Peer Effects, Financial Aid, and Selection of Students into Colleges and Universities: An Empirical Analysis*

Dennis Epple

Carnegie Mellon University and NBER

Richard Romano

University of Florida

Holger Sieg

Duke University and NBER

March 16, 2001

*We would like to thank Kerry Smith, an anonymous referee, and seminar participants at the 2000 American Economic Association Meetings and the January 2000 Brookings Conference on Empirical Analysis of Social Interactions for comments. We thank Gary Gates for expert research assistance. We would also like to thank the National Science Foundation and the MacArthur Foundation for financial support and the National Center for Educational Statistics and Petersons for providing us with the data used in this paper.

Abstract

This paper develops a model in which colleges seek to maximize the quality of the educational experience provided to their students. We deduce predictions about the hierarchy of schools that emerges in equilibrium, the allocation of students by income and ability among schools, and about the pricing policies that schools adopt. The empirical findings of this paper suggest that there is a hierarchy of school qualities which is characterized by substantial stratification by income and ability. The evidence on pricing by ability is supportive of positive peer effects in educational achievement from high ability at the college level.

Keywords: higher education, peer effects, school competition, non-linear pricing.

JEL classification: I21, C33, D58

1 Introduction

It is easy to think of many ways in which peer student quality might be important in educational settings. A student surrounded by able and motivated peers may benefit from higher quality in-class discussions, help outside of class in understanding difficult material, stimulating bull sessions, role models that encourage conscientious completion of homework assignments, and competition that fosters thorough preparation for examinations. More able and motivated students might also increase productivity of teachers. While it is quite plausible that such peer effects are present, it is not an easy matter to measure them. Fortunately, however, peer effects in education can be expected to have measurable effects on market outcomes. Our goals in this paper are to develop predictions regarding market consequences of peer effects in education and to offer empirical evidence about the extent to which those predictions are borne out in the data.

If peer quality does, in fact, provide educational benefits, then students and their parents can be expected to seek out schools where the student body offers high quality peers. Likewise, schools that wish to be ranked highly will attempt to attract students who contribute to improving peer quality. In higher education, schools have the latitude to choose price and admission policies to attempt to attract a high quality student body. We present a model in which schools seek to maximize the quality of the educational experience provided to their students. The quality of the educational experience depends on peer ability of the student body and on instructional expenditures per student. From this model we deduce predictions about the hierarchy of schools that emerges in equilibrium, the allocation of students by income and ability among schools, and about the pricing policies that schools adopt. These predictions are the subject of our empirical investigation.

Related research has investigated normative and positive consequences of competition in primary and secondary education, and the likely effects of policy changes including vouchers, public school choice, and changes in education financing.[1] Related research on higher education has investigated the payoffs associated with higher education (see Dale and Krueger (1998) and references therein), coordinated behavior (Carlton, Bamberger, and Epstein, 1995; Netz, 1998), market evolution (Goldin and L. Katz, 1998); market structure and tuition (Hoxby, 1997, 1999), and pricing, peer effects, and efficiency (Rothschild and White, 1995). The works by Hoxby (1997), Rothschild and White (1995), and Epple and Romano (1998, 1999, 2000) are closest in spirit to this research. Hoxby (1997) considers the effects of changing market structure of higher education on tuition over the period from 1940 to 1991. Hoxby (1999) continues the investigation of market structure, with emphasis on the way in which pricing by selective colleges and universities changed in response to antitrust action brought against private colleges for price fixing. We consider pricing and selection among colleges and universities throughout the school hierarchy with an emphasis on investigating the implications of our model of school competition and pricing.

Our work shares with Rothschild and White an interest in pricing in the presence of peer effects. Our work differs in its: consideration of students differentiated by both ability and

[1] See among others, Arnott and Rowse (1987), Bergstrom, Rubinfeld, and Shapiro (1982), Caucutt (1998), deBartolome (1990), Epple and Romano (1998, 1999, 2000), Evans, Murray, and Schwab (1998), Fernandez and Rogerson (1996, 1998), Fuller, Manski, and Wise (1982), Glomm and Ravikumar (1992), Hanushek, Kain, Markman, and Rivkin (2000), Hoxby (1996), Manski (1991), Nechyba (1998, 1999, 2000), Toma (1996), Venti and Wise (1982), and Zimmer and Toma (1998).

household income, focus on the implications of quality maximization, development of implications for allocation of students by ability and income across colleges, and development of implications about the ordering of endowments, inputs, student ability, and household income across colleges. The most significant differences from Epple and Romano's (1995, 1998, 1999) work on primary and secondary education are the alternative objectives of quality rather than profit maximization, the introduction of endowments, and the empirical analysis of colleges and universities

The rest of the paper is organized as follows. Section 2 outlines the equilibrium model and relates a number of results which characterize allocations in equilibrium. Section 3 discusses our data sources and provides some descriptive statistics of our sample. Sections 4 and 5 present the main empirical finding of this study. Section 6 concludes the analysis.

2 A Theoretical Model of Higher Education

In this section, we sketch our theoretical model of provision of undergraduate education. The model is developed in more detail in Epple, Romano, and Sieg (1999) where proofs of the results are provided.

2.1 Preferences and Technologies

There is a continuum of students who differ with respect to their income, y, and their ability level, b. Each student chooses among a finite set of J schools. The quality of school j is given by

$$q_j \;=\; I_j^{\omega}\theta_j^{\gamma} \qquad \omega,\ \gamma > 0 \tag{1}$$

where θ_j is the peer-student measure, equal to mean ability level in the student body, and I_j is the expenditure per student in excess of custodial costs. The schooling cost function is

$$C(k_j, I_j) \;=\; F \;+\; V(k_j) \;+ k_j\; I_j \qquad\qquad V', V'' > 0 \tag{2}$$

where k_j is the size of the school j's student body. In the empirical implementation of the model we assume that variable custodial costs are quadratic in k_j, i.e. $V(k_j) \;=\; V_1\; k_j \;+\; V_2\; k_j^2$.

We assume that the decision to attend college is made by the student's household. Household utility from attendance at school j is given by:

$$U(y - p_j, a(q_j, b_j)) \;=\; (y - p_j)\; a_j \tag{3}$$

where $a_j \;=\; q_j b^{\beta}$ is achievement of the household's student and p_j is tuition. Below we introduce non-school financial aid into the analysis. A choice of not attending school is equivalent to q being equal to a given low value q_0, and with $p = 0$. The joint distribution of income and ability is continuous with joint density $f(b, y)$.

Schools are assumed to maximize their quality. They must satisfy a profit constraint, with revenue equal to the sum of all tuition from students and earnings on exogenous endowment. Denote the latter earnings R_j. R_j also includes any other non-tuition revenues like state subsidies. While schools will condition tuition on student characteristics, we presume that school j charges a maximum tuition denoted p_j^m. We do not have an explicit theory to explain or determine the magnitude of p_j^m so we treat it as exogenous. Our motivation for introducing price caps is empirical. We interpret the price maximum as the school's marketed tuition, with lower tuition framed as financial aide, a scholarship, or, perhaps, a fellowship. Having a price cap below the maximum tuition chosen by an unconstrained quality maximizer might help a school market itself. We have assumed, however, that households observe all prices relevant to them, so our argument for price caps is somewhat incomplete. One can also conceive of the self-imposed price cap as reflecting some limit on revenue making, whether motivated by altruism or, again, related to marketing. One can also link the price cap to a school's cost as further discussed below. While treating the price caps as exogenous is not ideal, it is comforting that they lead to empirically realistic predictions as we will see.

2.2 School Optimization

Schools take types' (b, y) utilities as given. School j's optimization problem may be written:

$$\max \quad q_j = \theta_j^\gamma I_j^\omega \tag{4}$$

$$s.t. \quad [y - p_j^r]\, \theta_j^\gamma I_j^\omega b^\beta = U^a(b, y) \quad \forall\, (b, y) \tag{5}$$

$$p_j(y, b) = \min\{p_j^m, p_j^r(b, y)\} \quad \forall\, (b, y) \tag{6}$$

$$\int\int p_j(y,b)\ \alpha_j(y,b)\ f(b,y)\ db\ dy\ +\ R_j\ =\ F\ +\ V(k_j)\ +\ k_j\ I_j \tag{7}$$

$$\alpha_j(y,b)\ \in\ [0,1]\ \ \forall\ (b,y) \tag{8}$$

$$k_j\ =\ \int\int\ \alpha_j(y,b)\ f(y,b)\ dy\ db \tag{9}$$

$$\theta_j\ =\ \frac{1}{k_j}\int\int\ b\ \alpha_j(y,b)\ f(y,b)\ dy\ db \tag{10}$$

where here and henceforth integrals are over the support of (b, y) unless otherwise indicated. In (5), $U^a(b, y)$ is the maximum alternative utility available to type (b, y) in equilibrium, and thus $p_j^r(b, y)$ is type (b, y)'s reservation price for attending school j.[2] In expressions (6) and (7) we have built into the problem the obvious result that all types who attend school j will pay the minimum of their reservation price or the tuition cap. The function $\alpha_j(y, b)$ is an admission function that indicates the proportion of type (b, y) that school j admits. The upper bound of 1 on $\alpha_j(b, y)$ requires that schools can admit no more of a type than exists.[3] In equilibrium, admission sets and attendance sets will coincide. The constraints (9) and (10) define k_j, school size, and θ_j, the peer group measure, respectively.

Key first-order conditions that describe school j's admission policy may be written:[4]

$$\alpha_j(b,y)\begin{pmatrix} =1 \\ \\ \in[0,1] \\ \\ =0 \end{pmatrix}\quad \text{as}\ \ p_j(b,y)\begin{pmatrix} > \\ \\ = \\ \\ < \end{pmatrix}\ \text{EMC}_j \tag{11}$$

[2] Note that the reservation price depends on q_j, but we use the more compact notation subscripting the function with j.

[3] Hence it is innocuous to have specified that (6) holds for all students, i.e. , including those that will not be admitted.

[4] See Epple et al. (1999) for details.

where

$$\text{EMC}_j \; = \; V'(k_j) \; + \; I_j \; + \; \frac{\partial q_j/\partial\theta}{\partial q_j/\partial I}\,(\theta_j - b) \tag{12}$$

Equation (12) defines the "effective marginal costs (EMC)" of admitting a student of ability b to school j. EMC is the sum of the marginal resource cost of educating the student and the cost of maintaining quality due to the student's impact on the peer group. The latter effect is captured by the last term in (12), which equals the peer measure change from admitting a student of ability b, multiplied by the expenditure change that maintains quality. Note that this term is negative for students with ability above the school's mean, and EMC_j itself can be negative. Students whose maximum feasible tuition exceeds EMC_j permit quality increases and are all admitted, and the reverse for students who cannot be charged a tuition that covers their EMC_j (see (11).

Define

$$\alpha_j^r(b,y) \;\; = \;\; \begin{cases} \alpha_j(b,y) & \text{if} \;\; p_j(b,y) = p_j^r(b,y) \\ 0 & \text{otherwise} \end{cases} \tag{13}$$

Let, then, $A_j^r \;=\; \{(b,y) \mid \alpha_j^r(b,y) \;>\; 0\}$ denote the set of students that attend school j and pay their reservation price, and $A_j^m \;=\; \{(b,y) \mid \alpha_j(b,y)\;\; 0 \text{ and } \;(b,y) \;\notin\; A_j^r\}$ denote the remaining students that attend j. Let $A_j = A_j^r \cup A_j^m$ denote school j's admission and attendance sets of student types.[5]

[5]The market-clearing condition presented below can be used to show that schools' attendance sets do not overlap with positive measure in the support of (b,y).

2.3 Properties of Market Equilibrium

In market equilibrium, households choose among the J schools or choose no school, taking school qualities and tuition and admission policies as given. The J schools choose admission and tuition policies to maximize quality, taking as given their endowment and students' alternative utility possibilities.[6] The model is closed with the market clearing condition: $\sum_{j=1}^{J} \alpha_j(b, y) \leq 1 \quad \forall\ (b, y)$, where types for whom the inequality is strict are attending no school.

We now describe key properties of equilibrium.[7] One assumption we place on price caps is that schools of equal quality have the same price caps:

Assumption 1 $q_i = q_j \;\Rightarrow\; p_i^m = p_j^m$

With this assumption, we have proved the following proposition:

Proposition 1 *Strict Hierarchy:*

Equilibrium has a strict hierarchy of schools: $q_1 < q_2 < ... < q_{J-1} < q_J$*, that follows the endowment hierarchy:* $R_1 < R_2 < ... < R_{J-1} < R_J$.

Proof: See Epple et al. (1999).

---

[6] The maximum alternative utility is computed using Lemma 1 below. The assumption of utility taking is a generalization of price taking that has been utilized in the competitive club goods literature. See, for example, Gilles and Scotchmer (1997).

[7] We have not developed a general existence proof, but we have shown existence in some examples. Here we assume existence and focus on necessary properties of equilibrium.
Using (7) and (11), one can show that schools will be below the scale that minimizes average cost and that it is likely schools with higher endowment will be smaller. These results depend, however, on our presumption that schools have the same cost function., in particular, the same "efficient scale." Because the latter assumption is made for convenience and is not realistic, we do not take seriously the size predictions. hence we make no attempts to explain size empirically.

The intuition of the strict hierarchy traces to increased willingness of households to pay for quality as income rises. If equivalent quality schools were to exist, then either school could engineer a quality increase by reformulating the school with student body consisting of relatively higher income and higher ability types from the two schools' initial student bodies. Not only could this be done while maintaining budget balance, but, because the school would be richer and have better peer group, tuition could be set to relax the profit constraint. This would allow quality increases beyond the improved peer group by increasing expenditures on educational inputs. The implied quality hierarchy must follow the endowment hierarchy. All schools maximize quality and better endowed ones can spend more on inputs and can give steeper discounts to higher ability students.[8]

School j's price cap implies a minimum ability necessary for admission to j, i.e., regardless of a type's income and reservation price. Using (6), (11) and that EMC_j is decreasing in b, quality maximization dictates that $b \geq b_j^m$ is necessary for admission, where $p_j^m = \text{EMC}_j(b_j^m)$ defines b_j^m. Using (12) and that $\frac{\partial q_j/\partial \theta}{\partial q_j/\partial I} = \frac{\gamma I_j}{\omega \theta_j}$, we can write:

$$b_j^m = \theta_j \left[1 + \frac{\omega \left(I_j + V_j' - p_j^m \right)}{\gamma I_j} \right] \tag{14}$$

In most cases, θ_j and I_j will rise with q_j as discussed below. It is then likely that b_j^m will increase with school quality. For example, a plausible approximation is that p_j^m is a fixed

[8]Another result is that quality maximization leads schools to spend more than Pareto efficient amounts on inputs. Schools can get away with this because their equilibrium differentiation leads them to have some market power.

mark up over the marginal resource cost of educating a student, i.e., $p_j^m = M\ (I_j + V')$ with $M > 1$.[9] Then, from (14), b_j^m is an increasing function of θ_j and I_j. It facilitates the analysis to simply assume:

Assumption 2 $\qquad b_1^m < b_2^m < ... < b_{J-1}^m < b_J^m$

Assumption 2 will allow us to show equilibrium is characterized by stratification by income and ability across the school hierarchy. Two lemmas useful for this are presented next.

Lemma 1 *Among the set of schools $S(b)$ for which a student (b, y) qualifies, $S(b) = \{j \mid b \geq b_j^m\}$, equilibrium school attendance conforms to that if schools in $S(b)$ set tuition equal to $EMC_j(b)$.*

Proof: See Epple et al. (1999).

Lemma 1 essentially follows from the admission condition (11). Since any school wants a student who is willing to pay their EMC, equilibrium (i.e. market clearing) requires that students have access to schools that they do not attend at EMC. The school j that the student attends may set tuition exceeding EMC_j as discussed below, but this is due to strict preference if $p_j = \text{EMC}_j$ for school j. Perhaps clearer intuition is that the price discrimination over income that occurs in equilibrium is "perfect" or of the "first degree," hence the allocation is consistent with social marginal cost pricing (income effects aside).

For part of Lemma 2 below, it is convenient to assume that the coefficients on $(\theta_j - b)$

[9] We still do not, however, let schools optimize over p_j^m.

in EMC_j (see (12)) weakly ascend along the school hierarchy. This may be written:

Assumption 3 $\frac{I_1}{\theta_1} < \frac{I_2}{\theta_2} < .... < \frac{I_J}{\theta_J}$

While Assumption 3 is not immediately intuitive, there are several ways it can be defended. One interpretation is that the assumption implies that student expenditure rises more quickly than the quality of the peer group as one moves up the hierarchy. The latter is consistent with the data using conventional measures of peer quality (i.e. average SAT scores). As will be evident in several tables presented in Section 3 below, per student expenditure rises more rapidly than mean SAT as one moves up the hierarchy of schools. A theoretical interpretation of Assumption 3 that one can glean from the expression for EMC is that it conforms to a rising marginal value of peer group improvements along the school hierarchy. Not only is this intuitive, but we have consistently found this in related computational analysis of equilibria. Last, we will see in the Proof of Lemma 2 (that we then retain) that the Assumption is much stronger than needed for the theoretical results.

Lemma 2 *Let $S(b)$ denote the set of schools for which student of ability b qualifies, $S(b) = \{j \ |b \geq b_j^m\}$. Let $P(S_0)$ denote a set of students who qualify for the same set of schools, $P = \{(b, y) \ |S(b) = S_0\}$. (Note that (A2) implies P contains all types with $b \in [b_j^m, b_{j+1}^m)$ for some $j \in 1, .., J-1$ and that there are J sets $P(S_0)$.) P is characterized by income and ability stratification across S_0.*[10]

Proof:

[10]By income stratification we mean that, for any fixed b, $y_2 > y_1$ implies the school (b, y_2) attends is of weakly higher quality than the school type (b, y_1) attends, and strictly if different schools are attended. Ability stratification is defined analogously for fixed y.

By Lemma 1, the partition of P into schools is as though $p_j = \mathrm{EMC}_j$ for all $j \in S_0$. Let $U_j = (y - \mathrm{EMC}_j(b))\ q_j b^\beta$ denote utility from attending school j under EMC pricing. Consider any schools j and i in S_0 such that $q_j > q_i$. Then:

$$\frac{\partial(U_j - U_i)}{\partial y} = (q_j - q_i)\, b^\beta > 0 \tag{15}$$

which implies income stratification. Similarly:

$$\frac{\partial(U_j - U_i)}{\partial b} = \frac{\beta}{b}(U_j - U_i) + (\eta_j\, q_j - \eta_i q_i)\, b^\beta \tag{16}$$

where $\eta_j = (\partial q_j / \partial \theta_j)\ /\ (\partial q_j / \partial I_j)$. Using (A3), for b_1 such that $U_j > U_i$, it follows that $U_j > U_i$ for all $b > b_1$. This implies ability stratification. Q.E.D.

We can now establish two key properties of equilibrium allocations:

Proposition 2 *Stratification by Income and Ability: The equilibrium allocation is characterized by income and ability stratification across the hierarchy of schools.*

Proof: See Epple et al. (1999).

Figure 1 depicts an equilibrium allocation in a case with three schools. The solid lines are boundary loci separating student bodies (which never meet in this example). The lower, downward-sloping segments of the boundary loci are indifference loci between adjacent schools under EMC pricing (as justified below). The vertical portions follow the ability minima in each school. The dashed lines demarcate the subsets of each school's student body that pay the price cap or their (lower) reservation price.

Insert FIGURE 1 here

Proposition 3 reports central properties of the equilibrium tuition structure.

Proposition 3 *Along boundary loci, tuition at the school attended (either adjacent school) equals EMC and thus depends only on ability. In the interior of admission sets, tuition at the school attended exceeds EMC, and depends then in part on the student's household income. In any school, for given income, tuition decreases weakly in ability and strictly if the student does not pay his school's price maximum and has another school (rather than no college) as his best alternative.*

Proof: See Epple et al. (1999).

Insert FIGURE 2 here

Figure 2 provides additional information relevant to the determination of prices for the equilibrium in Figure 1. Here the downward sloping portions of the dashed lines are indifference loci for the alternative schools under EMC pricing by a corollary to Lemma 1. The numbers within the admission spaces indicate the student's best alternative choice of school (or 0 for no college). One can employ stratification implications of preferences and Lemma 1 to establish the latter properties of the equilibrium partition. Take the subset of students with $b \geq b_3^m$, for example. They face no restriction on school attended and prices at alternative schools equal EMC. It is straightforward to establish in a case like that in Figures 1-2 that those who attend school 3 all have school 2 as their best alternative. Then using (3) and (6), one can compute the dashed boundary in Figure 1 that separates students in school 3 who pay the price cap from those who pay the reservation price. The equation of the latter boundary is

$$y = \frac{q_3}{q_3 - q_2} p_3^m - \frac{q_2}{q_3 - q_2} \text{EMC}_2(b) \tag{17}$$

with slope $\frac{q_2 \eta_2}{q_3 - q_2}$. Those in A_3^m pay tuition of p_3^m and those in A_3^r pay p_3^r which is easily computed to be $p_3^r = \frac{q_3 - q_2}{q_3} y + \frac{q_2}{q_3} \text{EMC}_2(b)$. Those who qualify for school 3 but choose school 2 have school 1 or school 3 as their best alternative. Use the latter to determine their equilibrium price. Then proceed recursively by considering the set of types who qualify for all schools but 3.[11] And so on.

The model's central predictions for pricing can be summarized as follows.

P1 *Relatively higher income and lower ability types in any school (i.e., students in A_j^m) will pay the school's maximum tuition.*

P2 *Among those students not paying the maximum tuition (i.e., students in A_j^r), tuition will decline with ability.*[12]

P3 *Among those students not paying the maximum tuition, tuition's dependence on*

[11] Price will decline discretely within school 2 as b rises above b_3^m for those students who then have school 3 as their strictly preferred best alternative.

[12] An exception is students in A_1^r (in the lowest quality school) who have no college as their best alternative. Their tuition does not decline with ability because their competing alternative of no college does not discount to ability.

household income will be relatively weak except in the top schools where tuition will rise with income.

Prediction (P1) is an obvious consequence of the model. Prediction (P2), from Proposition 3, is due to competition for ability. More specifically, access to competing schools at EMC, which declines with ability, requires that the school increase financial aid with ability. Prediction (P3) is more subtle. In A_j^r, tuition equals $p_j^r(b, y)$, and the issue is how it depends on y. Consider Figure 2 and students in any other school than 3, say school 2. Fix the student's ability slightly (ϵ) above b_3^m and ask how $p_2^r(b_3^m + \epsilon, y)$ varies as income rises from its minimum. So long as school 1 is the student's best alternative, then p_2^r increases with income. But when income reaches the point such that school 3 becomes the best alternative, then p_2^r decreases with income, because the relative preference for higher school quality increases with income. In short, competition for students from "both sides" curtails the scope for discrimination over income. The highest quality school is exceptional because there is no competition for students "from the top". Within A_J^r, tuition rises with income.

To the extent that quality differentiation among schools that compete for students is minimal, as if there are many small colleges in an educational market, tuition cannot differ much from effective marginal cost. If on the other hand, a set of top schools essentially act as one leading school (among which students then are largely indifferent), the model predicts substantial income discrimination within this set of schools. It is no secret that an active cartel of elite private schools existed, although the consequences of their collusion is more controversial.[13] More generally, the extent to which markets for higher education are relatively insulated, e.g., regional, will permit relatively more income discrimination. A primary goal of our empirical analysis is to investigate the dependence of tuition on ability and income, and to examine how this varies along the quality hierarchy of schools.

Proposition 1 indicates that endowments rise along the quality hierarchy. Under our assumptions, we have also shown that ("non-custodial") per student expenditures rise along the hierarchy as well (Epple et al., 1999). Because the partition of students into schools is fairly complicated and because we have placed almost no restrictions on $f(b, y)$, we cannot show generally that the peer measure must ascend along the hierarchy. In some special cases the latter can be shown.[14], and we have found such ascension in related computational analysis. Given income and ability stratification (Proposition 2), and a positive correlation between b and y in the population, it is clear that an exception to ascension of the θ_j's would be pathological. Hence we take the model to predict:

$$\mathbf{P4} \quad R_1 < R_2 ... < R_J, \quad I_1 < I_2 ... < I_J \quad \text{and} \quad \theta_1 < \theta_2 ... < \theta_J$$

As we have discussed, tuition discrimination with respect to income will be quite limited if schools have close substitutes in the quality hierarchy with the exception of the top school. In other than the top school, tuition will then be close to effective marginal cost. The average of effective marginal cost in a school equals the marginal resource cost $(V' + I)$, implying average tuition, $\bar{p}_j$, will approximate the same. Conditional on intense competition

[13]See Carlton et al. (1995), Netz (1998), and Hoxby (1997).

[14]A trivial case, for example, assumes perfect positive correlation of b and y in the population.

for students, we then expect:

$$\mathbf{P5} \quad \bar{p}_1 < \bar{p}_2 ... < \bar{p}_J,$$

2.4 Extensions

2.4.1 Diversity

Schools and households may value diversity of their peer group along racial or other dimensions. Here we show how the model can be extended to accommodate such preferences, using race as the example. Students are members of one of N races. Schools observe race and can condition their admission and tuition policies on race.[15] Let Γ_j^r denote the proportion of race r, $r \in \{1, 2, ..., N\}$, in school j, and Γ_{pop}^r denote the proportion of race r in the economy. We assume quality in school j is given by:

$$q_j \quad = \quad q(\theta_j, I_j, \Gamma_j^1, ..., \Gamma_j^N); \qquad \frac{\partial q}{\partial \Gamma_j^r} > (=) 0 \quad \text{for} \quad \Gamma_j^r < (\geq) \Gamma_{\text{pop}}^r \tag{18}$$

Hence, we characterize diversity as placing value on increasing the attendance of under-represented races.

We have shown that the admission policy is race dependent for under-represented races, given by (11) but with admission function $\alpha()$ dependent on r and with effective marginal

[15] Such practices have, of course, recently come under constitutional challenge.

cost given by:[16]

$$\text{EMC}_j^r \;=\; V'(k_j) \;+\; I_j \;+\; \frac{\partial q_j/\partial\theta}{\partial q_j/\partial I}\,(\theta_j - b) \;-\; \frac{\partial q_j/\partial\Gamma_j^r}{\partial q_j/\partial I}\,(1-\Gamma_j^r) \tag{19}$$

Effective marginal cost is now augmented by the last term which equals the cost saving from increasing school quality by admitting a student from an under-represented race. Because this term with sign is negative for under-represented races, their effective marginal cost is lower.

To the extent that schools compete for students so that tuitions are bid down to effective marginal cost, within school tuitions of students of the same ability are predicted to be lower for under-represented races. Even for a monopoly provider of schooling, for given ability, the minimum tuition paid by race is predicted to be lower for under-represented races. Schools will admit lower-income types that have lower reservation prices if members of under-represented races.

2.4.2 Non-institutional Financial Aid

Substantial financial aid to many undergraduates in the form of grants, loans, and work-study funding is provided by the federal government and to a lesser extent by other entities that are independent of the student's school. We refer to such aid as non-institutional aid and briefly discuss here how such aid affects our model. Much of this aid is based on the federal government's calculation of expected family contribution, denoted EFC(y),

[16]Details are available on request.

which is an increasing function of household income (actually wealth). Let $G(y)$ denote non-institutional grants which are generally need based and hence written as a function of income. We presume that the amount of non-institutional aid to the student at school j, denote D_j, is given by:

$$D_j = \begin{cases} G(y) & \text{if} \quad p_j - G \leq \text{EFC}(y) \\ G(y) + \Omega[p_j - G(y) - \text{EFC}(y)] & \text{if} \quad p_j - G \geq \text{EFC}(y) \end{cases} \tag{20}$$

for some $\Omega \in [0, 1)$. Our specification presumes that aid in the form of subsidized loans and/or work study support is given to cover any gap between tuition and grant plus EFC, which is then discounted by $\Omega < 1$. Note that $D_j = 0$ for any student for whom $G(y) = 0$ and $p_j \leq \text{EFC}(y)$, which will include all "rich' students.

Continuing to define p_j^r as in (5), i.e. not entering non-institutional aid in the left-hand side, one finds that the optimal tuition to admitted students is simply:

$$p_j = min\{p_j^r + D_j, p_j^m\} \tag{21}$$

The intuition is that schools will continue to take away any surplus the students obtain from attending their school if tuition is below the price cap. Note, however, that p_j^r will be lower for students that have access to non-institutional aid if a competing school were attended, since they will continue to have such access at EMC but with the aid.[17] Using

[17] These income effects will affect the equilibrium allocation. The fundamental properties of the allocation will be unaffected, but this is not to suggest insignificant quantitative consequences, especially for poorer students.

(20) and (21), we have that for $p_j < p_j^m$:

$$p_j = \begin{cases} G + p_j^r & \text{if} \quad p_j - G \leq \mathrm{EFC}(y) \\ G + \frac{p_j^r}{1-\Omega} - \frac{\Omega}{1-\Omega}\mathrm{EFC} & \text{if} \quad p_j - G \geq \mathrm{EFC}(y) \end{cases} \tag{22}$$

At the same time, p_j must equal EMC_j along the boundary loci, and p_j cannot much deviate from EMC_j if there is intense competition among schools. Loosely, with intense competition, the consequences of non-institutional aid are predicted to affect mainly the allocation, but not tuitions. Alternatively, if schools have substantial market power, non-institutional aid is predicted to be reflected mainly in higher tuitions (and school expenditures). These findings provide additional scope for testing the intensity of competition.

3 Data Sources and Summary Statistics

In the empirical analysis, we use both university-level data and data for a representative sample of students. Our primary data source is the National Postsecondary Student Aid Study (NPSAS) obtained from the National Center for Education Statistics (NCES). The NPSAS contains extensive information for a sample of students. Of particular relevance for our work, the NPSAS contains the student's performance on standardized tests (either SAT or ACT), information about income of the student's family, and information about the financial aid received by the student. We have secured from the NCES a restricted-use version of the NPSAS that contains student-level data for 1995-96 and links each student in the sample to the school the student attended in academic year 1995-96.

We study four-year colleges and universities. For a given wave of the NPSAS survey, the NCES chooses a set of colleges and universities. It then selects a sample of students from within each of those institutions. For the most recent survey, the NCES chose 497 colleges and universities. Table 1 contains descriptive statistics of the main variables of the 1995-96 NPSAS used in this paper. The NPSAS sample consists of 11489 students attending 4 year colleges and universities. We have two measures of ability in this sample. The first one is based on SAT and ACT test scores and hence measures ability relative to the pool of applicants. The students in our sample have SAT scores (or imputed SAT scores) ranging from 400 to 1540 with a mean SAT score of 929. We also observe the GPA in the first semester of college which measures ability relative to the pool of students in a given college. The mean GPA in our sample is 2.70 with a standard deviation of 0.81.

Most of our analysis focuses on investigating whether students with higher ability levels pay lower tuitions in equilibrium because of the positive externality they have on other students through the peer group effect. Tuitions are directly related to the amount of financial aid received by a student. Financial aid is measured by the institutional grant amount. Hence it only includes aid received from the institution which is being attended. Later we factor non-institutional aid into the empirical analysis. The grant amount received by an individual in the sample ranges from 0 to 26278 with a mean of $2108 and a standard deviation of $3675. Approximately 75 % (25%) of the students in our sample receive a grant from the private (public) institution they attend.

The NPSAS also contains data about the financial position of the student's family. Mean family income is $46089 in the sample with a standard deviation of $35,689. Furthermore, the NPSAS collects demographic information which is useful for our analysis. More than 75 percent of the students in the sample attend colleges within the state in which the family resides. The composition by race of our sample is as follows: 12.1 percent classify themselves as black, 6.9 as Hispanic, 6.5 as Asian and 1.2 percent as other nonwhite race.

Table 1: Descriptive Statistics: NPSAS Sample

Variable	mean	std. deviation	min	max
Grant Amount	2108	3675	0	26278
SAT Score	929	207	400	1540
First year GPA	2.70	.819	0	4
Family Income	46089	35689	0	417388
Same state	.758	.427	0	1
Black	.121	.327	0	1
Black-private	.045	.209	0	1
Hispanic	.069	.253	0	1
Hispanic-private	.030	.170	0	1
Asian	.065	.248	0	1
Asian-private	.029	.170	0	1
Other race	.012	.113	0	1
Other race- private	.004	.067	0	1

In addition to data for individual students, we use data for colleges and universities. Peterson's conducts a survey of all colleges and universities, obtaining information on faculty resources, financial aid, the distribution of standardized test scores, and a host of other variables. We have purchased their database. We have supplemented this with information on educational expenditures and endowments from the NSF Web accessible Computer-Aided Science Policy Analysis and Research (WebCASPAR) database.

Table 2: Descriptive Statistics: Peterson's Sample

Variable	q90	q75	median	q25	q10
Mean SAT Score	1174	1100	1030	950	860
Gross Tuition	16300	12480	8990	3535	2337
Net Tuition	11763	9632	6901	3080	1911
Endowment	56488	16046	4916	685	22
Expenditures	9956	6637	4323	3267	2527
Salaries	58993	49256	42429	36345	31653
Enrollment	10474	5042	1957	1087	665

One of the most important variables in our analysis is the average standardized (ACT or SAT) test score, which provides a good measure for the quality of the peer group of each college. For the vast majority of colleges in the sample, Peterson's reports the empirical distributions of verbal and quantitative SAT scores. Alternatively, Peterson's reports the distribution of the ACT composite score which can be converted into the distribution of the composite SAT score using standard conversion tables. For these colleges, we can easily compute the average test score and the standard deviation of scores within the college. If the distributions of neither SAT nor ACT scores are available in Peterson's database, we try to approximate the mean SAT scores based on information available in a number of other publications. While the imputations of the SAT scores for these colleges are admittingly less precise for these colleges, it is still a useful exercise. It increases the sample size, and, more importantly, adds more colleges in the lower tiers to our sample which helps alleviate potential self-selection problems.

Constructing the remaining variables is a straightforward exercise based on the information in our sample. Gross tuition in Table 2 is the weighted average between full-time tuition for in-state students and full-time tuition for out-of-state students. Net tuition is gross tuition minus the average amount of need based and non-need based scholarships. Enrollment is measured by total undergraduate enrollment. All these variables are taken from Peterson's. Total educational expenditures and endowments are taken from the NSF WebCaspar database and converted to a per-capita basis using the enrollment variable from Peterson's. Average faculty salary is reported by Caspar. A public college indicator is also taken from Peterson's.

Peterson's database contains a total number of 1868 four year colleges and universities within the United States. We eliminate colleges from our sample which are highly specialized, do not have a regular accreditation and have missing values for the most interesting variables that our analysis focuses on. This leaves us with a sample of 1241 universities and colleges. Table 2 provides some descriptive statistics of the main variables in the sample.

We find that the median composite SAT score is 1030, the 90 percentile is 1174 and the 10 percentile is 860. Colleges also differ significantly in undergraduate enrollment ranging from 66 to 35475 with a median enrollment level of 1957. Roughly one third of the colleges in our sample are public universities, with the remaining two thirds private. Gross annual tuition ranges from $230 dollars to $22000. The median gross tuition is about $8990. Financial aid is quite significant. Median net tuition is \$6901, almost \$2100 less than gross tuition. Note that net tuition per student is greater than expenditure per student in all but the lowest quantile reported in Table 2. The expenditure measure we use is instructional expenditures reported in the NSF WebCaspar database. Thus, costs of administration and custodial services are not included in this measure of educational expenditures. Median endowment per undergraduate is approximately \$4916. However the distribution of endowment is quite skewed, with a small number of colleges having very large endowments, and the majority of colleges having only small endowments. The 99 percentile is \$423,077, the 75 percent quantile is only \$16,047 and the 25 percent quantile is \$686.

There is also a large amount of variation in educational expenditures per student. Unfortunately, the expenditure variable includes components which have nothing to do with undergraduate education. For example, colleges with large medical colleges have much higher expenditures per capita than comparable colleges. The colleges in the top 2 or 3 percent in our sample have expenditure levels which are dominated by expenditures which are unrelated to educational expenditures. For example the 99 percentile is \$51,388. For the remaining sample, expenditure per capita are more informative about educational expenditures. For example, the 95 percent quantile is \$13,731, the 75 percent quantile is \$6,637, the median is \$4,323 and the 25 percent quantile is \$3,267. Another measure which captures different expenditures is average faculty salary. This variable does not have the drawbacks associated with the expenditure variable. Median average salary is \$42,429. The 75 percent quantile is \$49,256 and the 25 percent quantile is \$36,345. The college with the highest salaries in our sample is Cal Tech with an average salary of \$112,401.

Finally, we report the correlation structure between tuition rates, mean SAT scores, endowments, expenditures, and some other measures in Table 3. Not surprisingly, we find that mean SAT scores are strongly positively correlated with tuition, endowment, expenditures and faculty salary. Also endowment is positively correlated with both expenditures and salaries. The correlation table also suggests that public universities are typically larger and have lower tuition and expenditure levels than private universities.

4 Evidence Regarding Hierarchy and Stratification

4.1 Evidence Regarding Hierarchy Predictions

The theoretical model implies a hierarchy in which university endowment and college quality have the same ordering. For most parameter sets the model will further imply that mean student ability and input per student will be similarly ordered. Investigation of predictions about ordering of variables across colleges provides a natural first test of the model. Of course the ordering predictions will not be satisfied perfectly by the data.

Before we can analyze whether our data provide some evidence in favor of the hierarchical predictions of the underlying equilibrium model, we need to define the appropriate choice set faced by individual households. A natural starting point of the analysis is to treat each college as a differentiated product. The relevant choice set is then the total number of colleges in our sample which is 1241. While this approach is appealing, it has some obvious limitations which arise to due the large number of potential choices. One of the main drawbacks of this approach relates to the fact that we need to observe the complete choice set faced by the individuals in order to test the predictions of our model. The NPSAS, however, does not sample all colleges in the population, but only a representative sample of colleges. For example, the most recent NPSAS sample only contains students of 497 colleges of the 1241 colleges in the Peterson's sample.

Furthermore, we do not expect that the strong predictions of our underlying equilibrium model hold at the college level. There are likely to be many idiosyncratic factors which influence college choice and which are omitted from our theoretical model. However, we expect that the model will be more successful in explaining patterns of choice, and admission and pricing behavior on a more aggregate level. The basic idea is that most of the idiosyncrasies are irrelevant on the aggregate level. For example, our model is better suited to explain whether a student with given income and ability attends a top private college or mediocre private college than whether a student attends Yale or Stanford. By aggregating colleges with similar observed characteristics, we thus abstract from a number of factors such as regional preferences which are important at a disaggregate level, but are likely to be less important in a suitably aggregated model.

Finding appropriate algorithms for aggregating colleges with similar characteristics and defining an appropriate choice set faced by the individuals is a challenging task. Aggregation should be based on the principle of substitutability. Public colleges may behave differently than private colleges for a number of reasons from which we have largely abstracted in our theoretical model. Public colleges may have somewhat different objective functions and face different financial constraints than private colleges. Then public colleges may use different pricing and admission policies than their private counterparts. We should therefore differentiate in our empirical analysis between private and public colleges.

Our model also suggests that colleges which are close substitutes should have similar pricing and admission policies as well as expenditure choice. Similarly, our model suggests that these differences are largely determined by differences in financial endowments. Colleges with similar levels of endowment per student are, therefore, likely to be close substitutes.

Table 4: Means by Difficulty Level: Private Colleges

	observations	mean sat	net tuition	endowment	expenditure
I	33	1350	14474	43820	5729
II	98	1208	11688	7996	1171
III	569	1035	8641	1625	570
IV	100	902	7203	987	475
V	27	890	4821	628	323

We address this aggregation problem using two approaches. The first approach draws on classification schemes which are frequently used in practice. For example, Peterson's classifies colleges based on their selectivity in admitting new students. Peterson's distinguishes between five types of colleges. Difficulty level I includes colleges such that more than 75 percent of the freshmen were in in top 10 percent of their high school class, scored more than 1250 on the SAT or 29 on the ACT, and admitted fewer than 30 percent of applicants. Difficulty level II includes colleges such that more than 50 percent freshmen were in the top 10 percent of their high school class, scored more than 1150 on the SAT or 26 on the ACT, and admitted fewer than 60 percent of applicants. Difficulty level III includes colleges in which more than 75 percent of the freshmen were in the top half of their high school and scored over 950 on the SAT or 18 on the ACT, and admitted fewer than 85 percent of all applicants. Difficulty level IV contains colleges with minimal admission standards, while colleges in difficulty level V have virtually no admission standards.

Table 4 reports the means of the most important variables by difficulty level for the sample of private colleges. We find that all four quality related variables are monotonic functions of the degree of difficulty. More selective colleges have higher endowments and expenditures than less selective colleges as predicted by our equilibrium model.

Table 5: Means by Difficulty Level: Public Colleges

	observations	mean sat	net tuition	endowment	expenditure
II	30	1168	3699	1414	866
III	278	1017	2675	199	501
IV	65	865	2142	42	347
V	41	913	2258	86	328

We repeat this exercise for the set of public colleges and the results are shown in Table 5.[18] We find that public colleges show similar patterns to private colleges. The main difference is that there does not seem to be an obvious differences between type IV and type V colleges. We conclude that, at least for a very coarse aggregation rule, the most interesting variables in the sample – mean SAT score, tuition, endowment and expenditures

[18]There are no difficulty level I public schools.

– satisfy the hierarchical predictions of the underlying equilibrium model.

Peterson's classification scheme is still very coarse. The largest groups still contains more than 550 colleges. It is not reasonable to assume that the colleges within each type are homogeneous with respect to their admission and pricing policies. We can refine our classification scheme and construct a finer grid of colleges using cluster analysis. The basic idea behind K-means cluster analysis is to find a clustering or grouping of the observations so as to minimize the total within-cluster sum of squares.

We perform a cluster analysis for the sample of private colleges using mean SAT scores, net tuition and endowment as the three main variables used to define the clusters. We implement the analysis using the standardized variables and assigning equal weights to each of the three variables. We perform the cluster analysis for a number of different choices of the number of clusters. Table 6 reports the cluster means for an analysis with 25 groups. We rank clusters by the mean SAT score. We find that the correlation between mean sat scores and mean net tuition across the sample of 25 groups is 0.6. The correlation between mean SAT and mean endowment (expenditures) is 0.47 (0.40) which is somewhat lower largely due to a couple of outliers.

We repeat this exercise using the sample of public colleges. The only difference is that we only use mean SAT scores and net tuition as variables in the analysis since endowments are less important for public colleges. Table 7 reports our findings based on an analysis of 15 clusters.

Table 6: Clustering Analysis: Private Colleges

	observations	mean sat	net tuition	endowment	expenditure
1	14	655.29	4707.85	410	254
2	18	782.11	9414.07	365	424
3	20	826.65	5237.70	1363	515
4	40	858.42	8052.52	693	460
5	25	902.64	11235.94	765	656
6	37	951.11	6561.54	1070	380
7	47	954.02	9556.43	697	603
8	24	968.75	2669.77	1288	337
9	47	975.22	8172.79	767	465
10	44	1016.05	5013.22	778	340
11	56	1023.82	10883.50	1140	535
12	50	1043.80	6754.57	1407	401
13	17	1056.59	13931.01	2218	824
14	90	1061.31	8927.41	1035	448
15	44	1118.05	7077.73	2008	439
16	60	1134.28	11531.91	2764	756
17	21	1139.48	4396.60	2706	470
18	52	1139.81	9601.22	3799	839
19	23	1187.35	15424.00	4656	1005
20	26	1191.48	7327.28	7237	1195
21	3	1277.90	12259.16	215547	27978
22	31	1278.63	11813.95	10777	2017
23	10	1318.55	11890.36	42099	2966
24	23	1323.63	15854.73	13714	2341
25	5	1442.27	14467.85	77885	6949

Table 7: Clustering Analysis: Public Colleges

	observations	mean sat	net tuition	endowment	expenditure
1	12	661.58	2336.22	27	334
2	15	824.47	1349.56	86	396
3	29	830.34	2682.54	75	385
4	17	922.29	840.56	62	412
5	43	928.98	1882.35	35	333
6	47	954.70	3102.30	62	404
7	14	954.93	5077.31	88	436
8	35	1005.09	1394.25	135	399
9	51	1017.32	2209.04	107	429
10	10	1032.30	302.89	222	531
11	46	1053.24	3813.50	279	569
12	43	1077.85	2740.08	305	581
13	15	1152.33	1477.06	1573	751
14	11	1177.77	6550.98	1442	964
15	26	1188.77	3716.69	609	760

4.2 Evidence Regarding Stratification

The theoretical model predicts that we should observe a certain amount of stratification in income and ability among universities. One way to measure the amount of ability stratification is to decompose the variance of ability in the set of universities into within-universities and between-universities components. A simple calculation shows that the following decomposition of the variance of ability, b_j, holds:

$$\mathrm{Var}(b) \;=\; \sum_{j=1}^{J} Pr(C_j)\,\mathrm{Var}(b_j) \;+\; \sum_{j=1}^{J} Pr(C_j)\,[E(b_j) \;-\; E(b)]^2 \tag{23}$$

where $Pr(C_j)$ is the proportion of sampled students in school j. Dividing both sides of the equation by $\mathrm{Var}(b)$ yields the decomposition of the variance measured in percent. If there is perfect stratification by ability, the first component will be small. Alternatively, if preference heterogeneity were large, the first component would be large. Hence, the magnitude of the first and the second components helps us evaluate the importance of ability stratification in the sample.

Table 8: Stratification of Test Scores

	sample size	within college variance	across college variance
ACT score	846	62.3 %	37.7 %
quantitative SAT score	710	60.6 %	39.4 %
verbal SAT score	712	58.3 %	41.7 %

We compute the variance decomposition for the colleges in the sample for which we have data on the distribution of the verbal SAT score, the quantitative SAT score or the ACT score. The results of these computations are shown in Table 8. We find that the second component of the variance is fairly large for all three measures of ability. This indicates that there is a large amount of stratification across schools as predicted by our equilibrium model.

Our model implies that in the (y, b) plane, the student population will be partitioned into schools by boundary loci as illustrated in Figure 1. If the kind of stratification depicted in Figure 1 is present, then the correlation of income and ability within schools will be less than the correlation of income and ability in the overall student population. Testing this prediction, we find that the correlation of income and ability in the student population is .263. By contrast, the partial correlation of income and ability when controlling for school fixed effects is .128. Thus, the within-school correlation is half the correlation in the overall student population. These correlations are based on a sample of 9,024 students, so the difference in the estimated correlations is highly significant. Thus, the prediction of stratification is also supported by these correlations.

5 Evidence Regarding Pricing and Financial Aid

The equilibrium model discussed in Section 2 has a variety of predictions regarding pricing policies that can be tested without estimating the structural parameters of the model. If colleges have close substitutes, then price approximately equals effective marginal cost for all students. Equation (12) then approximately characterizes net tuition paid by all students.

We assume that price is measured with an additive error and ability is measured without error. For student s with ability b in school i, equation (12) and the assumption that schools have close substitutes imply:

$$p_{is} \quad = \quad \alpha_{0,i} \; + \; \alpha_{1,i}\, b_s \; + \; \epsilon_{is} \tag{24}$$

The α's in this equation are school-specific intercepts and school-specific slope coefficients on student ability (b), and ϵ_{is} is the error in measuring tuition net of financial aid. The model and the assumption that schools with close substitutes also implies absence of income as a variable in equation (24). This can be tested by adding income terms in the above equation.

$$p_{is} \quad = \quad \alpha_{0,i} \; + \; \alpha_{1,i}\, b_s \; + \; \alpha_{2,i}\, y_s \; + \; \epsilon_{is} \tag{25}$$

We anticipate that pricing by income will be found in the top-ranked schools who face no competition from above.

So far in this section we have ignored the existence of price caps. We observe that a large fraction of students do not receive financial aid and hence must pay the regular tuition rate that the university charges. It is convenient to focus directly on the financial received by the students. Let p_i^m be the posted tuition in school i, and let g_{is} be the measured grant received by student s in school i. Then the price paid by student i in school s can be written as

$$p_{is} \;=\; p_i^m \;-\; g_{is} \tag{26}$$

Posted tuition is the same for all students in a school and is thus impounded in the fixed effect in (25). Using grants as the dependent variable reverses the signs of the coefficients in (25). Thus, our competitive model predicts $\alpha_{1i} > 0$ and $\alpha_{2i} = 0$. From an econometric perspective, the price caps give rise to censoring. Consequently, we estimate the above equation using a Tobit procedure.

Equations (24) and (25) can in principle be estimated separately for each school, provided one has a sufficient number of observations for each school in the sample. Unfortunately sample sizes for individual colleges are small in the NPSAS. Therefore, we need to impose more structure on the underlying regression function. To simplify the analysis, we create the following ranking variable:

$$r_i \;=\; \frac{h - m_i}{h - l} \tag{27}$$

where m_i is the median SAT score in school i among the students in the sample who attend school i, h is the highest value of m_i in the sample, and l is the lowest. Thus, for the highest ranked school, r_i takes on a value of zero, and for the lowest ranked school, r_i takes on a value of 1. We then assume that college specific intercepts of ability and income satisfy the following assumption:

$$\alpha_{1,i} = \alpha_{11} + \alpha_{1,2}\, r_i + \alpha_{1,3}\, d_i \tag{28}$$

$$\alpha_{2,i} = \alpha_{21} + \alpha_{2,2}\, r_i + \alpha_{1,3}\, d_i$$

where d_i is a dummy indicating whether the college is private or public. Substituting equation (28) into equation (25) and focusing on financial aid yields our preferred model and which can be expressed (with a slight abuse of notation) as follows:

$$\begin{aligned} g_{is} = & \ \alpha_{0,i} + \alpha_{11}\, b_s + \alpha_{12}\, r_i\, b_s + \alpha_{13}\, d_i\, b_s \\ & + \alpha_{21}\, y_s + \alpha_{22}\, r_i y_s + \alpha_{23}\, d_i y_s + \epsilon_{is} \end{aligned} \tag{29}$$

Motivated by our analysis of diversity in Section 2.4, we also include indicator variables for African American, Hispanic, Asian, and other non-white in our Tobit model. We also include an indicator variable for whether a student lived within the state where the college is located. State schools generally have a higher posted tuition for out-of-state students, and our aid measure is relative to the particular student's posted tuition. Nevertheless, schools may perceive out-of-state applicants as more mobile, hence participating in a more competitive market, and this motivates inclusion of this dummy variable.[19] The results of our estimation are reported in Table 9.[20]

To interpret Table 9, recall that RANK equals 0 for the top ranked school and 1 for the

[19] Another possibility is that geographic diversity in student body is an element of school "quality."

[20] We have also estimated the model with athletic scholarships removed from institutional grants. Results are very similar.

Table 9: Tobit: Financial Aid Amounts

	Coefficient	Standard Error	t-value
SAT	-1.68	1.23	-1.359
SAT × private	2.08	0.69	3.015
SAT × rank	11.09	2.40	4.618
GPA	821.32	303.09	2.710
GPA × private	-130.10	167.19	-0.778
GPA × rank	310.17	580.93	0.534
Income	-.0949	.0057	-16.600
Income × private	-.0021	.0035	-0.589
Income × rank	.1577	.0110	14.234
Same state	-2346.29	230.69	-10.171
Same state × private	1520.04	274.74	5.532
Black	1760.17	286.45	6.145
Black × private	293.97	418.69	0.702
Hispanic	1792.75	330.45	5.425
Hispanic × private	-771.32	446.35	-1.728
Asian	746.43	336.67	2.217
Asian × private	-1349.67	431.44	-3.128
Other race	500.74	641.50	0.781
Other race × private	-1522.62	945.80	-1.610

$N = 8497$. $\chi^2(424) = 7424.57$. $Prob > \chi^2 = 0.00$. Log Likelihood = -41587.706.

lowest ranked school, and PRIVATE equals 0 for public schools and 1 for private schools. In the discussion that follows, we will interpret the results for private schools. The findings are not markedly different for public schools, and the differences are easily seen by scanning the results in Table 9.

The estimates in Table 9 imply that the top-ranked schools give negligible discounts to more able students. For the top-ranked school, the coefficient on SAT is only .4 (=2.08-1.68) and not significantly different from zero (p=.71). For the lowest-ranked private school, the coefficient on SAT is 11.49 and highly significant, implying an \$11.49 increase in financial aid for each unit increase in SAT. We include first-semester GPA as another measure of ability, in an effort to measure elements of "ability," like motivation, that may exhibit little correlation with SAT.[21] The coefficient of the GPA variable indicates that private schools give discounts of roughly \$690 per unit increase in GPA, and this is statistically significant (p=.01). This discount is estimated to increase somewhat as school rank declines, though the coefficient (310.17) is not significantly different from zero. A plausible generalization of our specification of the combined utility-achievement function (recall (3)) may explain the relatively limited discounting to ability we find at top schools. Our Cobb-Douglas specification has the property that for fixed alternative school and tuition there, the marginal willingness to pay for quality elsewhere is independent of ability. Our model predicts discounting to ability because peer ability is valued, implying higher-ability students have more attractive alternatives. Suppose instead that for fixed alternative (including tuition),

[21] It is interesting to note that first-semester GPA and SAT have a correlation of .4 in the NCES data. Thus, there are clearly factors other than SAT that play an important role in determining academic performance.

the marginal willingness to pay for quality elsewhere rises with ability as, for example, a CES specification of the combined utility-achievement function allows. This would reinforce ability discounting in equilibrium among students whose best alternative is a lower-quality school, and reduce or reverse ability discounting among students whose best alternative is a higher-quality school (without changing the qualitative pattern of preferences and the allocation from that in Figures 1-2). Because the top schools face no competition "from above," the latter predictions are quite consistent with the observed pattern of pricing by ability. Another possibility that might explain the lack of observed discounting to ability at top schools is omitted variable bias as further discussed below.

The coefficients on income imply that the top-ranked private school reduces grant aid by \$96 for each \$1,000 increase in household income. This premium declines with rank as the interaction with rank and income indicates. Taken literally, the results would imply that there is actually a discount to the highest income households in the lowest quality schools. However, this is largely an artifact of the functional form. We will see in Table 10 that the tuition does not vary significantly with income at the lowest ranked schools.

The same-state variable equals 1 if the student is from the state in which the school is located and zero otherwise. The negative coefficient of this variable may be an indication that schools price discriminate to some degree against students located nearby. If schools have some geographically based market power, this would not be an implausible outcome as discussed above. However, we know that some schools, particularly public schools, charge lower tuition to in-state residents. Thus, the negative coefficient on the same-state variable might be picking up a tendency of schools to give lower grants to students who are already receiving a tuition discount. In results not shown in Table 9, we added the within-state discount to the dependent variable. We then found that the coefficient of the same-state variable was reduced to roughly half the value in Table 9, and there was then no significant difference in the same-state variable between the public and private schools. Thus, the evidence suggests some degree of price discrimination against nearby residents.[22]

The coefficients on race are largely self-explanatory and consistent with our theoretical predictions about the value of diversity. African American and Hispanic students receive significant tuition discounts and are under-represented in colleges relative to their population shares, especially in private colleges. The results for Asian students are affected by differential pricing to in-state residents. When discounts to in-state residents are added to the dependent variable, the results suggest that Asian students do not receive significant financial aid in either public or private schools. In addition, the results then indicate that both African American and Hispanic students receive significantly greater financial aid in private than in public schools.

We have estimated Tobits similar to those in Table 9 using rankings based on our cluster analysis in place of the SAT ranking variable used in Table 9. The results are not qualitatively different when the alternative ranking variables are used.

As an alternative to the ranking variable in (27), we also interacted the SAT and income measures with the Peterson's selectivity measures. As we discussed in the previous section, Peterson's selectivity measures are based on a combination of criteria that includes percent

[22]We noted above the possibility too that schools value geographic diversity in their student body.

Table 10: Tobit: Financial Aid Amounts and Selectivity

	Coefficient	Standard Error	t-value
Sat sel1	-6.560723	1.690119	-3.882
Sat sel2	4.689578	.7299503	6.425
Sat sel3	6.1664	.395808	15.579
Sat sel4	7.435139	1.189917	6.248
Sat sel5	8.398905	1.917852	4.379
Income sel1	-.0920505	.0059214	-15.545
Income sel2	-.0570317	.003355	-16.999
Income sel3	-.0244991	.0020838	-11.757
Income sel4	.0074151	.0072152	1.028
Income sel5	-.0146154	.0182821	-0.799

N = 8548. $\chi 2(424) = 7341.4$. $Prob > \chi^2 = 0.00$. Log Likelihood = -42268.338.

of freshmen in top 10 percent of their high school class, composite SAT (or ACT score), and percent of applicants accepted. The result is shown in Table 10. This Tobit also includes the same state and race variables and interactions with private as appear in Table 9, but, in the interest of space, we do not report coefficients of those variables. School fixed effects are also included in the Tobit in Table 10, as in Table 9. Our findings suggest that the amount of pricing by income is lower for lower ranked schools (and not significant for the bottom two selectivity groups). The results also show that lower ranked schools also give more to more able students. The coefficient on SAT is actually negative for the highest ranked schools, which may be due to omitted variables.

In estimating the pricing equation, we have not taken account of selection effects. We envision doing this in future structural estimation of the model. We believe that selection effects, if present, will bias the coefficients of SAT and income estimated in this paper toward zero. To see why, suppose that, in addition to SAT scores, schools use measures of ability that we do not observe. They will use such measures if, in the applicant population, such measures have value in predicting ability beyond that provided by SAT scores. Such measures may be correlated with SAT, but only components of such measures that are orthogonal to SAT provide predictive power beyond that provided by SAT. Hence, suppose schools have access to measures of ability that, in the applicant population, are orthogonal to SAT.

A school i might admit a student with relatively low SAT scores if the other measures of ability for the student were more favorable. A student with relatively high SAT scores might attend i because other measures of the student's ability were sufficiently low that schools ranked higher than school i did not admit the student. This problem is also likely to be worse in the top ranked schools where there is less variation in SAT. Thus, within a school, the component of ability that schools observe and we do not observe will tend to be negatively correlated with SAT. This implies that the coefficient of measured ability (SAT) will be biased toward zero. We have actually used first-semester GPA in addition to SAT as ability measures in our pricing equations, but the same logic applies if schools use predictors of ability that are orthogonal to both SAT and first-semester GPA. Thus, if selection effects are present, the extent of pricing by ability is greater than we estimate in this paper. The same logic also applies to income; if selection effects are present, the extent of pricing by income is greater than we have estimated in this paper.

Our empirical analysis of non-institutional aid suggests a need for more study, and we just summarize our findings here. Based on the theoretical analysis above and using (2.22), we included total federal grants (G) as a covariate in our basic Tobit specification, and also interacted it with the private dummy variable and our SAT school-rank measure. We also created a dummy variable equal to 1 (0) if $p_j - G - EFC > \ (\leq)\ 0$ that was interacted with EFC, income, and, again, the private dummy variable and rank measure. The latter was to test for the potential effects in the lower line of (2.22) of non-institutional aid beyond grants (like subsidized loans) provided to cover the ability-to-pay gap perceived by the federal government. Total federal grants and EFC are in the NPSAS data base. As discussed above, the non-institutional aid variables should have no direct effects if the environment is highly competitive as tuitions are bid down to effective marginal cost.[23] If substantial market power is present, then schools capture non-institutional aid according to the theory, and the coefficients should have the opposite signs of those in (2.22) since we use institutional aid, rather than tuition, as the dependent variable. Generally, G had no significant effects which is consistent with a competitive environment throughout the ranks of schools. At lowly ranked schools, neither did the EFC-related variables differ substantially from zero. However, at highly ranked schools, the EFC-related variables were significant but with the reverse signs of those predicted by the market-power model! At this point we can only speculate about the meaning of these unexpected results.[24] We take these

[23]The effects of federal aid are then essentially income effects that impact the allocation of students into schools but not the fundamentals of pricing.

[24]For example, our finding that institutional aid declines at highly ranked schools as EFC rises in the range where $p_j - G - EFC > 0$ may be due to omitted variables. The federal government's calculation of EFC may take account of important wealth variables not in our household income measure.

results to indicate a need for more theoretical and empirical investigation of the interaction of institutional and non-institutional aid.

6 Conclusions

Our empirical results provide support for several aspects of our model. They also raise interesting puzzles for future research. We find evidence that there is a hierarchy of school qualities, as our model predicts. We also find correlation of SAT scores, endowment per student, and expenditure per student across the hierarchy as our model predicts.

We see the evidence on pricing as supportive of the view that the more highly ranked schools exercise some degree of market power. This is reflected in the substantial variation of price with income coupled with discounts to more able students that are modest at best. Lower ranked schools exhibit behavior that is closer to the predictions of the competitive model. They charge lower tuitions to more able students while charging much lower premiums to income than the more highly ranked schools. While the pattern of pricing at middle- and lower-ranked schools accords reasonably well with the predictions of the competitive model, the extent of pricing by income at middle-ranked schools remains something of a puzzle and an interesting stimulus for further research.

The evidence on pricing by ability is supportive of positive peer effects in educational achievement from high ability at the college level. An alternative hypothesis is that there are no peer effects but colleges value higher-ability students since they increase prestige, signal educational quality, and so on. This environment would also lead to discounting to ability so long as schools have market power. However, to the extent schools compete for students, tuitions will be bid down to marginal educational cost, here independent of student ability.[25] We find that as we move down the hierarchy of schools, the evidence on pricing by income indicates little market power, while there is much stronger evidence of discounting to ability. Hence, our findings support the existence of ability driven peer effects in higher education.

The evidence of pricing by ability is supportive of the prediction that schools value the improved peer quality that results from attraction of more able students. Thus, the evidence is consistent with the presence of peer group effects. The evidence does not, however, establish the presence of peer effects. Similar pricing would be predicted in a model where households obtain utility from having their students at schools with a more able peer group even if that more able peer group does not convey any increase in educational benefits. Nonetheless, our evidence is encouraging for further work on peer effects in education. Had there not been evidence of pricing by ability, support for the hypothesis that peers convey educational benefits would have been considerably weakened.

[25] One might also believe that a unit of educational quality costs less to provide to higher-ability students. This would require that the same educational inputs are cheaper when students are better (since the model already has higher-ability students achieve more highly for given inputs). Cursory examination of the evidence on teacher salaries is not supportive of this in higher education, though, obviously, this may be confounded by variation in teacher quality. But we also find in the data that a significant proportion of students get a free ride. It is implausible that the resource educational cost of students is negative, so we are lead back to peer effects.

References

Arnott, R. and Rowse, J. (1987). Peer Group Effects and Educational Attainment. *Journal of Public Economics*, *32*, 287–305.

Bergstrom, T., Rubinfeld, D., and Shapiro, P. (1982). Micro-Based Estimates of Demand Functions for Local School Expenditures. *Econometrica*, *50*, 1183–1205.

Carlton, D., Bamberger, G., and Epstein, R. (1995). Antitrust and Higher Education: Was There a Conspiracy to Restrict Financial Aid?. *Rand Journal of Economics*, *26*, 131–147.

Caucutt, E. (1998). Educational Policy When There Are Peer Group Effects. Working Paper.

Dale, S. and Krueger, A. (1998). Estimating the Payoff to Attending a More Selective College: An Application of Selection on Observables and Unobservables. Working Paper.

deBartolome, C. (1990). Equilibrium and Inefficiency in a Community Model with Peer Group Effects. *Journal of Political Economy*, *98(1)*, 110–133.

Epple, D. and Romano, R. (1998). Competition Between Private and Public Schools, Vouchers and Peer Group Effects. *American Economic Review*, *88*, 33–63.

Epple, D. and Romano, R. (1999). Educational Vouchers and Cream Skimming. Working Paper.

Epple, D. and Romano, R. (2000). Neighborhood Schools, Choice and the Distribution of Educational Benefits. NBER Working Paper 7850.

Epple, D., Romano, R., and Sieg, H. (1999). Theory and Evidence of Competition in Higher Education. Working Paper.

Evans, W., Murray, S., and Schwab, R. (1998). Education Reform and the Distribution of Education Resources. *American Economic Review*, *88(4)*, 789–812.

Fernandez, R. and Rogerson, R. (1996). Income Distribution, Communities, and the Quality of Public Education. *Quarterly Journal of Economics*, *111 (1)*, 135–164.

Fernandez, R. and Rogerson, R. (1998). Public Education and Income Distribution: A Dynamic Quantitative Evaluation of Education-Finance Reform. *American Economic Review*, *88(4)*, 813–833.

Fuller, W., Manski, C., and Wise, D. (1982). New Evidence on the Economic Determinants of Postsecondary Schooling Choices. *Journal of Human Resources*, *17*, 477–498.

Gilles, R. and Scotchmer, S. (1997). Decentralization in Replicated Club Economies with Multiple Private Goods. *Journal of Economic Theory*, *72*, 363–387.

Glomm, G. and Ravikumar, B. (1992). Public Versus Private Investment in Human Capital, Endogenous Growth and Income Inequality. *Journal of Political Economy*, *100*, 813–834.

Goldin, C. and L. Katz, L. (1998). The Origins of State-Level Differences in the Public Provision of Higher Education: 1890-1940. *American Economic Review*, *88(4)*, 303–308.

Hanushek, E., Kain, J., Markman, J., and Rivkin, S. (2000). Do Peers Affect Student Achievement?. Working Paper.

Hoxby, C. (1996). Are Efficiency and Equity in School Finance Substitutes or Complements?. *Journal of Economic Perspectives*, *10(4)*, 51–72.

Hoxby, C. (1997). How the Changing Market Structure of U.S. Higher Education Explains College Tuition. Working Paper.

Hoxby, C. (1999). Benevolent Colluders? The Effect of Antitrust Action on College Financial Aid and Tuition. NBER Working Paper 7754.

Manski, C. (1991). Educational Choice (Vouchers) and Social Mobility. *Economics of Education Review*, *11 (4)*, 351–369.

Nechyba, T. (1998). Fiscal Federalism and Private School Attendance. Working Paper.

Nechyba, T. (1999). School Finance Induced Migration Patterns: The Case of Private School Vouchers. *Journal of Public Economic Theory*, *1*, 1–46.

Nechyba, T. (2000). Mobility, Targeting and Private School Vouchers. *American Economic Review*, *90*, 130–146.

Netz, J. (1998). Non-Profits and Price-Fixing: The Case of the Ivy-League. Working Paper.

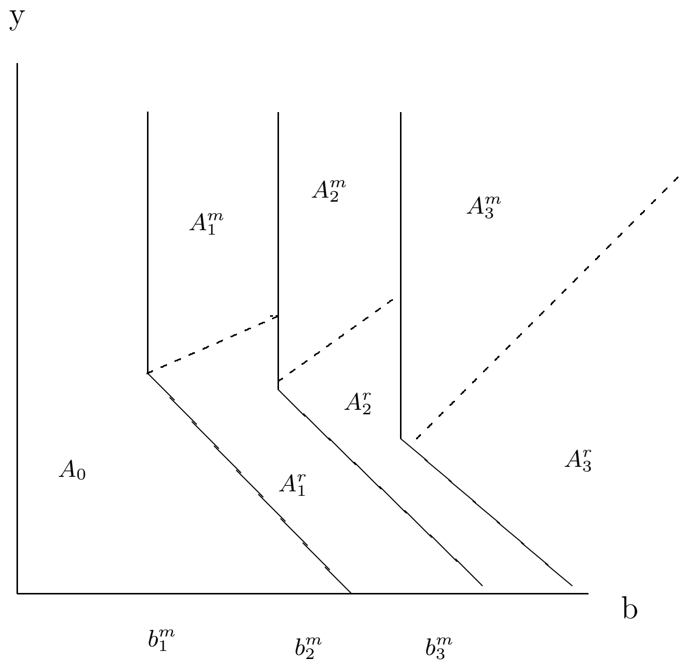
Rothschild, M. and White, L. (1995). The Analytics of the Pricing of Higher Education and Other Services in Which the Customers are Inputs. *Journal of Political Economy*, *103*, 573–623.

Toma, E. (1996). Public Funding and Private Schooling Across Countries. *Journal of law and Economics*, *39(1)*, 121–148.

Venti, S. and Wise, D. (1982). Test Scores, Educational Opportunities, and Individual Choice. *Journal of Public Economics*, *18*, 35–63.

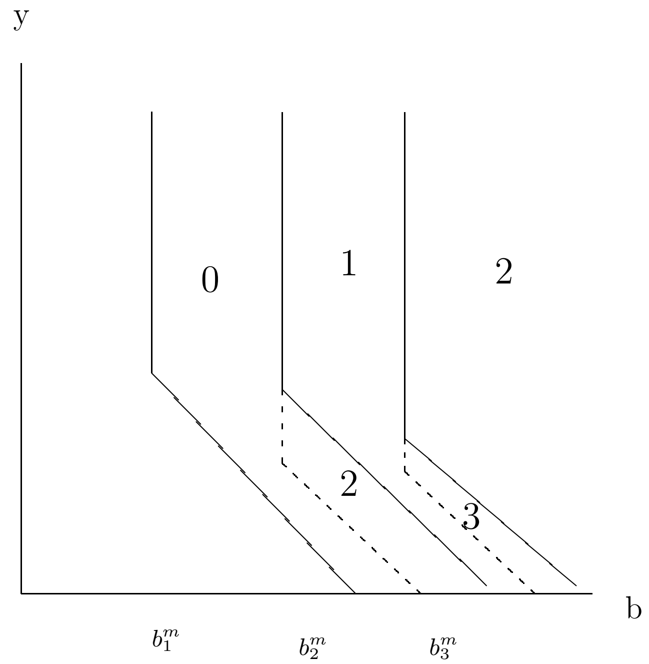
Zimmer, R. and Toma, E. (1998). TPeer Effects in Private and Public Schools Across Countries. Working Paper.

Figure 1: Boundary Loci in Equilibrium



This figure illustrates the boundaries of schools' admission sets.

Figure 2: Best Alternatives in Equilibrium



This figure illustrates the best alternative choices for students in equilibrium.