.09)	(22.43)		
(s13-s15)*ParentEd			-4.01	-1.15		
			(4.91)	(2.40)		
s16*ParentEd			9.97	3.94		
			(9.85)	(4.52)		
Constant	223.52	112.62	134.05	72.50	184.34	218.81
	(93.60)	(44.27)	(91.50)	(43.26)	-(153.82)	-(172.47)
Observations	1,982	1,982	1,982	1,982	1,352	1,352
R-squared	0.30	0.33	0.30	0.34	0.34	0.33

Notes: Robust standard errors in parentheses. Specifications (1) to (4) use all male high school graduates in the NLSY with non-missing covariates, including the poor white, black, and hispanic supplemental samples. Specifications (5) and (6) use only male high school graduates in the cross-section sample.

Table A- V: Predicted Lifetime Income and Incremental Returns by Years of Continuous Enrollment

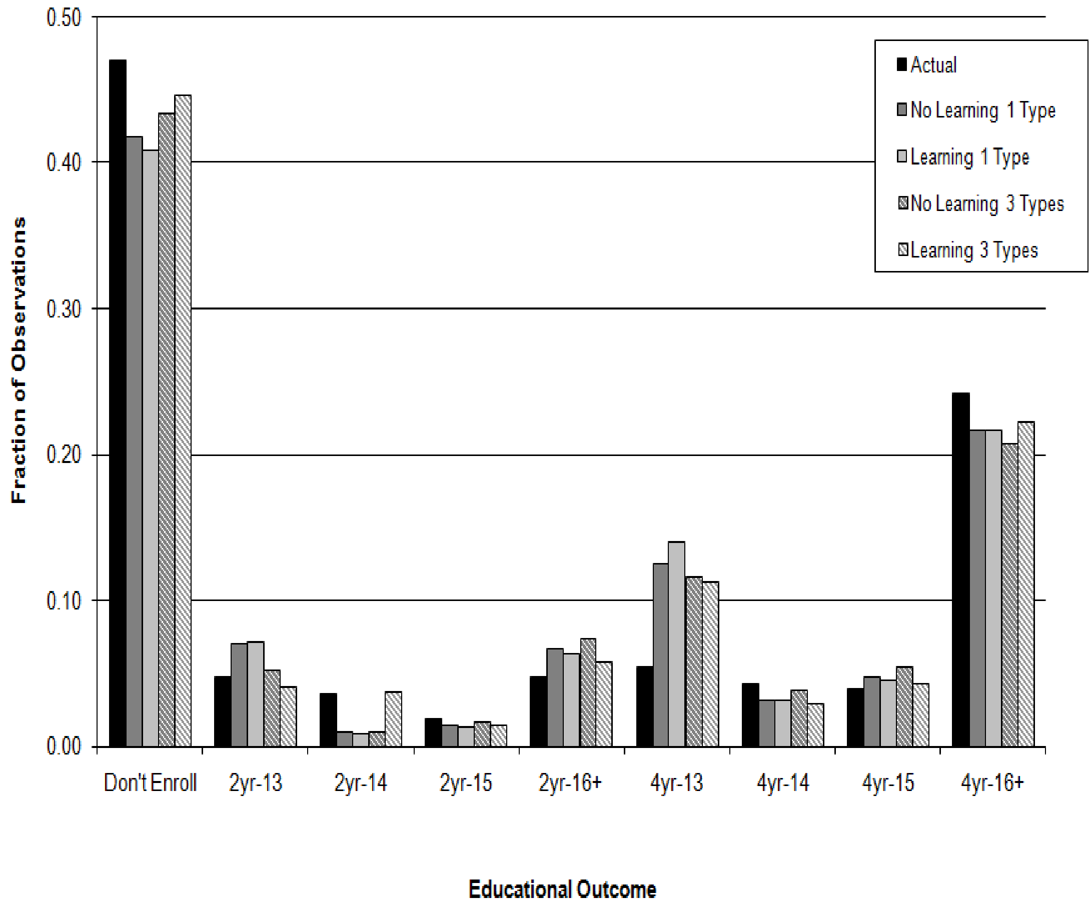
			Men in NELS Sample								
			Predicted Present Value of Lifetime Income (,000)					Predicted Incremental Income Increase (,000)			
Model	Discount rate		12	13	14	15	16	13	14	15	16
(1)	5%	mean	481	516	537	647	809	35	21	110	162
		stdev	89	89	89	89	89	0	0	0	0
(2)	10%	mean	244	263	276	333	428	19	13	57	95
		stdev	39	39	39	39	39	0	0	0	0
(3)	5%	mean	473	506	537	649	748	33	31	112	99
		stdev	72	66	66	66	151	11	0	0	94
(4)	10%	mean	241	259	279	336	401	18	20	58	65
		stdev	33	26	26	26	67	10	0	0	47

Notes: Parameters of lifetime income model were estimated using the data from the NLSY and fitted to men in the NELS sample. See Table A-IV for parameter estimates and model specifications.

Table A- VI: Fraction of Sample that Return to College After Dropping Out

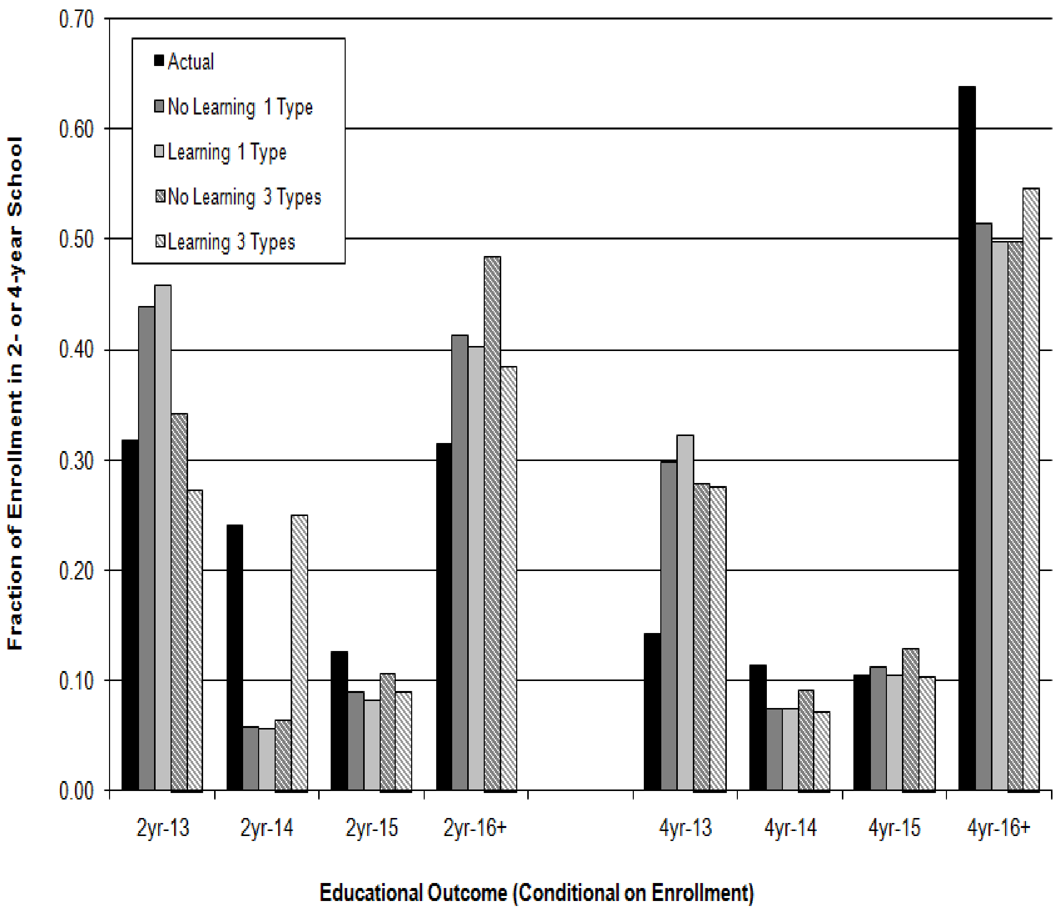
Years of Continuous Educ.	Freq	Fraction Who Enroll in Year t: 1	2	3	4	5	6	7	Fraction Who Earn BA
12	966	0.00	0.16	0.14	0.13	0.12	0.10	0.09	0.08
13	210	1.00	0.00	0.20	0.24	0.19	0.14	0.08	0.14
14	163	1.00	1.00	0.00	0.29	0.18	0.19	0.10	0.21
15	121	1.00	1.00	1.00	0.00	0.27	0.24	0.15	0.36
16+	595	1.00	1.00	1.00	1.00	0.50	0.22	0.15	0.87

Figure A- I: Actual vs. Simulated Educational Outcomes



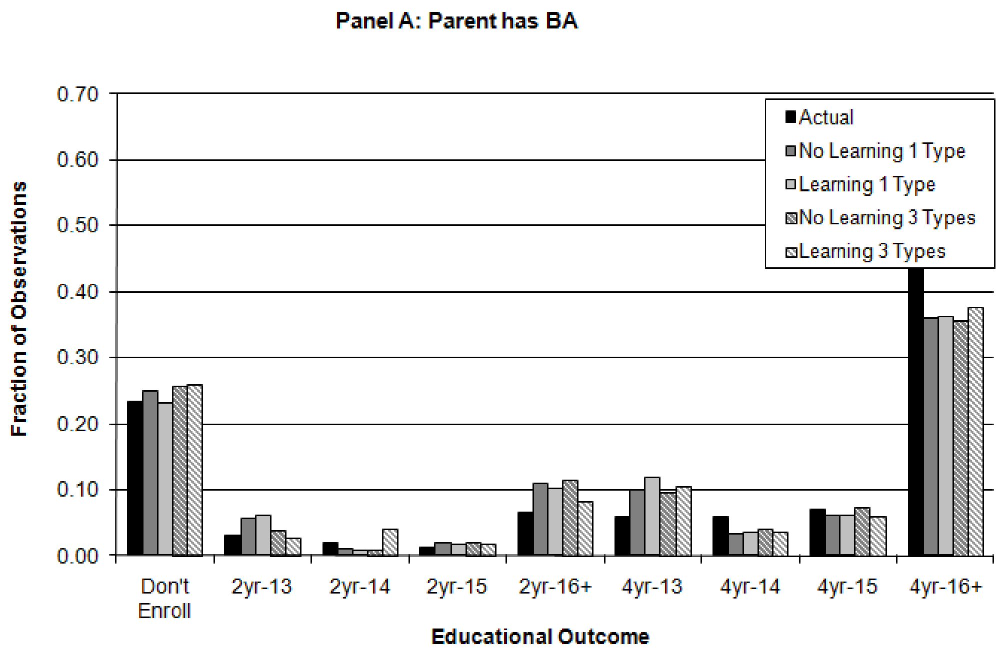
Notes: To generate simulated outcomes, the unobserved type, grade and preference shocks, and choices of each observation is simulated 100 times, assuming individuals follow the choice models described in the text with parameter values equal to those in specifications (1)-(4) of Table I.

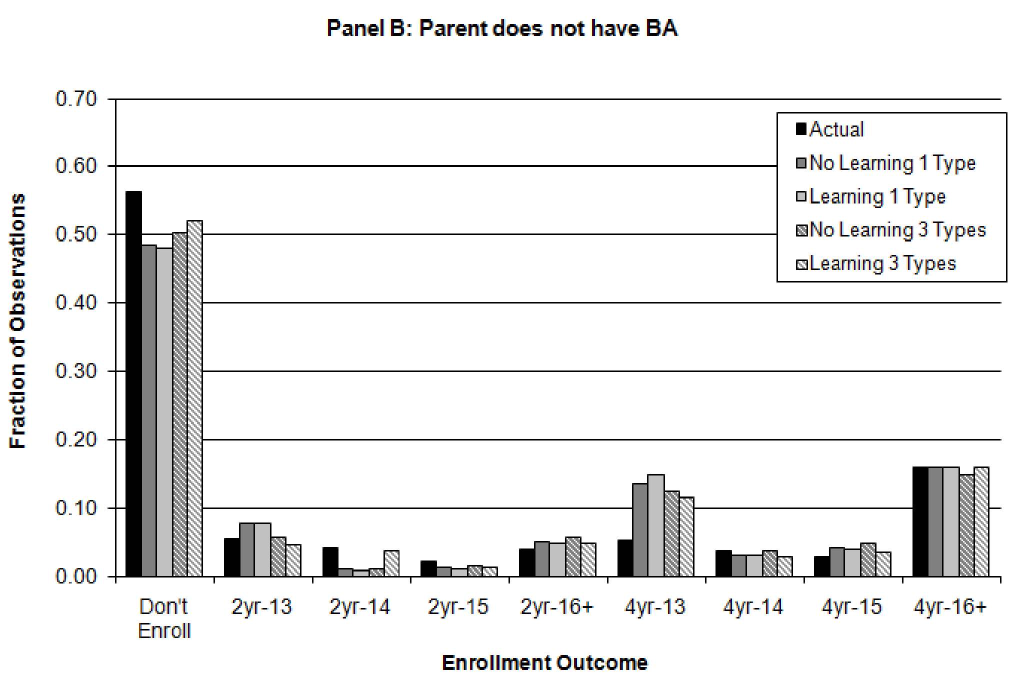
Figure A- II: Actual vs. Simulated Outcomes Conditional on Enrollment



Notes: To generate simulated outcomes, the unobserved type, grade and preference shocks, and choices of each observation is simulated 100 times, assuming individuals follow the choice models described in the text with parameter values equal to those in specifications (1)-(4) of Table I.

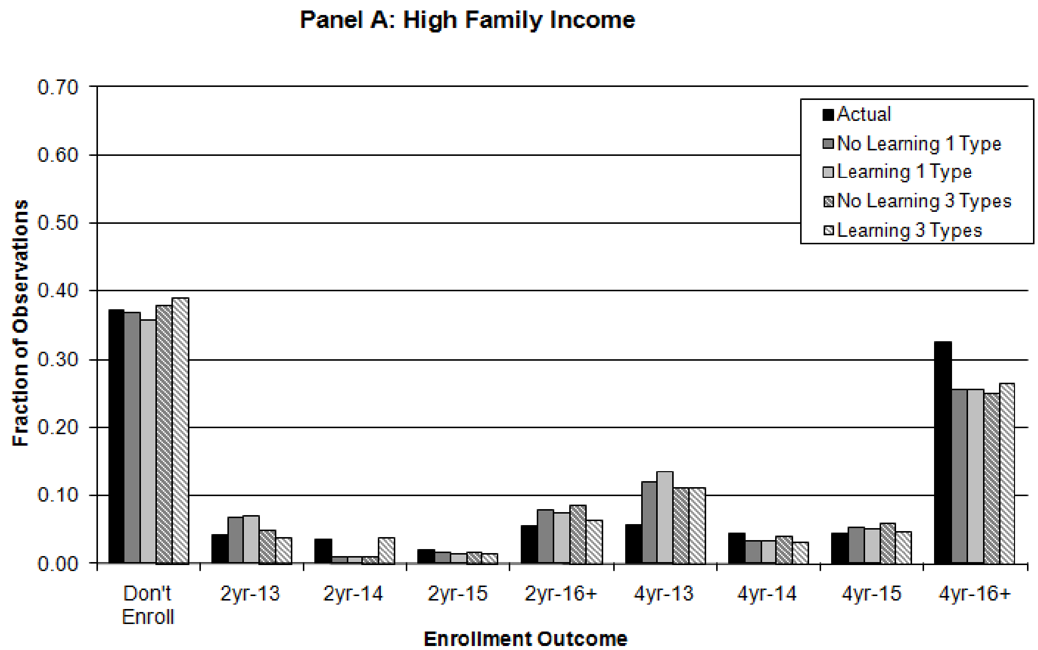
Figure A- III: Actual vs. Simulated Educational Outcomes, by Parent Education

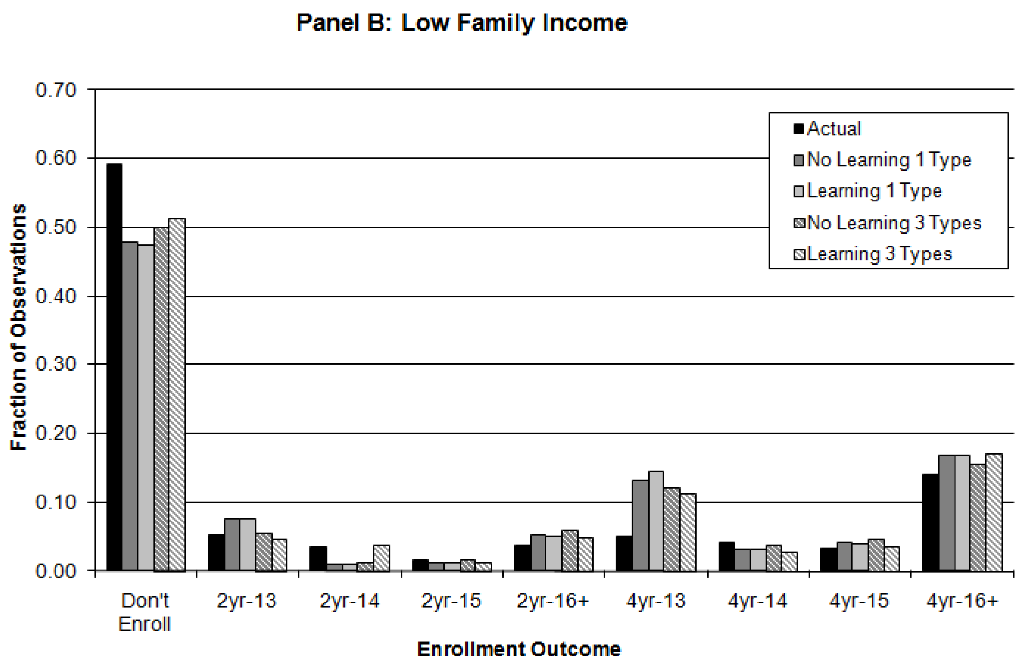




Notes: To generate simulated outcomes, the unobserved type, grade and preference shocks, and choices of each observation is simulated 100 times, assuming that individuals follow the choice models described in the text with parameter values equal to those in specifications (1)-(4) of Table I.

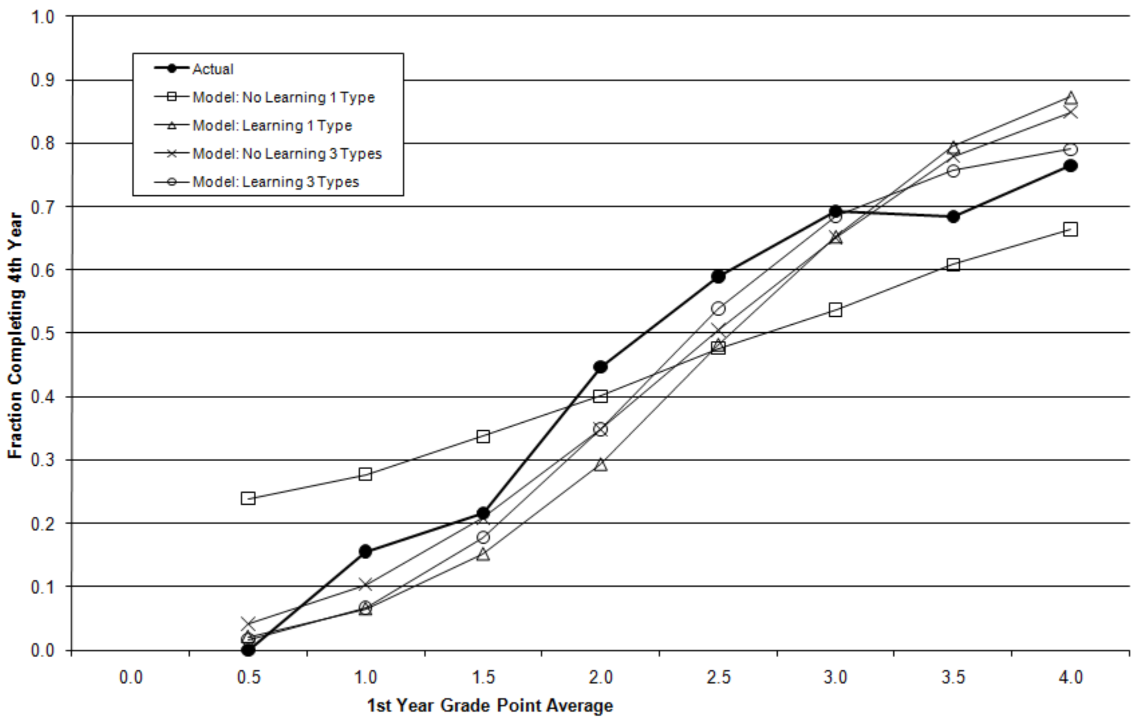
Figure A- IV: Actual vs. Simulated Educational Outcomes, by Family Income





Notes: To generate simulated outcomes, the unobserved type, grade and preference shocks, and choices of each observation is simulated 100 times, assuming that individuals follow the choice models described in the text with parameter values equal to those in specifications (1)-(4) of Table I.

Figure A- V: Actual vs. Simulated Graduation Rate by 1st Year GPA



Notes: To generate simulated outcomes, the unobserved type, grade and preference shocks, and choices of each observation is simulated 100 times, assuming that individuals follow the choice models described in the text with parameter values equal to those in specifications (1)-(4) of Table I.