Stepping Stone and Option Value in a Model of Postsecondary Education*

Nicholas Trachter
EIEF

July 4, 2012

Abstract

A stepping stone arises in risky environments with learning and transferable capital. An example is the role played by academic 2-year colleges in postsecondary education: students, as they learn about the uncertain educational outcomes, can drop out or transfer up to harder and more rewarding schools, carrying a fraction of the accumulated human capital. A theory of education is built and contrasted empirically to find i) option value explains a large part of returns to enrollment, ii) enrollment in academic 2-year colleges is driven by the option to transfer up, and iii) the value of the stepping stone is small.

JEL classification:

Keywords:

*The online appendix of this paper can be found at my website: https://sites.google.com/site/trachter/. I thank Fernando Alvarez for his guidance and comments. I also benefited from comments by Gadi Barlevy, Marco Bassetto, Christian Broda, Jeremy Fox, Willie Fuchs, Aspen Gorry, James Heckman, Mariano Lanfranconi, Francesco Lippi, Rodolfo Manuelli, Derek Neal, Juan Pablo Nicolini, Ezra Oberfield, Rob Shimer, Marcelo Veracierto, Andy Zuppann and seminar participants at EIEF, Federal Reserve Bank at Chicago, EERTEC 2010, University of Chicago, Universitat Pompeu Fabra, UTDT, and conference participants at the 19th CEPR ESSIM, and LACEA/LAMES 2008 at Rio de Janeiro. All errors are my own.
1 Introduction

Option value is a feature of environments where an investment, or a decision, needs to be made under uncertainty, as in Pindyck (1982), Abel (1983) or Dixit (1989). When there is learning and capital is at least partially transferable, a stepping stone arises as in Jovanovic and Nyarko (1997). Notably, academic 2-year colleges constitute one of these stepping stones; as they provide a cheaper learning technology, they act as a safer path towards more rewarding and demanding environments, namely, 4-year colleges. This paper, by the means of counterfactual analysis, quantifies this stepping stone and the different sources of option value using a micro-estimated model of education.

Evidence from the National Longitudinal Study of the High School Class of 1972 (NLS-72) is compelling both in terms of transitions and returns. Dropout rates are high in both academic 2- and 4-year colleges. Graduation rates are 6 times higher in 4-year colleges and transfer rates are mostly one-sided from academic 2- to 4-year colleges. For students enrolling in academic 2-year colleges, the return to graduation is negligible while that to transferring is large. Academic 2-year colleges are cheaper, easier and some fraction of the accumulated credits can be transferred to a 4-year college. Students sort across institution types with respect to how uncertain is their likelihood of success.\footnote{Similar to the likelihood of investing by firms facing uncertain demand as in Pindyck and Solimano (1993), Caballero and Pindyck (1996), or Leahy and Whited (1996).} For students with low expectations, academic 2-year colleges are an ideal training ground. In the educational ladder, 4-year colleges play the role of the step above academic 2-year colleges.

This paper produces a theory of postsecondary education where academic 2-year colleges act a stepping stone following from the availability of dropout and transfer options in both steps of the ladder. An estimated version of the model is then used to evaluate the value-added of each option to each step, how substitutable the steps are, and how does the provision of full insurance affects total enrollment and its composition across types of colleges.

The model incorporates academic 2-year colleges together with 4-year col-
leges and work. High-school graduates are uncertain about their ability to accumulate human capital. Pessimistic agents join the workforce, optimistic agents enroll in 4-year colleges and those with intermediate beliefs enroll in academic 2-year colleges. During their tenure as students, agents are presented with exams, which govern the accumulation of credits and provide information that updates beliefs about ability, inducing dropouts and transfers.

Academic 2-year colleges are ideal for students with aspirations regarding graduation at 4-year colleges but with low expectations about their ability to accumulate human capital. Depending on the evolution of their beliefs and accumulation of credits, students can decide to transfer to 4-year colleges and carry with them a proportion of their stock of credits, implying that academic 2-year colleges act as a stepping-stone, similar to Jovanovic and Nyanko (1997) where workers move up in the work ladder once they acquired the necessary skills. Further, the model has features of bandit models as students learn about their innate ability to accumulate human capital, similar to Johnson (1978) and Miller (1984). Jovanovic and Nyarko (1997) evaluate the predictive power of bandit and stepping stone models in terms of job mobility and find that there is some evidence favoring a combination of both. Post-secondary education also presents features of bandit and stepping stone models.

The model relates to that one of Miao and Wang (2007), that studies the problem of an investor facing two investment projects, one risky and one riskless. The riskiness follows from uncertainty regarding the true quality of the project. For agents investing in the risky project, the arrival of information might induce re-balancing the investment portfolio towards the riskless one. The model developed here has, as a limiting case, a configuration similar to the model in Miao and Wang (2007). When grades provide only information and graduation is random with poisson arrival rate, the model can be casted in a very similar fashion to Miao and Wang (2007) with the only difference that in this model agents face 3 investment opportunities with two that are not absorbing, while in Miao and Wang (2007) they only face two with only one non absorbing. Even in this simplified setup, academic 2-year colleges can be used to learn in a cheaper and less risky environment.
This paper also relates to Groes, Kircher, and Manovskii (2010). They document, using a large panel of workers from Denmark, that for a given occupation, high and low wage earners are more likely to switch occupations, the first to more rewarding occupations and the second to occupations with lower mean wage. Later, they produce a theory of learning about the worker’s own ability that explains the sorting and the transitions. In postsecondary education, and in the model developed here, the same pattern of sorting and transitions occurs.

The paper is also related to some empirical research that recognizes education as a sequential process. Altonji (1993) computes Internal Rates of Return for a simple sequential model where agents are uncertain about future income flows and thus evaluation of expectations induces dropout behavior. Stange (2012) and Heckman-Urzua estimate models of educational choice where students are allowed to transfer and drop out. This paper departs from these two in two important dimensions. First, both Stange (2012) and the work in progress by Heckman-Urzua assume that students learn about their own ability and non-pecuniary costs of education. To make the model consistent with the data, their estimations provide that learning about non-pecuniary costs of education is important in explaining dropout behavior. In my paper students learn only about their own ability and the model is still able to capture the main features of the data. Moreover, my modeling assumption is is consistent with Stinebrickner and Stinebrickner (2009) which found using a panel study that learning about ability explains dropout behavior in deterrence of learning about non-pecuniary costs of education. Second, my model is highly tractable, allowing for a clean characterization of the optimal policy of students.

As briefly commented before, an important feature of the model is its tractability, which allows for a clear characterization of the optimal policy that governs enrollment, dropout and transfer behavior. The model is parameterized using data from NLS-72 by assuming that observable measures of ability are correlated with the high-school graduate initial beliefs. The model is consistent with the following facts: (1) among those initially enrolled in aca-

\[\text{See Comay, Melnik, and Pollatschek (1973) for some early work on the topic.}\]
ademic 2-year colleges, more able agents are less likely to graduate, more likely to transfer, and less likely to dropout; (2) among those initially enrolled in 4-year colleges, more able agents are more likely to graduate and less likely to dropout or transfer; (3) there is a higher concentration of high ability students among transferees.

Using the parameterized version of the model, a counterfactual analysis of the value added of each option (dropout and transfer), the importance of the lack of full insurance, and the value added of academic 2-year colleges as a whole are performed. To do this not only changes in the enrollment pattern are analyzed but also how the return to enrollment, relative to joining the workforce directly after high school graduation, are affected by each of these options.

The decomposition of returns to enrollment shows that the dropout option accounts for 31% of the full average return to enrolling in an academic 2-year college while the transfer option accounts for 69%. No one would enroll in an academic 2-year college if these options were not available. For 4-year colleges, the dropout option accounts for 87% of the full average return to enrollment while the rest is explained by simple human capital accumulation that follows from enrollment until graduation. Full insurance would reduce enrollment in academic 2-year colleges from nearly 16% to 10%, while enrollment in 4-year colleges would rise from 27% to nearly 50%. The interaction of risk and option value proves to be an important force in postsecondary education.

Finally, the paper then turns to evaluate how close of a substitute are academic 2-year colleges for 4-year colleges to find a high degree of substitutability. In accordance with this high degree of substitutability, the welfare effect of the availability of academic 2-year colleges is at most moderate and is primarily driven by an increase in participation of nearly 7%. This result contrasts with the common wisdom that academic 2-year colleges are a key part of the educational system as they train the students at the margin.
2 Evidence: patterns and returns

This section presents statistics on postsecondary educational patterns and returns based on the NLS-72. The unit of analysis are high-school graduates that join the workforce directly after leaving school (with no spells of postsecondary education) or join a postsecondary institution with no discontinuities in their educational spells. NLS-72 follows the educational histories of the senior class of 1972 up to 1980. A final wave in 1986 was performed to acquire long-run job market information.

The postsecondary educational system broadly consists of 4-year colleges, academic 2-year colleges and vocational/technical school. The goal of academic 2-year colleges is to prepare students for transferring to 4-year colleges while vocational schools offer specialized education through terminal programs that require either 2 or 3 years to complete. In the model, vocational schools will not be present as the model will be a story of a stepping stone, which will be given by academic 2-year colleges. For the estimation of the model (see Section 4), agents that enroll in a vocational school or program will be pooled with those that never pursued higher education. It will be clear in this section that individuals that enroll at a vocational school or program do not transfer to other types of institutions, and that the opposite is also true. Moreover, returns to enrollment are below returns to enrollment in the other types of schools or programs, and individuals that enroll in vocational schools or programs are sorted below those that enroll in the other types of institutions or programs.

Table 1 can be used to evaluate the dynamic pattern of postsecondary educational histories. Individuals are faced with an initial enrollment choice between 4-year colleges, academic 2-year colleges, vocational schools or joining the workforce. The first spell of education can end in three different ways. First, a student can dropout ($D$) and join the workforce. Second, a student can transfer ($T$) to a different type of educational institution. Third, a student

\[\text{Some institutions provide both academic and vocational programs. In this paper, each of these programs is understood as a separate entity.}\]

\[\text{Within-type transfers (e.g. 4-year college to 4-year college) are not understood here as}\]
can graduate (G) and join the workforce. The proportion of students that transfer more than once is negligible and therefore the analysis that follows is reduced to only account for students that transfer at most once. Students that transfer can end their second spell of education in two ways: obtaining a degree or becoming a dropout.

<table>
<thead>
<tr>
<th>Table 1: Transitional dynamics</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>voc. school</td>
</tr>
<tr>
<td>ac. 2-year c.</td>
</tr>
<tr>
<td>4-year c.</td>
</tr>
</tbody>
</table>

The transitions are for first and second educational spells after high-school graduation. T stands for Transfer, G for graduation and D for dropout. Share: share of high-school graduates that enroll either at vocational school, academic 2-year colleges or 4-year colleges. Source: NLS-72.

Only half of the sample pursues higher education directly after high-school graduation. Among them, nearly 20% enroll in vocational school, around 30% enroll in academic 2-year colleges and the rest enroll in 4-year colleges.

Dynamics (i.e. dropout, graduation and transfer behavior) differ for students depending on their initial enrollment choice. Dropout rates are high in the three types of institutions but are higher in vocational schools and academic 2-year colleges than in 4-year colleges. Transfer rates are important in academic 2-year colleges: approximately 32% of students that initially enroll in this type of institution eventually transfer to 4-year colleges; Only 4-year colleges graduate a large percent of their students. Note that the graduation rate at 4-year colleges is similar for those initially enrolled at 4-year colleges and for those that transferred from academic 2-year colleges. This fact favors the idea that the initial enrollment choice does not hinder the probability of transfers.

\footnote{A student that transfer holding a degree is counted as a transfer.}

\footnote{Dropout rates at vocational schools are inflated since vocational schools have students that enroll in particular classes such as Pottery, learning to use Excel, etc. Once they acquire the particular skill, these students leave the school and return to the workforce. These students don’t get terminal degrees and so they are recorded as dropouts.}
graduation at 4-year colleges.\footnote{Similar patterns are present in NELS:88 that correspond to a cohort that graduates from high-school in 1992.}

Low enrollment and high attrition rates can be associated with the risk (possibly due to heterogeneity in returns) and costs attached to education. Costs include foregone earnings (income stream that a student ’loses’ by attending school) and direct costs of education that include tuition, fees, and housing. Table 16 shows that 4-year colleges cost twice as much as academic 2-year colleges, providing one reason why students might enroll in academic 2-year colleges.

\subsection{Returns to Education}

Table 2 presents the Internal Rate of Return for the average student with a particular educational path relative to joining the workforce directly after high-school graduation. The Internal Rate of Return takes into account direct and indirect costs of education together with its benefits. Controlling for observable measures of ability, a mincer and growth regressions are used to back out the wages for different educational histories (benefit) relative to the foregone wage due to education (indirect cost). The regressions are in Appendix B. The direct cost of education includes tuition, books and, sometimes, room and board.

Few agents join the workforce after graduating from academic 2-year colleges providing low confidence in the estimated coefficients in the mincer and growth regressions. In fact, Table 13 presents the results of a test that evaluates whether the estimated wage differential of graduating at academic 2-year colleges differs from the estimate for dropouts. The test can not reject the equality of the coefficients. With this in mind, Internal Rates of Return are calculated in two different ways. The first uses direct results from both mincer and growth regressions. The second calculates Internal Rates of Return by assuming that graduates at academic 2-year colleges enjoy similar wage profiles than dropouts.

The results in Table 2 do not represent the true internal rate of return
of each alternative since it relies on estimates of wage differentials that do not adjust for self-selection but nonetheless provide intuition regarding the value attached to each of the possible educational paths.

### Table 2: Internal Rates of Return

<table>
<thead>
<tr>
<th></th>
<th>Voc. school</th>
<th>Ac. 2-year C.</th>
<th>4-year C.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Graduation</td>
<td>9.37</td>
<td>-0.63 / 4.06</td>
<td>9.93</td>
</tr>
<tr>
<td>Dropout</td>
<td>2.07</td>
<td>6.4</td>
<td>8.55</td>
</tr>
<tr>
<td>Transfer to V.S.</td>
<td>-</td>
<td>1.44</td>
<td>0</td>
</tr>
<tr>
<td>Transfer to Ac. 2-year</td>
<td>3.71 / 4.74</td>
<td>-</td>
<td>3.37 / 3.46</td>
</tr>
<tr>
<td>Transfer to 4-year</td>
<td>8.49</td>
<td>8.84</td>
<td>-</td>
</tr>
<tr>
<td>Enrollment</td>
<td>2.75 / 2.78</td>
<td>6.63 / 6.87</td>
<td>8.93 / 8.94</td>
</tr>
</tbody>
</table>

The cost of education includes Room and Board for 4-year colleges only. All the values are in percentage points. Appendix B presents the details of the calculations and show how to obtain the wage estimates that are imputed using mincer and growth regressions. Statistics for the direct cost of education can be found in Table 16. The excluded group are agents that join the workforce directly. Finite lifetime of 47 years is assume throughout the calculations. Returns to graduation at academic 2-year colleges are computed in two different ways: (1) using the parameters obtained by both mincer and growth regressions and (2) imposing a similar wage profile for graduates than for dropouts.

Table 2 shows that enrollment in 4-year colleges provides the highest return and is mostly driven by the return for graduates. Among agents that enroll in academic 2-year colleges the results show that the best educational path is to eventually transfer to 4-year colleges rather than staying and eventually graduating, reinforcing the ‘transfer function’ associated with academic 2-year colleges. Finally, note that the internal rates of return conditional on initial enrollment choice are ordered: low in vocational school, average in academic 2-year colleges and high in 4-year college.

### 2.2 Sorting in Initial Enrollment

Students that enroll in 2-year colleges have observable measures of ability that lie between those of high-school graduates that join the workforce directly and those of students that enroll in 4-year colleges as noted by Grubb (1993) and

---

8The results are in line with empirical studies. See, for example, Belzil and Hansen (2002), or Cunha, Heckman, and Navarro (2005).
Kane and Rouse (1999). Table 3 presents summary statistics for measures of ability correlated with enrollment decisions tabulated by initial enrollment choice, extending the analysis of Grubb (1993) and Kane and Rouse (1999) by splitting 2-year colleges between vocational schools and academic 2-year colleges. From left to right the table shows that there is evidence of an ordered enrollment choice. For example, see the Rank in the senior year at high-school class. The rank decreases monotonically with the enrollment choice.

Table 3: Summary Statistics for Measures of Ability Conditional on Initial Enrollment

<table>
<thead>
<tr>
<th></th>
<th>work</th>
<th>V.S.</th>
<th>Ac. 2-year</th>
<th>4-year C.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Male</strong></td>
<td>0.492</td>
<td>0.4</td>
<td>0.54</td>
<td>0.524</td>
</tr>
<tr>
<td></td>
<td>(0.5 )</td>
<td>(0.49)</td>
<td>(0.5)</td>
<td>(0.5)</td>
</tr>
<tr>
<td><strong>Black</strong></td>
<td>0.108</td>
<td>0.126</td>
<td>0.085</td>
<td>0.105</td>
</tr>
<tr>
<td></td>
<td>(0.31)</td>
<td>(0.33)</td>
<td>(0.28)</td>
<td>(0.3)</td>
</tr>
<tr>
<td><strong>Socio. Status:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Low</strong></td>
<td>0.415</td>
<td>0.294</td>
<td>0.192</td>
<td>0.166</td>
</tr>
<tr>
<td></td>
<td>(0.49)</td>
<td>(0.45)</td>
<td>(0.39)</td>
<td>(0.37)</td>
</tr>
<tr>
<td><strong>Medium</strong></td>
<td>0.506</td>
<td>0.582</td>
<td>0.55</td>
<td>0.407</td>
</tr>
<tr>
<td></td>
<td>(0.5 )</td>
<td>(0.49)</td>
<td>(0.49)</td>
<td>(0.49)</td>
</tr>
<tr>
<td><strong>High</strong></td>
<td>0.078</td>
<td>0.122</td>
<td>0.253</td>
<td>0.426</td>
</tr>
<tr>
<td></td>
<td>(0.27)</td>
<td>(0.33)</td>
<td>(0.43)</td>
<td>(0.49)</td>
</tr>
<tr>
<td><strong>Education of Father:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;HS</td>
<td>0.516</td>
<td>0.391</td>
<td>0.289</td>
<td>0.215</td>
</tr>
<tr>
<td></td>
<td>(0.5 )</td>
<td>(0.48)</td>
<td>(0.45)</td>
<td>(0.41)</td>
</tr>
<tr>
<td>HS</td>
<td>0.323</td>
<td>0.398</td>
<td>0.344</td>
<td>0.288</td>
</tr>
<tr>
<td></td>
<td>(0.46)</td>
<td>(0.49)</td>
<td>(0.47)</td>
<td>(0.45)</td>
</tr>
<tr>
<td>4-year C. (no degree)</td>
<td>0.108</td>
<td>0.145</td>
<td>0.211</td>
<td>0.19</td>
</tr>
<tr>
<td></td>
<td>(0.31)</td>
<td>(0.35)</td>
<td>(0.41)</td>
<td>(0.39)</td>
</tr>
<tr>
<td>4-year C. graduate</td>
<td>0.051</td>
<td>0.064</td>
<td>0.155</td>
<td>0.305</td>
</tr>
<tr>
<td></td>
<td>(0.22)</td>
<td>(0.24)</td>
<td>(0.36)</td>
<td>(0.46)</td>
</tr>
<tr>
<td><strong>Rank</strong></td>
<td>0.495</td>
<td>0.435</td>
<td>0.395</td>
<td>0.271</td>
</tr>
<tr>
<td></td>
<td>(0.31)</td>
<td>(0.28)</td>
<td>(0.27)</td>
<td>(0.24)</td>
</tr>
</tbody>
</table>

Rank=rank in high-school class. Socio-Status: Socioeconomic Status of Family at moment of high-school graduation. Source: NLS-72.

2.3 Academic 2-year colleges as a stepping stone

Ladders have been associated with the growth of skill as discussed in Jovanovic and Nyarko (1997). Lower steps of the ladder are characterized as stepping
stones because they provide a less risky environment to learn compared to higher steps. As agents acquire the necessary skills, they move upwards in the ladder. The process that starts after high-school graduation and culminates with 4-year college graduation is a ladder with two steps. The first step, the stepping stone, is academic 2-year colleges. The second step is 4-year colleges. But there also ways of going down. First, a student at 4-year college can decide to transfer to academic 2-year colleges. Second, a student can decide to dropout.

In contrast with the characterization of the ladder discussed above where agents move on once they acquire the necessary skills, this ladder also present features of bandit models as those discussed in Johnson (1978), Miller (1984) and Jovanovic and Nyarko (1997). Models of skill accumulation imply that agents should enroll first in the lower step of the ladder, as it provides a less riskier environment for learning and experimentation. Bandit models suggest that students should enroll in the harder step - the last step - since the learning technology provides more information about innate ability.

3 Model

The economy is populated by agents that, upon high-school graduation, decide whether to join the labor force or pursue a degree at a postsecondary educational institution. At $t = 0$ agents graduate from high-school endowed with asset level $a_0$. Agents differ in their ability to accumulate human capital at college, that can either be low or high. Let $\mu \in \{0, 1\}$ denote the ability level, with $\mu = 0$ denoting low ability. The ability level $\mu$ is not observable by the agent. Instead, a high-school graduate inherits a signal about her true type denoted by $p_0 \in [0, 1]$, where $p_0 = Pr(\mu = 1)$.

At any period in time an agent can either be working, studying at 4-year colleges (or $C$) or at academic 2-year colleges (or $A$). Let $i \in I = \{A, C\}$ denote the type of institution. The cost of education per period of schooling is denoted by $\tau^i$, with $\tau^C > \tau^A$. A student graduates from institution $i$ after accumulating $T^i$ credits, with $T^C > T^A$. The evolution of credits is closely
tied to signals that arrive during an agent’s tenure as student that are labeled by $\eta$.

Work is assumed to be an absorbing state with constant wage function $h(GS, i, \mu)$, where the first argument accounts for the graduation status of the agent, the second for the type of institution where the agent graduated from (highest degree) and the third for her true ability level.\footnote{In the current setup dropouts do not enjoy higher wage profiles. That is, increases in wages only occur after graduation. The model can be easily extended to include this feature by making the function $h(\cdot)$ to depend on amount of credits $s$ but it will lose much of its tractability.} Further, the wage function $h(GS, i, \mu)$ is specified as follows:

$$h(GS, i, \mu) = \begin{cases} h^w & \text{if } GS=0; \\ h^i(\mu) & \text{if } GS=1. \end{cases}$$

with $h^i(1) \geq h^i(0) > h^w$ for all $i$ and $h^C(\mu) \geq h^A(\mu) > h^w$ for all $\mu$. That is, for any ability level, graduation at a 4-year college implies higher wage profiles than graduation at an academic 2-year college and, for any institution $i$, wage profiles of graduates are increasing in their ability level.

The evolution of the asset level $a$ is given by

$$a_{t+1} = \begin{cases} (1 + r)a_t - \tau^i - c_t & \text{if enrolled at } i; \\ (1 + r)a_t + h(GS, i, \mu) - c_t & \text{if working.} \end{cases}$$

where no borrowing constraints are present.\footnote{The assumption of no borrowing constraints is consistent with several empirical studies. See, for example, Cameron and Taber (2004) and Stinebrickner and Stinebrickner (2008).}

During tenure as students agents receive signals in form of exams labeled as $\eta$. Let $\eta$ denote the signal with PDF given by $f_i(\eta|\mu)$.

**Assumption 1** The ratio of densities $\frac{f_i(\eta|\mu=1)}{f_i(\eta|\mu=0)}$ satisfies the Monotone Likelihood Ratio Property (MLRP). That is, for any $\eta_1 > \eta_0$, $\frac{f_i(\eta|\mu=1)}{f_i(\eta|\mu=0)}$ is well defined (full support) and

$$\frac{f_i(\eta_1|\mu = 1)}{f_i(\eta_1|\mu = 0)} \geq \frac{f_i(\eta_0|\mu = 1)}{f_i(\eta_0|\mu = 0)}$$
The assumption states that high ability students are prone to receiving better signals than low ability students.

The evolution of credits is a function of current signal $\eta$ and amount of accumulated credits $s$,

$$s' = s + \tilde{\Omega}(\eta, s)$$

with

$$\tilde{\Omega}(\eta, s) = \begin{cases} 
\Omega(\eta) & \text{if } s < T_i; \\
0 & \text{if } s \geq T_i. 
\end{cases}$$

(1)

with $\Omega'(\eta) \geq 0$ so that the evolution of credits is a non-decreasing function on the received signal. Equation (1) states that, while the amount of current credits is less than the necessary amount for graduation, accumulation of credits is only a function of the received signal $\eta$, which is positively correlated with a student’s ability level through Assumption 1.

Students are allowed to transfer across different types of schools and can carry with them part of the credits earned in the current institution. Let $\theta^i$ denote the operator that maps credits $s$ in institution $i$ to credits $s$ in institution $j$, $j \neq i$. Formally,

$$\theta^i(s) : \mathbb{R}^+ \times I \rightarrow [0, T^j]$$

where it is assumed that $\theta^i(s)$ is non-decreasing in credits $s$.

A high-school graduate, endowed with her prior $p_0$ and initial asset level $a_0$ chooses her consumption stream $\{c_t : t \geq 0\}$ and whether to enroll in, dropout or transfer in $A$ or $C$, in order to maximize her time-separable expected discounted lifetime utility derived from consumption,

$$\mathbb{E} \left\{ \sum_{t=0}^{\infty} \left( \frac{1}{1+r} \right)^t \left( \frac{e^{-\gamma c_t} - 1}{-\gamma} \right) \right\ | F_0 \right\}$$

where $F_0 = \{p_0, a_0\}$ and $\gamma$ is the coefficient of Constant Absolute Risk Aversion (CARA).
Let \( V_i(a, s, p) \) denote the value for a student currently enrolled in institution type \( i \) with asset level \( a \), amount of credits accumulated \( s \) and prior \( p \). Also, let \( W(a; h(GS, i, \mu)) = -\frac{1+r}{\gamma r} e^{-\gamma (ra + h(GS, i, \mu))} + \frac{1+r}{\gamma r} \) denote the value for a worker with asset level \( a \) and wage profile \( h(GS, i, \mu) \). Finally, let \( \Lambda(a_0, p_0) \) denote the value for a high-school graduate with asset level \( a_0 \) and prior \( p_0 \).

The value for a high-school graduate \( \Lambda(a_0, p_0) \) equals
\[
\Lambda(a_0, p_0) = \max (W(a_0; h^w), V_A(a_0, 0, p_0), V_C(a_0, 0, p_0))
\]
as the agent chooses whether to join the workforce or pursue higher education (either in academic 2-year colleges or 4-year colleges) by comparing the value of each alternative.

Let \( R_i(p) \) denote return to enrollment at institution \( i \) relative to joining the workforce,
\[
R_i(p) \equiv \frac{\Sigma_i(p)}{\frac{1+r}{r} h^w}
\]
where \( \Sigma_i(p) \), which solves \( V_i(a - \Sigma_i(p), 0, p) = W(a; h^w) \), describes the payoff of the education option and \( \frac{1+r}{r} h^w \) its price (i.e. the opportunity cost of becoming a student).

\[\text{3.1 The problem of a student}\]

The student’s beliefs are updated by the stream of information that arrive through the signal \( \eta \). Let \( p' = b(\eta; p) \) denote the posterior that depends on the prior \( p \) and the signal \( \eta \). For a given institution \( i \), Bayes’ rule is
\[
b(\eta; p) = \frac{1}{1 + \frac{f_i(\eta|\mu=0)}{f_i(\eta|\mu=1)} \frac{1-p}{p}}
\]

\[\text{11} \text{ The value function for a worker is the solution to the problem } W(a; h) = \max_{c, a'} \frac{e^{-\gamma c} - 1}{\gamma r} + \frac{1}{\gamma r} W(a'; h) \text{ where } a' = (1 + r) a + h - c. \]
\[\text{12} \text{ The lack of available assets to diversify the risk that comes from education (as income flows are unknown) imply that standard techniques to value the option to pursue postsecondary education can not be applied.} \]
The evaluation of expectations about future income flows depends on the likelihood of the signals. Any new signal can be produced by either $f_i(\eta|\mu = 1)$ or $f_i(\eta|\mu = 0)$ so that expectations about the governing CDF have to be accounted for. Define

$$H_i(\eta, p) \equiv pF_i(\eta|\mu = 1) + (1 - p)F_i(\eta|\mu = 0)$$

as the CDF that accounts for this uncertainty.

The problem faced by a student in institution type $i$ can be written as

$$V_i(a, s, p) = \max_{c, a'} \frac{e^{-\gamma c} - 1}{-\gamma} + \frac{1}{1 + r} \int_{\eta} \tilde{V}_i(a', s', p') H_i(d\eta, p)$$

with

$$\begin{align*}
a' &= (1 + r)a - \tau_i - c \\
s' &= s + \Omega(\eta) \\
p' &= b(\eta; p)
\end{align*}$$

In any given period a student that accumulated $s'$ credits faces alternatives. If $s' < T_i$ she can decide to stay in the current institution, transfer or drop. If $s' = T_i$ graduation is a fact and the options are reduced to graduation or transfer as dropping out is dominated.\(^\text{13}\) Let $\mathbb{I} = 1$ if $s' < T_i$ and $= 0$ otherwise. $\tilde{V}_i(a', s', p')$ is equal to

$$\max \left\{ W(a'; h^w), \mathbb{I}V_i(a', s', p') + (1 - \mathbb{I})[pW(a'; h^1_i) + (1 - p')W(a'; h^0_i)], V_{-i}(a', \theta'(s'), p') \right\}$$

The timing of the problem, presented in Figure 1, is as follows. For a given institutional choice $i$ at the beginning of a period a student chooses her consumption and level of assets for the next period given her expectations about future income streams. Next, she receives the signal $\eta$, producing bayesian updating of prior $p' = b(\eta, p)$, and the amount of credits accumulated for next period $s'$. When the new period begins the student chooses whether

\(^{13}\)As $h^1_i > h^0_i > h^w$ and $W(a; h)$ increasing in wage $h$, it follows directly that $pW(a; h^1_i) + (1 - p)W(a; h^0_i) > W(a; h^w)$ as $p \in [0, 1]$. Then, the dropout option is strictly dominated by the graduation option.
to dropout or remain a student and whether to transfer to another institution.

Figure 1: Timeline

At the beginning of the period, $V_i(a, s, p)$ denotes the value of enrollment at institution type $i$ with wealth level $a$, accumulated credits $s$ and prior $p$. The student chooses her consumption level producing the accumulated wealth for next period $a'$. Later, she takes an exam, receiving grade $\eta$. This produces a reevaluation of beliefs $p' = b(\eta; p)$ and accumulation of credits $s' = s + \Omega(\eta)$. The student then chooses between staying in institution type $i$ (or graduating if enough credits were accumulated), transferring or dropping out.

The solution to the student’s problem is

$$V_i(a, s, p) = -\frac{1 + r}{\gamma r} e^{-\gamma (r a + v_i(s, p))} + \frac{1 + r}{\gamma r}$$

where $v_i(s, p)$ solves a recursive equation,

$$v_i(s, p) = \frac{\tilde{v}_i(s, p) - r \tau^i}{1 + r}$$

with

$$\tilde{v}_i(s, p) = -\frac{1}{\gamma} \ln \left[ f_\eta \max \left\{ -e^{-\gamma \omega_{0}}, -e^{-\gamma \omega_{1}} + (1 - p' e^{-\gamma \omega_{2}}) + (1 - 1) e^{-\gamma \omega_{3}} \right\} \right]$$
being the consumption equivalent of the continuation value and where \( p' = b(\eta, p) \) and \( s' = s + \Omega(\eta) \). The details of the calculations can be found in Appendix C. Also, the value for a student \( V(a, s, p) \) is both increasing and convex in credits \( s \) and prior \( p \) (see Appendix D). This results will prove useful in characterizing the optimal policy.

### 3.2 Characterization of Solution

In this section I impose some restrictions on the model parameters to analyze a particular pattern of education: students with high priors enroll in 4-year colleges, average priors enroll in academic 2-year colleges and low priors join the workforce directly. Later on, on the quantitative exploration of the model, none of these restrictions are imposed.

The next assumption is useful to characterize the optimal policy under the desired pattern of initial enrollment.

**Assumption 2** The primitives of the model are such that

\[
\begin{align*}
\frac{\tilde{v}_C(0,0) - r \tau_C}{1+r} & \leq \frac{\tilde{v}_A(0,0) - r \tau_A}{1+r} & \leq h^w \\
\frac{\tilde{v}_C(0,1) - r \tau_C}{1+r} & \geq \frac{\tilde{v}_A(0,1) - r \tau_A}{1+r} & \geq h^w
\end{align*}
\]

The assumption states that high-school graduates with low ability to accumulate human capital are better off joining the workforce and, in the eventuality of enrollment, they are better off in academic 2-year colleges than in 4-year colleges. The opposite idea applies for high ability agents. They are better off by pursuing higher education and the best enrollment choice for them are 4-year colleges.

**Assumption 2** has an interesting interpretation. The existence of academic 2-year colleges in this model is driven by the learning mechanism and the option value. **Assumption 2** is useful on characterizing the optimal policy, which is described in the next proposition.
**Proposition 1** For a given institution $i$ and amount of credits $s$, the optimal policy is a collection of dropout and transfer thresholds $\{p_{id}(s), p_{it}(s)\}_{s \in [0, T_i]}$ independent of the current wealth level $a$. The optimal policy for students enrolled in institution $i$ is

\[
\text{Enrolled in } i \text{ today: } \begin{cases} 
\text{join workforce tomorrow} & \text{if } p < p_{id}(s) \\
\text{enrolled in } A \text{ tomorrow} & \text{if } p \in [p_{id}(s), p_{it}(s)] \\
\text{enrolled in } C \text{ tomorrow} & \text{if } p > p_{it}(s)
\end{cases}
\]

The proof is straightforward and therefore not presented. Independence of the optimal policy function with respect to current wealth level $a$ is a result of the choice of CARA utility function and the dependence of the policy function on credits $s$ follows as the amount of accumulated credits affect both the amount of credits that can be transferred and also the likelihood and time to graduation. Assumption 2 drives the optimal policy not only at time 0 but also as credits accumulate. Consider the case where $T_i$ is large so that in terms of distance until graduation an agent with $s = 0$ and one with $s = 1$ are very similar. It follows that a similar condition holds for $s = 1$ but the difference in the value functions should decrease as students get closer to graduation.

The next proposition evaluates the interaction of accumulated credits $s$ and the evolution of the thresholds for the case where the support of credits $s$ is $\mathbb{R}^+$. 

**Proposition 2** If (i) $F_i(\eta | \mu)$ is continuous and differentiable for all $i$ and $\mu$, (ii) $\Omega(\eta)$ is continuous and differentiable, and (iii) $\theta^i(s)$ is continuous, differentiable and concave with $\frac{\partial \theta^i(0)}{\partial s} < 1$, it is the case that

\[
\frac{\partial p_{id}}{\partial s} \leq 0 , \quad \frac{\partial p_{it}}{\partial s} \geq 0 , \quad \frac{\partial p_{it}}{\partial s} \leq 0
\]

**Proof.** See Appendix F. ■
The proposition states that as credits accumulate, the likelihood of transferring or dropping out decreases as the terminal payoff at institution $i$ is getting closer. In other words, as credits accumulate, the inaction bands in a given institution widen out.

4 Parameterization

The model explores the dynamic interaction of academic 2-year colleges and 4-year colleges. The evidence obtained from NLS-72 shows that vocational school (excluded from the model analysis) can be merged with the workforce as little interaction occurs between vocational school and other types of institutions (see Table 1) and the sorting in initial enrollment, presented in Table 12 and Table 11, places vocational school below academic 2-year colleges. Version B in all of the tables accounts for the case where work and vocational school are merged. Further, as in the model, column 2 on the tables accounts for the cases where increases in wages only occur upon graduation.

The operator that maps credits $s$ in institution $i$ to credits that remain after transferring $\theta^i(s)$ is simplified to be of the multiplicative form,

$$\theta^i(s) = \begin{cases} \theta^i s & \text{if } \theta^i s < T^i \\ T^i & \text{if } \theta^i s \geq T^i \end{cases}$$

The signal $\eta$ plays two different roles in the model. First, it updates beliefs $p$ as the signal conveys information regarding the likelihood of the true talent level of the student. Under this definition, the signal $\eta$ accounts for grades in exams, subjects, problem sets, overall experience as a student, etc. The second role of the signal $\eta$ is to generate accumulation of credits through the function $\Omega_i(\eta)$, which suggests that the signal is closely tied to grades in subjects.

To simplify the model, think that the signal $\eta$ is the mean of the grades in a quarter obtained by a student. The set of possible values of $\eta$ is simply the set of possible grades. For simplicity assume three possible grades: $\{F, N, E\}$. That is, a student can fail, get a neutral grade or excel in a particular exam.
Let \( q_i^\mu(\eta) \) denote the probability of each event. Further, assume that \( q_1^A(F) = q_0^C(\eta) = 0 \). That is, high ability students never fail an exam at academic 2-year colleges and low ability students never excel at 4-year colleges.

The function \( \Omega(\eta) \), that maps grade \( \eta \) into credits \( s \) is chose to be as follows:

\[
\Omega(\eta) = \begin{cases} 
0 & \text{if } \eta = F \\
1 & \text{if } \eta = N \\
1 & \text{if } \eta = E 
\end{cases}
\]

so that only students that fail in an exam do not accumulate credits.

The time period is chosen to be a quarter, therefore \( T^A = 8 \) and \( T^C = 16 \) (a student needs to accumulate \( T^i \) quarters of accumulated credits at institution \( i \) to graduate). The risk-free interest rate \( r \) is set to be 0.45% which implies a yearly interest rate of 1.81% and a yearly discount factor of 0.9822. All the monetary values in the model are measured in logs and further standarized by the wage of agents with no degrees \( h^w \), so that \( h^w = 1 \). Academic 2-year colleges are located in every city and town while 4-year colleges are scarce. This way, the cost of education includes housing for 4-year colleges and does not include housing for academic 2-year colleges because students there can live with their parents. The standarized cost of education is then \( \tau^A = 0.1152 \) for academic 2-year colleges and \( \tau^C = 0.3205 \) for 4-year colleges (see Table 16).

The risk aversion parameter, \( \gamma \), is chosen to be equal to 8. The risk aversion parameter \( \gamma \) is hard to identify and the literature didn’t spend much time estimating risk aversion parameters using CARA utility functions (for natural reasons). There is a whole string of literature in asset pricing that argues that the CRRA risk aversion parameter \( \sigma \) lies between 4 and 10. Using the definition of relative risk aversion it is possible to relate \( \sigma \) and \( \gamma \), \( \gamma c = \sigma \). Here, the consumption level \( c \) has a lower bound given by \( ra + h^w \geq 1 \) so \( \gamma < 10 \).

Students that initially enroll in an academic 2-year college and transfer later on to a 4-year college do so, on average, after the completion of the first year of education (see Figure 8). Then, \( \theta^A \) is chosen to be \( \frac{1}{2} \). The
evidence for students that transfer from 4-year colleges to academic 2-year colleges is less revealing as the fraction of students that transfer is very low (see Table 1). Figure 8 shows that students transfer during their first year of education. Among academic 2-year college graduates, those that started their educational career at 4-year colleges spend more time in school prior to graduation (4.5 years vs. 3.84 years). Then, \( \theta^C = 0 \).

The remaining parameters, \( q_i^{\mu}(\eta) \) and \( h^i(\mu) \) are estimated using a two-stage Simulated Method of Moments. In the first stage an Ordered Probit regression on the initial enrollment choice is used to produce estimates of the initial prior \( p_0 \). In the second stage, conditional on the distribution of priors obtained in the first stage, enrollment and transition choices are simulated for each individual on the sample to match an over-identified set of moments.

Let
\[
p_0 = \left( 1 + e^{-(X'\beta + \varepsilon)} \right)^{-1}, \varepsilon \sim N(0, 1)
\] (7)
where \( X \) is a vector that includes all the observable characteristics of high-school graduates that are correlated with the ability level of the agent and \( \beta \) is the vector of factor loadings, identified by an Ordered Probit regression for the initial choice of agents. The estimates for \( \beta \) are presented in Table 11. It is assumed that the value of the random variable \( \varepsilon \) is observed by the agent but unobserved by the econometrician. Still, by observing the initial enrollment choice of the agent, the econometrician can compute bounds to the value of \( \varepsilon \) consistent with the action of the agent. This is useful in computing the distribution of priors, which are needed to calibrate the model through indirect inference, that are consistent with the vector of observable characteristics \( X \) and initial enrollment choice of each agent.

Students with high priors join 4-year colleges, with average priors join academic 2-year colleges and with low priors join the workforce. The thresholds are those of the optimal policy considered above for \( s = 0 \) as agents that graduate from high-school didn’t acquire any credits yet. Further, monotonicity of \( p_0 \) as a function of \( X'\hat{\beta} + \varepsilon \) implies that \( \beta \) can be estimated by an Ordered Probit regression on the initial choice (Table 11 - Version B) and then upper and lower bounds for \( \varepsilon \) can be computed using \( X'\hat{\beta} \) and the enrollment choice of
the agent. Further, the cutoffs shown in Table 11 (version B) are monotonic transformations of the threshold for \( p \).

Conditional indirect inference, through Simulated Method of Moments, is used to estimate the remaining 10 parameters (the 6 learning parameters, and the four wages). The estimation process attempts to match the educational histories of each individual in sample, which are characterized by an educational history and long run wage, and a distribution of possible initial priors \( p_0 \) that were produced using an Ordered Probit regression as discussed in the previous paragraph. The estimation procedure computes the long run wage and the probability of dropout, graduation, and transfer for each agent, and contrast it with its empirical counterpart. To clarify, the educational history of each individual consists of an initial enrollment choice (either enrolling in an academic 2-year college, a 4-year college, or joining the workforce), transition behavior (dropout, graduation, or transfer to a different type of college), and a long run wage upon joining the workforce. Because the educational histories greatly differ across students (see Section 2), and in sample I have 3462 individuals, the model is over-identified.

Before presenting and discussing the estimation results it is important to point out that Assumption 1 nor Assumption 2 are imposed in the estimation process.

Estimation of the model requires the next steps: (1) pick values for the learning parameters and wages, (2) solve the model numerically, (3) for every value of \( X'\beta \) (that is, for every individual on sample - using the results of the Ordered Probit regression in Table 11 and the initial choice of each individual) - produce the censored distribution of initial priors \( p_0 \), (4) for every \( p_0 \), use the

---

17Therefore, for each individual on sample I have a distribution of possible priors \( p_0 \).
18The probability of each event is computed using the Kolmogorov Forward Equation (see Appendix G for details).
19The wages in the data are measured in 1985, several years after college graduation, and dropouts (see specification (2) of Version B in Table 13). The idea is that this measure captures part of the return to experience in the labor market, even though this channel is not present in the model. Naturally, the type of education should affect not only the level of wages, as in Table 13, but also the growth rate of wages. This is explored in some detail in Table 14.
Kolmogorov Forward Equations to produce the probabilities of each different event, (5) using step (3) compute the probabilities of each different event and wages upon graduation for every individual on sample, and (6) repeat the procedure until the convergence criterion is reached (objective is to minimize the squared difference of the simulated and true educational histories from data).

Table 4 shows the wage differential to graduation at both academic 2- and 4-year colleges relative to joining the workforce without holding a degree. These wages in the data are measured in 1985, around 8 years after the average graduation time so that the estimated wages account for part of the premium due to experience in the labor market. For high ability graduates the differential is almost three times larger in 4-year colleges than in academic 2-year colleges suggesting that students with high expectations currently enrolled in academic 2-year colleges should be eager to transfer to 4-year colleges. As students can transfer credits, many enroll in academic 2-year colleges as individuals are risk-averse and therefore care about the high volatility of wages.

<table>
<thead>
<tr>
<th>Table 4: Estimated Wage Differentials</th>
<th>estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>academic 2-year colleges</td>
<td></td>
</tr>
<tr>
<td>low ability</td>
<td>( \frac{h^a(0) - h^w}{h^w} )</td>
</tr>
<tr>
<td>high ability</td>
<td>( \frac{h^a(1) - h^w}{h^w} )</td>
</tr>
<tr>
<td>4-year colleges</td>
<td></td>
</tr>
<tr>
<td>low ability</td>
<td>( \frac{h^c(0) - h^w}{h^w} )</td>
</tr>
<tr>
<td>high ability</td>
<td>( \frac{h^c(1) - h^w}{h^w} )</td>
</tr>
</tbody>
</table>

Wages are measured in 1985 (see specification (2) of Version B in Table 13), around 8 years after the mean graduation time, so that the wages account for part of the experience premium in the labor markets.

Table 5 shows the estimated learning parameters. Trivially, by assumption, failing an exam at academic 2-year colleges signals low ability while an excelling at 4-year colleges signals high ability as the probability of failing an

\[20\] The idea is that the discounted flow of constant wages in the model is similar to the discounted flow of wages with no constant wage profiles as happens in the data. This channel is shut down in the model to keep things as simple as possible.
exam at academic 2-year colleges was set to be zero for high ability students and the probability of excelling in an exam at 4-year college was set to zero for low ability students. The likelihood of obtaining a neutral or an excellent at academic 2-year colleges is higher for high ability students and thus receiving these signals improves the expectations about the innate ability. Similar idea happens for a neutral grade at 4-year colleges while a fail lowers the expectations. Overall, notice that exams at 4-year colleges produce more information than academic 2-year colleges: academic 2-year colleges are easy and thus provide information in the left tail while 4-year colleges are hard and thus provide information in both the left and right tails.

Table 6 aggregates the educational histories implied by the model and compare them with the data counterpart. The model does a good job on fitting most of these aggregate moments but underestimates the amount of dropouts in academic 2-year colleges.\(^\text{21}\)

<table>
<thead>
<tr>
<th></th>
<th>academic 2-year colleges</th>
<th>4-year colleges</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>low ability $\mu = 0$</td>
<td>high ability $\mu = 1$</td>
</tr>
<tr>
<td>Fail</td>
<td>0.23</td>
<td>0(^*)</td>
</tr>
<tr>
<td>Neutral</td>
<td>0.695</td>
<td>0.87</td>
</tr>
<tr>
<td>Excel</td>
<td>0.075</td>
<td>0.13</td>
</tr>
</tbody>
</table>

\(^*\) by assumption.

The evolution of thresholds as a function of accumulated credits are presented in Figure 2. In a broad way, as discussed in Proposition 2, the inaction region in both academic 2- and 4-year colleges increases with credits. Opposed to Proposition 2 the inaction does not increase monotonically provided that signals are discrete and, as a result, the amount of credits $s$ can only take a finite amount of values and thus the proof does not go through: the discreteness of the accumulated credits $s$ together with the transfer function $\theta_i(s)$ being increasing, imply that non-monotonicity can arise naturally in the

\(^\text{21}\)A perfect match of these moments is possible is we construct an arbitrary set of priors $p_0$. In the estimation process, the priors are disciplined with the Ordered Probit regression.
Table 6: Average value of educational histories in data and model

<table>
<thead>
<tr>
<th>% of High-school graduates that</th>
<th>Data</th>
<th>Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>join Workforce</td>
<td>59.4</td>
<td>56.9</td>
</tr>
<tr>
<td>enroll in A</td>
<td>15.2</td>
<td>15.69</td>
</tr>
</tbody>
</table>

% of those initially enrolled in A that

<table>
<thead>
<tr>
<th></th>
<th>Data</th>
<th>Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>drop at A (1st spell)</td>
<td>63.2</td>
<td>48.63</td>
</tr>
<tr>
<td>transfer from A to C (1st spell)</td>
<td>32.2</td>
<td>43.79</td>
</tr>
</tbody>
</table>

% of those initially enrolled in C that

<table>
<thead>
<tr>
<th></th>
<th>Data</th>
<th>Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>drop at C (1st spell)</td>
<td>40.9</td>
<td>42.1</td>
</tr>
<tr>
<td>graduate at C (1st spell)</td>
<td>57.1</td>
<td>57.79</td>
</tr>
</tbody>
</table>

Mean Wage Differential for Graduates after (1st spell)

<table>
<thead>
<tr>
<th></th>
<th>Data</th>
<th>Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>graduated from A</td>
<td>0.045</td>
<td>0.045</td>
</tr>
<tr>
<td>graduated from C</td>
<td>0.212</td>
<td>0.1482</td>
</tr>
</tbody>
</table>

Moments to Discipline Priors

<table>
<thead>
<tr>
<th></th>
<th>Data</th>
<th>Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>FOCs from Ordered Probit (for $\beta$)</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>distribution of X</td>
<td>YES</td>
<td></td>
</tr>
</tbody>
</table>

To simulate educational histories first it is necessary to produce initial priors. Let $p_0 = (1 + e^{-(X'\beta+\varepsilon)})^{-1}$ where $X$ is a vector of observable measures of ability and $\varepsilon \sim N(0, 1)$. $\beta$ is estimated using an Ordered Probit regression on the initial enrollment choice. Bounds for $\varepsilon$ are also produced by looking at the choices of the individuals after high-school graduation. The estimation procedure attempts to match the educational histories of every individual on the sample. The sample consists of 3462 individuals.

In the data, the Mean Wage Differential for Graduates after (1st spell) is measured in 1985 (see specification (2) of Version B in Table 13), around 8 years after the mean graduation time, so that the wages account for part of the experience premium in the labor markets.

framework.\textsuperscript{22}

To evaluate the fit of the model in Appendix H I compute the distribution of dropouts, transferred students and graduation rates for both academic 2- and 4-year colleges implied by the parameterized model and contrast it with its empirical counterpart. I find that the transition patterns implied by the model resemble those in the data.

\textsuperscript{22} To fully understand this results consider the case where $\theta^i(s) = 0$. Let $s_1 > s_0$ be two different amounts of credits. Further, let $p_A^{i}(s_0)$ denote the transfer threshold at $s_0$ at academic 2-year colleges. By definition, $v_A(s_0, p_A^{i}(s_0)) = v_C(0, p_A^{i}(s_0))$. As both functions are increasing and convex in $p$ and $v_A(0, 1) < v_C(0, 1)$, the fact that $v_A(s_1, p_A^{i}(s_0)) > v_A(s_0, p_A^{i}(s_0)) = v_C(0, p_A^{i}(s_0))$ implies that $p_A^{i}(s_1) > p_A^{i}(s_0)$.
Left panel: academic 2-year colleges. Right panel: 4-year colleges. The dropout and transfer thresholds at both academic 2- and 4-year colleges are as a function of amount of accumulated credits $s$.

5 Insurance and Option Value

High dropout and transfer rates are features commonly associated with risk and thus the availability of transfer and dropout options should be highly valued by agents as they provide lower bounds to the risk of the investment. In terms of risk, keeping the primitives unaltered the model is solved again letting $\gamma$ tend to zero. This case maps to full insurance. Comparing the benchmark model (i.e. $\gamma = 8$) with the risk-neutral case provides insights regarding the interaction of risk and insurance with the optimal policy and returns in this economy.\textsuperscript{23} A similar strategy is followed to evaluate the size of the option value. Still keeping the primitives unaltered, the model is solved two more times. The first time, the transfer option is discarded and the second time eliminates both the transfer and dropout options. The value-added of each option is then evaluated using a decomposition of returns.

\textsuperscript{23}The analysis will abstract from Moral Hazard that can potentially arise from credit provision.
5.1 Insurance

For a high-school graduate with any given prior $p_0$, Figure 3 presents the returns for the benchmark model ($\gamma = 8$) and the model where $\gamma \to 0$. The comparison is important since risk aversion is tightly connected to market completeness. The more complete the markets, the lower the value for $\gamma$. Figure 3 therefore compares the benchmark model with an economy where markets are complete. When risk aversion decreases, the enrollment thresholds shift to the left as the risk implied by education is discounted less heavily by agents. The fact that the shift to the left is stronger in the threshold between 4-year colleges and academic 2-year colleges than for the one between academic 2-year colleges and work is not casual: enrollment at 4-year colleges is more risky than enrollment at academic 2-year colleges (simple comparison of the ratio of wages). It follows that a decrease in risk aversion has a stronger effect on 4-year colleges than in academic 2-year colleges. Figure 3 also shows that risk aversion hinders the returns to education in an important way, and the effect is stronger the more uncertain the prior is.

Table 7 presents the mean return for the cross-section of agents that initially enroll at either academic 2-year colleges and 4-year colleges for both the benchmark and the risk-neutral models using the estimated distribution of priors for the NLS-72 data. The provision of insurance increases returns unambiguously for every prior $p_0$ but decreases measured returns in academic 2-year colleges through the compositional change that follows the provision of full insurance.

<table>
<thead>
<tr>
<th></th>
<th>ac. 2-year college</th>
<th>4-year college</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\gamma = 8$</td>
<td>0.41</td>
<td>3.09</td>
</tr>
<tr>
<td>$\gamma \to 0$</td>
<td>0.29</td>
<td>3.42</td>
</tr>
</tbody>
</table>

Table 7: The Effect of Full Insurance on Average Measured Returns

All the numbers in the table are in percentage points. 10,000 samples of priors $p_0$ were generated using the measures of ability $X'\hat{\beta}$, the functional form $p_0 = \frac{1}{1 + e^{-(X'\hat{\beta} + \epsilon)}}$ and the initial enrollment choice of each individual.

Providing insurance not only increases returns for every prior but also
The full insurance case is when $\gamma \to 0$. The vertical lines define the indifference prior for enrollment between work and academic 2-year colleges and between academic 2-year colleges and 4-year colleges. Return: measures the return at high-school graduation of enrollment at institution $i$ relative to joining the workforce.

affects enrollment decisions (this can be seen in Figure 3 where the vertical dotted lines denote the enrollment thresholds). Table 8 computes the distribution of initial enrollment for both cases. As expected, full insurance increases total enrollment by nearly 40% and enrollment in 4-year colleges - where risk matters the most as the wedge in wages and cost of education are higher and time until graduation longer -by around 81%. Finally, the mass of students still enrolling in academic 2-year colleges with full insurance highlights the importance of the learning channel as a feature of academic 2-year colleges.

Insurance affects the enrollment distribution both at the extensive and intensive level. At the extensive level, providing full insurance increases total enrollment. At the intensive level, the provision of full insurance affects the composition of enrollment as risk (through differential learning), tuition and wages upon graduation differ across types of institutions.
Table 8: The Effect of Full Insurance on the Initial Enrollment workforce ac. 2-year college 4-year college

| γ = 8 | 56.9 | 15.68 | 27.42 |
| γ → 0 | 39.82 | 10.42 | 49.76 |

All the numbers in the table are in percentage points. 10,000 samples of priors \( p_0 \) were generated using the measures of ability \( X'\beta \), the functional form \( p_0 = \frac{1}{1+e^{-(X'\beta+c)}} \) and the initial enrollment choice of each individual.

5.2 How much Option Value?

The model is solved again once more to evaluate the size of the option value, this time reducing the amount of options. First, the transfer option is discarded and therefore the only available alternative after the initial enrollment choice is to dropout. Second, the dropout option is discarded thus no action, other than consumption decisions, is possible during tenure as student. Let \( R_{i}^{E+D+T}(p_0) \), \( R_{i}^{E+D}(p_0) \), and \( R_{i}^{E}(p_0) \) denote the value of enrollment at institution \( i \): with both options available to the agent; with only the dropout option available; and with no dropout or transfer options available.\(^{24}\) The following decomposition of returns applies,

\[
1 = \frac{R_{i}^{E+D+T}(p_0) - R_{i}^{E+D}(p_0)}{R_{i}^{E+D+T}(p_0)} + \frac{R_{i}^{E+D}(p_0) - R_{i}^{E}(p_0)}{R_{i}^{E+D+T}(p_0)} + \frac{R_{i}^{E}(p_0)}{R_{i}^{E+D+T}(p_0)}
\]

The first term in the right hand side is the value-added to total returns \( R_{i}^{E+D+T}(p_0) \) by the transfer option, the second term provides the value-added by the dropout option and the third term accounts for the value accrued through enrollment. Figure 4 shows that returns at 4-year colleges are explained by the dropout option and by simply having the enrollment choice, in accordance with high graduation and dropout rates observed for 4-year college students at NLS-72. Also, Figure 4 accounts for the importance of the transfer option in explaining returns to academic 2-year college enrollment. As observed in NLS-72 (see Table 1), the value-added by the enrollment option that accounts for the simple human capital accumulation story has zero share

\(^{24}\)An alternative decomposition would be first to discard the dropout option and later discard the transfer option.

28
Table 9 produces the same decomposition this time for the mean return of
the population distribution of priors \( p_0 \). The transfer option is very valuable
in academic 2-year colleges, accounting for 69% of total value. The dropout
option is valuable in both types of institutions but more so in 4-year colleges.

<table>
<thead>
<tr>
<th>Enrollment</th>
<th>ac. 2-year C.</th>
<th>4-year C.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dropout Option</td>
<td>30.81</td>
<td>87.04</td>
</tr>
<tr>
<td>Transfer Option</td>
<td>69.19</td>
<td>0</td>
</tr>
</tbody>
</table>

All the numbers in the table are in percentage points. 10,000 samples of priors \( p_0 \) were
generated using the measures of ability \( X'\hat{\beta} \), the functional form
\( p_0 = \frac{1}{1 + e^{-(X'\hat{\beta} + \epsilon)}} \) and the
initial enrollment choice of each individual.
6 How close of a Substitute are Academic 2-year Colleges for 4-year Colleges?

The main purpose of enrollment in academic 2-year colleges is to learn in a less demanding and cheaper environment and eventually transfer to 4-year colleges with a fraction of the already accumulated credits. Still, agents have the choice to enroll at 4-year colleges. Given that a set of agents prefer to enroll in academic 2-year colleges, the question that arises is how much value academic 2-year colleges provide to these agents relative to initial enrollment at 4-year colleges. That is, how close of a substitute are these institutions? To answer this, Figure 5 evaluates how the elimination of academic 2-year colleges affects the return for an agent with prior $p_0$. For students that enroll in 4-year colleges in the benchmark model, the availability of academic 2-year colleges provides no value since the transfer option has no value for them. For students that enroll in academic 2-year colleges in the benchmark model what happens when academic 2-year colleges are eliminated is: (1) most students simply enroll in 4-year colleges and the difference in value is very small, and (2) the rest join the workforce. Overall, academic 2-year colleges are very close substitutes for 4-year colleges.

In order to characterize the welfare effect of academic 2-year colleges, an analysis of population aggregates is needed. Specifically, it is important to evaluate how participation in postsecondary education and welfare change due to the availability of academic 2-year colleges. Recall that $R_i(p)$ is a monotonous transformation of $v_i(0, p)$ that in turn is the certainty equivalent of enrollment at institution $i$ abstracting from the wealth level $a$. As a result, $R_i(p)$ is a good measure of utility and therefore a comparison of returns provides a good approximation for welfare losses or gains. Table 10 compares total participation and measured returns for the benchmark model where academic 2-year colleges are available and the only enrollment option are 4-year colleges. The availability of academic 2-year colleges increases participation by nearly 7% and generates a drop in measured returns of around 6% mostly due to the compositional change in total participation.
Figure 5: Value added by the availability of Academic 2-year colleges

Table 10: The Effect of Academic 2-year Colleges on Enrollment and Returns

<table>
<thead>
<tr>
<th></th>
<th>total participation</th>
<th>return</th>
</tr>
</thead>
<tbody>
<tr>
<td>4-year colleges + ac. 2-year colleges</td>
<td>43.11%</td>
<td>2.12%</td>
</tr>
<tr>
<td>Only 4-year colleges</td>
<td>40.3%</td>
<td>2.25%</td>
</tr>
</tbody>
</table>

10,000 samples of priors $p_0$ were generated using the measures of ability $X'\hat{\beta}$, the functional form $p_0 = \frac{1}{1+e^{-(X'\hat{\beta})}}$ and the initial enrollment choice of each individual.

7 Conclusion

This paper proposes a simple, highly tractable, model of postsecondary education, that incorporates together the work decision, academic 2- and 4-year colleges. In the model, students are allowed to drop out and transfer to more/less rewarding institutions as credits are accumulated and expectations are adjusted. The decision is not a trivial one. Higher rewards imply higher costs of education and more risk.

The parameterized version of the model is consistent with the data in terms of transitions, sorting, and returns. Therefore, it sheds some light on the role played by academic 2-year colleges. These institutions act as a stepping
stone towards more rewarding, and demanding, environments, namely, 4-year colleges.

In line with the literature of investment under uncertainty, the drop out and transfer options, and the availability of academic 2-year colleges, have an attached option value. A novel feature of this paper is that presents structural estimates of the different option values.

The drop out option explains a large share of ex-ante returns on both academic 2- and 4-year colleges. In fact, very few students would pursue higher education if dropping out is not allowed. The transfer option is not valued at 4-year colleges but explains nearly 70% of ex-ante returns at academic 2-year colleges, consistent with the idea of a stepping stone.

Interestingly, academic 2-year colleges as a whole are not highly valued, even though there are 4 students enrolling in this institutions per 10 enrolling in a 4-year college. Their value comes mostly by a modest increase in participation from the marginal students. All the other, would simply enroll in a 4-year college given that they are still allowed to drop out.

References


Appendix

A Ordered Probit regression on Initial Enrollment Choice

Ordered returns to enrollment together with the evidence presented in Table 3 suggest that the initial enrollment choice is ordered as follows: work, vocational school, academic 2-year colleges, and 4-year colleges. Table 11 in Appendix A presents the results of an Ordered Probit regression of the initial enrollment choice on a vector $X$ of observable measures of ability. Let $\beta$ denote the vector of factor loadings. Relative to Kane and Rouse (1999), the analysis is extended here to consider vocational school and academic 2-year colleges as separate institutions. The reference column in Table 11 is Version A (Version B pools vocational school and work together).

<table>
<thead>
<tr>
<th>Table 11: Ordered Probit Regression</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variable</td>
</tr>
<tr>
<td>-------------------</td>
</tr>
<tr>
<td>Male</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Black</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Socio. Status:</td>
</tr>
<tr>
<td>Low</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Medium</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Education of Father:</td>
</tr>
<tr>
<td>&lt;HS</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>HS graduate</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>4-year graduate</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Rank</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Cut 1</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Cut 2</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Cut 3</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td># of observations</td>
</tr>
</tbody>
</table>

The value $X'\hat{\beta}$ is a composite measure of ability, consistently estimated by $X'\hat{\beta}$. To evaluate a measure of the degree of sorting in initial enrollment, Table 12 produces the mean and standard deviation (in the cross-section) of $X'\hat{\beta}$ across the different alternatives. See Version A in the first row of the table. The measure of ability $X'\beta$ is unitless as it is just an ordinal representation of ability measures. Start with high-school graduates that join the workforce - labeled as work in the table - and move upwards across enrollment options. The mean value for the measure of ability increases monotonically with the enrollment options.

<table>
<thead>
<tr>
<th>Version</th>
<th>work</th>
<th>vocational school</th>
<th>academic 2-year college</th>
<th>4-year college</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>-1.782</td>
<td>-1.609</td>
<td>-1.348</td>
<td>-0.972</td>
</tr>
<tr>
<td></td>
<td>(0.567)</td>
<td>(0.571)</td>
<td>(0.604)</td>
<td>(0.649)</td>
</tr>
<tr>
<td>B</td>
<td>-1.426</td>
<td>-1.013</td>
<td>-0.633</td>
<td>-0.633</td>
</tr>
<tr>
<td></td>
<td>(0.572)</td>
<td>(0.609)</td>
<td>(0.654)</td>
<td></td>
</tr>
</tbody>
</table>

Table 12: Measure of Ability $X'\hat{\beta}$

Constructed from Ordered Probit Estimation (see Table 11). In version B work and vocational school are collapsed together. Source: NLS-72.

### B Internal Rates of Return

#### B.1 Backing Out Wage Profiles

The typical Mincer regression evaluates the effect of educational histories on lifetime earnings by estimating a wage regression on years of education and work experience. There is a significant ongoing literature that accounts for non-linearities in years of education (see, for example, Grubb (1993), Heckman, Lochner, and Todd (2008) and Kane and Rouse (1995)). The typical example in favor of non-linearities is the graduation premium. This literature has treated years of education (or amount of credits earned) in different type of institutions as perfect substitutes. Instead, it is now assumed that different educational histories affect lifetime earnings in different ways. Further, as has been already discussed in the literature, this analysis breaks the additive form (in the log version in the Mincer regression) of years of education and experience by estimating a growth equation.

Table 13 presents the results of the extended mincer regression, accounting for the different types of education and graduation premium. See Version A. Graduation in both vocational schools and 4-year colleges is associated with
Table 13: Mincer Regression

<table>
<thead>
<tr>
<th>Description</th>
<th>Version A</th>
<th>Version B</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td>$\omega_{DC}$ drop at 4-year C.</td>
<td>0.24</td>
<td>0.22</td>
</tr>
<tr>
<td></td>
<td>(0.042)</td>
<td>(0.041)</td>
</tr>
<tr>
<td>$\omega_{DA}$ drop at Ac. 2-year C.</td>
<td>0.09</td>
<td>0.077</td>
</tr>
<tr>
<td></td>
<td>(0.045)</td>
<td>(0.044)</td>
</tr>
<tr>
<td>$\omega_{DV}$ drop at Voc. school</td>
<td>0.072</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>(0.046)</td>
<td>-</td>
</tr>
<tr>
<td>$\omega_{GC}$ graduation at 4-year C.</td>
<td>0.304</td>
<td>0.286</td>
</tr>
<tr>
<td></td>
<td>(0.04)</td>
<td>(0.04)</td>
</tr>
<tr>
<td>$\omega_{GA}$ graduation at Ac. 2-year C.</td>
<td>0.015</td>
<td>0.0149</td>
</tr>
<tr>
<td></td>
<td>(0.1)</td>
<td>(0.104)</td>
</tr>
<tr>
<td>$\omega_{GV}$ graduation at Voc. school</td>
<td>0.284</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>(0.15)</td>
<td>-</td>
</tr>
</tbody>
</table>

Prob > F Prob > F

Test: $\omega_{DC} = \omega_{DA}$

<table>
<thead>
<tr>
<th>Prob &gt; F</th>
<th>Prob &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5698</td>
<td>0.5735</td>
</tr>
</tbody>
</table>

Test: $\omega_{DV} = \omega_{GC}$

<table>
<thead>
<tr>
<th>Prob &gt; F</th>
<th>Prob &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.9092</td>
<td>-</td>
</tr>
<tr>
<td>0.9992</td>
<td>-</td>
</tr>
</tbody>
</table>


higher wages relative to dropping out. The same idea does not apply to academic 2-year colleges as the return to becoming a dropout is higher than the return from graduation. The low number of students that graduate at academic 2-year colleges raises questions about the significance of the difference. The significance test presented in Table 13 shows that there is not enough evidence to reject that the return to graduation is similar (or even higher) to the return to dropping out.

Table 14 presents the results of a growth regression where the dependent variable is the average growth rate of wages between 1979 and 1985. The growth rate for vocational school graduates, $\alpha_{VG}$, is around half the growth rate of 4-year college graduates $\alpha_{CG}$. This fact, together with the results from Table 13 reads as follows: graduation at vocational schools provide a higher wage and 4-year colleges provide steeper profiles of wages, while graduation at academic 2-year colleges is dominated.

Backing out the initial wage upon graduation can not be done directly from the data because wage information prior to 1979 is scarce. To produce the initial wage, it is possible to combine the results of Table 13 and Table 14.
### Table 14: Growth Regression

<table>
<thead>
<tr>
<th>Description</th>
<th>Version A</th>
<th>Version B</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td>$\alpha_C^G$ growth rate for C. grads.</td>
<td>0.0447</td>
<td>0.0457</td>
</tr>
<tr>
<td></td>
<td>(0.0144)</td>
<td>(0.011)</td>
</tr>
<tr>
<td>$\alpha_A^A$ growth rate for A. grads.</td>
<td>-0.0036</td>
<td>-0.0071</td>
</tr>
<tr>
<td></td>
<td>(0.0351)</td>
<td>(0.018)</td>
</tr>
<tr>
<td>$\alpha_V^V$ growth rate for V. grads.</td>
<td>0.0268</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0536)</td>
<td></td>
</tr>
<tr>
<td>$\alpha_C^D$ growth rate for C. dropouts</td>
<td>0.024</td>
<td>0.0236</td>
</tr>
<tr>
<td></td>
<td>(0.0165)</td>
<td>(0.0164)</td>
</tr>
<tr>
<td>$\alpha_A^D$ growth rate for A. dropouts</td>
<td>0.0103</td>
<td>0.0091</td>
</tr>
<tr>
<td></td>
<td>(0.0196)</td>
<td>(0.0194)</td>
</tr>
<tr>
<td>$\alpha_V^D$ growth rate for V. dropouts</td>
<td>-0.0071</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0202)</td>
<td></td>
</tr>
<tr>
<td>$\alpha^0$</td>
<td>-0.009</td>
<td>-0.006</td>
</tr>
<tr>
<td></td>
<td>(0.0166)</td>
<td>(0.0164)</td>
</tr>
</tbody>
</table>

Dependent Variable: average growth rate of wages between 1979 and 1985. Source: NLS-72

### B.2 Computing Internal Rates of Return

Let $w_0$ denote the wage for an agent that joins the workforce at $t = 0$. Let $L$ denote the lifetime of an agent, $S_i$ the proportion of time spent at institution $i$, $\tau_i$ the flow cost of attendance, $\varpi_i^D$ the increase in wages due to dropping out at institution $i$, $\varpi_i^G$ the graduation premium, $\alpha_i$ the increase in wages due to experience, $G_i$ a dummy that accounts for graduation at institution $i$ and $D_i$ a dummy that accounts for dropping out at institution $i$. Further, assume that students transfer only once.\(^{25}\)

The initial wage for a student with history $H = \{S, G, D\}_{i \in \{V, A, C\}}$ when joining the workforce is

$$w_0(H) = w_0 e^{\sum \varpi_i^D D_i + \varpi_i^G G_i}$$

The present value of costs $K(H, r)$ attached to history $H$ is

$$K(H, r) = \int_0^{S_i} e^{-r s_i} \tau_i ds_i + \int_{S_i}^{S_i + S_{-i}} e^{-r s_i} \tau_{-i} ds_{-i}$$

\(^{25}\)Very few students transfer more than once in NLS-72. Also, the assumption makes the presentation of the methodology much easier.
that can be reduced to

\[ K(H, r) = \left(1 - e^{-rS_i}\right) \frac{\tau_i}{r} + e^{-rS_i} \left(1 - e^{-rS_{-i}}\right) \frac{\tau_{-i}}{r} \]

The internal Rate of Return is the interest rate \( r(H) \) that solves,

\[
\int_{S_i+S_{-i}}^L e^{-(r(H)-\alpha_i)t} w_0(H) dt - K(H, r(H)) = \int_0^L e^{-(r(H)-\alpha_0)t} w_0 dt - K(0, r(H))
\]

The variables \( S_i, D_i \) and \( G_i \) can be obtained directly from inspection of the dynamic patterns of education. \( L \) is chosen so that agents are alive until they are 65 years old. Then, \( L = 47 \).

Table 13 shows the results of an extended Mincer regression using the log of wages in 1985 as dependent variable and years of education at a particular institution and graduation dummies as explanatory variables. The coefficients \( \varpi_i^D \) and \( \varpi_i^G \) can be obtained from that table (see Version A).

Table 14 present the results of a growth regression on the graduation status and type of every individual. The results of this table are interpreted here as estimates of \( \alpha_i \) (see Version A).

The average time spent in each institution for a given educational path used in the calculations can be found in the Supplementary Material. The direct cost of education, \( \tau_i \) can be found in Table 16.

C Computing the Value for a Student

Solving for \( c \) from the budget constraint and substituting back into equation (2) reduces the problem to a single variable problem.

The lowest level for \( a' \) that can be chosen is 0 and the highest is \( (1+r)a - \tau_i \). Define \( \Gamma(a) = [0, (1+r)a - \tau_i] \) so that the choice variable \( a' \) belongs to the graph \( \Gamma(a) \).

The next claim shows that there exists a unique solution to equation (2).

Claim 1 \( V_i(a, s, p) \) is single-valued.

Proof. First note that the only choice variable is \( a \), that \( p \) evolves stochastically and \( s' \) is a function of \( p \) and signals.

Signals \( \eta \) that arrive produce updating in the state \( p \), which is thus stochastic. \( ^{26} \) Let \( P \) be such that \( p \in P \). Trivially, \( P = [0, 1] \) and thus \( P \) is compact.

\( ^{26} \)In principle the function \( p' = b(\eta, p) \) can be inverted to produce a stochastic process for the evolution of \( p \).
Also \( s \in [0, T^i] \) so the set for \( s \) is compact. The union of compact sets is compact. Further, the transition from \( p \) to \( p' \) satisfies the feller property.

Next, note that \( \Gamma(a) \) is non-empty, compact and continuous. Also, as \( c > 0 \), \( e^{-\gamma c - 1} \) is bounded.

Then, Theorem 9.6 of Lucas and Stokey (1989) is satisfied and thus the proposition holds.

After substituting the first order condition into equation (2) provides the maximized value function,

\[
V_i(a, s, p) = \frac{1}{1 - \gamma(1 + r)} \int \frac{\tilde{V}_i(a', s', p') H(d\eta, p)}{da'} + \frac{1}{1 + r} \int \tilde{V}_i(a', s', p') H(d\eta, p) + \frac{1}{\gamma} \tag{8}
\]

Conject that \( \int \tilde{V}_i(a', s', p') H(d\eta, p) = -\frac{1 + r}{\gamma} e^{-\gamma (ra + \tilde{v}_i(s, p))} + \frac{1 + r}{\gamma^r} \) and substitute into equation (8) together with \( a' = a - \frac{r^i}{1 + r} - \frac{\tilde{v}_i(s, p)}{1 + r} \) to obtain

\[
V_i(a, s, p) = -\frac{1 + r}{\gamma^r} e^{-\gamma (ra + v_i(s, p))} + \frac{1 + r}{\gamma^r}
\]

where \( v_i(s, p) \) solves the recursive equation

\[
v_i(s, p) = \frac{\tilde{v}_i(s, p) - r \tau_i}{1 + r}
\]

Further, applying the conjecture and using equation (3) reads,

\[
\tilde{v}_i(s, p) = -\frac{1}{\tau_i} \ln \left[ \int \max \left\{ -e^{-\gamma h'}, -e^{-\gamma v_i(s', p')}, -\left( 1 - p' e^{-\gamma h_i} + (1 - p') e^{-\gamma h_0} \right) + (1 - 1) e^{-\gamma v_i(s', p')} \right\} H_i(d\eta, p) \right]
\]

with \( s' = s + \Omega(\eta) \) and \( p' = b(\eta, p) \).

Finally, note that the conjecture for \( \tilde{v}_i \) holds as a result of the functional form of \( V_i(a, s, p) \).

\section{Properties of Value Function of Students}

The proof for \( p \) follows by induction. The ultimate goal in postsecondary education is graduation at 4-year colleges so start with a student that accumulated \( s = T^C - 1 \) credits. \( \frac{\partial}{\partial p} V_i(a, T^C - 1, p) > 0 \) as (1) the wage upon graduation is increasing in the agent’s true ability level, (2) the prior \( p \) measures the probability of high ability, (3) the pdf of grades satisfy the Monotone Likelihood Ratio property and (4) \( \Omega(\eta) \) non-decreasing. For a student enrolled in academic 2-year colleges with \( s = T^A - 1 \) the same proof applies but it is necessary to add
that the continuation value (through the transfer option) is increasing in the prior $p$. For $s = T^C - 2$ and any institution $i$ the proof follows as properties (2)-(4) still hold and the continuation values are increasing in $p$. Convexity follows directly as, (a) for any $p$ the continuation value is bounded below by the dropout option, (b) the continuation value of transferring or remaining at institution $i$ increasing in $p$, (c) the function max() being convex.

The proof for $s$ is very similar and simpler so it is left as an exercise to the interested reader.

### E Computing the Value for a Risk Neutral Student

Taking the limit for the value of a student (equation (4)) when $\gamma$ approaches 0 produces the value under risk neutrality,

$$V_i(a, s, p) = \frac{1 + r}{r} (ra + v_i(s, p))$$

with

$$v_i(s, p) = \lim_{\gamma \to 0} \tilde{v}_i(s, p) - \frac{r \tau_i}{1 + r}$$

where

$$\lim_{\gamma \to 0} \tilde{v}_i(s, p) = \int \max \left\{ \left\{ h^w, v_i(\theta^i(s + \Omega(\eta)), b(\eta, p)) \right\}, \left\{ 1 - b(\eta, p) b^i_1 + (1 - b(\eta, p)) b^i_0 \right\} + (1 - 1) v_i(s + \Omega(\eta), b(\eta, p)) \right\} H_i(d\eta, p)$$

I now prove the statement.

See equation (6). For any given prior $p$, each of the exponentials inside the max{} as function of $\eta$ are bounded as the set of attainable $b(\eta, p)$ being compact ($b(\eta, p) \in [0, 1]$) and the set of attainable payoff being $[e^{-\gamma h^w}, e^{-\gamma h^C}]$, also compact. It follows that each of the functions inside the max{} can be arbitrarily well approximated by a Taylor expansion (each of these functions is also differentiable). As an example, see that

$$e^{-\gamma v_i(\theta^i(s + \Omega(\eta)), b(\eta, p))} \approx 1 - \gamma v_i(\theta^i(s + \Omega(\eta)), b(\eta, p)) + O(-\gamma v_i(\theta^i(s + \Omega(\eta)), b(\eta, p)))$$

where the term $O(-\gamma v_i(\theta^i(s + \Omega(\eta)), b(\eta, p)))$ is a sequence of terms of order higher than one. Moreover, each of the functions $O(-\gamma v_i(\theta^i(s + \Omega(\eta)), b(\eta, p)))$ is a convergent series and thus bounded.

The previous argument can be used to approximate the elements of max{}.
For example,
\[ b(\eta, p)e^{-\gamma h_1} + (1 - b(\eta, p))e^{-\gamma h_1} \approx 1 - \gamma(b(\eta, p)h_1 + (1 - b(\eta, p))h_0) + O(\gamma, h_1, h_0, b(\eta, p)) \]
where \( O(\gamma, h_1, h_0, b(\eta, p)) = b(\eta, p)O(-\gamma h_1) + (1 - b(\eta, p))O(-\gamma h_0) \) is bounded and convergent by composition of bounded and convergent series.

Rewrite the original expression as
\[
\tilde{v}_i(s, p) = -\frac{1}{\gamma} \ln \left( \int_{\eta} \left( 1 - \max \left\{ \gamma \left( h_0^+ - \frac{O(\gamma, h_0^+)}{\gamma} \right), \gamma \left( v_i - \frac{O(-\gamma v_i)}{\gamma} \right) + (1 - \Pi\gamma) \left( p'h_i^+ + (1 - p')h_i^0 - \frac{O(\gamma, h_i^0, h_i^0, p')}{\gamma} \right) \right) \right) H_i(d\eta, p) \]
\]
where \( p' = b(\eta, p) \) and the state of \( v_i \) and \( v_{-i} \) is omitted to ease notation.

The claim follows by taking the limit when \( \gamma \) approaches 0. L’Hopital is required as \( \lim_{\gamma \to 0} \tilde{v}_i(s, p) \approx 0 \). The issue is whether the function \( \max \{ \} \) is differentiable with respect to \( \gamma \) and, in the affirmative case, how to characterize the derivative.

Recall that both \( v_i \) and \( v_{-i} \) depend of credits \( s \) and prior \( p \). Define
\[
J_1(\gamma, p', s') = \gamma \left( h_0^+ - \frac{O(\gamma, h_0^+)}{\gamma} \right) \\
J_2(\gamma, p', s') = \gamma \left( v_i - \frac{O(-\gamma v_i)}{\gamma} \right) \\
J_3(\gamma, p', s') = \Pi\gamma \left( v_i - \frac{O(-\gamma v_i)}{\gamma} \right) + (1 - \Pi)\gamma \left( p'h_i^+ + (1 - p')h_i^0 - \frac{O(\gamma, h_i^0, h_i^0, p')}{\gamma} \right)
\]
so that the previous expression for \( \tilde{v}_i(s, p) \) can be written as
\[
\tilde{v}_i(s, p) = -\frac{1}{\gamma} \ln \left( 1 - \int_{\eta} \max \left\{ J_1(\gamma, b(p, \eta), s + \Omega(\eta)), J_2(\gamma, b(p, \eta), s + \Omega(\eta)), J_3(\gamma, b(p, \eta), s + \Omega(\eta)) \right\} H_i(d\eta, p) \right)
\]

**Claim 2**
\[
\int_{\Omega} \max \{ J_1(\gamma, b(p, \eta), s + \Omega(\eta)), J_2(\gamma, b(p, \eta), s + \Omega(\eta)), J_3(\gamma, b(p, \eta), s + \Omega(\eta)) \} H_i(d\eta, p)
\]
differentiable with respect to \( \gamma \).

**Proof.** \( J_1, J_2 \) and \( J_3 \) are continuous functions and under the assumptions discussed in the paper, there is a unique threshold value for the signal \( \eta \) (that depends on \( s, p, \gamma \), and other parameters) that equates \( J_1 \) with \( J_2 \) and \( J_2 \)
with $J_3$. Let $\eta^L(\gamma, p, s)$ and $\eta^H(\gamma, p, s)$ denote these thresholds. Note that, as $J_1$, $J_2$ and $J_3$ are differentiable with respect to $\gamma$, these thresholds are also differentiable by construction. It follows that the expression of the claim can be rewritten as

$$\int_{-\infty}^{\eta^L(\gamma, p, s)} J_1(\gamma, p', s') H_i(d\eta, p) + \int_{\eta^L(\gamma, p, s)}^{\eta^H(\gamma, p, s)} J_2(\gamma, p', s') H_i(d\eta, p) + \int_{\eta^H(\gamma, p, s)}^{\infty} J_3(\gamma, p', s') H_i(d\eta, p)$$

(9)

where $p' = b(p, \eta)$ and $s' = s + \Omega(\eta)$. That is differentiable with respect to $\gamma$.

Let $Q(\gamma, p, s)$ denote the object in equation (2).

Claim 3

$$\frac{dQ(\gamma, p, s)}{d\gamma} = \int_{\eta} \max \left\{ \frac{\partial J_1(\gamma, p', s')}{\partial \gamma}, \frac{\partial J_2(\gamma, p', s')}{\partial \gamma}, \frac{\partial J_3(\gamma, p', s')}{\partial \gamma} \right\} H_i(d\eta, p)$$

Proof. Follows by applying Leibniz’s rule to equation (9) and by noting that

$$J_1(\gamma, b(p, \eta^L), s + \Omega(\eta^L)) \frac{\partial \eta^L}{\partial \gamma} = J_2(\gamma, b(p, \eta^L), s + \Omega(\eta^L)) \frac{\partial \eta^L}{\partial \gamma}$$
$$J_2(\gamma, b(p, \eta^H), s + \Omega(\eta^H)) \frac{\partial \eta^H}{\partial \gamma} = J_3(\gamma, b(p, \eta^H), s + \Omega(\eta^H)) \frac{\partial \eta^H}{\partial \gamma}$$

by construction of the thresholds $\eta^L$ and $\eta^H$. ■

Now, the result follows by applying L’Hopital’s rule for the case where $\gamma \to 0$ and by the results of Claim 3.

F Proof of Proposition 2

Pick two amount of accumulated credits $s_1$ and $s_0$. Without loss of generality assume that $s_1 > s_0$. Let $p^d_0(s_0) > 0$ be the dropout threshold associated with $s_0$ so that $V_i(a, s_0, p^d_0(s_0)) = W(a; h^w)$. As $V_i(a, s, p)$ increasing in credits $s$, $V_i(a, s_1, p^d_0(s_0)) > W(a; h^w)$. Finally, as $W(a; h^w)$ independent of $p$ and $V_i(a, s, p)$ increasing in $p$, $p^d_0(s_1) < p^d_0(s_0)$. Note that if $p^d_0(s_0) = 0$, then the same argument implies that $p^d_0(s_1) = p^d_0(s_0) = 0$.

Next the proof for $p^A_i(s_1) > p^A_i(s_0)$ is provided (the proof for $p^C_i(s_1) < p^C_i(s_0)$ is almost identical). Let $p^A_i(s_0) < 1$ be the dropout threshold associated with $s_0$ so that

$$V_A(a, s_0, p^A_i(s_0)) = V_C(a, \theta^i(s_0), p^A_i(s_0))$$
Consider the case where the value $V_i$ for $s = s_0$ do not include the transfer option. The next claim shows that, in this case, $V_A(a, s_1, p^A_i(s_0)) > V_C(a, \theta^A(s_1), p^A_i(s_0))$.

**Claim 4** $V_A(a, s_1, p^A_i(s_0)) - V_C(a, \theta^A(s_1), p^A_i(s_0)) > 0$.

**Proof.** Let $s_1 = s_0 + \epsilon$ where $\epsilon \in (0, T^A - s_0]$. Let (to ease on notation) $\Psi^A_j = \{a, s_j, p^A_i(s_j)\}$ and $\Psi^A_q = \{a, \theta^A(s_j), p^A_i(s_q)\}$. Applying a Taylor Expansion of second order to both $V_A(a, s_1, p^A_i(s_0))$ and $V_C(a, \theta^A(s_1), p^A_i(s_0))$ around $s_0$ provides (the assumptions on the proposition guarantees that the value function is continuous and differentiable with respect to $s$),

$$V_A(\Psi^A_1) \approx V_A(\Psi^A_0) + \frac{\partial V_A(\Psi^A_0)}{\partial s} \epsilon + \frac{1}{2} \frac{\partial^2 V_A(\Psi^A_0)}{\partial s^2} \epsilon^2$$

and

$$V_C(\Psi^A_1) \approx V_C(\Psi^A_0) + \frac{\partial V_C(\Psi^A_0)}{\partial s} \frac{\partial A(s_0)}{\partial s} \epsilon + \frac{1}{2} \frac{\partial^2 V_C(\Psi^A_0)}{\partial s^2} \left[ \frac{\partial A(s_0)}{\partial s} \right]^2 \epsilon^2$$

Note that, by definition of $p^A_i(s_0)$, $V_A(\Psi^A_0) = V_C(\Psi^A_0)$. Moreover, by the optimality of the policy, $\frac{\partial V_A(\Psi^A_0)}{\partial s} = \frac{\partial V_C(\Psi^A_0)}{\partial s}$ and $\frac{\partial^2 V_A(\Psi^A_0)}{\partial s^2} = \frac{\partial^2 V_C(\Psi^A_0)}{\partial s^2}$ as when $s$ is moved away from $s_0$ it is the case that both $V_A$ and $V_C$ coincide. Then, $V_A(a, s_1, p^A_i(s_0)) - V_C(a, \theta^A(s_1), p^A_i(s_0))$ reduces to

$$V_A(\Psi^A_1) - V_C(\Psi^A_1) = \frac{\partial V_A(\Psi^A_0)}{\partial s} \left[ 1 - \frac{\partial A(s_0)}{\partial s} \right] \epsilon + \frac{1}{2} \frac{\partial^2 V_A(\Psi^A_0)}{\partial s^2} \left[ 1 - \left( \frac{\partial A(s_0)}{\partial s} \right)^2 \right] \epsilon^2$$

Finally, as $V_A(a, s, p)$ increasing and convex in $s$, $\theta^A(s)$ increasing and concave and $\frac{\partial A(s_0)}{\partial s} < 1$, $V_A(\Psi^A_1) - V_C(\Psi^A_1) > 0$, which completes the proof.

**Claim 5** For any $p > \min \{ p_0^A(s), p_0^A(s) \}$, the difference $V_A(a, s, p) - V_C(a, \theta^A(s), p)$ satisfy the single-crossing property.

**Proof.** Follows from strict convexity of $V_A(a, s, p)$ and $V_C(a, s, p)$ as a function of $p$ together with $V_C(a, s, 1) > V_A(a, s, 1)$ and $V_C(a, s, 0) < V_A(a, s, 0)$.

Assume now that $p_i^A(s_1) \leq p_i^A(s_0)$ (the proof follows by contradiction). Then, by the single crossing property $V_A(a, s_1, p_i^A(s_0)) - V_C(a, \theta^A(s_1), p_i^A(s_0)) < 0$ (Claim 5) which violates Claim 4. Then, it follows that $p_i^A(s_1) > p_i^A(s_0)$.

Note that the only case where $p_i^A(s_1) = p_i^A(s_0)$ is when both thresholds are inactive (that is, equal to 1). The proof in this case is trivial.
G  Computing dropout, graduation, and transfer probabilities

To produce the different probabilities for each agent it is necessary to compute the probability of dropout, transfer and graduation at both academic 2-year colleges and 4-year colleges for each level of the initial prior $p_0$. A simple way of computing these values is via the Kolmogorov Forward Equation. To fixate ideas see the next example. Let $Pr^A(D|p, s)$ denote the dropout probability at $A$ with current prior $p$ and accumulated credits $s$. Also, define $\omega^A_\eta = pq^A_\eta(1) + (1 - p)q^A_\eta(0)$ as a set of weights. Finally, let $p'_\eta$ denote the posterior for grade $\eta$ and prior $p$. For $p \in [p^A_d(s), p^A_t(s)]$,

$$Pr^A(D|p, s) = \sum_\eta \omega^A_\eta \left[ \mathbb{I}\{p'_\eta > p^A_t(s')\} \ast 0 + \mathbb{I}\{p'_\eta < p^A_d(s')\} \ast 1 + \mathbb{I}\{p^A_d(s') \leq p'_\eta \leq p^A_t(s')\} \ast Pr^A(D|p'_\eta, s') \right]$$

with terminal condition $Pr^A(D|p, T^A) = 0$ for $p \in [p^A_d(T^A), p^A_t(T^A)]$.

Using this recursive equation it is possible to back-out $Pr^A(D|p_0, 0)$. This can be easily extended to all the other probabilities.

H  Fit of Model

The prior $p_0$ is positively correlated with the measure of ability $X'\beta$ obtained from the Ordered Probit regression (Table 11 and Table 12) so that the enrollment pattern generated by the model fits the empirical distribution obtained from NLS-72.

The model also has predictions regarding the dropout, transfer and graduation behavior of students. In particular, conditional on the initial enrollment choice, the model produces probabilities of different educational patterns as a function of the initial prior $p_0$, as shown in Figure 6. The initial prior $p_0$ affects the decision making of the student and the dynamic pattern in two different ways. First, it affects the likelihood of different educational histories as the distance to different threshold values changes with the prior. Second, the value of the prior is related to the likelihood of different signals as $p_0 = Pr[\mu = 1]$. In fact, the estimation of the model attempted to match the empirical marginal density of dropouts, transfers and graduation at both academic 2- and 4-year colleges.
Figure 6 has three different regions (the straight vertical lines separate the different regions). The first region, given for low values of the prior $p_0$, is for agents that join the workforce directly. The second region, the middle one, is for agents that enroll in academic 2-year colleges (average values for the prior) and the third region, the top one, is for agents that enroll in 4-year colleges. Conditional on the initial enrollment choice, Figure 6 presents the likelihood of each of the three possible events (i.e. drop, transfer or graduation) in the first spell of education for a student with a given initial prior $p_0$. The likelihood of graduation and dropping out in academic 2-year colleges are decreasing functions of the initial prior $p_0$ while the likelihood of transferring to 4-year colleges is increasing. For students that initially enroll in 4-year colleges, the likelihood of graduation increases with the prior while the likelihood of dropping out decreases with the prior. Another interesting aspect observed in Figure 6 is that students that transfer from 4-year colleges to 2-year colleges have above average priors (relative to students that enroll in 4-year colleges).\footnote{The shape of the figure is robust to different calibrations of the model. In particular, calibrations that match more closely dropout and transfers (in detriment of enrollment moments).}

Figure 6: Probability of Dropout, Transfer, and Graduation

The probabilities are computed according to the initial enrollment choice of an individual with initial prior $p_0$. The plot has three regions (from left to right). The first region is empty as individuals with low priors join the workforce. The second region is for individuals enrolling in academic 2-year colleges, while the third region is for individuals that enroll in 4-year colleges.
To evaluate the predictions of the model regarding the transition probabilities, Table 15 shows students by behavior (i.e. dropout, transfer, graduation) in the first spell of education and size of the measure of ability $X'\hat{\beta}$. For students initially enrolled in academic 2-year colleges, the pattern observed for dropout and transfer probabilities is similar to the model’s predictions. A similar thing happens for the dropout and graduation probabilities for students initially enrolled in 4-year colleges. Graduation probability in academic 2-year colleges and transfer probability in 4-year colleges are less revealing due to the low number of students included in these cells. Still, the evidence in these two cases does not conflict with the model’s predictions.

Table 15: Proportion of agents initially enrolled in $i$ with particular history

<table>
<thead>
<tr>
<th></th>
<th>Ac. 2-year C.</th>
<th>4-year C.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>D  T  G</td>
<td>D  T  G</td>
</tr>
<tr>
<td>Low $X'\beta$</td>
<td>77.3% 17.6% 5.1%</td>
<td>51.7% 1.7% 46.6%</td>
</tr>
<tr>
<td>Med $X'\beta$</td>
<td>66.1% 29.4% 4.5%</td>
<td>45.4% 2.4% 52.2%</td>
</tr>
<tr>
<td>High $X'\beta$</td>
<td>44.9% 50% 5.1%</td>
<td>35.6% 2.4% 62%</td>
</tr>
<tr>
<td># of obs.</td>
<td>332 192 26</td>
<td>392 47 474</td>
</tr>
</tbody>
</table>

These proportions are conditional on the initial enrollment status of the student. D: Drop, T: Transfer, G: Graduate. $X'\hat{\beta}$: estimator of the observable measures of ability. Recall that $p_0 = \frac{1}{1+e^{-X'\beta}}$ and that a student enrolls in 4-year college if her prior is high enough and in an academic 2-year college if it is average (so that she compares her initial prior with thresholds). Then, the enrollment problem can be estimated by using an Ordered Probit regression from where $\beta$ is estimated.

An alternative, and more involved way of evaluating the results on Figure 6 would be, conditional on the initial enrollment choice of agents, to do a nonparametric estimation of the densities of each of the different educational histories. Weighting these densities accordingly, it is possible to produce the empirical counterpart of Figure 6. First, the density associated with a given history is weighted by its share on initial enrollment in a given institution. For a given measure of ability $X'\hat{\beta}$, now it is possible to compute the proportion of agents that eventually end their first spell of education either by becoming dropouts, transferring or by graduation. Figure 7 presents the results. Notice how the patterns in Figure 6 (model) are very similar to those in Figure 7 (data).

I Other selected tables and figures
NLS-72 cohort. As a function of ability measure $X'\hat{\beta}$ and conditional on initial enrollment choice. First, a conditional (on the initial enrollment choice) nonparametric estimation of each event was performed. Next, each density was weighted by their share in initial enrollment. Finally, for every level of $X'\hat{\beta}$, the proportion of each event was constructed.

Table 16: Average Cost of Education

<table>
<thead>
<tr>
<th></th>
<th>Tuition</th>
<th>Tuition + R&amp;B</th>
</tr>
</thead>
<tbody>
<tr>
<td>vocational school</td>
<td>3803.84</td>
<td>5904.74</td>
</tr>
<tr>
<td>academic 2-year college</td>
<td>1811.36</td>
<td>2729.65</td>
</tr>
<tr>
<td>4-year college</td>
<td>3420.73</td>
<td>5038.80</td>
</tr>
</tbody>
</table>

R&B: Room and Board. Missing Values were imputed by running a cost regression and imputing missing values through observables. The values are measured in 1984 dollars. Source: NLS-72.
Figure 8: Transitions in ac. 2-year and 4-year colleges