

Behavioral and Statistical Models of Educational Inequality

Anders Holm and Richard Breen

Tuesday, February 23, 2016

Anders Holm, Department of Sociology, University of Copenhagen and SFI: The Danish
National Centre for Social Research, Copenhagen, Denmark.
email: ah@soc.ku.dk

Richard Breen, Department of Sociology and Nuffield College, University of Oxford, UK.
email: richard.breen@nuffield.ox.ac.uk

Keywords: Rationality, Behavioral Model, Educational Decisions, Identification

Behavioral and Statistical Models of Educational Inequality

Abstract

This paper addresses the question of how students and their families make educational decisions. We describe three types of behavioral model that might underlie decision-making and we show that they have consequences for what decisions are made. Our study thus has policy implications if we wish to encourage students and their families to make better educational choices. We also establish the conditions under which empirical analysis can distinguish between the three sorts of decision-making and we illustrate our arguments using data from the National Educational Longitudinal Study.

Behavioral and Statistical Models of Educational Inequality

How do students and their families make educational decisions, what consequences does the way in which they make decisions have for the decisions they make, and how can social scientists empirically identify different kinds of decision making? To address these questions we focus on the choice of whether to continue in education or leave, and we consider three possible models of decision-making: these are a forward-looking, imperfect foresight model, a myopic decision-making model, and an irrational or non-rational model. We show what consequences each of these kinds of decision making may have for the choice of whether to continue in education or not, and we derive the statistical model that each implies. Our paper therefore contributes to the discussion of how social scientists can empirically differentiate between different types of student behavior and our paper has policy relevance in so far as policy makers would like to encourage students to make the best educational choices (in the sense of making optimal use of available information).

As Breen and Jonsson (2005) noted in their review, there has been a growing interest in sociology in the use of behavioral models to explain aggregate patterns and trends in educational inequality – by which we mean the extent to which educational careers depend on a student's family circumstances. A behavioral model in this sense attempts to capture, in a simplified form, how students and their families (henceforth students for short) make educational decisions. The approach rests on a basis of methodological individualism and frequently, though not necessarily, students are modeled as rational decision makers (as in Breen and Goldthorpe 1997). Educational inequality is thus the aggregated outcome of individual decisions. These decisions are shaped, in part, by

institutional arrangements, which make particular choices more or less likely, depending on individual and familial circumstances. For example, one such institutional arrangement is the cost of a given course of educational study. We should expect that if the cost increased it would lead to a decline in the likelihood that students from poorer families would choose the particular course. This case has been studied (Rosenbaum 2001) but it serves to illustrate an important point: if we want to know how new institutional arrangements are likely to affect educational inequality we need to know how behavior responds to existing institutional arrangements and how behavior might be expected to respond if those arrangements changed.

However, there are many behavioral models to choose from and so, on the one hand, we would like to know not only what implications they have for the choices that students and their families will make but also what statistical model would be appropriate for the estimation of the parameters of a particular behavioral model of decision making and what would be needed to identify those parameters. Only if we can answer this last question can we hope to learn empirically how students and their families do indeed make decisions and what consequences their mode of decision-making may have for the choices they make. At the same time, although a particular statistical model, such as Mare's (1980, 1981) educational transition model, may be used in an analysis because it is believed to be the standard approach, or because it is thought useful for descriptive purposes, it nevertheless carries implicit assumptions about how students make educational decisions, and we should like to know what these assumptions are.

We address these, and other, issues in what we believe is a straightforward and simple way. We consider three general classes of behavioral model of educational decision making, drawing our chief distinction between rational and non-rational decision

making, and then further distinguishing, among the former, forward-looking from myopic decision-making. We define rational decision makers as those who take all relevant and available information into account in making their educational choices, whereas non-rational students do not. In making their educational choices, forward-looking rational students consider the option value of remaining in education but myopic students do not. The major questions we address are, first, do these different models lead to different decisions? For example, would a myopic decision maker terminate her education earlier than a forward-looking decision maker? Secondly, what statistical models do these three types of behavioral model imply? Most importantly, we seek to establish the conditions under which they give rise to different statistical models because, failing this, we will be unable to say anything about whether students are forward-looking or myopic or, indeed, irrational, in how they go about making their educational choices. Both these questions have obvious policy consequences because, if the ways in which decisions are made influence what decisions are taken, we would like to know how decisions are made – but we can only know this if different ways of decision-making have empirically identifiable consequences. If they do then we may be in a position to consider how policy can help students and their families to make better decisions by setting up institutional arrangements that change decision-making in the desired way.

The paper proceeds as follows. We first present the general set-up for our models, explain the three classes of behavioral model, and derive the estimating equations that would allow us to recover their structural parameters. Then we fit the models to US data to determine whether students and their families make rational decisions about whether or not to continue in education at crucial points in their educational career. The

paper concludes with a discussion of the implications of our work for understanding educational decision-making.

Explaining Educational Choices

Within empirical sociology, three major theories seek to explain both stability and change in inequality of educational outcomes with respect to social origins. These are Raftery and Hout's (1993) maximally maintained inequality theory (MMI), Breen and Goldthorpe's (1997) relative risk aversion theory (RRA), and Lucas's (2001) effectively maintained inequality theory (EMI; see Lucas 2009 for an assessment of the testability of these theories). Of these, only RRA has an explicit behavioral model. MMI posits a set of relationships that are deemed to hold at the aggregate level. Inequality in the likelihood of making a given educational transition will remain constant unless forced to change by increasing enrollments that cannot be accommodated by an increase in the share of advantaged students making that transition (Raftery and Hout 1993: 42). Hence inequality declines only when the demand for making that transition reaches saturation levels among the advantaged classes such that any further expansion must come from students from less advantaged classes. It is unclear what behavioral model of educational decision-making would be implied by MMI but one can suppose that, first, prior to its saturation by the advantaged classes, the costs of making the transition in question are too high, and/or the perceived payoff from doing so is too low for students from less advantaged backgrounds; and, secondly, that this calculus of costs and benefits changes, for whatever reason, once saturation by the advantaged classes is reached.

Although the original formulation of EMI (Lucas 2001) is not in the form of a behavioral model, it probably would be possible to develop one because the theory is couched in

terms of the different educational choices that students and parents make. The essential argument of the theory is that families who have the necessary resources will seek advantages in education for their children. Insofar as these are manifested through educational choices they will lead to differences in both the quantity and type of education. That is, there will be both vertical (educational level) and horizontal (educational type or track) differentiation according to social background, and the latter will be especially important within levels of education that are universal or close to universal.

By contrast, Breen and Goldthorpe's (1997) RRA theory is an explicit behavioral model, focusing on how families and students make their educational decisions. It rests on the assumption that all students want to secure a level of education sufficient to allow them to secure a class position at least as good as that of their parents. At points in the educational system below this minimum educational threshold students will continue in education almost come what may, whereas above it they will evaluate the costs and benefits of educational pathways. Because this threshold varies according to social origins, the likelihood of making particular educational transitions varies according to social origins, so giving rise to observed patterns of educational inequality. Students and their families are modeled as rational, forward-looking decision makers who have imperfect foresight. They are assumed to know the probabilistic relationship between all levels and types of educational attainment and the class positions to which they lead, but, as they progress in their educational careers, they update their beliefs about their likelihood of succeeding in particular educational pathways (see also Breen and Yaish 2006).

Cameron and Heckman (1998) present a related model of rational, forward-looking decision-making but this time with perfect foresight. In their model, each student knows, at the outset of her educational career, the returns to all levels of education and the costs she faces in attaining each level. The only uncertainty is on the part of the econometrician who does not know as much as the student.

Behavioral Model

In this paper we draw on Breen and Goldthorpe (1997) and Cameron and Heckman (1998) to develop a straightforward behavioral model in which students decide between educational alternatives on the basis of their perceived costs and benefits. To explain our arguments we consider the relatively simple case of an educational system comprising three levels, which we could think of as compulsory schooling, post-compulsory high school, and college.¹ We label these three levels $j = 0, 1, 2$. This restriction to three levels is made for expositional purposes only and nothing hinges on it. Students can end their educational career after completing any of these levels and they receive a net return of U_j . This includes such things as the expected lifetime earnings given that level of education, the direct and indirect costs of acquiring that level of education, and so on. It can also include non-monetary returns (indeed there is no restriction on what counts as a return or cost as yet). Net returns (and their components) are assumed to be heterogeneous over students (so we should really have subscripted U with i to index students but we do not do this to avoid notational clutter). In this set-up students and their families must make two decisions: whether to leave the

¹ We acknowledge that horizontal differences in educational choice (e.g. STEM versus non-STEM) are important but incorporating this issue into the current model would add a great deal of complication without any payoff in respect of the arguments we want to make about distinguishing different types of choice behavior.

educational system after the completion of compulsory education (level 0) or continue, and, if they continue, whether to leave the educational system after the completion of high school (level 1).²

A myopic behavioral model is one in which students choose to progress from educational level j to level $j+1$ if and only if $U_{j+1} > U_j$. In making this decision they do not consider the payoff to educational levels beyond the $j+1^{th}$, and this means that their choices may be less than optimal because, although the condition $U_{j+1} > U_j$ may not be met, it might be the case that $U_{j+2} > U_j$.

As far as forward-looking decision models are concerned, we should make a distinction between those that involve perfect or imperfect foresight. A model of perfect foresight assumes that the $U_j, j = 0, 1, 2$, are fully known to students and thus they can choose their optimal level of education at the very start of their educational career. To see this, consider the choice they will make at the end of high school (level 1). Here they will choose $\max(U_1, U_2) \equiv V$. At the end of compulsory schooling, they choose $\max(U_0, V) = \max(U_0, U_1, U_2)$. Hence, the choice they make at their first opportunity to choose includes the choice that they will make when (and if) they complete high school. An important corollary of this is that students never gain extra knowledge about U_j as they progress through the educational system because they are fully informed from the outset.

² In our set up, we do not differentiate between entering and completing an educational level. This is again for only for simplicity and without loss of generality and it is immaterial provided we assume that entering a given level requires a belief that it can be completed. In an empirical analysis of choice behavior at a given educational level one might need to take account of selection effects from choosing and completing the previous stage. We will return to this issue in the empirical application.

More plausible is a model of imperfect foresight, and this is the case that concerns us in this paper. We assume that students begin their educational career with some imperfect knowledge of U_j and, as they progress, they acquire information and revise their beliefs about U_j . To capture this we write $U_{ij} = W_{ij} + e_{ij}$. Now we include the student-specific subscript and we decompose U into a part known to the student, W , and an unknown part, e . In concrete terms, we could imagine that e_j captures two kinds of uncertainty: uncertainty about the returns to the j^{th} educational level, and uncertainty about the costs of attaining that level of education. In turn we could think of the costs as comprising not just the direct and indirect (in the shape of foregone opportunities) monetary costs but also the costs that stem from the degree of difficulty that the student experiences. So, learning how she has performed in education is one important element of learning the costs of education.

Given these two kinds of uncertainty, one reasonable assumption is that a student's uncertainty about the j^{th} educational level's costs is completely resolved once she has attained that level but that she remains uncertain about the returns to that level of education. Indeed, she can only learn the true returns to the j^{th} educational level if she leaves the educational system and realizes those returns. Therefore, under this assumption, the student's uncertainty about U_j is reduced but not eliminated once she has completed level j . In the model, this is captured through the student updating her belief about the distribution of e_j . An alternative assumption is that the returns to education are known and so e_j captures only the uncertainty about costs, implying that, when the j^{th} level of education has been completed, e_j becomes known.

In either case, if the e 's are correlated, then knowing e_j will give some more information about e_k , $k > j$. Thus the expected values of later e 's are conditional on the known values

of earlier e 's and these conditional expectations evolve as the student progresses. Note that the assumption here is that students know the joint distribution of all the e 's. This seems plausible: for example, students will learn from observing older students that there is a correlation between performance at earlier and later stages of education.

Forward-looking decision making with imperfect foresight

We begin with the forward-looking, imperfect foresight model because, although the most complicated, it is also the most general: the other models are special cases.

Assuming three educational levels, as before, the model, from the individual perspective, can be written:

$$U_0 = \beta_0 x_0 + e_0$$

$$U_1 = \beta_1 x_1 + e_1$$

$$U_2 = \beta_2 x_2 + e_2$$

We assume that the utility of each level is an additively separable function of a known (by the student) linear predictor, $\beta_j x_j$, and an error. We also assume that $e = (e_0, e_1, e_2)$ is multivariate normal with mean $\mathbf{0}$ and covariance matrix Σ .³

We derive the parameters of this model by starting from the final decision and working back to the first decision. At period 1 after the completion of high school, e_1 has been realized and so the student chooses:

$$\max(U_1, E(U_2 | U_1, U_0))$$

³ The assumption of normality is mainly for expository reasons in order to obtain closed form solutions for some of the expressions in our derivations. Thus, there is nothing essential in what follows that hinges on the parametric assumptions.

Consider now the student's decision after completing stage 0, at which time only e_0 is realized and hence e_1 is unknown:

$$\max(U_0, E(V_1 | U_0))$$

where $V_1 = \max(U_1, E(U_2 | U_1, U_0))$. Choosing the max of U_0 and $E(V_1 | U_0)$ corresponds to the choice when the student has completed compulsory school and so knows, for example, about labor market conditions without further education (e.g. from being offered a unskilled job or hearing about unskilled wages and work conditions from friends in unskilled occupations) but as yet has only incomplete information about the costs and benefits (including wages) of further education.

To proceed we need an analytical solution for $E(V_1 | U_0)$. In appendix A we show that

$$\begin{aligned} E(V_1 | U_0) &= E \max(\beta_1 x_1 + e_1 | e_0, \beta_2 x_2 + E(e_2 | e_1, e_0) | e_0) \\ &= \mu_{1|0} \Phi\left(\frac{\mu_{1|0} - \mu_{E(2|10)0}}{\theta}\right) + \mu_{E(2|10)0} \Phi\left(\frac{\mu_{E(2|10)0} - \mu_{1|0}}{\theta}\right) + \theta \varphi\left(\frac{\mu_{E(2|10)0} - \mu_{1|0}}{\theta}\right) \end{aligned} \quad (1)$$

Here the functions $\Phi(\cdot)$ and $\varphi(\cdot)$ are the standard normal distribution and density

respectively and $\theta = \sqrt{\sigma_{1|0}^2 + \rho_{21|0}^2 \sigma_{1|0}^2 - 2\rho_{21|0} \sigma_{1|0}^2} = \sigma_{1|0} (1 - \rho_{21|0})^2$. The interpretation of $\rho_{21|0}$

and several other parameters of (1) is based on the following. Students hold beliefs about future levels of education, given the information they possess at a particular time. Thus, $\rho_{21|0}$ is the belief the student holds, having completed compulsory education, about the relationship between, on the one hand, the costs and benefits of leaving the educational system after high school, and, on the other, the costs and benefits of leaving after college. Similarly, the parameters $\mu_{1|0}$ and $\mu_{E(2|10)0}$ are the expected values of U_1 at

stage 0 (completion of compulsory schooling) and the expected value of U_2 at stage 1 (completion of high school), evaluated at stage 0. $\mu_{E(2|10)0}$ minus $\mu_{1|0}$ is the expected value of continuing to college rather than leaving after high school, evaluated at the end of compulsory school.

In words, equation (1) tells us that $E(V_1 | U_0)$ -- the expected value of continuing after compulsory school -- is a weighted average of the expected (at that time) costs and benefits of leaving after high school, $\mu_{1|0}$, the expected costs and benefits of completing college, $\mu_{E(2|10)0}$, and an option value of continuing after compulsory school,

$\varphi\left(\frac{\mu_{E(2|10)0} - \mu_{1|0}}{\theta}\right)$ that depends on how different future choice are expected to be in their

costs and benefits. The smaller the difference in expected net benefits from college compared to high school, the smaller the option value of continuing after compulsory education.

We can gain some insight into the choice dynamics implied by (1) by considering some special cases.

(A) If the error terms, e_j are independent ($\sigma_{01} = \sigma_{02} = \sigma_{12} = 0$) and the predictor variables and their effects are the same for each educational level,

$\beta_0 x_0 = \beta_1 x_1 = \beta_2 x_2 = \beta x$, we find that $\theta = \sigma_1$ and $\mu_{1|0} = \mu_{E(2|10)0} = \beta x$, yielding:

$$E(V_1 | U_0) = \beta x \Phi(0) + \beta x \Phi(0) + \sigma_1 \varphi(0) = \beta x + \sigma_1 \varphi(0)$$

In this case, when the student has to evaluate the value of $\beta x + e_0$ (leaving after stage 0), against $\beta x + \sigma_1 \varphi(0)$ (continuing), the students chooses to continue if $\sigma_1 \varphi(0) > e_0$; that is,

even in the case of a positive draw of e_0 (recall that e_0 has a mean of zero) it is not always optimal for the student to drop out because of the option value of further draws (i.e. continuing).

(B) If the draws are positively correlated, continuing is even more likely because this increases the likelihood of further larger draws. For example, assuming that

$$\sigma_0 = \sigma_1 = \sigma_2 = 1 \text{ and also } \sigma_{01} = \sigma_{02} = \sigma_{12} = \rho, (0 < \rho < 1), \text{ and } \beta_0 x_0 = \beta_1 x_1 = \beta_2 x_2 = \beta x,$$

we find:

$$E(V_1 | U_0) = \beta x + \rho e_0 \Phi(0) + \beta x + \rho e_0 \Phi(0) + (1 - \rho)\varphi(0) = \beta x + \rho e_0 + (1 - \rho)\varphi(0)$$

Now, when the student compares continuing against dropping out, the value of continuing is -- in addition to the observed effects, βx -- a weighted sum of expected future utility ρe_0 and the optional value of continuing $(1 - \rho)\varphi(0)$. The larger is ρ (that is, the more information e_0 contains about future values of e_1 and e_2) the more the value of continuing is made up of expected future values of continuing. Conversely, the smaller is ρ the more the value of continuing consists of the option value, with $\rho = 0$ being the limiting case, A, studied above.

The model allows us to determine the effect of forward-looking compared to myopic decision-making. In choosing whether or not to continue to post-compulsory high school education (that is, to make the first transition), in case B forward-looking students compare $E(V_1 | U_0) = \beta x + \rho e_0 + (1 - \rho)\varphi(0)$ against $\beta x + e_0$ whereas the myopic student, who takes no account of the possibility of entering college after high school, compares $\beta x + E(e_1 | e_0)$ against $\beta x + \rho e_0$. Hence, in this case, for otherwise identical students the probability that the forward-looking student makes the first transition is:

$$P(\beta x + \rho e_0 + (1 - \rho)\varphi(0) > \beta x + e_0) = P(e_0 < \varphi(0)) = \Phi(\varphi(0)) \approx 0.60$$

while the probability for the myopic student is:

$$P(\beta x + E(e_1 | e_0) > \beta x + e_0) = P(\rho e_0 > e_0) = \Phi\left(\frac{e_0}{1 - \rho} < 0\right) = 0.5$$

Hence, all else equal, myopic students have lower educational transitions rates because they ignore the option value of continuing in the educational system.

Estimating the model of rational decision making

We have established that students will choose to leave the educational system after stage 0 if $\beta_0 x_0 + e_0 > E(V_1 | e_0)$. But, although the student may have observed e_0 when he or she made the decision, it remains unknown to the sociologist. So, to estimate the parameters of the model, we must form the probability that a student chooses to leave after stage 0. This is $P(\beta_0 x_0 + e_0 > E(V_1 | e_0)) = P(E(V_1 | e_0) - e_0 < \beta_0 x_0)$. When the student reaches stage 1 she has knowledge of both e_0 and e_1 and will leave after stage 1 if

$$\beta_1 x_1 + e_1 > E(U_2 | e_1, e_0) \Rightarrow \beta_1 x_1 + e_1 > \beta_2 x_2 + \rho_{21|0} e_1 + \rho_{20|1} e_0,$$

conditional on having not left after stage 0. In this case both e_0 and e_1 are unobserved by the sociologist and so we form the probability of students continuing after stage 0 and leaving after stage 1:

$$P(\beta_0 x_0 + e_0 < E(V_1 | e_0)) \cdot P(\beta_1 x_1 + e_1 > E(U_2 | e_1, e_0) | \beta_0 x_0 + e_0 < E(V_1 | e_0))$$

This can be rewritten more conveniently as a joint probability

$$P(e_0 - E(V_1 | e_0) < -\beta_0 x_0, (\rho_{21|0} - 1)e_1 + \rho_{20|1} e_0 < \beta_1 x_1 - \beta_2 x_2).$$

For notational convenience, we let a multinomial variable Y take the value 0 if the student leaves at stage 0, 1 if the student leaves at stage 1, and 2 if the student leaves after stage 2.

Because $E(V_1 | e_0)$ is a non-linear function of $\beta_1 x_1, \beta_2 x_2$ and e_0 we cannot get a closed form expression for $P(E(V_1 | e_0) - e_0 < \beta_0 x_0)$ despite the assumption of normality of the e 's.

But via a first order Taylor series approximation of $E(V_1 | e_0)$ around $e_0 = 0$ we get⁴

$$E(V_1 | e_0) \approx F_0(S; \omega) + f(S; \omega)e_0$$

where $S = \{x_1, x_2, x_3\}; \omega = \{\beta_0, \beta_1, \beta_2\}$, and where

$F_0(S; \omega) = E(V_1 | e_0 = 0); f(S; \omega)e_0 = E'(V_1 | e_0 = 0)(e_0 = 0)$. This yields the required probabilities for estimating the forward-looking model as follows:

$$\begin{aligned} P(Y = 0) &\approx P(\beta_0 x_0 + e_0 > F_0(S; \omega) + f(S; \omega)e_0) \\ &= \Phi\left(\frac{\beta_0 x_0 - F_0(S; \omega)}{f(S; \omega) - 1}\right) \\ P(Y = 1) &\approx P\left(e_0 < \frac{F_0(S; \omega) - \beta_0 x_0}{1 - f(S; \omega)}, (\rho_{21|0} - 1)e_1 + \rho_{20|1}e_0 < \beta_1 x_1 - \beta_2 x_2\right) \\ &= \Phi\left(\frac{F_0(S; \omega) - \beta_0 x_0}{\sigma_{e_0}(1 - f(S; \omega))}, \frac{(\beta_1 x_1 - \beta_2 x_2)}{\sqrt{(\rho_{21|0} - 1)^2 \sigma_1^2 + \rho_{20|1}^2 \sigma_0^2 + (\rho_{21|0} - 1)\rho_{20|1}\sigma_{10}}}; \rho\right) \end{aligned}$$

(2)

⁴ The derivations are in appendix B.

with $P(Y = 2) = 1 - P(Y = 0) - P(Y = 1)$ and where $\Phi(\cdot)$; $\Phi(\cdot, \cdot; \rho)$ are, respectively, the standard normal distribution function and the bivariate normal distribution function, with correlation coefficient ρ equal, in this case, to $\rho = \text{corr}(e_0, (\rho_{21|0} - 1)e_1 + \rho_{20|1}e_0)$.

Estimating the model under myopia

Recall that students who are myopic decision makers only pay attention to the expected returns from staying in the educational system in the next period. However, they are rational in that they use all available information to predict the future – so they still use e_0 or e_1 to predict e_2 . Therefore, for them we have:

$$\begin{aligned}
 P(Y = 0) &= P(\beta_0 x_0 + e_0 > \beta_1 x_1 + E(e_1 | e_0)) = \Phi \left(\frac{\beta_0 x_0 - \beta_1 x_1}{\frac{\sigma_{01}}{\sigma_0} - \sigma_0} \right) \\
 P(Y = 1) &= P(\beta_0 x_0 + e_0 < \beta_1 x_1 + E(e_1 | e_0), \beta_1 x_1 + e_1 > \beta_2 x_2 + E(e_2 | e_1, e_0)) \\
 &= P \left(\beta_0 x_0 + e_0 < \beta_1 x_1 + \frac{\sigma_{10}}{\sigma_0^2} e_0, \beta_1 x_1 + e_1 > \beta_2 x_2 + \rho_{21|0} e_1 + \rho_{20|1} e_0 \right) \\
 &= \Phi \left(\frac{\beta_1 x_1 - \beta_0 x_0}{\sigma_{1|0}}, \frac{\beta_1 x_1 - \beta_2 x_2}{\sqrt{\sigma_0^2 \sigma_1^2 - \sigma_{01}^2}}; \rho \right)
 \end{aligned} \tag{3}$$

$$\rho = \text{corr} \left(e_0 \left(1 - \frac{\sigma_{10}}{\sigma_0^2} \right), (\rho_{21|0} - 1)e_1 + \rho_{20|1}e_0 \right).$$

Comparing (2) with (3) we find that myopic and forward-looking behaviors both lead to a reduced form model in which the error terms are correlated across transitions.

Estimating the model under irrationality

Both forward-looking and myopic students have been assumed to be rational in the sense that they use all the available relevant information in making their decisions. One deviation from rationality would be a failure, by students, to use revealed information about the errors to help them form expectations about future errors. Thus, one possible model of irrational decision-making would eliminate updating on prior information, such that the student does not use revealed values of e_j to form expectations of e_k (where $k > j$). This would entail replacing the conditional expectations of the errors with their unconditional expectations. In this case the statistical model becomes:

$$\begin{aligned}
 P(Y = 0) &= P(\beta_0 x_0 + e_0 > \beta_1 x_1 + E(e_1)) = \Phi\left(\frac{\beta_0 x_0 - \beta_1 x_1}{\sigma_0}\right) \\
 P(Y = 1) &= P(\beta_0 x_0 + e_0 < \beta_1 x_1 + E(e_1), \beta_1 x_1 + e_1 > \beta_2 x_2 + E(e_2)) \\
 P(Y = 1) &= P(\beta_0 x_0 + e_0 < \beta_1 x_1, \beta_1 x_1 + e_1 > \beta_2 x_2) \\
 &= \Phi\left(\frac{\beta_1 x_1 - \beta_0 x_0}{\sigma_0}, \frac{\beta_1 x_1 - \beta_2 x_2}{\sigma_1}; \rho = \frac{\sigma_{10}}{\sigma_1 \sigma_0}\right) \quad (4)
 \end{aligned}$$

Yet, even under irrationality, transitions will be correlated across outcomes because the unobservables are correlated across outcomes. Thus, all the three behavioral models – irrational, forward-looking and myopic – imply the same reduced form statistical model. Thus, data alone will not suffice to test empirically which of the three different types of behavior are prevalent in the population, and even a Mare model with correlated errors across transitions, as suggested by Cameron and Heckman (1998) will not tell researchers whether educational choices are rational or not.

Identifying rational forward-looking behavior

It seems that, without further assumptions, there is no way of determining whether students are forward-looking or not (in line with the result in Taber 2000). Indeed, if

we assume that the same set of predictor variables applies to each educational level (that is, $x_0 = x_1 = x_2 = x$) it is obvious that none of the models - for forward-looking, myopic or irrational behavior - is identified.

So how might we try to determine whether students are rational and forward-looking, or rational and myopic, or non-rational? One way, which we adopt here, is to exploit the fact that different x variables might be relevant for choices at different educational levels. There are at least two circumstances in which this might arise. First, there are variables whose value is only revealed after a particular level of education has been completed, and so they can only be relevant after this point. Exam results would be an example. They can influence subsequent choices but not previous ones. Secondly, there may be variables whose values are known to students when they are at level j of the educational system but which are only relevant to later decisions (at level k , where $k > j$). An example is distance to college: this might directly affect the choice of whether or not to proceed to college but would likely have no direct effect on the utility of choosing whether to complete high school, even though its value is known at the time when that decision has to be taken (Stange 2012). But in the forward-looking model it would have an indirect effect on the probability of leaving in the first period because students would take it into account when deciding whether or not to complete high school. In other words, forward looking students' decisions about whether to complete high school will depend, in part, on their beliefs about the costs and benefits of college, and these costs will depend, to some degree, on distance to college. But myopic decision makers will not take these costs and benefits into account. Therefore, in the equation to estimate forward-looking behavior, the expression for $P(Y=0)$ would involve $F_0(S; \omega)$ which is a function of, among other things, x_2 . But x_2 would not enter into the

equivalent expression in the myopic case because myopic students, when deciding whether or not to complete high school, would not consider the potential payoffs to the choice of college education.

In what follows we use these two types of x variables to try to distinguish between our three behavioral models.

Identifying irrational behavior

We have defined rational decision makers as those students who use all available and relevant information in making their educational choices and irrationality is defined as any deviation from this. In terms of our model these deviations could be of two sorts. As noted above, one is a failure to use all the available information to form expectations of the error terms in future possible choices (for example that the student only uses grades from high school to form expectations about the prospect of entering college and ignores results from compulsory education). Another failure would lie in not using all available observed variables to estimate the utility from particular choices. Short of having a strong, and more specific, theory of irrationality, it is impossible to say which variables an irrational student would fail to take into account. Thus, irrational behavior can only be identified *ex post* from a statistical model through zero coefficients on variables that are known, or strongly suspected, to be relevant to a particular educational choice. We demonstrate this argument below when we turn to our examples.

Above we have shown that in the absence of observed explanatory variables that belong only to certain transitions it is very difficult to distinguish forward-looking and myopic behavior from irrational behavior. However, if some of the variables that are revealed

to the student sequentially are observed in the data then it is possible to distinguish rational -- myopic or forward-looking -- from irrational behavior. To do this a small modification is needed to the model. We have to assume that some of the x 's are unobserved, otherwise the model will be deterministic from the perspective of the sociologist. Now we expand our model of the utilities of reaching the different stages of the educational system into:

$$\begin{aligned} U_0 &= \beta_0 x_0 + \gamma_0 e_0 + v_0 \\ U_1 &= \beta_1 x_1 + \gamma_1 e_1 + v_1 \\ U_2 &= \beta_2 x_2 + \gamma_2 e_2 + v_2 \end{aligned}$$

Where $\gamma_j; j = 0, 1, 2$ is the effect of the information sequentially revealed to the student (but now observed by the sociologist) and $v_j; j = 0, 1, 2$ is the effect of variables unobserved by the sociologist but known to the student in all time periods. In this case

$$\begin{aligned} P(Y = 0) &= P(v_0 > E(V_1 | U_0; v_1, v_2) - \beta_0 x_0 - \gamma_0 e_0) \\ P(Y = 1) &= (1 - P(Y = 0)) \cdot P(\beta_1 x_1 + \gamma_1 e - \beta_2 x_2 - \gamma_1 \rho_{2|10} e_1 - \gamma_2 \rho_{20|1} e_0 > v_2 - v_1) \quad (5) \\ P(Y = 2) &= 1 - P(Y = 1) \end{aligned}$$

where

$$\begin{aligned} E(V_1 | U_0) &= E \max(\beta_1 x_1 + v_1 + \gamma_1 e_1 | e_0, \beta_2 x_2 + v_2 + \gamma_2 E(e_2 | e_1, e_0) | e_0) \\ &= \mu_{1|0} \Phi\left(\frac{\mu_{1|0} - \mu_{E(2|10)0}}{\theta}\right) + \mu_{E(2|10)0} \Phi\left(\frac{\mu_{E(2|10)0} - \mu_{1|0}}{\theta}\right) + \theta \varphi\left(\frac{\mu_{E(2|10)0} - \mu_{1|0}}{\theta}\right). \quad (6) \end{aligned}$$

Details of the derivation are in Appendix C. The important thing to notice is that, in (6), the expected values of future choices, $\mu_{1|0}$ and $\mu_{E(2|10)0}$ respectively, now depend on

terms that are unobserved by the sociologist but known to the student (v_0, v_1, v_2). Table 1 makes this clearer.

--- Table 1 here ---

Information sequentially revealed to the student – the e 's – should enter the model sequentially: first period information enters in the first period choice, second period information and first period information enter in the second period choice. Note that we are estimating models that include information known to the student at all times but unobserved by the sociologist (the v 's), and information that is revealed sequentially to the student and observed by the sociologist (the e 's). We do not explicitly consider the situation when some of the information sequentially revealed to the student is not observed by the sociologist. However, this is easily dealt with, as this situation will be a mixture of the model when sequential information is unobserved and the model when it is observed. In this situation, errors are correlated across choices not only due to unobservables (the v 's) but also, potentially, because of unobserved sequentially revealed information. But the important element of the model with observed sequential revealed information is that some sequentially revealed information is observed and available to the sociologist for analysis.

Estimating the model based on (6) is difficult. Assuming joint normality of v_0, v_1, v_2 does not lead to a closed form solution of the probabilities in (5). For example, the probability of $Y = 0$ depends on unobservables from all three transitions and it depends on v_1 and v_2 in a non-linear way. So $P(Y = 1)$ is not a probit even if v_0 is normal.

Furthermore, we cannot estimate $P(Y = 1)$ and $P(Y = 2)$ separately from $P(Y = 0)$ as this would ignore selection effects from making the first transition, even in the case where the unobservables are independent of each other. Selection effects appear because making the first transition depends on unobservables that appear not only in the equation for the first transition but also in the equations for subsequent transitions.

In sum, when variables are revealed sequentially we have two possible tests of rational behaviour. First, sequentially revealed variables should all enter later transitions: for example, grades from previous educational levels should enter the choice of current transitions. Secondly, if factors unobserved by the sociologist are determining the utility, choices should be correlated if individuals are choosing rationally.

Identifying myopic behavior

The existence and availability of one or more variables that should directly affect only later transitions provides a test for myopic decision-making. As explained earlier, they should be relevant for the earlier decisions of forward looking students in so far as they have an indirect effect on the probability of continuing in the educational system. But they should play no role in the earlier decisions of myopic students.

Empirical examples

We illustrate how estimating models where information is sequentially revealed enables us to infer whether students are rational. To this end, we use data from the National Educational Longitudinal Study of 1988. NELS is a nationally representative sample of eighth-graders, first surveyed in the spring of 1988. They were resurveyed through four follow-ups in 1990, 1992, 1994, and 2000. NELS provides detailed data on

student transitions, parental background characteristics, and performance by students at different points in their educational career.

Description of data

We provide summary statistics of the data used in our empirical example in Table 2. We use information on father's and mother's years of education, family income, family SES and dummy variables for whether the family is broken, whether the student is male, and whether the student's family lives in an urban or suburban area. The dependent variables are whether the student has completed high school (transition 1) and whether he or she entered a four year college education (transition 2).⁵

We constructed composite measures of cognitive outcomes for each student at different stages in the educational system. These use tests taken by the student which, we assume, provide her with knowledge about her cognitive skills. Our measures use the first principal component of math and reading test scores after 12th grade, after 10th grade (which we consider transition 0 because this is the end of compulsory schooling) and after 8th grade, respectively. All cognitive outcomes have been standardized to have mean zero and variance one before the principal component extraction.

We use composite measures to avoid the multicollinearity that would result from the high correlations between the test scores. Nevertheless, reducing the number of tests included in the estimations limits our ability to detect effects of lagged test scores on future educational decisions. Thus, when we find significant effects of lagged composite

⁵ The realizations of these two events, in most cases, come quite close together, but they reflect two different decisions. The event that a student completes high school is an aggregated effect of time and effort devoted to this accomplishment through the high school years, while the decision to enter college is taken in the light of information on ability acquired through high school and the expectation of the cost and benefits of completing college. To keep the model simple we do not address high school dropout.

test scores this should be taken as “lower bound” evidence that students use lagged information on test scores to make future educational decisions.

----- Table 2 here -----

Model estimation and results

The aim of the analysis is to investigate whether students use the full set of cognitive outcomes revealed sequentially during transitions through the educational system. If outcomes are serially correlated, students should use the full set of revealed outcomes up to the current transitions. If students are irrational, in the sense outlined in this paper, they will only use the information from the last period and ignore previous periods. We also have a variable that should only enter the utility function in later but not in earlier transitions: this enables us to distinguish between myopic and forward-looking behaviour. In other words, we can non-parametrically identify $E(V_1 | U_0; v_1, v_2)$ because we have a predictor variable – namely distance to the nearest four-year college -- that should only appear in the utility of the later periods. If this were significant in a regression for earlier educational choices it would illustrate forward-looking behaviour by the student. Our distance to college variable is measured as the distance between the residence of the student’s family and the nearest four-year college,⁶ calculated using restricted NELS data. The procedure is very similar to that used by Stange (2012).

⁶ We use the residence of the student when living with his parents and not residence after potential reallocation to a college because it is the cost of moving to a college that should affect to utility of entering college when completing high school.

Rational students should use the full set of sequentially revealed information only if it provides more information than they would otherwise have in the current period. To see if this is the case we begin with an analysis of the auto-correlation of cognitive outcomes across transitions. More specifically we examine whether cognitive outcomes at period one (used to infer the utility of transition two) are predicted by cognitive outcomes at previous periods or whether all the information in cognitive outcomes in period two is contained in cognitive outcomes at period one.

We also have to allow for the fact that unobservables are correlated across transitions, generating selection into subsequent transitions. Therefore, we estimate models that assume uncorrelated errors (standard probit models) and models that allow for selection on unobservables across transitions (the Heckman (1979) selection model for binary outcomes).⁷ Note that we cannot compare coefficients across the different model specifications because the models have different residual variation, depending on whether they allow for selection on unobservables (see Winship and Mare 1983; Karlson, Holm and Breen 2012).

Results

In Table 3 we show results from the regression of lagged cognitive outcomes on cognitive outcomes after transition one (completing high school). They show that one period lagged cognitive outcomes (at the end of 10th grade) are highly predictive of current cognitive outcomes – but so too are earlier period outcomes (scores at the end of 8th grade). Hence, rational students should use information about cognitive outcomes

⁷The Heckman selection model assumes joint normality of the errors. As mentioned above this is at best an approximation to the true model with sequentially revealed information. Hence, as a robustness check to test the empirical validity of the normal assumption, we have also estimated mixed logit models that allow for a non-parametric distribution of the errors (see Breen and Jonsson 2002 and Holm et al. 2013 for applications of the mixed logit model in the analysis of educational transitions). Results are available upon request.

lagged for more than one period when inferring how cognitive outcomes affect future educational choices.⁸

--- Table 3 here ----

Having established that rational students should use information on cognitive outcomes lagged for more than one period when making decisions about educational transitions, we now test whether they actually do this.

--- Table 4 here ----

Table 4 shows the results from two statistical models fitted to three groups. The two models are a probit and a probit allowing for selection into high school. They were fitted to all the sample observations (columns 1 and 2), to those students whose parents had an SES below the mean (columns 3 and 4), and to those with parents above the SES mean (columns 5 and 6). Furthermore, the upper rows of the table report the coefficient estimates for the transition into a four-year college, based only on those who complete high school, while the lower rows report those for the completion of High School among all students in the sample. Note that for the probit models allowing for selection we also report the correlation coefficient between the error terms of the unobservables across the two transitions. This is the estimated regression coefficient of the inverse mills ratio of selection into college from high school.

For the transition to a four-year college, cognitive scores after 12th grade are highly significant predictors. This finding is consistent across model specifications. Lagged

⁸One may argue that the correlation across test score over time is due to student fixed effects and not a true lagged effect. To test this we used the Anderson and Hsiao (1982), and Arrelano-Bond (1991) estimators, using the level of test scores as instruments. Both estimators find evidence of large one period lagged effects of test scores (with only three periods, this analysis does not allow us to test for more than one period lags).

cognitive outcomes from the end of 10th grade are also important. We conclude, therefore, that students are rational in the sense that they use both current and lagged information about cognitive ability when deciding whether or not to enrol in a four-year college. The distance to the nearest four-year college has a significant, negative coefficient.

In the equation predicting completion of high school, we find that cognitive outcomes at the end of 10th grade are powerful predictors whereas earlier cognitive outcomes are not. One interpretation of this is that, even though students are rational, the probability of failing to complete high school is very low once results at 10th grade are sufficiently high, excluding the need for the student to consider further information. More importantly, we find that the distance to a four-year college has significant negative affects on the probability of completing high school. Thus, the likelihood of completing high school is affected by the future utility of entering a four-year college, which itself depends on the distance to the nearest four-year college. This strongly indicates that students are forward looking. In the model that allows for selection bias, we find a positive correlation between errors across transitions, indicating that students who are more likely to complete high school are also more likely to enter a four-year college, over and above effects captured by our observed explanatory variables.

Overall, our results indicate that students behave rationally in the sense that they use both current and lagged information about cognitive outcomes and that they are forward looking when making future educational decisions. However, we might expect that rational behaviour is not common throughout all groups in society. For example, in forming their expectations of future performance, students whose parents are of lower socio-economic status might be less likely to believe that current performance is a

reliable predictor of future performance. To see whether this is indeed the case, we split our sample according to whether the parents are below or above the SES mean, and estimate the same model as before for the two groups. The results for these analysis are also shown in Table 4, columns 3 to 6.

The results for students with parental SES below the mean are very similar, in all respects, to the results for the whole sample. For students with parental SES above the mean, the one area of difference lies in the effect of distance to the nearest four-year college. In the probit this is not significant for either the choice of entry to college nor for high school completion, although it is significant, and in the predicted direction, in the probit with selection. Interpreting the non-significant effect in the model for completing high school as showing that higher SES families are not forward looking would seem to be incorrect, given the variable's non-significant status in the model for college entry. Rather, it seems that, for these families, distance to the nearest four-year college does not act as a constraint on whether or not a student should enter college. If this is really the case it implies that we require a different instrument to determine whether or not this group engages in forward-looking decision making.

Conclusions

In the empirical analysis of this paper we found evidence that students are both rational and forward-looking when making educational decisions. This means that educational choices are linked and that the attractiveness of further educational opportunities - here four-year college - will affect the likelihood of completing high school. Thus, students for whom the prospects of obtaining a college degree are low, or who believe that the returns to a college degree, compared to graduating high school, are meagre, are less likely to complete high school. High school dropout can therefore, at least partly, be

explained by poor educational future prospects: students drop out of high school, not because being in high school itself is unattractive but because college, to which completing high school is the pathway, does not seem attractive or feasible enough.

Our paper's main contribution has been to discuss how statistical models for educational transitions relate to behavioral models, and what information is required to distinguish between rational, myopic and forward-looking behaviour. It is important to do this because, as we showed, how students and their families make decisions is consequential for the choices they make. For example, myopic students, even though otherwise rational in their decision making, will, all else equal, be less likely to progress to higher educational levels than forward-looking students. We also showed that data alone cannot determine whether or not students and their families make rational educational decisions. All the behavioural models we considered in this paper give rise to statistical models in which the errors for different transitions are correlated: in short, they all lead to the same statistical model. This is a problem: for example, if we care about policies to help students make better decisions we first need to find out how they make decisions. We have shown that this cannot be done on the basis of statistical models alone. As our analysis demonstrated, to distinguish one type of decision making from others we require testable implications, which we can derive only on the basis of assumptions that need to be justified on the basis of an implicit or explicit theory. In our case we assumed that distance to college has no bearing on current utility when entering or completing high school, so that myopic decision makers would not take distance into account in the decision of whether to complete high school, whereas forward-looking decision makers would. Other testable implications might also be derived and, indeed, for students from high SES backgrounds for whom distance to

college appears to be not relevant for the choice of college entry, other testable implications would be needed to determine whether or not they are rational decision-makers . The general point with which we conclude, however, is that if sociologists want to go beyond descriptive models of educational career paths they need to use explicit theories of educational decision-making from which testable implications can be derived.

REFERENCES

- Anderson, T.W., and Cheng Hsiao (1982), "Formulation and Estimation of Dynamic Models Using Panel Data," *Journal of Econometrics*, 18, 47-82.
- Arellano, Manuel and Stephen Bond, (1991), "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations," *Review of Economic Studies*, 58, 277-297.
- Breen, Richard and Jan O. Jonsson (2005) "Inequality of Opportunity in Comparative Perspective: Recent Research on Educational Attainment and Social Mobility", *Annual Review of Sociology* Vol. 31, 2005, pp. 223-244.
- Breen, Richard and Meir Yaish (2005), "Testing the Breen-Goldthorpe model of educational decision making" (with Meir Yaish), in Stephen L. Morgan, David B. Grusky, and Gary S. Fields (eds.) *Mobility and Inequality: Frontiers of Research in Sociology and Economics*, Stanford: Stanford University Press, 2006, pp. 232-58.
- Breen, Richard and Jan O. Jonsson (2000) "A Multinomial Transition Model for Analyzing Educational Careers" (with Jan O. Jonsson) *American Sociological Review*, Vol. 65, No. 5, 2000, pp.754-72.
- Breen, Richard and John H. Goldthorpe (1997) "Explaining Educational Differentials: Towards a Formal Rational Action Theory" (with John H. Goldthorpe), *Rationality and Society*, Vol. 9, No. 3, 1997, pp. 275-305. Reprinted
- Cameron, Stephen V. and James J. Heckman (1998) "Life Cycle Schooling and Dynamic Selection Bias: Models and Evidence for Five Cohorts of American Males", *Journal of Political Economy*, 106(2), 262- 311.
- Holm, Anders, Kristian Karlson, Mads Jæger, David Reimer (2013) Incomplete Equalization: The Effect of Tracking in Secondary Education on Educational Inequality. *Social Science Research*, 42(6), 1431-1142.
- Karlson, Kristian B. , Anders Holm and Richard Breen (2012), Comparing Regression Coefficients Between Same-sample Nested Models Using Logit and Probit: A New Method *Sociological Methodology* 42, 286 - 313.
- Ker, Alan B. (2001) On the Maximum of Bivariate Normal Random Variables, *Extremes* 4: 185-190.
- Lucas, Samuel R. 2001. "Effectively Maintained Inequality: Education Transitions, Track Mobility, and Social Background Effects." *American Journal of Sociology* 106: 1642-1690.
- Lucas, Samuel R. 2009. "Stratification Theory, Socioeconomic Background, and Educational Attainment: A Formal Analysis." *Rationality & Society* 21: 459-511.

- Mare, R. D. 1981. "Change and Stability in Educational Stratification." *American Sociological Review* 46: 72-87.
- Mare, R. D. 1980. "Social Background and School Continuation Decisions." *Journal of the American Statistical Association* 75: 295-305.
- Raftery, Adrian E., and Michael Hout. "Maximally Maintained Inequality: Educational Stratification in Ireland." *Sociology of Education* 65 (January): 41-62.
- Rosenbaum, James. 2001. *Beyond College for All*. New York: Russell Sage Foundation.
- Stange, Kevin (2012). An Empirical Investigation of the Option Value of College Enrollment. *American Economic Journal: Applied Economics* 4 (1): 49-84.
- Taber, R. Christopher (2000) "Semiparametric identification and heterogeneity in discrete choice dynamic programming models." *Journal of Econometrics*, 96(2): 201-29.
- Winship, C., and R. D. Mare. 1983. "Structural Equations and Path Analysis for Discrete Data." *American Journal of Sociology*, 88: 54-110.

Appendix A: Derivation of a closed form expression for $E(V_1 | U_0)$ in equation (1)

We want to derive a closed form expression for $E(V_1 | U_0)$ where

$V_1 = \max(U_1, E(U_2 | U_1, U_0))$. To do this we need to find

$E(U_2 | U_1, U_0)$ and $E(V_1 | U_0) = E(\max(U_1, E(U_2 | U_1, U_0)) | U_0)$. Now:

$$\begin{aligned}
 E(U_2 | U_1, U_0) &= \beta_2 x_2 + E(e_2 | e_1, e_0) \\
 &= \beta_2 x_2 + (\sigma_{02} \quad \sigma_{12}) \begin{pmatrix} \sigma_0^2 & \sigma_{01} \\ \sigma_{01} & \sigma_1^2 \end{pmatrix}^{-1} \begin{pmatrix} e_0 \\ e_1 \end{pmatrix} \\
 &= \beta_2 x_2 + \frac{(\sigma_0^2 \sigma_{12} - \sigma_{02} \sigma_{01}) e_1 + (\sigma_1^2 \sigma_{02} - \sigma_{12} \sigma_{01}) e_0}{\sigma_0^2 \sigma_1^2 - \sigma_{01}^2} \\
 &= \beta_2 x_2 + \rho_{21|0} e_1 + \rho_{20|1} e_0
 \end{aligned}$$

We can then work out:

$$V_1 = \max(U_1, E(U_2 | U_1, U_0)) = \max(\beta_1 x_1 + e_1, \beta_2 x_2 + E(e_2 | e_1, e_0))$$

\Rightarrow

$$E(V_1 | U_0) = E \max(U_1 | U_0, E(U_2 | U_1, U_0) | U_0) = E \max(\beta_1 x_1 + e_1 | e_0, \beta_2 x_2 + E(e_2 | e_1, e_0) | e_0)$$

This is the expected maximum of two normal variables, the first is:

$$U_1 | U_0 = \beta_1 x_1 + e_1 | e_0 = \beta_1 x_1 + \left(\varepsilon_1 + \frac{\sigma_{10}}{\sigma_0^2} e_0 \right);$$

$$\varepsilon_1 : N\left(0, \sigma_1^2 - \frac{\sigma_{01}^2}{\sigma_0^2}\right)$$

with mean:

$$\mu_{1|0} = \beta_1 x_1 + \frac{\sigma_{10}}{\sigma_0^2} e_0$$

and variance

$$\text{Var}(U_1 | U_0) = \left(\sigma_1^2 - \frac{\sigma_{01}^2}{\sigma_0^2} \right) = \sigma_{1|0}^2$$

and the second is:

$$\begin{aligned} E(U_2 | U_1, U_0) | U_0 &= \beta_2 x_2 + E(e_2 | e_1, e_0) | e_0 \\ &= \beta_2 x_2 + (\rho_{21|0} e_1 | e_0 + \rho_{20|1} e_0) \\ &= \beta_2 x_2 + \left\{ \rho_{21|0} \left(\varepsilon_1 + \frac{\sigma_{10}}{\sigma_0^2} e_0 \right) + \rho_{20|1} e_0 \right\} \\ &= \beta_2 x_2 + \rho_{21|0} \varepsilon_1 + \left(\frac{\rho_{21|0} \sigma_{10}}{\sigma_0^2} + \rho_{20|1} \right) e_0 \end{aligned}$$

with mean:

$$\begin{aligned} \mu_{E(2|10)0} = E(E(U_2 | U_1, U_0) | U_0) &= E \left\{ \beta_2 x_2 + \rho_{21|0} \varepsilon_1 + \left(\frac{\rho_{21|0} \sigma_{10}}{\sigma_0^2} + \rho_{20|1} \right) e_0 \right\} \\ &= \beta_2 x_2 + \left(\frac{\rho_{21|0} \sigma_{10}}{\sigma_0^2} + \rho_{20|1} \right) e_0 \end{aligned}$$

and variance

$$\text{Var}(E(U_2 | U_1, U_0) | U_0) = \rho_{21|0}^2 \sigma_{1|0}^2$$

Finally we also need:

$$\begin{aligned}
& \text{cov}(U_1 | U_0, E(U_2 | U_1, U_0) | U_0) \\
&= \text{cov}\left(\beta_1 x_1 + \left(\varepsilon_1 + \frac{\sigma_{10}}{\sigma_0^2} e_0\right), \beta_2 x_2 + \rho_{21|0} \varepsilon_1 + \left(\frac{\rho_{21|0} \sigma_{10}}{\sigma_0^2} + \rho_{20|1}\right) e_0\right) \\
&= \rho_{21|0} \sigma_{1|0}^2
\end{aligned}$$

Note that $\mu_{1|0} - \mu_{E(2|10)0} = \beta_1 x_1 - \beta_2 x_2 + \left\{ \frac{\sigma_{10}}{\sigma_0^2} - \left(\frac{\rho_{21|0} \sigma_{10}}{\sigma_0^2} + \rho_{20|1} \right) \right\} e_0$.

Applying a result from Ker (2001) we find that:

$$\begin{aligned}
E(V_1 | U_0) &= E \max(\beta_1 x_1 + e_1 | e_0, \beta_2 x_2 + E(e_2 | e_1, e_0) | e_0) \\
&= \mu_{1|0} \Phi\left(\frac{\mu_{1|0} - \mu_{E(2|10)0}}{\theta}\right) + \mu_{E(2|10)0} \Phi\left(\frac{\mu_{E(2|10)0} - \mu_{1|0}}{\theta}\right) + \theta \varphi\left(\frac{\mu_{E(2|10)0} - \mu_{1|0}}{\theta}\right)
\end{aligned}$$

where $\theta = \sqrt{\sigma_{1|0}^2 + \rho_{21|0}^2 \sigma_{1|0}^2 - 2\rho_{21|0} \sigma_{1|0}^2} = \sigma_{1|0} (1 - \rho_{21|0})^2$

Appendix B. First order Taylor expansion of $E(V_1 | U_0)$

The first order Taylor expansion of $E(V_1 | U_0)$ is

$E(V_1 | e_0 = 0) + E'(V_1 | e_0 = 0)(e_0 - 0)$ where

$$\begin{aligned}
 E(V_1 | e_0 = 0) &= \beta_1 x_1 \Phi\left(\frac{\beta_1 x_1 - \beta_2 x_2}{\theta}\right) + \beta_2 x_2 \Phi\left(\frac{\beta_2 x_2 - \beta_1 x_1}{\theta}\right) + \theta \varphi\left(\frac{\beta_1 x_1 - \beta_2 x_2}{\theta}\right) \\
 E'(V_1) &= \left[\begin{aligned} &\mu'_{1|0} \Phi\left(\frac{\mu_{1|0} - \mu_{E(2|1)0}}{\theta}\right) + \mu_{1|0} \Phi'\left(\frac{\mu_{1|0} - \mu_{E(2|1)0}}{\theta}\right) \\ &+ \mu'_{E(2|1)0} \Phi\left(\frac{\mu_{E(2|1)0} - \mu_{1|0}}{\theta}\right) + \mu_{E(2|1)0} \Phi'\left(\frac{\mu_{E(2|1)0} - \mu_{1|0}}{\theta}\right) + \theta \varphi'\left(\frac{\mu_{E(2|1)0} - \mu_{1|0}}{\theta}\right) \end{aligned} \right] \\
 &= \left[\begin{aligned} &\frac{\sigma_{10}}{\sigma_0^2} \Phi\left(\frac{\mu_{1|0} - \mu_{E(2|1)0}}{\theta}\right) + \mu_{1|0} \left\{ \frac{\sigma_{10}}{\sigma_0^2} - \left(\frac{\rho_{21|0} \sigma_{10}}{\sigma_0^2} + \rho_{20|1} \right) \right\} / \theta \varphi\left(\frac{\mu_{1|0} - \mu_{E(2|1)0}}{\theta}\right) \\ &+ \left(\frac{\rho_{21|0} \sigma_{10}}{\sigma_0^2} + \rho_{20|1} \right) \Phi\left(\frac{\mu_{E(2|1)0} - \mu_{1|0}}{\theta}\right) + \mu_{E(2|1)0} \left\{ \frac{\sigma_{10}}{\sigma_0^2} - \left(\frac{\rho_{21|0} \sigma_{10}}{\sigma_0^2} + \rho_{20|1} \right) \right\} / \theta \varphi\left(\frac{\mu_{E(2|1)0} - \mu_{1|0}}{\theta}\right) \\ &+ \left\{ \frac{\sigma_{10}}{\sigma_0^2} - \left(\frac{\rho_{21|0} \sigma_{10}}{\sigma_0^2} + \rho_{20|1} \right) \right\} / \theta \varphi'\left(\frac{\mu_{E(2|1)0} - \mu_{1|0}}{\theta}\right) \end{aligned} \right] \\
 E'(V_1 | e_0 = 0) &= \left[\begin{aligned} &\frac{\sigma_{10}}{\sigma_0^2} \Phi\left(\frac{\beta_1 x_1 - \beta_2 x_2}{\theta}\right) + \beta_1 x_1 \left\{ \frac{\sigma_{10}}{\sigma_0^2} - \left(\frac{\rho_{21|0} \sigma_{10}}{\sigma_0^2} + \rho_{20|1} \right) \right\} / \theta \varphi\left(\frac{\beta_1 x_1 - \beta_2 x_2}{\theta}\right) \\ &+ \left(\frac{\rho_{21|0} \sigma_{10}}{\sigma_0^2} + \rho_{20|1} \right) \Phi\left(\frac{\beta_2 x_2 - \beta_1 x_1}{\theta}\right) + \beta_2 x_2 \left\{ \frac{\sigma_{10}}{\sigma_0^2} - \left(\frac{\rho_{21|0} \sigma_{10}}{\sigma_0^2} + \rho_{20|1} \right) \right\} / \theta \varphi\left(\frac{\beta_2 x_2 - \beta_1 x_1}{\theta}\right) \\ &+ \left\{ \frac{\sigma_{10}}{\sigma_0^2} - \left(\frac{\rho_{21|0} \sigma_{10}}{\sigma_0^2} + \rho_{20|1} \right) \right\} / \theta \varphi'\left(\frac{\beta_1 x_1 - \beta_2 x_2}{\theta}\right) \end{aligned} \right]
 \end{aligned}$$

Appendix C: Deriving the model with sequentially revealed information

$$\mu_{1|0} = \beta_1 x_1 + v_1 + \gamma_1 \frac{\sigma_{10}}{\sigma_0^2} e_0,$$

$$\begin{aligned} \mu_{E(2|10)0} &= E(E(U_2 | U_1, U_0) | U_0) = E \left\{ \beta_2 x_2 + v_2 + \rho_{21|0} \varepsilon_1 + \left(\frac{\rho_{21|0} \sigma_{10}}{\sigma_0^2} + \rho_{20|1} \right) e_0 \right\} \\ &= \beta_2 x_2 + v_2 + \gamma_2 \left(\frac{\rho_{21|0} \sigma_{10}}{\sigma_0^2} + \rho_{20|1} \right) e_0 \end{aligned}$$

So that

$$\mu_{1|0} - \mu_{E(2|10)0} = \beta_1 x_1 - \beta_2 x_2 + v_1 - v_2 + \left\{ (\gamma_1 - \gamma_2) \frac{\sigma_{10}}{\sigma_0^2} - \gamma_2 \left(\frac{\rho_{21|0} \sigma_{10}}{\sigma_0^2} + \rho_{20|1} \right) \right\} e_0$$

And hence

$$\begin{aligned} E(V_1 | U_0) &= E \max(\beta_1 x_1 + v_1 + \gamma_1 e_1 | e_0, \beta_2 x_2 + v_2 + \gamma_2 E(e_2 | e_1, e_0) | e_0) \\ &= \mu_{1|0} \Phi \left(\frac{\mu_{1|0} - \mu_{E(2|10)0}}{\theta} \right) + \mu_{E(2|10)0} \Phi \left(\frac{\mu_{E(2|10)0} - \mu_{1|0}}{\theta} \right) + \theta \varphi \left(\frac{\mu_{E(2|10)0} - \mu_{1|0}}{\theta} \right) \end{aligned}$$

But where $\mu_{1|0}$ and $\mu_{E(2|10)0}$ now depends on unobservables (by the sociologist), v_0 and v_1, v_2 respectively. As a consequence we change the notation of the expected value of continuing after stage 0: $E(V_1 | U_0; v_1, v_2)$ to emphasize that this now depends on unobservables.

Tables:

Table 1: Information available to the student and the sociologist

	Known to the student at period 0	Revealed to the student at period 1	Revealed to the student at period 2
Observed by the sociologist	$x_0, x_1, x_2,$ e_0	e_1	e_2
Unobserved by the sociologist	v_0, v_1, v_2		

Table 2. Descriptive statistics, NELS data

	Mean	Std. dev.
Score composite at age 16/12 th grade test score	0	1
Score composite at age 11/10 th grade test score	0	1
Score composite at age 7/end of 8 th grade	0	1
Ln(distance to four year college – in radians)	-2.341	1.224
Fathers years of education	13.324	1.843
Mothers years of education	13.119	1.664
SES	0	1
Family income (thousands of dollars)	10.16	2.354
Broken family	0.229	-
Male	0.475	-
Completed High school	0.949	-
Entered Higher Education/Four year college	0.359	-
Non-rural	0.650	
N	5948	

Table 3. Regression of score composite at age 16/12th grade

Score composite at end of 10 th grade	0.749*** (0.012)
Score composite at end of 8 th grade	0.152*** (0.012)
Broken family	-0.010 (0.017)
Male	0.018 (0.013)
Family income (NELS/1000)	0.010** (0.003)
Mothers years of education (NELS/10)	-0.002 (0.005)
Fathers years of education (NELS/10)	0.023*** (0.005)
R ²	0.835
N	6260

***: $p < 0.001$; **: $p < 0.01$; *: $p < 0.05$.

Table 4. Probit models of educational choice

	All sample		Parental SES below mean.		Parental SES higher than mean.	
	(1)	(2)	(3)	(4)	(5)	(6)
	<i>Probit</i>	<i>Probit with selection into High School</i>	<i>Probit</i>	<i>Probit with selection into High School</i>	<i>Probit</i>	<i>Probit with selection into High School</i>
<i>Entering four year college</i>						
Score composite at end of 12 th grade	0.388*** (0.035)	0.319*** (0.038)	0.439*** (0.048)	0.344*** (0.048)	0.171** (0.070)	0.399*** (0.049)
Score composite at end of 10 th grade	0.173*** (0.035)	0.213*** (0.044)	0.089* (0.047)	0.157*** (0.046)	0.347*** (0.069)	0.157*** (0.046)
Ln(distance to four year college)	-0.089*** (0.018)	-0.090*** (0.018)	-0.102*** (0.023)	- 0.113*** (0.022)	-0.064 (0.035)	-0.133*** (0.022)
<i>Completing high school</i>						
Score composite at end of 10 th grade	0.361*** (0.038)	0.394*** (0.040)	0.326*** (0.047)	0.355*** (0.048)	0.427*** (0.066)	0.355*** (0.048)
Score composite at end of 8 th grade	-0.020 (0.040)	-0.024 (0.042)	0.036 (0.051)	-0.032 (0.051)	-0.104 (0.066)	0.032* (0.051)
Ln(distance to four year college)	-0.041*** (0.018)	-0.045*** (0.019)	-0.053** (0.027)	-0.058* (0.028)	0.027 (0.039)	-0.058* (0.027)
Correlation between unobservables		0.664*** (0.250)	-	0.778*** (0.536)	-	0.815* (0.404)
n	5833	5833	3854	3854	1979	1979

***: $p < 0.001$; **: $p < 0.01$; *: $p < 0.05$; #: $p < 0.10$. In the models for entering four-year college and for completing high school, we control for family income, SES, mother's and father's years of education, student gender, and whether the family is broken.