Sorting of Students into Colleges: Inefficiencies and Policy Implications

Lutz Hendricks, Tatyana Koreshkova and Oksana Leukhina*

February, 2018

How efficient is the sorting of college students into colleges of varying quality? We develop a general equilibrium lifecycle model of human capital accumulation that allows us to tackle this question. Our framework explicitly features the variation in college quality, which we measure in the data as the average test score of the freshmen class. Higher quality colleges provide access to a superior human capital accumulation technology, but charge higher tuition and impose stricter standards on their students. We discipline this model by matching college quality choices, college credit accumulation histories, dropout and college transfer behavior, as well as earnings histories for different types of students in NLSY 1997 cohort. These data, when viewed through the lens of the student decision making, help us identify the human capital accumulation technologies and provides insight into student sorting among colleges of varying quality. We employ the calibrated model to quantify the importance of financial constraints in generating sorting inefficiencies and compare their impact on the evolution of student sorting among colleges between the 1979 and 1997 NLSY cohorts. We then ask which policy is most effective at improving upon outcomes, focusing on merit-based and need-based financial aid.

JEL Codes: E24, J24 (human capital), I21 (analysis of education).
*J*Keywords: Education, college, quality, college premium

*Hendricks: University of North Carolina, Chapel Hill, NC; Koreshkova: Concordia University, Montreal, Canada; Leukhina: Federal Reserve Bank of St. Louis, St. Louis, MO.

1. Introduction

How efficient is the sorting of college students among colleges of varying quality? We develop a general equilibrium lifecycle model of human capital accumulation that allows us to tackle this question. Our framework explicitly features the variation in college quality, which we measure in the data as the average test score of the freshmen class. We classify all colleges into 4 types according to this measure of quality, with the lowest type representing 2 year schools. Higher quality colleges provide access to a superior human capital accumulation technology, but charger higher tuition and impose stricter standards on their students: It takes a greater increment of human capital to pass a given course. Human capital accumulated in college translates into labor market productivity and therefore affects student earnings. We think of the main sources of student heterogeneity being human capital upon high school graduation, learning ability and financial variables.

We discipline this model by matching college quality choices, college credit accumulation histories, dropout and college transfer behavior, as well as earnings histories for different types of students from the NLSY 1997 cohort. To do so, we employ the restricted NLSY97 data which provides information on specific colleges attended and augmented with college transcript data. These data, when viewed through the lens of the student decision making, help us identify the human capital accumulation technologies and provides insight into how students of different types select into colleges of varying quality. More specifically, the important moments that we target are college entry rate by High School GPA and family income, the joint distribution of High School GPA and college quality among the college entrants, college graduation rates by High School GPA and college quality, lifetime earnings by GPA and by college quality for both graduates and dropouts, work hours while in college by High School GPA and college quality, college credits attempted and college credits earned by High School GPA and college quality.

Our quantitative model implies that high ability students have the most to gain from high quality colleges. The more stringent standards ensure they exert sufficient study effort and accumulate the optimal amount of human capital. Financial constraints and uncertainty about own learning ability create inefficiencies in sorting outcomes. For example, a financially constrained high ability student may choose to start at a low quality college because it is cheaper and allows him to work while studying without sacrificing the pace of credit accumulation. This choice will hamper his human capital accumulation and delay entry into the labor market. Likewise, a financially unconstrained weaker student may start at a college of higher than optimal quality, learn that he is unable to make sufficient progress and end up dropping out, thereby wasting time, financial resources and forgoing human capital accumulation at a better matched school.

We employ the calibrated model to examine the importance of financial constraints in generating inefficiencies in matching between students and colleges. We also perform a counterfactual experiment where we feed in the financial constraints and borrowing opportunities of the NLSY

79 cohort into the benchmark model in order to analyze the evolution of student-college sorting between the two NLSY cohorts. We then ask what kind of policies could improve the efficiency of matching, focusing specifically on merit-based financial aid and need-based financial aid.

The rest of the paper is organized as follows. In Section 2, we introduce the model. In Section 3, we explain the calibration strategy, discuss the model fit and explain which empirical moments help identify the human capital production technology for colleges of varying quality. We present our main results in Section 4. We conclude in Section 5.

2. Model

2.1. Endowments and Timing

The economy is populated by overlapping generations of individuals who enter the economy at high school graduation (age $t = 1$) and live until age T. During the first $T^c = 6$ periods, individuals may pursue college education. They have to retire at age T^R.

Colleges Human capital is accumulated in college stochastically but can be affected by study effort. Before the start of period 1, individuals decide whether to study in college or forego college education altogether. There are four types of colleges, $q \in \{1, 2, 3, 4\}$ that differ by quality. The first type represents 2-year academic programs at community colleges. This college type does not grant a college degree but allows students to work full time. Students can study there for a maximum of two years. Four-year colleges grant degrees after a minimum of four years of studies. Only part-time employment is feasible during studies at a 4-year college. Graduation from a 4-year college requires that the student's human capital attains a minimum level. College types differ in (1) tuition fees, 2) quality, represented by the productivity term in human capital production, and (3) credit-pass standards, modeled as the minimum human capital increment required to pass a given course.

Students are allowed to transfer from one college type to another at the end of the second year of studies ($t = 2$).

Student Endowments An individual is born with a permanent type j. The type determines learning ability a, parental transfers conditional on student choice of entering college (C) or working as a high school graduate (H), denoted by $\{b_j^H(t), b_j^C(t)\}_{t=1}^{T^c}$ and the type-specific tuition component $w_j(t)$. In addition, individuals differ in their initial human capital h_0 and assets k_0. A high school graduate of type j knows his type except the true ability a. He observes ability signal, $\tilde{a}$, labor market productivity draw, $\boldsymbol{\chi}$, and preference shocks for each college type $\{p_q\}_{\forall q}$ and work choice p_w. He then decides whether to work or enter a college of type q.

High school graduates who choose work learn their true ability upon labor market entry. Workers draw productivity shocks in the beginning of each period and make consumption, saving and labor supply decisions.

High school graduates that enter college are committed to staying in college for at least one full period (a year). Upon college entry, the student learns his true ability a. In the beginning of each period, college students choose their course loads, study effort and make their consumption/savings decisions. At the end of every period, students observe their human capital shock which together with the effort choice determine the new stock of human capital, they draw their labor market productivity and college preference shocks, and decide whether to drop out of college or continue. If the individual drops out from college, he becomes a worker with a schooling level $s = D$. The student graduates as soon as he accumulates the required number of credits after a minimum 4 years of study ($s = C$).

Timing of events during college stage:

1. In the beginning of the first period of college, the true ability is observed by the student;
2. Course loads, study effort, work time, savings and consumption decisions are made;
3. The human capital shock is observed. The new human capital stock and cumulative college credits are then determined;
4. Preference shocks and labor market productivity shocks are observed;
5. Schooling decisions (stay/transfer/drop out) are made.

The model features college drop-out risk and labor market risk. College dropout risk occurs due to incomplete information about ability and stochastic human capital accumulation in college. There are no insurance markets for these risks.

Borrowing Constraints Individuals are allowed to borrow up to a limit.

2.2. Worker's Problem

The current version of the model assumes that worker's labor supply and experience accumulation are exogenous but conditional on schooling level. Worker's wage rate per hour of work is a function of schooling, human capital, experience (given by age)[1] and is subject to shocks.

$$\ln y(t, h, s, \boldsymbol{\chi}) = \ln w^s + \gamma^s h + x(t, s) + z_t + u_t$$

[1] We assume that students do not forego experience accumulation on the job while in college. This eliminates the need to track labor supply of students and the age of college exit.

where w_s is skill-specific wage rate, γ^s is return to human capital, $x(s,t)$ is education-specific return to experience, $\boldsymbol{\chi} \equiv (z_t, u_t)$ is a productivity shock consisting of a persistent component z_t and a transitory component u_t. The persistent component follows an age-invariant Markov process with transition probabilities $\pi_{zz'}$ and initial distribution Π_z; the transient component is iid with probability distribution Π_u.

Worker's state is given by his age t, savings k, human capital h, and productivity shocks $\boldsymbol{\chi}$, schooling level s and permanent type j.

The worker chooses savings by solving

$$V^W(t,k,h,\boldsymbol{\chi};s,j) = \max_{k'} \left\{ u(c, 1-v^s) + \beta \mathbf{E}_{\boldsymbol{\chi}} V^W(t+1,k',h,\boldsymbol{\chi};s,j) \right\} \tag{1}$$

subject to

$$c = (1+r(k))k - k'^s - \tau^w(r(k)k, yv^s) + b_j(t) \tag{2}$$

$$y = w^s \exp \gamma_s h + x(t,s) + z + u$$

$$r(k) = \begin{cases} r^K & \text{if} \quad k \geq 0 \\ r^D & \text{if} \quad k < 0 \end{cases}$$

$$k' \geq \min\{\underline{k}^w, k\} \tag{3}$$

$$v \in [0,1]$$

given $\{k,h\} = \{k_j, h_j\}$ at $t=1$.

Interest rate on debt, $r(k<0) = r^D$, may differ from the interest on savings, $r(k \geq 0) = r^K$.

Constraint (3) makes borrowing opportunities differ for college students and workers. The tax function $\tau^w(rk, yv^s)$ may include tax deduction for student debt. The current version of the code sets $r^D = r^K$ and the borrowing limit for all individuals (i.e. constraint (3) does not apply).

At $t = T^R - 1$, the expected value term in (1) is replaced with the value of the retiree.

2.3. College

Individual state: $(t,i;j)$ where $i \equiv (k,h,n,q)$ is the dynamic state.

Each period a student attempts to pass $\bar{n}$ credits, chooses study effort $\varepsilon \in [0,1]$ and work time ν. After the student makes his decisions, the human capital shock ξ is realized, which together with ability, college quality and current human capital h determines the increment of human capital attained per credit attempted, $F(\varepsilon, h, q, a)$. The new human capital level h' is then

$$\begin{aligned} h' &= h + \Delta h \\ n' &= n + \Delta n \\ \Delta h &= F(\varepsilon, h, q, a)\bar{n} \end{aligned}$$

We assume that course passing rates (which is what we observe in college transcript data) depend on human capital accumulation:

$$\begin{aligned} Pr(pass) &= P(\Delta h, h, \bar{n}) \\ \Delta n &= round\left(\bar{n}Pr(pass)\right) \end{aligned}$$

We make the following functional form assumptions:

$$\begin{aligned} \frac{\Delta h}{\bar{n}} &= A(h\varepsilon)^{\alpha}\left(\frac{\varepsilon}{\tilde{n}} - \bar{\varepsilon}\right)^{\beta} e^{\phi(a+\xi)+(1-\phi)q} M(a,q), \\ Pr(pass) &= \left[1 + exp(-\nu(\frac{\Delta h}{\bar{n}} - \bar{h}_q))\right]^{-1}, \end{aligned}$$

where $M(a, q) = 1 - \gamma(a - q)^2$ is ability-college mismatch penalty.

College Quality There are q types of college quality. Each quality type corresponds to a combination of human capital productivity $A(q)$, the minimum human capital required for graduation $h^{grad}(q)$, and a tuition scale $\lambda(q)$. Type $q = 1$ is designated for two-year community colleges. Students are allowed to work full time only in community colleges (hours are constrained by $\hat{v}(q)$). Since in the model college choice can only be made at entry and then after 2 years, the set of available college types depends on the age of the student: $Q_0 = \{1, 2, 3, 4\}$ and $Q_2 = \{2, 3, 4\}$. In other periods, the college type remains the same ($Q_t = q$).

We assume that college quality preference shocks ξ shocks are iid across agents and time. They are drawn from distribution Π_ξ.

The value of studying at type$q$ college is given by

$$V^C(t, i; j) = \max_{c,k',\tilde{n},v,\varepsilon} \left\{ u(c, 1 - v - \varepsilon) + \beta \sum_{\xi} \Pi_\xi V^{EC}(t + 1, i'; j) \right\} \quad (4)$$

subject to

$$
\begin{aligned}
c &= (1+r(k))k - k' + y_c(v) - \tau(r(k)k, y_c(v)) + b_j(t) - \lambda\omega(t), && (5)\\
h' &= h + \Delta h,\\
\Delta h &= H(q, \tilde{n}, h, \varepsilon, a, \xi),\\
\tilde{h} &= \varphi h + (1-\varphi)\Delta h,\\
n' &= n + N(q, \tilde{n}, \tilde{h}),\\
1 &\geq v + \varepsilon,\\
\bar{v}(q) &\geq v,\\
k' &\geq \underline{k}^c, && (6)\\
i' &= (k', h', n', q),
\end{aligned}
$$

and $k = k_0$ at $t = 1$.

For $t \neq 2$, the expected value at the start of the next period is

$$
\begin{aligned}
&V^{EC}(t+1, i'; j) = && (7)\\
&\mathbf{E}_{\boldsymbol{\chi}_0}\mathbf{E}_{\mathbf{p}} \max\left\{V^W(t+1, k', h', \boldsymbol{\chi}_0; D, j) + \pi p^w, V^C(t+1, i'^c\right\} - \pi\bar{\gamma},
\end{aligned}
$$

where p^w and p^c are iid preference shocks drawn from an extreme value type I distribution.

At $t = 2$, the next-period value is

$$
\begin{aligned}
&V^{EC}(t+1, i'; j) = && (8)\\
&\mathbf{E}_{\boldsymbol{\chi}_0}\mathbf{E}_{\mathbf{p}} \max\left\{V^W(t+1, k', h', \boldsymbol{\chi}_0; s, j) + \pi p^w, \max_{q' \in Q_t} V^C(t+1, i''^c_q - m\mathbf{I}_{q' \neq q}\right\} - \pi\bar{\gamma},
\end{aligned}
$$

where

$$
\begin{aligned}
i'' &= (k', h', n', q')\\
\hat{k} &= k' - \frac{\theta}{1+r(k')}\mathbf{I}_{q' \neq q}
\end{aligned}
$$

reflects the college transfer costs. In the current code, only the utility cost of moving is used (m). p^w and $p^c_{q \in Q}$ are iid preference shocks drawn from an extreme value type I distribution.

There are three possible cases:

1. Graduate regardless of $(\boldsymbol{\chi}_0, \mathbf{p})$ shocks: If $n \geq n^{grad}$, then $s = C$ and $V^C = -\infty$

2. Drop out regardless of $(\boldsymbol{\chi}_0, \mathbf{p})$ shocks: If $t = T^c$ and $n < n^{grad}$, then $s = D$

3. Choose between dropping out and staying in college.

In the latter case, the probability of dropping out of college at the start of period t is

$$Pr(s = D|t, k, h, q, \boldsymbol{\chi}_0; j) = \frac{\exp\{V^W(t, k, h, \boldsymbol{\chi}_0; D, j)/\pi\}}{\exp\{V^W(t, k, h, \boldsymbol{\chi}_0; D, j)/\pi\} + \sum_{q' \in Q_t} \exp\{V^C(t, \hat{k}, h, q'; j)/\pi\}},$$

where $Q_{t=2} = \{2, 3, 4\}$ and $Q_{t \neq 2} = q$.

At $t = 2$, the probability of transfer to college quality $q' \in Q$

$$Pr(q'|t, k, h, q, \boldsymbol{\chi}_0; j) = \frac{\exp\{V^C(t, \hat{k}, h, q'; j)/\pi\}}{\exp\{V^W(t, k, h, \boldsymbol{\chi}_0; CD, j)/\pi\} + \sum_{q' \in Q_t} \exp\{V^C(t, \hat{k}, h, q'; j)/\pi\}}.$$

The value in college at the start of the period t for case 3:

$$V^{EC}(t, k, h, q; j) = \mathbf{E}_{\boldsymbol{\chi}_0} \pi \ln \left[\exp\{V^W(t, k, h, \boldsymbol{\chi}_0; CD, j)/\pi\} + \sum_{q' \in Q_{t-1}} \exp\{V^C(t, \hat{k}, h, q'; j)/\pi\} \right].$$

2.4. College Entry Decision

A newborn who believes his type to be $\tilde{j}$, observes his college and work preference shocks (p_w, p_c) and his labor market shock $\boldsymbol{\chi}_0$ and decides whether to go to college and, if yes, of what quality. The life-time utility of type-$\tilde{j}$ individual is

$$V_0(k, h, \boldsymbol{\chi}_0, \tilde{j}) = \mathbf{E}_{\mathbf{p}} \max \left\{ \mathbf{E}_{\tilde{j}} V^W(1, k, h, \boldsymbol{\chi}_0; H, j) + \pi p^w, \max_{q \in Q_0} \mathbf{E}_{\tilde{j}} V^C(1, k, h, 0, q; j) + \pi p^c_q \right\} - \pi \bar{\gamma}, \tag{9}$$

where the first expectation is taken over preference shocks, the second expectation is taken over true ability types, conditional on signal, and $Q_0 = \{1, 2, 3, 4\}$. Dynamic state in college is $i = (k, h, q, 1)$.

The choice probability of no college can be analytically derived to be

$$Pr(s = H|t = 0, k, h, \boldsymbol{\chi}_0; \tilde{j}) = \frac{\exp\{\mathbf{E}_{\tilde{j}} V^W(1, k, h, \boldsymbol{\chi}_0; H, j)/\pi\}}{\exp\{\mathbf{E}_{\tilde{j}} V^W(1, k, h, \boldsymbol{\chi}_0; H, j)/\pi\} + \sum_{q \in Q_0} \exp\{\mathbf{E}_{\tilde{j}} V^C(1, i; j)/\pi\}}.$$

Likewise, the probability of entering college of quality q is given by

$$Pr(q|t = 0, k, h, \boldsymbol{\chi}_0; \tilde{j}) = \frac{\exp\{\mathbf{E}_{\tilde{j}} V^C(1, i; j)/\pi\}}{\exp\{\mathbf{E}_{\tilde{j}} V^W(1, k, h, \boldsymbol{\chi}_0; HS, j)/\pi\} + \sum_{q \in Q_0} \exp\{\mathbf{E}_{\tilde{j}} V^C(1, i; j)/\pi\}}$$

The expected welfare of a high-school graduate of type $\tilde{j}$ is given by

$$V_0^E(\tilde{j}) = \mathbf{E}_{\boldsymbol{\chi}_0} \pi \ln \left[\exp\{\mathbf{E}_{\tilde{j}} V^W(1, k, h, \boldsymbol{\chi}_0; H, j)/\pi\} + \sum_{q \in Q_0} \exp\{\mathbf{E}_{\tilde{j}} V^C(1, i; j)/\pi\} \right].$$

2.5. Retirement

Individuals retire exogenously at age T^R. There is no uncertainty. State of a retired individual includes age, asset and schooling. Retired individuals derive income from a social security benefit $y^R(s)$ and an interest on savings, pay income taxes, and allocate the remainder between consumption and savings:

$$V^R(t, k; s) = \max_{k', v} \left\{ u(c, 1) + \beta V^R(t+1, k'; s) \right\} \tag{10}$$

subject to

$$\begin{aligned} c &= (1 + r(k))k - k'^R(s) - \tau(k, y^R(s)) \\ k' &\geq \min\{0, k\}. \end{aligned} \tag{11}$$

Borrowing constraint in the current code: $k' < \underline{k}$.

3. Calibration

Specifically, we target college entry rate (by High School GPA and family income), the joint distribution of High School GPA and college quality among the college entrants, graduation rates by High School GPA and college quality, lifetime earnings by GPA and by college quality for both graduates and dropouts, work hours while in college by GPA and college quality, credits attempted and earned by GPA and college quality, among others.

4. Results

To be finalized and written up.

5. Conclusions

To be finalized and written up.