HCEO WORKING PAPER SERIES

Price Discrimination and Public Policy in the U.S. College Market

Ian Fillmore

Working Paper

2021-028

06/2021

HUMAN CAPITAL AND
ECONOMIC OPPORTUNITY
GLOBAL WORKING GROUP

The University of Chicago
1126 E. 59th Street Box 107
Chicago IL 60637

www.hceconomics.org

Price Discrimination and Public Policy in the U.S. College Market

Ian Fillmore*
Economics Department, Washington University in St. Louis

Abstract

In the United States, the federal government grants colleges access to a student's Free Application for Federal Student Aid (FAFSA) which facilitates substantial price discrimination. This paper is the first to estimate the consequences of allowing colleges to use the FAFSA in their pricing decisions. I build and estimate a structural model of college pricing and simulate counterfactuals wherein some or all of the FAFSA information is restricted. I find that if FAFSA information were restricted, 13 percent of students attending elite colleges would be inefficiently priced out of the elite market. Nevertheless, student welfare would rise as colleges charged the majority of students lower prices. Colleges do use the FAFSA to transfer resources from high- to low-income students on average, but this redistribution is highly imprecise: allowing colleges to use the FAFSA harms one-third of low-income students while one in seven high-income students actually benefit.

KEYWORDS: Price discrimination, higher education, first-price auction, Bayes-Nash equilibrium, financial aid, Free Application for Federal Student Aid

*Assistant Professor of Economics. *Address*: Campus Box 1208, 1 Brookings Drive, St. Louis, MO, U.S.A. 63130, *Email*: ianfillmore@wustl.edu

1 Introduction

In the United States, the federal government plays a major role in helping students pay for college, with two-thirds of full-time undergraduates receiving some sort of federal aid and roughly half of those receiving federal grants.[1] To receive any federal aid, a student must first complete the Free Application for Federal Student Aid, or FAFSA. The FAFSA requests detailed financial information as well as a list of colleges the student is considering attending. The government uses a fixed formula to determine eligibility for federal aid, but it does not directly dispense any of the aid itself. Rather, the government forwards the information to the colleges listed on the FAFSA and enlists them as partners in distributing federal aid. The partnership between colleges and the government is well known to anyone who has personally been through the financial aid process. The government must distribute aid to millions of students across the country, and enlisting the help of colleges seems like an obvious solution. However, colleges do more than simply distribute federal aid on the government's behalf. They also receive access to each student's FAFSA information. This partnership has been treated as a mere administrative detail by students, parents, policymakers, and even economists. It is not. This paper is the first to estimate the consequences of allowing colleges to use the FAFSA in their pricing decisions. I find that sharing the FAFSA with colleges enables them to engage in substantial price discrimination, with widespread repercussions for the cost of a college education as well as the equilibrium sorting of students into colleges.

Colleges routinely offer discounts of varying sizes to their students, with many colleges making the FAFSA a prerequisite for being considered for a discount. These discounts can be sizable and are intended to influence the student's choice of which college to attend. According to the College Board, "in 2013-14, institutions provided $37.9 billion in grant aid to undergraduate students. This constituted 21% of total undergraduate aid and 36% of undergraduate grant aid" (Baum et al. 2014). The $37.9 billion in institutional grants surpassed the size of the entire federal Pell grant program ($33.7 billion). Among freshmen in 2007-2008, 69 percent of students at private and very selective public colleges received discounts with the average discount equal to 36 percent of the average sticker price (see Table 1). As a result, the *transaction price*—sticker price minus institutional discount—varies tremendously across students at the same college.[2]

The FAFSA is valuable to colleges because it amounts to a source of low-cost, high-quality information about a student's willingness to pay. It is low-cost because the federal government bears the burden of collecting this information, and it is high-quality because the government

[1]See Tables 353 and 355 of the 2011 Digest of Education Statistics.

[2]The transaction price deliberately ignores outside sources of aid, including federal grants like Pell grants. Although federal Pell grants do make college more affordable *relative to not attending college*, they are portable across colleges, and as long as the student is planning to attend *some* college, the Pell grant does not incentivize her to choose one college over another (assuming that income effects are small). Moreover, Pell grants can also be used to pay for consumption, so if a student received a $2,000 Pell grant and a college were to offer her a transaction price of $1,500, she could keep the remaining $500 of the Pell grant as cash. Thus, the transaction price determines how much the student will have left to spend out of her available resources (including outside financial aid) as well as how much revenue the college will earn from enrolling that student.

imposes penalties, in the form of fines or jail time, for misreporting. Perhaps even more importantly, the FAFSA comes bundled with a convenient monitoring technology for ensuring that its information is reliable. Thirty percent of FAFSA forms are cross-checked against a variety of government databases, including tax records, in a process called verification. If a student's FAFSA is not randomly selected for verification by the government, then that student's college has full discretion to flag her FAFSA for verification anyway. Indeed, many colleges simply verify *all* of their students' FAFSA forms.[3] Effectively, the FAFSA grants colleges generous access to tax records and other government databases and allows them to use that information to learn about a student's willingness to pay.

What would happen if we restricted the colleges' ability to use some or all of the FAFSA information in their pricing? To answer this question, I build a structural model of college pricing and show that the model is identified from data on student-level transaction prices and student characteristics. Using student-level data from the 2008 wave of the National Postsecondary Student Aid Study (NPSAS), I test the qualitative predictions of the model using a reduced-form analysis. Finally, I estimate the structural model and simulate several counterfactuals, wherein colleges are restricted from using some or all of the FAFSA in their pricing.

The structural model incorporates institutional features of the U.S. college market while highlighting the competitive forces that shape equilibrium pricing behavior among colleges. In the model, each student invites (via her college applications) a set of colleges to make offers. If a college chooses to participate, it admits the student and makes a take-it-or-leave-it price offer. Students care about both price and other college characteristics, so a student may be willing to pay more to attend a particularly attractive college. The student attends the college that, in her judgment, makes the best offer. Colleges care about both tuition revenue and enrolling desirable students, so a college may be willing to forgo some tuition revenue in order to increase its chances of attracting a particularly desirable student. The model captures the tradeoff colleges face between attracting students and maximizing tuition revenue as well as the competition between colleges as they vie for students. I show how to reformulate the model in terms of a first-price auction in utility bids, which allows me to leverage both theory and empirical methods from the auctions literature. My approach offers the additional benefit of allowing me to remain fairly agnostic about the precise objective functions of students and colleges. Using an identification strategy in the spirit of Guerre et al. (2000), I show that the model is identified from data on student-level transaction prices and student characteristics.

In the reduced-form analysis, I find that, among elite colleges, the reduced form pricing patterns are consistent with the model's predictions. I estimate my structural model using data on freshmen at elite colleges and find that colleges successfully capture an average of 70 percent of the total match surplus through price discrimination, leaving the remaining 30 percent to the students. On average, students at elite colleges value attending their current college $18,181 (per year) more than their outside option of attending a nonelite college, but their average consumer

[3] It appears to be public knowledge that many colleges verify all of their FAFSA forms (Grant 2006; Weston 2014).

surplus falls to \$5,023 after paying the transaction price. Students who list more colleges on the FAFSA reap the benefits of intensified competition; those who list six colleges receive an average of 42 percent of the surplus, while those who list only one college receive an average of only 17 percent.

For simplicity, I consider three pieces of information that the FAFSA provides colleges with: 1) family finances (summarized by parent adjusted gross income), 2) the number of colleges the student listed on the FAFSA, which provides some indication of the amount of competition the college faces, and 3) the fact that the student chose to complete the FAFSA at all. Using my structural estimates, I simulate five separate counterfactuals wherein colleges are no longer permitted to use some or all of this FAFSA information in their pricing. I find that restricting FAFSA information would cause elite colleges to inefficiently price up to 13 percent of students out of the elite market because they would be unable to tailor their transaction prices as precisely and on occasion would inefficiently charge a student more than she is willing to pay. These missed matches lower total surplus by as much as \$234 per student (per year), depending on the counterfactual. Nevertheless, restricting FAFSA information raises student welfare by up to \$827 per student because, among those students who remain at elite colleges, their transaction prices would fall by an average of up to \$986 per year.[4] Restricting income information on the FAFSA tends to benefit middle- and high-income students and hurt low-income students, although these effects differ substantially across students. For example, 33 percent of low-income (bottom tercile) students would actually *benefit* from restricting FAFSA information while 14 percent of high-income (top tercile) students would be harmed. On the other hand, restricting the number of colleges listed on the FAFSA would benefit students in a more income neutral way. Taken as a whole, the results indicate that allowing colleges to use FAFSA information increases efficiency by promoting better student-college matches. However, while some students also enjoy lower transaction prices, the majority of students are harmed by higher prices.

The paper proceeds as follows. Section 2 reviews the previous literature and discusses how this paper both complements and differs from prior research. Section 3 presents a structural model of college pricing and price discrimination and discusses the intuition and qualitative predictions provided by the model. Section 4 describes the data and tests these qualitative predictions with a reduced form analysis. Section 5 proves that the structural model is nonparametrically identified, outlines the empirical strategy, and presents the baseline estimates. Section 6 models the counterfactuals and presents the counterfactual estimates. Section 7 offers concluding thoughts.

[4]Tuition revenue per student falls slightly although total tuition revenue may go up or down depending on how many students are inefficiently priced *into* the elite market. The welfare consequences of raising or lowering tuition revenues depend crucially on whether colleges are spending their marginal revenue on socially valuable activities, such as providing a valuable public good. In this paper, I do not take a stand on this issue.

2 Review of the Literature

This paper is the first to examine the consequences of allowing colleges to use a student's FAFSA information in their pricing. It is most closely related to two papers that have looked at the determinants of equilibrium pricing behavior in the U.S. college market. Fu (2014) estimates a matching model using data from the 1997 National Longitudinal Survey of Youth (NLSY97). For one of her counterfactuals, she estimates the effect of eliminating student ability measures, like test scores, on equilibrium prices and student-college matching. However, to keep the model tractable, she is unable to incorporate price discrimination into the model. Epple et al. (2006) use primarily college-level data[5] to estimate a structural model of U.S. colleges and simulate a counterfactual wherein price discrimination is banned and all colleges must price at "cost" for each student. In essence, this counterfactual simulates what would happen if colleges lost all of their market power. They find that such a drastic change would significantly affect the sorting of students into colleges as well as the market shares of different colleges. In contrast to both of these papers, I estimate the effect of restricting colleges' ability to use some or all of the FAFSA information in their pricing, while still allowing colleges to use other student characteristics (demographics, test scores, etc.) to proxy for the restricted FAFSA information. Thus, colleges are still permitted to price discriminate, but they cannot do so based directly on a student's FAFSA information.

Several economists have interpreted tuition discounts as price discrimination (Dynarski 2002; Lawson and Zerkle 2006; Tiffany and Ankrom 1998), and several papers have found that these discounts are an effective tool that colleges use to attract students (Avery and Hoxby 2004; Long 2004; van der Klaauw 2002). However, the literature on price discrimination makes a distinction between price differences driven by differences in willingness to pay and price differences driven by differences in cost (Varian 1989). Rothschild and White (1995) make essentially the same point in the context of the college market. Some price differences among students at the same college could be because some students are more desirable (for instance, because they have higher test scores) and hence are effectively less "costly" to enroll. Thus, when I structurally estimate my model of price discrimination among colleges, I also account for the possibility that students may differ in how attractive they are to colleges. Although economic theory clearly predicts that price discrimination raises producer surplus, theory can say little in general about the consequences for welfare. Indeed, Bergemann et al. (2013) obtain an "almost anything can happen" result in the case of a price-discriminating monopolist. The outcome in a given situation will depend on the nature of the supply and demand curves as well as the information that is being restricted. Hence, the consequences of restricting colleges' use of FAFSA information remain an empirical question.

[5]They use student-level data from the 1995–1996 wave of the NPSAS in one portion of their estimation procedure. However, their primary data set for most of the estimation is at the college level.

3 Modeling the U.S. College Market as a First-Price Auction

3.1 Institutional Detail

During her junior year of high school (and possibly even earlier) a student will typically take at least one national standardized test: either the ACT or the SAT. Colleges place a great deal of weight on these tests, particularly because they provide a way of comparing students from different high schools in different parts of the country. In the fall of her senior year, the student applies to one or more colleges. As early as January 1st, the student can complete the FAFSA which she must do if she wishes to be considered for federal aid. Moreover, colleges typically require students to complete the FAFSA before they will consider offering their own discounts. By April, the student receives an acceptance decision from each college to which she applied. If accepted she will also learn about any federal financial aid she might qualify for as well as any price discounts the college is offering her. With all of her offers in hand, she weighs the pros and cons of each option and selects a college. If none of the offers are satisfactory, then she always has the option of enrolling in a nonelite college. These colleges, which enroll the majority of college students in the United States, have minimal admissions requirements and will usually allow students to enroll at any point in time.

In the U.S. college market, students evaluate colleges but colleges also evaluate students. While students want to attend the best college (as judged by them) at the lowest possible price, colleges want to attract the best students (as judged by the college) at the highest possible price. At the time colleges are making price offers, their non-price characteristics are already set, but colleges can, and do, adjust their prices on a student-by-student basis. In offering a discount, they trade off lower potential tuition revenue in exchange for a higher probability of enrollment. A college performs this balancing act with each student, trying to determine how much of a discount it needs to offer the student in order to persuade her to choose it over its competitors.[6]

Students also receive aid from other sources, particularly the federal government. Students qualify for federal aid on the basis of several factors such as income, assets, and family size, but a student's federal aid is largely independent of which college she chooses to attend. In other words, the money follows the student, and, as long as she attends *some* college, her federal aid award does not incentivize her to choose one college over another.[7] The same is true for most private grants and scholarships—they are typically portable and do not directly affect the student's preferences over individual colleges.

[6]Colleges are technically limited in the size of discount they can offer by the cost of attendance (COA). A college's COA includes tuition and fees, room and board, books, and a few other expenses. If a student receives any federal aid, her total aid from all sources (including loans) may not exceed the college's COA. However, in order for this constraint on the college's discount to be binding, the student's total *grant* aid would have to be equal to the COA. In practice, less than half a percent of freshmen in my data receive such generous grant aid. It is true that a larger fraction of students, although still less than 10 percent, have a binding COA constraint once loans are taken into account. But in this case the college could have still offered a larger discount, thereby reducing the student's loans dollar for dollar.

[7]Again, as long as income effects are small. See footnote 2

3.2 The Model

Colleges fall into two groups—elite and nonelite. Elite colleges use discounts to compete with each other for students while nonelite colleges do not and operate in a competitive fringe. The nonelite sector constitutes each student's outside option. I model the matching process between students and elite colleges as a bidding game with endogenous entry, where each student is a separate "auction" and the colleges are bidders. It is important to note that my definition of elite colleges is quite broad. It includes all private four-year (non-profit) colleges and very selective public colleges.[8] I focus on elite colleges because the reduced form evidence points toward a limited role for tuition discounting outside of this group (see Figure 1). Anecdotal evidence also supports the view that tuition discounting is primarily a phenomenon of elite colleges.[9] Thus, the market for elite colleges, with its intensive application process and personalized financial aid offers, appears to differ substantially from the market for nonelite colleges. I interpret this as evidence that nonelite colleges lack the market power to successfully price discriminate.

In the model, a student invites a college to participate in her auction by submitting an application (in the fall). The college chooses to participate by admitting the student and making a take-it-or-leave-it offer (in the spring). Student i evaluates her offer from college j using the utility function $u_{ij} = v_{ij} - p_{ij}$, where v_{ij} represents her valuation, in dollars, of attending the college and p_{ij} is the price college j offers her.[10] Let $F_{V|X}$ denote the distribution of valuations conditional on a vector of student characteristics X_i that are observed to both the colleges and the econometrician. The student's valuation, v_{ij}, depends on college j's characteristics as well as how the student values those characteristics. I normalize the utility of a student's outside option, attending a nonelite college, to be zero. If none of the student's offers provide her with positive utility, then she can always enroll in a nonelite college. One strength of my model is that I am able to remain agnostic about the preferences of students. That is, rather than placing structure on the components of v_{ij}, I identify v_{ij} from equilibrium behavior.

Elite colleges compete for students on price. The college "wins" the auction—the student enrolls—if it makes the best offer, as judged by the student. Colleges care about maximizing both the quality of their students as well as tuition revenue. Let $\mathbf{\Pi}$ denote the space of college payoffs for enrolling a student. College j's payoff from enrolling student i is $\pi_{ij} = w_{ij} + p_{ij}$, where $w_{ij} = z_j + \omega(X_i) + \gamma I_{ij}$ represents college j's valuation, in dollars, of enrolling the student. I_{ij} is a dummy equal to one if college j is a public college and i is from out of state. The vector X_i denotes characteristics of student i that are observed to both the college and the econometrician, while z_j is observed to the college only. w_{ij} may be positive or negative and captures the value the student would contribute on campus as well as the cost to the college (including opportunity

[8] Very selective public colleges roughly correspond to research-oriented state colleges.

[9] While nonelite colleges typically do have a financial aid office, that office is primarily devoted to helping students qualify and apply for federal student aid programs. The admissions requirements at many of these colleges are relatively minimal, and students can and do enroll in classes right up until the school year starts.

[10] Because federal and private grants are portable from college to college, they cancel out and do not affect the student's decision. This also means that colleges will be unable to extract federal aid from students by, for instance, charging students more when they receive a larger Pell grant from the government.

cost) of enrolling her. I remain agnostic about why colleges value some students more than others. Perhaps colleges value having high ability students on campus, or maybe they anticipate that such students are likely to give large alumni donations in the future. One could imagine several other reasons why a college may prefer some students over others, but I do not place any structure on those preferences. Rather, w_{ij} serves as a sufficient statistic for how college j evaluates student i's characteristics, and I will show how to identify w_{ij} directly from data on transaction prices. Lastly, note that since w_{ij} is college j's valuation, $-w_{ij}$ is j's willingness-to-receive. That is, $-w_{ij}$ represents the lowest price that the college would be willing to offer student i because charging her less than $-w_{ij}$ would give the college a negative payoff.[11]

In this model, colleges may have market power for two reasons: 1) if the college knows it faces few competitors for a given student, then it does not need to make as generous an offer (i.e., it can bid less aggressively), and 2) if the college learns that v_{ij} is relatively high, then it has some room to extract surplus from the student. College j knows z_j and learns v_{ij} and X_i, and by extension w_{ij}, during the application process, but it does not know the v's or w's of the other bidders. It also does not know the number of bidders, n_i, but it does observe X_i which includes a noisy signal $\widetilde{n}_i$, which is the number of colleges listed on the FAFSA. The college knows the probability of n bidders conditional on the characteristics of the student, $\rho(n|X)$.

At this point, a discussion is in order regarding the assumption that colleges perfectly observe v_{ij}. This assumption is central to the identification argument later in the paper. In essence, this assumption says that the college perfectly knows how it compares with nonelite colleges in the competitive fringe, so that all of the college's uncertainty about the student's willingness to pay is driven by uncertainty about the v's and w's of its elite competitors. This assumption allows me to infer student preferences from the behavior of the colleges while remaining completely agnostic about precisely what students value in a college or in the college experience. For example, I do not need to take a stand on which college characteristics students value, nor do I need to parse out how much of the college experience is investment and how much is consumption. In practice, colleges learn about v_{ij} in a variety of ways: student essays, campus visits, whether a family member is also an alumnus, etc. Thus, a college's uncertainty about whether the student will accept its offer is likely driven by uncertainty about the behavior of its competitors. And it is precisely this strategic element that my model is designed to capture.

After observing w_{ij} and v_{ij}, college j makes a take-it-or-leave-it price offer, p_{ij}, to student i to maximize its expected payoff $\pi_{ij}\,\mathbb{P}[j \text{ wins}] = (w_{ij} + p_{ij})\,\mathbb{P}[u_{ij} \geq u_{i\ell}\ \forall \ell \neq j | X_i]$. Up to this point, we have been thinking about the college's decision in terms of price offers. However, if we recast the college's problem in terms of utility bids, we can express the model in a way that lends itself to empirical estimation. Define $s_{ij} \equiv u_{ij} + \pi_{ij} = v_{ij} + w_{ij}$ to be the total surplus from matching student i with college j. Then $w_{ij} + p_{ij} = (v_{ij} + w_{ij}) - (v_{ij} - p_{ij})$, and the college's objective can

[11] Empirically, the willingness-to-receive, $-w_{ij}$, is positive for most students. That is, colleges would be unwilling to charge a negative price to most students.

be rewritten as

$$\max_{u_{ij}} \quad \{s_{ij} - u_{ij})\} \, \mathbb{P}[u_{ij} \geq u_{i\ell} \; \forall \ell \neq j | X_i] \tag{1}$$

College j offers student i a portion of the match surplus, and student i accepts j's offer if it is the highest offer she receives. Mathematically, this is equivalent to a first-price auction in utility bids. To estimate the model, I additionally assume a symmetric independent private values (IPV) information environment. That is, conditional on student covariates X_i, the match surpluses are assumed to be drawn independently from the same distribution $F_{S|X}$ with support $\boldsymbol{S} = [\underline{s}, \overline{s}]$ ($\underline{s} \leq 0 < \overline{s}$). [12] Put differently, for a given student type, given by the covariates X_i, all variation in match surpluses is driven entirely by idiosyncratic differences in student tastes v_{ij} and college valuations w_{ij}.[13] Under this assumption, equation (1) can be written as

$$\max_{u_{ij}} \quad (s_{ij} - u_{ij}) \left\{ \sum_{n=1}^{\overline{n}} F_{S|X_i}^{n-1}\left(\beta^{-1}(u_{ij}|X_i)\right) \rho(n|X_i) \right\}, \tag{2}$$

where $\beta(s|X_i)$ is the equilibrium bid function conditional on covariates. As is standard in the auction literature, I assume that the density $f_{S|X}$ is strictly positive over the entire support. While this setup is very similar to a canonical first-price auction, it differs from the usual model in that the private values have been replaced by match surpluses s_{ij}, the bids are now in terms of student utility u_{ij} (expressed in dollars), and the number of bidders is uncertain.[14] The reserve price, now a reservation utility, is known to all the bidders. Because the student always has the option of attending a nonelite college and receiving zero utility, college j will not bother to enter the auction unless $s_{ij} \geq 0$. Intuitively, the college offers a portion of the match surplus s_{ij} to the student and keeps the rest for itself, and the transaction price p_{ij} is the means by which surplus gets transferred from one party to another. But if $s_{ij} < 0$, then there is no surplus for the college to offer the student, and the college doesn't bother to participate in the auction.

The assumption of independent private values (IPV) is restrictive, although it is used extensively in the empirical auctions literature. This assumption could be violated if colleges observed a dimension of student quality that is unobserved to the econometrician (known as unobserved auction heterogeneity). Or it could be violated if students differed along a dimension unaccounted for by X_i that is unobserved to the colleges as well (which would introduce affiliation, a strong form of correlation, into the match surpluses). In sections F.3 and F.4 of the online appendix, I extend the model to accommodate unobserved auction heterogeneity and affiliation. I find that neither extension alters the qualitative conclusions of the paper while the effect on the quantitative findings is modest.[15] Therefore, for ease of exposition I maintain the IPV assumption

[12]It is true that some colleges may be thought of as "strong" or "weak" bidders due to their realized values of z_j. For instance, if $z_j > z_{j'}$, then there is a sense in which college j is a stronger bidder than college j'. But, since z_j and $z_{j'}$ are private information and colleges cannot see who their competitors are, a symmetric model is still appropriate.

[13]After conditioning on X_i, idiosyncratic variation in w_{ij} only occurs through z_j.

[14]The uncertainty in the number of bidders complicates the algebra but does not pose a problem as long as I can separately identify and estimate $\rho(n|X)$.

[15]This suggests that the observed student characteristics X_i are sufficient to account for the component of match

throughout the rest of the paper.

Taking the first order condition for the utility bid u_{ij} in equation (2) yields the ODE

$$\beta'(s|X_i) = (s - \beta(s|X_i)) \frac{\sum_{n=1}^{\overline{n}} (n-1) F_{S|X_i}^{n-1}(s) f_{S|X_i}(s) \rho(n|X_i)}{\sum_{n=1}^{\overline{n}} F_{S|X_i}^{n}(s) \rho(n|X_i)}, \tag{3}$$

with the initial condition $\beta(0|X_i) = 0$, which says that if $s_{ij} = 0$, then the college can only just match the student's outside option.[16] The equilibrium bid function is strictly increasing. When $s_{ij} > 0$, the college offers the student some, but not all, of the positive match surplus.[17]

Intuitively, we might expect for the slope of the equilibrium bid function to be less than one so that, on the margin, an increase in match surplus is split between the college and the student. Indeed, this intuition holds "on average" because $\frac{\beta(s|X)}{s} < 1$, and it is easy to show that $\beta'(0|X) < 1$. However, it is possible for $\beta'(s|X)$ to exceed one in a region of the support of S.[18] The following assumption is required to guarantee that this does not happen.

Assumption 3.1. *Let $\mathcal{G}(s|X) = \sum_{n=1}^{\overline{n}} F_{S|X}^{n}(s)\rho(n|X)$ denote the cdf of the distribution of the highest match surplus among a college's competitors. Then $\mathcal{G}(s)^2 > \mathcal{G}'(s) \int_0^s \mathcal{G}(y)dy$ for all $s \in \mathbf{S}$.*

Lemma 3.1. *Under assumption 3.1, the slope of the bid function $\beta'(s|X)$ lies between zero and one for any value $s \in \mathbf{S}$.*

Proof. It is straightforward to show that $\beta'(s|X) > 0$ as long as $\mathcal{G}'(s|X) > 0$. To show that $\beta'(s|X) < 1$, first note that $\beta'(s|X) = (s - \beta(s|X)) \frac{\mathcal{G}'(s|X)}{\mathcal{G}(s|X)}$. The solution to this ODE is $\beta(s|X) = s - \int_0^s \frac{\mathcal{G}(y)}{\mathcal{G}(s)} dy$ (see Krishna (2010)). It follows that $\beta'(s|X) = 1 - \frac{\mathcal{G}^2(s|X) - \mathcal{G}'(s) \int_0^s \mathcal{G}(y)dy}{\mathcal{G}^2(s)}$ which will be less than one if Assumption 3.1 is satisfied. ■

As shown in Lemma 3.1, Assumption 3.1 guarantees that the equilibrium bid function has a slope that is everywhere less than one, a property that will be important for identification of the model. In essence, the assumption says that the density $\mathcal{G}'(s)$ cannot spike "too quickly" and is satisfied by many common parametric distributions. See section B in the online appendix for a further discussion about this assumption. Standard empirical auction methods do not require this assumption because they are able to achieve identification from observed bids based on the monotonicity of the equilibrium bid function. Because I cannot observe the utility bids, I must

surpluses that is common across bidders.

[16]One could also view this as an individual rationality constraint. $\beta(0) > 0$ would be irrational for the college while $\beta(0) < 0$ would never be accepted by the student.

[17]Incidentally, this implies that participating bidders, the only ones who ever submit bids, are really drawn from $F_{S|X_i}$ truncated from below at zero, and I can only identify this truncated distribution rather than the full distribution, as is always the case in auctions with a binding reservation price. For ease of notation, I will nevertheless suppress this issue and retain the notation $F_{S|X_i}$.

[18]This can happen if $f_{S|X}(s)$ "spikes quickly" in a particular way. See section B in the online appendix for more details.

identify the model from observed *ex post* college payoffs, and therefore I need Assumption 3.1 to guarantee that the *ex post* college payoff function is monotone so that the model is identified (see section 5).

3.3 Qualitative Predictions of the Model

Under assumption 3.1, the slope of the bid function $\beta'(s|X)$ lies between zero and one, so a one-dollar increase in the match surplus translates into an increase in the utility bid of β'. In other words, when the value of the match rises, the college offers a fraction β' of that gain to the student and keeps the remainder, $1 - \beta'$, for itself. But s_{ij} could rise because of either an increase in w_{ij} or an increase in v_{ij}. If w_{ij} rises, then the college values the student more and will make a higher utility bid by lowering its price offer. On the other hand, if v_{ij} rises, then the student values the college more. The college still increases its utility bid, but this time it actually *raises* its price offer. The intuition is that if the student values the college one dollar more, then the college will extract some of that dollar by raising its price, but by less than a dollar, so that its utility bid still rises. In short, the college charges the student more when it learns that the student has a higher willingness to pay.

In the model, colleges do not know precisely how many other colleges are bidding on a student. Rather, they see the student's characteristics, including a noisy signal $\widetilde{n}$ about the actual number of bidders, n. A college will respond to a higher signal by making a more aggressive bid through offering a lower price. However, holding constant the noisy signal and other student characteristics, the college's offer should be unrelated to the actual number of bidders.

4 Data Description and Reduced Form Tests of Model Predictions

4.1 Data Description

In this section I provide a brief description of the data and test some of the qualitative predictions of the model using a reduced-form analysis. The data come from the 2007–2008 wave of the National Postsecondary Student Aid Study (NPSAS). This data set contains a large, nationally representative cross section of U.S. college students enrolled during the 2007–2008 school year. As its name suggests, the study is focused on financial aid and contains an extremely rich set of variables on all aspects of student expenses and aid, including the items on each student's FAFSA form. The NPSAS also contains information on ACT/SAT scores, high school GPA, and other measures of student quality as well as information about the college the student is attending. The data provide a comprehensive picture of a student's expenses and financial aid. Unlike many other data sets, the NPSAS collects information at the student level from several different sources: government records, college administrative records, third-party organizations (e.g., ACT and the College Board) and a student interview. For example, a student's federal aid awards are pulled from federal databases, her tuition discounts come from her college's administrative records, and her SAT scores are obtained from the College Board.

	Freshmen Sample				Elite Sample			
	Mean	Std Dev	Min	Max	Mean	Std Dev	Min	Max
Sticker price	$10,168	($9,737)	$275	$39,289	$20,591	($9,676)	$1,687	$39,289
Tuition discount	$2,914	($5,722)	$0	$37,000	$7,432	($7,739)	$0	$37,000
Transaction price	$7,254	($7,170)	-$11,771	$38,678	$13,159	($8,811)	-$11,771	$38,678
Student received discount	0.38	(0.49)	0	1	0.69	(0.46)	0	1
Parent adjusted gross income	$65,400	($55,502)	$0	$496,347	$78,103	($64,883)	$0	$490,000
ACT score	21.2	(4.7)	11	36	23.5	(4.6)	11	36
High school GPA	3.42	(0.57)	0.90	4.00	3.63	(0.47)	1.40	4
Earned AP credit	0.22	(0.42)	0	1	0.31	(0.46)	0	1
Parents with college degree	0.68	(0.80)	0	2	0.91	(0.84)	0	2
Completed FAFSA	0.82	(0.38)	0	1	0.85	(0.35)	0	1
Additional colleges listed on FAFSA	1.2	(1.7)	0	5	1.9	(1.9)	0	5
Age as of 12/31/07	18.6	(0.9)	17	23	18.4	(0.7)	17	23
Female	0.54	(0.50)	0	1	0.56	(0.50)	0	1
White	0.65	(0.48)	0	1	0.68	(0.47)	0	1
Black	0.13	(0.34)	0	1	0.11	(0.31)	0	1
Hispanic	0.12	(0.33)	0	1	0.10	(0.31)	0	1
Asian	0.05	(0.22)	0	1	0.07	(0.25)	0	1
Other	0.04	(0.20)	0	1	0.04	(0.20)	0	1

Sticker prices come from NPSAS variable TUITION2. Tuition discounts come from NPSAS variable INGRTAMA. The transaction price is the difference between sticker price and the student's discount. The prices reported apply to the college the student actually attended. Additional Colleges Listed on FAFSA is only calculated for those students who completed the FAFSA. Students can list up to 6 college on the FAFSA, so the number of *additional* colleges listed runs from 0 to 5. The elite sample is a subset of the freshmen sample and consists of freshmen at private and very selective public four-year colleges. No sample weights were used.

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Postsecondary Student Aid Study, 2008.

Table 1: Descriptive Statistics

The higher education market is extremely diverse, and the NPSAS sampling scheme reflects that diversity by sampling students at a wide variety of postsecondary institutions, ranging from cosmetology programs to Ivy League universities. I restrict the sample to "traditional" college students who are less than 31 years old, are U.S. citizens or residents, and attended college in the 50 states (plus D.C.) during the 2007–2008 school year.[19] I exclude athletes because their tuition discounts are determined in a very different way from those of the general population. I also exclude students with tuition waivers because of a parent's employment at the college because those waivers are not really discounts; rather, they represent a nonwage benefit to the student's parent. The sample is restricted to U.S. citizens and residents because foreign students are not eligible to complete the FAFSA.[20]

I refer to students who satisfy the above criteria as the *full sample*. I further restrict the full sample to dependent freshmen and call this the *freshman sample*. Finally, I restrict the freshman sample to those attending private and very selective public colleges and call this the *elite sample*. The full sample consists of 33,180 students at 1,210 colleges. In the reduced-form analysis that follows I focus on the freshman sample. Table 1 contains descriptive statistics for both the freshman and elite samples. Women constitute 54 percent of the freshman sample, and the average age is 18.6 years. Freshmen received an average ACT score of 21.2 and had mean parent adjusted

[19]See section A.1.3 of the online appendix for complete details about the sample selection criteria.

[20]In the fall of 2007, nonresident aliens accounted for 2.2 percent of undergraduate enrollment (see Table 237 of the 2011 *Digest of Education Statistics*).

gross income of just over $65,000.

4.2 Testing Qualitative Predictions of the Model

The model makes four qualitative predictions. First, all else equal, elite colleges will offer a higher price to students whom it believes are willing to pay more. Second, all else equal, elite colleges will offer a lower price to students whom it views as more desirable. Third, all else equal, elite colleges will offer a lower price to students when it believes it is facing stiffer competition. And fourth, holding constant student characteristics and the number of colleges listed on the FAFSA, the price a student is offered will be unrelated to the actual number of colleges a student applies to. To test these predictions, Table 2 reports the estimates from the following regression

$$\text{Tuition Discount}_{ij} = X_i\beta + a_j + e_{ij}, \tag{4}$$

where i indexes students, j indexes colleges, X_i represents student covariates, and a_j is a college fixed effect. By including college fixed effects, I am isolating variation across students at the same college. I also include in X_i a dummy for whether the student is an out-of-state student at a public college. Ideally, I would regress tuition discount *offers* on student characteristics. Unfortunately, I only observe the actual tuition discount student i receives at the college she attends.[21] Nevertheless, the qualitative predictions listed above should still hold, even though I only observe the winning college.

Consistent with the first prediction of the model, a student's willingness-to-pay, proxied for by parent income, is associated with a higher price. A $10,000 increase in parent adjusted gross income is associated with a $124 decrease in the student's tuition discount. Moreover, the first panel of Figure 1 demonstrates that this relationship is driven by elite colleges rather than nonelite colleges. Consistent with the second prediction of the model, a student's desirability to the college, proxied for by ACT scores and high school GPA, is associated with a lower price. Consistent with the third prediction, students with a higher signal of competition—that is, with more colleges listed on the FAFSA—pay lower prices. Listing one more college on the FAFSA is associated with a $373 increase in the student's discount.[22]

Do these discounts simply amount to colleges' efforts to direct need-based aid to low-income students? It is useful to contrast tuition discounts with federal need-based aid. To illustrate the difference, I re-estimate the regression specification in (4) with federal grants as the dependent variable. Figure 1 contrasts the relationship between parent income and tuition discounts vs federal grants. In both regressions I include parent income as a cubic spline. Federal grants

[21]In other words, the observed tuition discount is the "accepted" discount, to borrow from the labor economics literature.

[22]Of course, these estimates are consistent with other interpretations as well. Perhaps colleges simply prefer lower-income students over higher-income students. And perhaps ACT scores are also correlated with a student's willingness-to-pay. Although reduced form estimates cannot distinguish between these interpretations, the structural estimates do allow us to separately identify price variation due to college preferences from price variation due to willingness-to-pay.

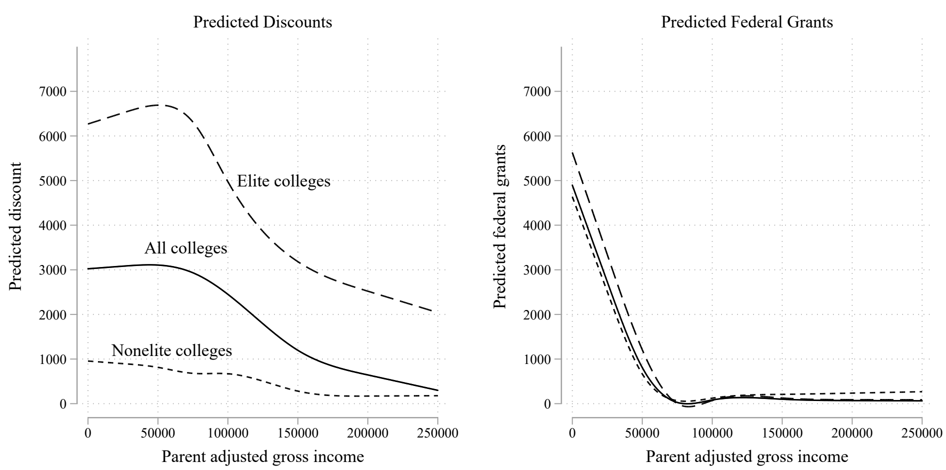
Dependent Variable	*Tuition Discount*		*Federal Grants*
Parent AGI (in $10,000's)	-123.5 (15.22) ***	*(cubic spline)*	*(cubic spline)*
Number of additional colleges listed on FAFSA	372.7 (56.64) ***	366.3 (56.09) ***	39.1 (15.30) *
Completed FAFSA	850.0 (164.3) ***	870.3 (163.6) ***	522.3 (55.43) ***
ACT score	133.4 (18.43) ***	130.5 (18.53) ***	-3.5 (6.062)
High school GPA	394.1 (101.1) ***	384.7 (100.8) ***	-5.0 (45.29)
Earned AP credit in high school	277.7 (174.2)	292.1 (174.2)	17.2 (51.18)
Number of parents with college degree	93.8 (90.59)	114.6 (91.07)	-67.5 (27.45) *
Age as of 12/31/07	-37.5 (58.41)	-18.6 (58.25)	-86.9 (30.02) **
Female	113.3 (115.6)	111.1 (115.5)	47.9 (40.93)
Black	580.2 (232.0) *	649.8 (232.8) **	340.6 (93.32) ***
Hispanic	581.0 (245.8) *	635.2 (245.7) **	228.0 (86.48) **
Asian	44.1 (391.9)	91.2 (390.2)	278.4 (104.4) **
Other / Multiple	549.7 (360.7)	517.0 (358.1)	100.5 (108.5)
Out-of-State, Public	1135.5 (295.4) ***	1162.4 (295.3) ***	61.7 (97.12)
College fixed effects	Yes	Yes	Yes
Parent AGI included as cubic spline	No	Yes	Yes
Observations	5640	5640	5640
R-squared	0.670	0.672	0.718

The regressions include students from the freshmen sample. The omitted race category is "white." For students who completed the FAFSA, "number of additional colleges listed" ranges from zero to 5 (students can list up to six colleges on the FAFSA). For those who did not complete the FAFSA, "number of additional colleges listed" is set to zero and the dummy "completed FAFSA" is included. In columns 2 and 3, Parent Adjusted Gross Income is included as a cubic spline with knots at $25k, $50k, $75k, $100k, $150k, and $200k, although the coefficients are not reported here (see Figure 1). Robust standard errors are in parentheses. Sampling weights were used (NPSAS variable WTA000).

* p<0.05, ** p<.01, *** p<.001

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Postsecondary Student Aid Study, 2008.

Table 2: Within-College Discounting and Federal Aid Patterns



The fitted values plotted here come from regressing discounts and federal grants on student covariates. Regressions were estimated on the full freshman sample and on elite and nonelite colleges separately. In all six regressions, the fitted values represent a white female student with all other covariates set to their sample means. Parent adjusted gross income is included as a cubic spline with knots at $25k, $50k, $75k, $100k, $150k, and $200k. Sample weights were used in all regressions (NPSAS variable WTA000).

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Postsecondary Student Aid Study, 2008 and Beginning Postsecondary Students, 2004.

Figure 1: Relationship between Parent AGI and Tuition Discounts and Federal Grants

decline steeply with parent income, starting at about $5,000 for the poorest students and reaching zero at about $75,000 of adjusted gross income. After that, federal grants are essentially flat at zero. In contrast, tuition discounts at elite colleges begin at about $6,000 for the poorest students. They are relatively flat at first until they begin declining, slowly, at about $75,000 in adjusted gross income. While students with parent income above $75,000 receive almost no federal grants, even those with parent income in excess of $200,000 receive on average more than $2,000 in discounts.

Tuition discounts at elite colleges increase with the number of colleges listed on the FAFSA. Among students at the same very selective private college, all else being equal, those who listed one more college on the FAFSA tended to enjoy a $910 larger tuition discount. Column one of Table 3 replicates the regression from Table 2. In column two, I interact the number of colleges listed with college type and selectivity, which shows that the coefficient in column one is driven by elite colleges, especially very selective private colleges.

Dependent Variable: Tuition Discount

Data Set			**NPSAS 07-08**	**NPSAS 07-08**	**BPS 03-04**	**BPS 03-04**	**BPS 03-04**
Number of additional colleges listed on FAFSA			372.7 (56.64) ***		269.4 (37.36) ***	267.3 (39.59) ***	
Number listed on FAFSA ×	Very selective	Public		131.8 (90.02)			246.9 (84.39) **
		Private		909.8 (239.9) ***			344.0 (149.4) *
	Moderately selective	Public		56.0 (62.38)			87.9 (49.34)
		Private		249.3 (155.8)			148.5 (104.3)
	Not selective	Public		44.9 (104.6)			-114.3 (87.57)
		Private		264.5 (307.7)			157.0 (196.6)
	Two-year	Public		-0.5 (36.80)			-7.2 (52.15)
		Private		-463.8 (204.5) *			91.8 (411.1)
Number of colleges applied to						3.9 (26.64)	
Number of colleges applied to ×	Very selective	Public					-31.5 (43.38)
		Private					67.0 (75.13)
	Moderately selective	Public					86.0 (36.85) *
		Private					-53.4 (77.07)
	Not selective	Public					75.5 (71.52)
		Private					23.5 (163.0)
	Two-year	Public					-4.7 (30.39)
		Private					20.0 (280.1)
College fixed effects			Yes	Yes	Yes	Yes	Yes
Observations			5640	5640	5290	5290	5260
R-squared			0.670	0.714	0.640	0.640	0.678

See notes to Table 2. The regressions reported in columns 1 and 2 are identical to the first regression in Table 2 except that in column 2 the number of additional colleges listed on the FAFSA has been interacted with college selectivity. The remaining covariates were included but not reported here. Column 3 reports estimates from the same regression specification as in column 1, but using data from BPS:2003-2004. The reported coefficient in column 3 is smaller than in column 1 partly because the dependent variable is expressed in current dollars (no adjustment for inflation). In column 4, the number of colleges the student actually applied to was included as an additional control (this variable is available in BPS but not in NPSAS). In column 5, the number of additional colleges listed on the FAFSA and the number of college applications were both interacted with college type. The number of colleges listed on the FAFSA is set to zero for students who did not complete the FAFSA and a dummy for whether the student completed the FAFSA is included. In columns 2 and 5, when the number of colleges listed on the FAFSA is interacted with college type, the dummy for completing the FAFSA is also interacted with college type. Robust standard errors are in parentheses. Sampling weights (NPSAS and BPS variable WTA000) were used in all regressions.

* p<0.05, ** p<.01, *** p<.001

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Postsecondary Student Aid Study, 2008, and Beginning Postsecondary Students, 2004.

Table 3: The Number of Colleges Listed on a Student's FAFSA

We might worry that the positive relationship between the number of colleges listed on the FAFSA and a student's tuition discount may reflect a "selection effect" rather than a "competition effect.” For example, suppose students chose colleges based purely on price. If this were true, then students who list more colleges on the FAFSA, and tend to apply to more schools, would also tend to have larger tuition discounts. In this case, the selection effect would mechanically introduce a positive relationship between the number of colleges listed and the student’s tuition discount. To address this concern, I add the number of actual applications as a control in columns four and five of Table 3. I use data from the 2003–2004 wave of Beginning Postsecondary Students (BPS) which directly asks freshmen how many colleges they applied to.[23] The number of actual applications should proxy for the selection effect, if it exists, while the number listed on the FAFSA should proxy for the competition effect. If a selection effect were mechanically producing the positive correlation between colleges listed on the FAFSA and discounts, then adding actual applications to the regression should reduce the coefficient on the number listed on the FAFSA. In fact, consistent with the fourth prediction of the model, the reverse occurs—the coefficient on number listed on the FAFSA is virtually unchanged while the coefficient on number of actual applications is close to zero. This result suggests that the positive correlation between discounts and the number of colleges listed on the FAFSA is not driven by a selection effect but by a competition effect.[24] The timing of when students complete the FAFSA provides another piece

Dependent Variable: Tuition Discount

Number of additional colleges listed on FAFSA	January	530.6	(144.2)	***
	February	290.7	(81.7)	***
	March	11.3	(91.5)	
	April or later	77.4	(95.1)	
College fixed effects		Yes		
Observations		5630		
R-squared		0.673		

See notes to Table 2. The regression specification here is identical to column 1 of Table 2 except that the month the student's FAFSA was received is interacted with the number of colleges listed on her FAFSA. Dummies for month were included but not reported (the omitted category is those students who did not complete the FAFSA). The remaining covariates from Table 2 were included but not reported here. Robust standard errors are in parentheses. Sampling weights were used (NPSAS variable WTA000).

* p<0.05, ** p<.01, *** p<.001

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Postsecondary Student Aid Study, 2008.

Table 4: Colleges Listed and Month FAFSA Completed

of evidence in favor of a competition effect. If the correlation between discounts and the number of colleges listed on the FAFSA were purely due to a selection effect, then the timing of when the

[23]BPS is an offshoot of NPSAS. The 2003–2004 wave of BPS is the only wave of data which asked students about the number of colleges they applied to.

[24]One way to interpret these regression results is that the competition effect is substantial while the selection effect is not. A small selection effect could arise because students care about both price and quality, and higher quality may be more expensive. A student’s preferred college among her choice set may offer a larger or smaller discount than her other options, and it so happens that the discount tends to be about the same. This is not to say that students don’t care about discounts. Rather, it is just to say that, in equilibrium, a student’s preferred college offer tends to offer a discount that is about the same as the average discount offers of the competing colleges.

FAFSA was actually completed should not matter. On the other hand, if the relationship between tuition discounts and listing more colleges on the FAFSA were really due to a competition effect, then we would expect the relationship to disappear for students who complete the FAFSA after admissions season is over. In Table 4 I find that the positive correlation between discounts and the number listed on the FAFSA is strongest for those who complete the FAFSA early, before colleges have made their offers. For those who complete the FAFSA after the market has already cleared, there is little or no relationship between the number of colleges listed on the FAFSA and the student's tuition discount.[25]

There are multiple interpretations of the positive relationship between the number of colleges a student lists on the FAFSA and her tuition discount. Colleges may be responding to increased competition from other colleges. On the other hand, perhaps the number of colleges on the FAFSA is correlated with an omitted variable that is also correlated with a student's discount. For instance, students who list more colleges on the FAFSA may also tend to be more price sensitive.[26] If colleges knew this, or if they had some indication of a student's price sensitivity, then they may offer more generous discounts to these students. Thus, the finding that students who list more colleges on the FAFSA receive larger discounts may reflect differences in price sensitivity rather than increased competition. In the structural model, I will allow for both possibilities. Colleges will respond to increased competition, and they can use the number of colleges listed on the FAFSA to make inferences about student preferences.

In summary, the reduced form evidence is consistent with the structural model. Although the reduced form evidence alone cannot distinguish between college preferences and price discrimination, the structural model does allow me to disentangle price variation due to the preferences of colleges from price variation due to price discrimination.

5 Identification and Estimation

Standard empirical auction methods combine economic theory with observed bids to estimate the structural primitives of the auction model. We could use such methods to estimate the structural model from section 3.2, except for one problem: the bids are expressed in terms of student utility and are thus never directly observed in the data. Rather, I observe the tuition offer, p, which is only one component of the utility bid $\beta(s) = v - p$. While this appears to be a serious problem, in this section I show how we can still identify the model using data on student characteristics and transaction prices as well as an additional restriction on the distribution of

[25]Of course, the timing of when a student completes the FAFSA is itself an endogenous choice, and we would not be surprised to find that students with larger discounts tended to complete the FAFSA more quickly. But it is less clear whether or in which direction this endogeneity would bias the coefficients estimated in Table 4.

[26]Bettinger et al. (2012) run a field experiment where customers at H&R Block were randomly selected to receive assistance with completing the FAFSA. They find that assistance with the FAFSA increased college attendance from 34 percent to 42 percent one year later. The results suggest that some students find the FAFSA daunting enough to deter them from attending college. Students may also differ in the costs of applying to colleges, and it is possible that students who apply to more colleges may be more price sensitive than those who apply to fewer colleges.

match surpluses (Assumption 3.1).The identification strategy is similar in spirit to that used by Guerre et al. (2000) for first-price auctions. They show how to transform the bidders' first order condition to express unobservables (bidder valuations) in terms of observables (bids and the equilibrium bid distribution). My approach is similar, except I must first deal with the fact that the bids themselves are not directly observed.

The identification strategy is as follows. Define the *ex post* college payoff function $\pi(s|X_i) \equiv s - \beta(s|X_i)$, which gives the college payoff as a function of the total surplus s. Note that under Assumption 3.1 $\pi(s|X_i)$ is monotone in s with $\pi'(s|X_i) \in (0,1)$.[27] Denote the distribution of college payoffs, π, conditional on X_i by $F_{\pi|X_i}$. The derivative of the *ex post* payoff function $\pi(s|X_i)$ is $\pi'(s|X_i) = 1 - \beta'(s|X_i)$. By substituting equation (3) in for β' and using the fact that $F_{S|X_i}(s) = F_{\pi|X_i}(\pi(s|X_i))$ and therefore $f_{\pi|X_i}(\pi(s|X_i)) = \frac{f_{S|X_i}(s)}{\pi'(s|X_i)}$, we can derive the following ODE for π

$$\pi'(s|X_i) = \left(1 + \pi(s|X_i)\frac{\sum_{n=1}^{\overline{n}}(n-1)F^{n-1}_{\pi|X_i}(\pi(s|X_i))f_{\pi|X_i}(\pi(s|X_i))\rho(n|X_i)}{\sum_{n=1}^{\overline{n}}F^{n}_{\pi|X_i}(\pi(s|X_i))\rho(n|X_i)}\right)^{-1}, \quad 0 \leq s \quad (5)$$

$$\pi(0|X_i) = 0. \quad (6)$$

Notice that, since $\pi : \boldsymbol{S} \to \boldsymbol{\Pi}$ is monotone, its inverse $\psi : \boldsymbol{\Pi} \to \boldsymbol{S}$ exists and has a derivative that is simply the reciprocal of π'. So we can write

$$\psi'(\pi|X_i) = 1 + \pi\frac{\sum_{n=1}^{\overline{n}}(n-1)F^{n-1}_{\pi|X_i}(\pi)f_{\pi|X_i}(\pi)\rho(n|\widetilde{n}_i)}{\sum_{n=1}^{\overline{n}}F^{n}_{\pi|X_i}(\pi)\rho(n|\widetilde{n}_i)} \qquad 0 \leq \pi \quad (7)$$

$$\psi(0|X_i) = 0. \quad (8)$$

Finally, Equation (7) can be solved by simply integrating from 0 to π. $\psi(\pi|X)$ is the equilibrium inverse payoff function; it maps from the space of college payoffs, $\boldsymbol{\Pi}$, to the space of match surpluses, $\boldsymbol{S}$. $\psi(\pi|X)$ depends only on the equilibrium distribution of colleges payoffs, $F_{\pi|X}$, and the distribution of potential bidders, $\rho(n|\widetilde{n})$. Equation (7) provides the key to identifying the model.

In section 5.1, I prove that the model can be identified from equilibrium transaction prices, but the intuition behind the proofs is not complicated. From Equation (7) we can see that the model is identified if we observe the distribution of college payoffs, $F_{\pi|X}$. Since college payoffs $\pi_{ij} = w_{ij} + p_{ij}$, these payoffs really just amount to a location shift of prices. Therefore, the distribution of college payoffs, $F_{\pi|X_i}$, is just a shifted version of the distribution of transaction prices, $F_{p|X_i}$, with the shift equal to w_{ij}. It turns out that w_{ij} can also be identified from data on transaction prices. $-w_{ij}$ will be equal to the lowest price that college j would ever charge a student like student i (i.e., with covariates equal to X_i). So data on transaction prices and student

[27] This is where the additional restriction imposed by Assumption 3.1 allows us to identify the model. We need this additional assumption because, rather than identifying match surpluses from bids (which are not observed), we must identify them from *ex post* college payoffs. Thus, we need the *ex post* college payoff function to be monotone.

covariates can be used to estimate the conditional distribution of college payoffs. The conditional distribution of payoffs, $F_{\pi|X}$, and its density, $f_{\pi|X}$, can then be combined with $\rho(n|X)$ in Equation (7) to identify the primitives of the model.

5.1 Identification

Lemma 5.1. *The distribution of winning payoffs $G_{\pi|X}$ is identified from $G_{p|X,j}$, the distribution of winning transaction prices conditional on student characteristics X and the identity of the college j.*

Proof. Recall that college j's payoff from enrolling student i is $\pi_{ij} = \omega(X_i) + z_j + p_{ij}$. Thus, holding constant the college's identity and student covariates X_i, the distribution of college payoffs $\pi_{ij}|X_i, j$ is simply the distribution of transaction prices $p_{ij}|X_i, j$ shifted by the constant $w_{ij} = \omega(X_i) + z_j$. Define the function $y(X, j) \equiv \inf\{\text{supp}(G_{p|X,j})\}$ to be the greatest lower bound of the support of transaction prices, conditional on student covariates and the identity of the college. Now define the random variable $\pi|X, j \equiv p|X, j - y(X, j)$ and call its distribution $G_{\pi|X,j}$. Integrating over j yields the distribution $G_{\pi|X}$.

It remains to show that $y(X, j)$ exists and is equal to college j's willingness to receive $-w_{ij}$ when $X = X_i$. By definition, $-w_{ij}$ is a lower bound of $\text{supp}(G_{p|X,j})$.[28] But is $-w_{ij}$ the greatest lower bound? Suppose, by way of contradiction, that it is not; that is, suppose $-w_{ij} < y(X, j)$. Recall from Equation (6) that $\pi(0|X) = 0$; that is, the college receives zero payoff from matches with zero surplus. Since, by assumption, the density $f_{S|X} > 0$ everywhere, including at 0, a positive mass of winning college payoffs will lie in the interval $[-w_{ij}, y(X, j)]$. But this means that $y(X, j)$ is not a lower bound after all. Thus, $y(X, j) = -w_{ij}$. ∎

Lemma 5.2. *The distribution of college payoffs $F_{\pi|X}$ is identified from the distribution of winning payoffs $G_{\pi|X}$ and the distribution of actual bidders $\rho(n|\widetilde{n})$.*

Proof. $G_{\pi|X}$ is the distribution of *winning* payoffs. Since colleges are using a monotone bidding function, the college with the highest match surplus will win the auction. Furthermore, under assumption 3.1 the payoff function $\pi(s|X)$ is monotone so that the winning college will also have the highest payoff. In other words, the observed winning payoff is just the first order statistic of all payoffs in the auction. Thus, $G_{\pi|X}$ and $F_{\pi|X}$ are related according to $G_{\pi|X}(z) = \sum_{n=1}^{\overline{n}} \rho(n|X_i) F^n_{\pi|X}(z)$. Define the transformation $T(\alpha) \equiv \sum_{n=1}^{\overline{n}} \rho(n|\widetilde{n})\alpha^n$ and note that T is monotonically increasing for $\alpha \in [0, 1]$. Thus, T can be inverted and $F_{\pi|X}(z) = T^{-1}(G_{\pi|X}(z))$ for all z. ∎

Theorem 1. *The distribution of match surpluses $F_{S|X}$ is identified from the distribution of college payoffs $F_{\pi|X}$ and the distribution of actual bidders $\rho(n|\widetilde{n})$.*[29]

[28] The college would never be willing to charge a student less than $-w_{ij}$.

[29] To be more precise, I actually identify the truncated distribution of $F_{S|X_i}$ conditional on $S \geq 0$, which is the standard result in the presence of a binding reserve price.

Proof. Combining $F_{\pi|X}$ and $\rho(n|\widetilde{n})$ with Equations (7) and (8) defines a unique inverse payoff function $\psi(\pi|X)$ that maps college payoffs into match surpluses. $\psi(\pi|X)$ links the payoff distribution $F_{\pi|X}$ to the match surplus distribution $F_{S|X}$ according to $F_{\pi|X}(z) = F_{S|X}(\psi(\pi|X)|X)$ and $f_{\pi|X}(z) = f_{S|X}(\psi(\pi|X)|X)\psi'(\pi|X)$.

■

5.2 Empirical Strategy

I estimate the model using student-level data from the 2007–2008 wave of the National Postsecondary Student Aid Study (NPSAS). NPSAS contains information on various student characteristics as well as detailed information about prices and discounts. For the structural estimation I focus on freshmen at elite colleges—private and very selective four-year colleges. To estimate the model, I first directly estimate the distribution of the number of bidders $\rho(n|X)$.[30] Then, I follow a two-step empirical procedure in the spirit of Guerre et al. (2000). In the first step, I estimate w_{ij} and, by extension, π_{ij}. In the second step, I estimate the distribution of college payoffs $F_{\pi|X}$ and combine this with the equilibrium conditions in Equations (7) and (8) to recover $F_{S|X}$, the equilibrium bid function $\beta(s|X)$, and an estimate of the match surplus, $\widehat{s}_{ij}$.

5.2.1 Estimating $\rho(n|X)$

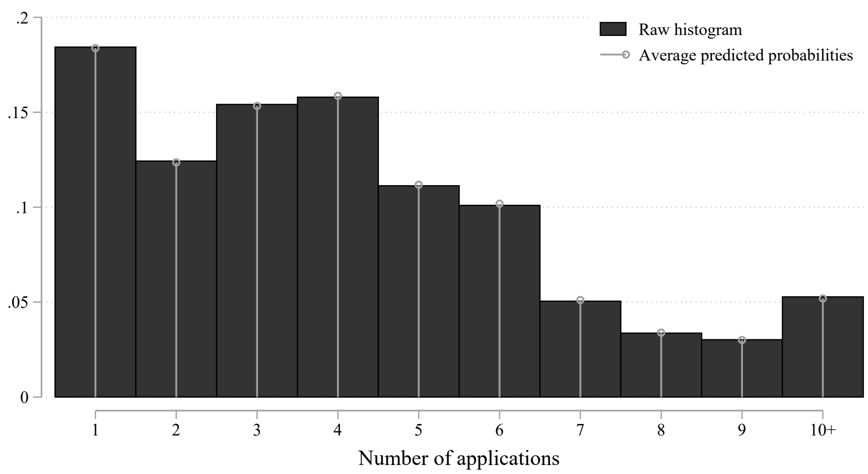
The number of colleges a student lists on her FAFSA is related to, but not necessarily the same as, the number of colleges to which she applies. NPSAS does not ask students about applications, but Beginning Postsecondary Students (BPS), a related study that focuses on first-time freshmen in NPSAS, does ask about applications.[31] Figures 2 and 3 use the BPS data to look at the relationship between the number of colleges a student applies to and the number she lists on her FAFSA. In Figure 2 we can see that, although some students apply to many colleges, most students apply to only a handful. Although colleges do not necessarily know how many colleges a student has applied to, they can see how many colleges she has listed on her FAFSA. In Figure 3, I plot the distribution of number of applications separately by the number of colleges listed on the FAFSA. We see that the number listed on the FAFSA is an informative, although imperfect, signal about a college's actual number of (potential) competitors. Indeed, the mode of the distribution of applications is always equal to the number of colleges listed on the FAFSA.

To estimate $\rho(n|X)$, I regress the number of colleges a student applied to, n_i, on student covariates, X_i, using a generalized ordered probit regression (Williams 2006). The sample for this regression comes from the 2003–2004 wave of BPS. The regression specification is identical to a standard ordered probit model, except that the number of colleges listed on the FAFSA (and whether the student completed the FAFSA) are interacted with the cut points. Row 6 of

[30]For the structural estimation, I treat students who completed the FAFSA after March 31 as though they did not complete the FAFSA. I code their dummy for completing the FAFSA to zero and the number of colleges listed on the FAFSA to zero, just as for those students who never complete the FAFSA.

[31]See section A.2 of the online appendix for more information about the BPS data.

Table 7 reports the pseudo-R^2 and root mean squared error from this regression. By far, the number of colleges listed on the FAFSA provides the single best predictor of the number of colleges a student actually applies to. In fact, including only the number of colleges listed on the FAFSA and whether the student completed the FAFSA yields a pseudo-R^2 of 27.2 percent (see row 7 of Table 7). Adding the remaining covariates to the regression raises the pseudo-R^2 to 30.2 percent. Although the student covariates do provide information about the number of college applications, the root mean squared error of the ordered probit model never falls below 2.1, leaving colleges with a significant degree of uncertainty about how many competitors they face for any given student. After estimating the ordered probit model on the BPS data, I then estimate $\rho(n|X_i)$ as the predicted probability of n applications for student i (from the 2008 NPSAS sample). In Figure 3 I average those predicted probabilities separately by the number of colleges listed on the FAFSA. These average predicted probabilities match closely with the raw histograms in Figure 3.[32]



The raw histogram is calculated based on the counts of students in each cell. The average predicted probabilities simply average the predicted probability of applying to a given number of schools, obtained from the baseline generalized ordered probit model, over all students.

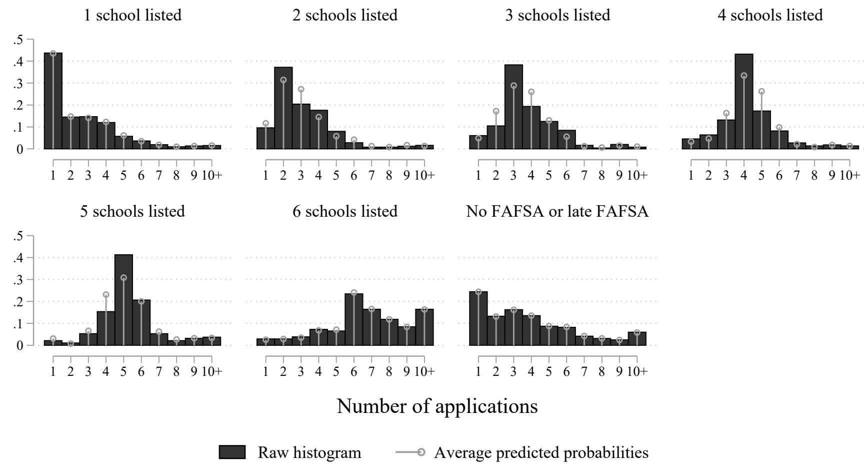
SOURCE: U.S. Department of Education, National Center for Education Statistics, Beginning Postsecondary Students, 2004.

Figure 2: Histogram of Applications

5.2.2 Two-step empirical procedure

I begin by estimating w_{ij}. Recall that $-w_{ij}$ is college j's willingness to receive for student i and is identified by $\underline{p}(X,j) \equiv \inf\{\text{supp}(p|X,j)\}$. Unfortunately, although w_{ij} is nonparametrically identified, the identification proof does not immediately suggest an estimator that could be used in a finite data set. Therefore, I adopt a parametric assumption about the distribution of $p|X,j$

[32]As a robustness check, I also estimate the model assuming that the number of colleges listed on the FAFSA perfectly reveals the number of potential bidders (see section F.6 in the online appendix). The main consequence of this alternative assumption is to magnify the results I report in the paper.



The raw histogram is calculated based on the counts of students in each cell. The average predicted probabilities simply average the predicted probability of applying to a given number of schools, obtained from the baseline generalized ordered probit model, over all students in each group. A FAFSA is late if it is received after March 31.

SOURCE: U.S. Department of Education, National Center for Education Statistics, Beginning Postsecondary Students, 2004.

Figure 3: Histogram of Applications Given Schools Listed on FAFSA

and assume that the left tail of the cdf $F_{p|X,j}$ follows the parametric form

$$\widehat{F}_{p|X,j} = \alpha_1(X,j)(p - \underline{p}(X,j)) + \alpha_2(X,j)(p - \underline{p}(X,j))^2, \tag{9}$$

where $\alpha_1(X,j) > 0$, $\alpha_2(X,j) > 0$, and $\underline{p}(X,j)$ are all parameters to be estimated. This quadratic form for the left tail of the cdf implies a linear and nondecreasing density in the left tail.[33] We can think about this assumption in two ways. The first is to treat it as an assumption about the precise functional form of $F_{p|X,j}$. The second is to treat this assumption as a polynomial approximation of the left tail. Whichever perspective we take, in order to estimate α_1, α_2, and $\underline{p}$, I estimate several quantiles of the price distribution $F_{p|X,j}$ using the quantile regression[34]

$$F^{-1}_{p|X,j}(t) = X\gamma^t + a^t_j \qquad t = .05, .10, \ldots, .40, \tag{10}$$

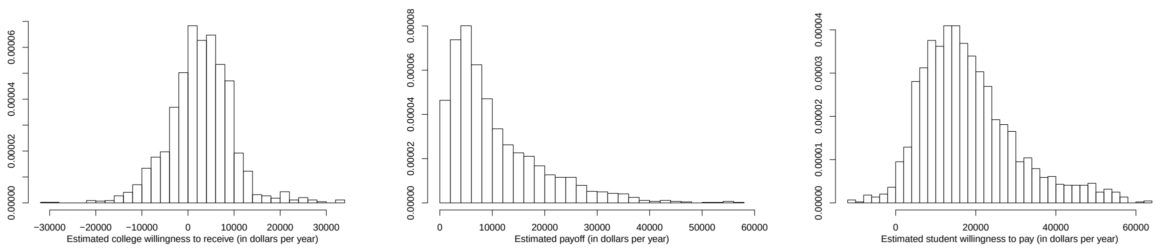
where the a^t_j are college fixed effects.[35] For each observation, I obtain the fitted values from these quantile regressions, giving me eight (estimated) points on the left tail of $F_{p|X,j}$. Then, separately for each observation, I fit the curve in (9) to these points by estimating the parameters α_1, α_2,

[33] I have also estimated a cubic specification for equation (9), but the results were essentialy unchanged (see appendix section F.5).

[34] I include all observations from the full sample, including upperclassmen, in these quantile regressions. This is the only place in the structural estimation where I use data on students other than freshmen. I interpret the prices of upperclassmen as a continuation of the offers they received when they were freshmen. However, I don't observe whether these upperclassmen completed the FAFSA when they were freshmen, nor do I observe the number of colleges they listed on the FAFSA. So I impose the exclusion restriction that neither variable enters into a college's valuation for a student w_{ij}, and I exclude both variables from the quantile regressions.

[35] I also run the estimation using only $t = .05, .10, \ldots, .25$, but the results are essentially unchanged (see appendix section F.5).

and $\underline{p}$ via nonlinear least squares, subject to the constraint that $\underline{p}(X_i, j) \leq p_{ij}$.[36] Figure 4a plots a histogram of the estimated willingness to receive for the students in the elite sample. For the majority of students, the willingness to receive is positive, although typically not above $10,000. For some students, the willingness to receive is negative, meaning that the college would be willing to offer that student discounts in excess of sticker price (which could be used to cover books or room and board).



(a) Willingness to Receive ($-w_{ij}$) (b) College Payoffs (π_{ij}) (c) Willingness to Pay (v_{ij})

Figure 4: Histograms of estimated college willingness to receive, college payoffs, and student willingness to pay
SOURCE: U.S. Department of Education, National Center for Education Statistics, National Postsecondary Student Aid Study, 2008, and Beginning Postsecondary Students, 2004.

College valuations are simply the negative of willingness to receive, $\widehat{w}_{ij} = -\widehat{\underline{p}}(X, j)$, and I estimate the college payoff as $\widehat{\pi}_{ij} = \widehat{w}_{ij} + p_{ij}$. Figure 4b plots a histogram of estimated college payoffs.

The second step in the estimation process begins by estimating the distribution of winning college payoffs, conditional on student covariates. Students in the elite sample vary considerably in their characteristics. In order to justify the independent private values auction framework, I condition on student covariates that would be observable to colleges and estimate $G_{\pi|X_i}$ which is the distribution of *winning* college payoffs conditional on student covariates.[37] $G_{\pi|X_i}$ is related to the parent distribution of college payoffs $F_{\pi|X_i}$ by the expression $G_{\pi|X}(z) = \sum_{n=1}^{\overline{n}} \rho(n|X) F^n_{\pi|X}(z)$, which I invert to solve for $F_{\pi|X}$. Substituting $F_{\pi|X}$ and $\rho(n|X)$ into equation (7), I numerically solve for the inverse college payoff function $\psi(\pi|X)$. For each student, I calculate $\widehat{s}_{ij} = \psi(\widehat{\pi}_{ij}|X_i)$, which is the estimated match surplus from matching student i with the college she ended up attending as well as $\widehat{v}_{ij} = \widehat{s}_{ij} - \widehat{w}_{ij}$. Finally, the $\widehat{s}_{ij}$ are also used to estimate $F_{S|X_i}$, the distribution of match surpluses conditional on student covariates.

5.3 Baseline Structural Estimates

On average, colleges extract 69.9 percent of the match surplus through their individualized prices while students receive the remaining 30.1 percent (see Table 9). However, a student's share of the

[36]This constraint binds for about three percent of the students in my sample.
[37]I provide more details on this in section C in the online appendix.

surplus depends on how much competition the college believes it faces. Table 5 illustrates how the student share of the match surplus rises with the number of colleges listed on the FAFSA. Those who list six colleges receive 41.9 percent of the match surplus, while students who list only one college receive just 16.9 percent.

		Average student share of match surplus
Number of schools listed on FAFSA	No FAFSA	25.7%
	1	16.9%
	2	32.6%
	3	37.7%
	4	40.2%
	5	40.4%
	6	41.9%

Each cell reports the average student share of the match surplus for students in that cell. Students who completed the FAFSA after March 31 are included in the "No FAFSA" cell. No sample weights were used.

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Postsecondary Student Aid Study, 2008.

Table 5: Student Share of Match Surplus

The first three rows of column one of Table 9 contain a summary of the structural and counterfactual estimates. Total surplus per student, net of attending a nonelite college, averages \$15,417 per year. This is the total value to both the college and the student of having the student attend her observed college rather than a nonelite college. On average, students receive \$5,023 of consumer surplus while the remainder accrues to the colleges. Note that these estimates *do not* imply that students value attending college at a mere \$5,023 per year. Rather, this number represents the student surplus above and beyond what the student would have received if she had attended a nonelite college. Adding the average student surplus of \$5,023 with the average transaction price of \$13,158 implies that, on average, students at elite colleges are willing to pay \$18,181 to attend their current colleges rather than a nonelite college (their outside option).

6 The Counterfactuals

The FAFSA collects over 70 pieces of information about a student's background, although much of the information turns out to be largely redundant. For instance, in the freshmen sample, parent adjusted gross income alone can explain 73 percent of the variation in a student's Expected Family Contribution.[38] Indeed, Dynarski and Scott-Clayton (2006) argue that much of the information collected on the FAFSA could be ignored with negligible effects on the targeting of federal aid. With this in mind, I assume that the FAFSA provides colleges with three pieces of information on a student: 1) family finances (summarized by parent adjusted gross income), 2) a noisy signal about the number of competitors the college faces (the number of colleges listed on the FAFSA), and 3) the fact that the student chose to complete the FAFSA at all.

[38] A student's expected family contribution (EFC) is the federal government's assessment of how much the student and her family should be expected to contribute toward her education. The government uses this number as a sufficient statistic for determining a student's federal aid eligibility.

I simulate five counterfactuals wherein colleges are restricted from using some or all of this information. In the first counterfactual, colleges must use a coarsened version of parent income in lieu of parent adjusted gross income. This counterfactual is meant to capture the possibility that colleges may use characteristics like a student's neighborhood or home value to proxy for parent income. Regressions based on census data (described below), indicate that neighborhood income and home value can explain about one-third of the variation in household income, which is very close to how much can be explained by a dummy variable for being above or below the poverty line for a family of four. Thus, in this counterfactual (and in counterfactual four below) I restrict colleges to use a dummy variable for being above or below the poverty line in lieu of parent income. In the second counterfactual, colleges cannot use parent income at all. In the third counterfactual, colleges cannot use the number of colleges listed on the FAFSA. In the fourth counterfactual, colleges cannot use any FAFSA information, including whether the student completed the FAFSA, but they can use the dummy variable for whether parent AGI is below the poverty line. Finally, in the fifth counterfactual colleges cannot use any FAFSA information nor can they use the dummy for being below the poverty line. I assume that colleges can always use basic demographic characteristics such as age, gender, and race, as well as indicators of student quality like ACT score and high school GPA. $\widetilde{X}_i$ denotes the limited set of student covariates that the college can use in the counterfactual. That is, $\widetilde{X}_i$ contains the full set of student covariates X_i minus the restricted FAFSA information (which varies depending on the counterfactual). Colleges always make the best use of the information they have, so they can use available covariates to proxy for those that they cannot use.

It is possible that when FAFSA information is restricted colleges may turn to other student covariates that provide information about a student's income such as the student's neighborhood or home value. To get a sense for how well neighborhood and housing values might proxy for income, I use data from the 1970 census to regress family income on the median family income of a respondent's neighborhood and the respondent's home value. This regression yields an R-squared of 32 percent, and I consider this as a benchmark for thinking about how well colleges may be able to infer a student's income from readily available information about her. NPSAS does not contain detailed information on where students live, so I cannot include neighborhood income and housing values as student covariates. However, as Table 6 demonstrates, regressing parent income on the core student covariates I use in the paper gives an R-squared of 18.8 percent. To simulate colleges obtaining a bit more information about student income (i.e. by looking at neighborhood or house values), I assume in counterfactuals 1 and 4 that colleges can use a coarsened version of a student's income. In particular, I assume that, in lieu of using parent income directly, colleges are instead permitted to use a dummy for whether parent income is above or below the federal poverty line for a family of four. In other words, colleges cannot easily distinguish middle- and upper-income students, but they are able to identify low-income students. When I use the NPSAS data to regress parent income on this poverty dummy and the core student covariates, I get an R-squared of 33.4 percent, remarkably close to the R-squared from the census data. In light of this, I interpret counterfactuals 1 and 4 as representing a counterfactual where income information is restricted, but colleges are able to use outside proxies, such as neighborhood and house value, to infer a student's poverty status. In contrast, counterfactuals 2 and 5 represent counterfactuals where colleges are legally prohibited from directly using parent income as well as using outside proxies, although they are still permitted to use student characteristics like ACT score and race.

Colleges use all available student covariates to proxy for the restricted FAFSA information, to the extent possible. Table 6 reports some summary measures of how well those characteristics are able to proxy for the restricted information. In general, student covariates do help proxy for the restricted FAFSA information, but they are far from perfect. For example, regressing parent income on the core student covariates alone results in an R-squared of only 18.8 percent and a root mean squared error of more than \$50,000. I also estimate the distribution of the number of applications $\rho(n|\widetilde{X}_i)$. Table 7 reports goodness of fit measures for the estimates of $\rho(n|\widetilde{X}_i)$. The most informative piece of information about a student's applications is the number of colleges she lists on the FAFSA. Without this information, the pseudo-R^2 falls from 30 percent to less than ten percent and the root mean squared error rises by fourteen percent.

	Information available to colleges				Dependent variable		
Counterfactual	Core covariates	Parent income	Number of colleges listed	Whether completed FAFSA	Parent income	Number of colleges listed	Whether completed FAFSA
1	Yes	Poverty Dum	Yes	Yes	34.5% (46,094)	100.0% (0)	100.0% (0)
2	Yes	No	Yes	Yes	21.2% (50,554)	100.0% (0)	100.0% (0)
3	Yes	Yes	No	Yes	100.0% (0)	8.4% (1.64)	100.0% (0)
4	Yes	Poverty Dum	No	No	33.4% (46,239)	--- ---	7.6% (0.372)
5	Yes	No	No	No	18.8% (51,032)	--- ---	6.0% (0.375)

This table reports the R^2 from regressing parent income, the number of colleges listed on the FAFSA, and whether a student completed the FAFSA on the remaining student covariates. The root mean squared error is reported in parentheses. The table provides a sense for the degree to which other student covariates are able to serve as proxies for the FAFSA variables. In rows 1 and 4, I assume that colleges, in lieu of using actual income, are instead able to use whether income is above or below the federal poverty line for a family of four. In rows 2 and 5, I assume that colleges are unable to use any direct measures of income. In row 3, the regression for number of colleges listed is restricted to students who completed the FAFSA. In rows 4 and 5, I assume that colleges are unable to use any information from the FAFSA, and, since colleges must pool the FAFSA and non-FAFSA students together, in rows 4 and 5 I do not report an R^2 for the number of colleges listed. No sample weights were used.

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Postsecondary Student Aid Study, 2008.

Table 6: Student Covariates Are Imperfect Proxies For Missing Information

Given that colleges are intimately involved with administering federal aid, are the counterfactuals I simulate practical from a policy perspective? At one extreme, it is relatively straightforward to restrict information about the number of colleges listed on the FAFSA.[39] This information is not used in any federal financial aid calculation nor is it used in distributing aid to students. At the other extreme, it would be impossible for colleges to distribute federal aid without also

[39]Beginning with the 2016–2017 academic year, colleges are no longer allowed to see the number of colleges a student lists on her FAFSA.

	Information used in prediction				Goodness of fit	
Counterfactual	Core covariates	Parent income	Number of colleges listed	Whether completed FAFSA	Pseudo R^2	root-MSE
1	Yes	Poverty Dum	Yes	Yes	30.0%	2.147
2	Yes	No	Yes	Yes	29.9%	2.146
3	Yes	Yes	No	Yes	8.6%	2.440
4	Yes	Poverty Dum	No	No	7.1%	2.455
5	Yes	No	No	No	7.1%	2.455
Baseline	Yes	Yes	Yes	Yes	30.2%	2.143
	No	No	Yes	Yes	27.2%	2.184

The table reports results from ordered probit regressions of the number of college applications on student covariates using data from BPS:2003-2004. Rows 1-5 correspond to the five counterfactuals in the structural estimation. Row 6 represents the baseline prediction. Row 7 represents the prediction using only the number of colleges listed on the FAFSA and FAFSA completion. Comparing rows 6 and 7, we see that including other student covariates modestly improves the prediction of the number of applications. In rows 1-2 and 6-7, a separate ordered probit regression was run for each level of the number of college listed on the FAFSA as well as for students who did not complete the FAFSA. In row 3, separate regressions were run for students who completed the FAFSA and those that did not. In rows 4-5, all students were pooled in the same regression. The pseudo R^2 is computed by calculating each observation's predicted (expected) number of applications and taking the sum of squared differences between the observed and predicted number of applications. The same quantity is calculated for a "null" model in which no student covariates were used. Finally, the pseudo R^2 is computed as one minus the ratio of these two quantities. The root-MSE is computed by dividing the sum of squared differences by the residual degrees of freedom and taking the square root. Sample weights were not used.

SOURCE: U.S. Department of Education, National Center for Education Statistics, Beginning Postsecondary Students, 2004.

Table 7: Predicting the Number of Applications With an Ordered Probit Regression

knowing whether a given student applied for federal aid. Moreover, to the extent that colleges knew the types and amounts of federal aid that different students received, they would have a significant amount of information about family income. Certainly, firms in many industries are prohibited by regulation from basing prices on certain customer characteristics that the firms can observe (e.g. insurance, lending, etc.). Of course, the efficacy of such regulations is likely to vary based on the context. In the counterfactual simulations, I assume that the government is able to restrict colleges from directly using FAFSA information, either by withholding it, by directly regulating the use of the information, or both. But the government is unable to prevent colleges from using other student characteristics such as ACT score or race as proxies for the restricted information.

6.1 Modeling the Counterfactuals

In all five counterfactuals, I model the restriction of information as a pair of forecast errors in the college's beliefs about v_{ij} and w_{ij}. The first forecast error, e_i, represents each college's uncertainty about the value that student i places on the college. That is, $v_{ij} = \widetilde{v}_{ij} + e_i$, where v_{ij} is student i's true valuation for college j (which is known to the college in baseline), $\widetilde{v}_{ij}$ is college j's best forecast of v_{ij} given the limited information $\widetilde{X}_i$, and e_i represents a mean zero forecast error due to the restricted information.[40] e_i has distribution function F_e, with density f_e strictly positive over the support $[\underline{e}, \overline{e}]$. Importantly, e_i is not college-specific because the same information is restricted for all colleges. Similarly, the second shock, ξ_i, represents each college's

[40]Note that e_i is orthogonal to the observable student covariates $\widetilde{X}_i$.

uncertainty about its own valuation for student i. Thus, college j's true valuation can be written as $w_{ij} = z_j + \widetilde{\omega}(\widetilde{X}_i) + \xi_i = \widetilde{w}_{ij} + \xi_i$. Again, $\widetilde{w}_{ij}$ is college j's best forecast of w_{ij}, given the limited information $\widetilde{X}_i$, and ξ_i is the associated (mean zero) forecast error.[41]

Define $\widetilde{u}_{ij} \equiv \widetilde{v}_{ij} - p_{ij} = u_{ij} - e_i$ to be college j's forecast of the utility offer that it is making to student i. (In reality, j's utility offer is $u_{ij} = \widetilde{u}_{ij} + e_i$.) Similarly, $\widetilde{\pi}_{ij} \equiv \widetilde{w}_{ij} + p_{ij} = \pi_{ij} - \xi_i$ is college j's forecast of the payoff that it will get if it enrolls the student at price p_{ij}. Finally, the college's best forecast of the match surplus is $\widetilde{s}_{ij} = \widetilde{\pi}_{ij} + \widetilde{u}_{ij}$ with distribution $F_{\widetilde{S}}$. The college makes a tuition offer to maximize its expected payoff in the event it wins, which is equal to the probability it wins, $\mathbb{P}[u_{ij} \geq 0 \ \cap \ u_{ij} > u_{i\ell}, \ \ell \neq j]$, times the expected value of its payoff conditional on winning, $\mathbb{E}[\pi_{ij} | u_{ij} \geq 0 \ \cap \ u_{ij} > u_{i\ell}, \ \ell \neq j]$.[42] Substituting for u_{ij} and π_{ij}, college j's objective function can be written as $\mathbb{E}[\widetilde{\pi}_{ij} + \xi_i | \widetilde{u}_{ij} + e_i \geq 0] \, \mathbb{P}[\widetilde{u}_{ij} + e_i \geq 0 \ \cap \ \widetilde{u}_{ij} > \widetilde{u}_{i\ell}, \ \ell \neq j]$ so that college j's problem is

$$
\begin{aligned}
& \max_{\widetilde{u}_{ij}} \quad \left(\widetilde{s}_{ij} - \widetilde{u}_{ij} + \mathbb{E}[\xi_i | e_i \geq -\widetilde{u}_{ij}]\right) \mathbb{P}[e_i \geq -\widetilde{u}_{ij} \ \cap \ \widetilde{u}_{ij} > \widetilde{u}_{i\ell}, \ \ell \neq j] \\
\to \quad & \max_{\widetilde{u}_{ij}} \quad \left(\widetilde{s}_{ij} - \widetilde{u}_{ij} + \mathbb{E}[\xi_i | e_i \geq -\widetilde{u}_{ij}]\right) \left(1 - F_e(-\widetilde{u}_{ij})\right) \sum_{n=1}^{\overline{n}} F_{\widetilde{S}}^{n-1}(\widetilde{\beta}^{-1}(\widetilde{u}_{ij}))\widetilde{\rho}(n) \qquad (11)
\end{aligned}
$$

where $\widetilde{\beta}(\cdot)$ is the equilibrium bid function in the counterfactual and $\widetilde{\rho}(\cdot) = \rho(n|\widetilde{X})$ is the distribution of the number of competitors conditional on the information available in the counterfactual. For ease of notation I have suppressed the conditioning on $\widetilde{X}_i$. Taking the first order condition for equation (11), we can derive an ODE similar to (3)

$$
\begin{aligned}
\widetilde{\beta}'(\widetilde{s}) & \left[1 - h_e(-\widetilde{\beta}(\widetilde{s}))(\widetilde{s} - \widetilde{\beta}(\widetilde{s}) + \mathbb{E}[\xi | e = -\widetilde{\beta}(\widetilde{s})])\right] = \\
& \left(\widetilde{s} - \widetilde{\beta}(\widetilde{s}) + \mathbb{E}[\xi | e \geq -\widetilde{\beta}(\widetilde{s})]\right) \frac{\sum_{n=1}^{\overline{n}} (n-1) F_{\widetilde{S}}^{n-1}(\widetilde{s}) f_{\widetilde{S}}(\widetilde{s}) \widetilde{\rho}(n)}{\sum_{n=1}^{\overline{n}} F_{\widetilde{S}}^{n}(\widetilde{s}) \widetilde{\rho}(n)} \qquad (12)
\end{aligned}
$$

where $h_e(\cdot)$ is the hazard function of the forecast error e and $\mathbb{E}[\xi | e = -b] \equiv 0$ for $b \geq -\underline{e}$. The initial condition is $\widetilde{\beta}(-\overline{e} - \mathbb{E}[\xi | e = \overline{e}]) = -\overline{e}$, which says that the lowest-surplus bidder is the college with a forecasted match surplus—conditional on winning—so low that, even if it offered all of the forecasted surplus to the student, the most favorable draw of e possible would only just make the student indifferent between the college and the competitive fringe. Notice that the right hand side of Equation (12) is analogous to Equation (3), with the addition of the term $\mathbb{E}[\xi | e \geq -\widetilde{\beta}(\widetilde{s})]$ which reflects the fact that learning that it has beat the nonelite sector may provide some information about ξ_i. The expression in brackets on the left hand side of Equation (12) contains additional terms reflecting the uncertainty about whether the college will beat out the nonelite sector. However, once $\widetilde{\beta} \geq -\underline{e}$, these additional terms disappear because the college

[41] I estimate these forecast errors by projecting the estimated $\widehat{v}_{ij}$ on X_i and on $\widetilde{X}_i$ and taking the difference between these projections. See section E in the online appendix for details.

[42] Unlike in baseline, college j now faces the additional uncertainty about whether its offer will exceed the student's outside option.

is now making an offer large enough that it no longer needs to worry about losing to the nonelite sector and therefore learning that it has beaten the nonelite sector is uninformative about its valuation for the student.[43]

Counterfactual bids (and prices) will differ from those in baseline for five reasons: 1) $F_{\widetilde{S}|\widetilde{X}_i}$ will differ, perhaps substantially, from $F_{S|X_i}$; 2) the distribution $\rho(n|\widetilde{X}_i)$ will also differ from $\rho(n|X_i)$, especially when colleges cannot use the number of colleges listed on the FAFSA; 3) colleges now face uncertainty over their position relative to the nonelite sector; 4) learning that it has beat out the nonelite sector ($e_i \geq -\widetilde{u}_{ij}$) is informative to the college about the forecast error in its own valuation ξ_i; and 5) each college's forecasted match surplus, $\widetilde{s}_{ij}$, will differ from the true (baseline) match surplus, s_{ij}, by the realization of the forecast error $e_i + \xi_i$. For example, when e_i is higher, colleges will incorrectly believe that the student has a low willingness to pay and will lower their price offers accordingly. Since e_i and ξ_i do not vary by college, the relative rankings of colleges are unaffected. However, the relative ranking of the nonelite sector *could* be affected. On occasion, the winning elite college will not beat the nonelite sector because e_i turned out to be unexpectedly low (negative). That is, the student appeared to have a high willingness to pay when she actually did not, so the colleges made higher price offers and inefficiently lost the student to a nonelite college. Because the elite colleges face more uncertainty, they are unable to tailor prices as precisely and inefficient matches arise.

6.2 Counterfactual Estimates

In order to simulate the counterfactuals, I estimate the colleges' two forecast errors, e_i and ξ_i, for v_{ij} and w_{ij}, respectively. We can write student i's willingness to pay as $v_{ij} = X_i\widehat{\gamma} + \widehat{a}_{ij}$ where $\widehat{\gamma}$ are OLS coefficients and $\widehat{a}_{ij}$ is the residual. Both $\widehat{\gamma}$ and $\widehat{a}_{ij}$ are known to the college, since the college knows v_{ij}, and can be estimated by the econometrician. After restricting the set of student covariates to $\widetilde{X}_i$, college j's best forecast of v_{ij} is now $\widetilde{v}_{ij} = \mathbb{E}[v_{ij}|\widetilde{X}_i, \widehat{a}_{ij}] = \widetilde{X}_i\widetilde{\gamma} + \widehat{a}_{ij}$. Notice that, although $\widetilde{\gamma}$ will generally differ from $\widehat{\gamma}$, $\widehat{a}_{ij}$ is unaffected by restricting the contents of X_i because, by construction, $\widehat{a}_{ij}$ is orthogonal to X_i. Then I can use the fact that $e_i = v_{ij} - \widetilde{v}_{ij}$ to estimate e_i, and an analogous procedure can be used to estimate ξ_i. The results in Table 8 indicate that most of the college's forecast error in match surplus is coming from its uncertainty about student preferences, e_i, rather than from uncertainty about its own valuation for the student, ξ_i. I then estimate the marginal distributions of the forecast errors F_e and F_ξ, the function $\mathbb{E}[\xi|e]$, and the distribution of the college's forecast of match surplus $F_{\widetilde{S}|\widetilde{X}_i}$. Finally, after plugging these into equation (12), I simulate counterfactual bids.[44]

6.2.1 Prices and Misallocation

Table 9 compares the baseline structural estimates with the five counterfactuals. Restricting FAFSA information affects the average transaction price through two channels. First, due to

[43] $h_e(b) = 0$ and $\mathbb{E}[\xi|e \geq -b] = 0$ for $b \geq -\underline{e}$.

[44] See section E in the online appendix for more details.

Information available to colleges				Standard deviation of forecast error	
Core covariates	Parent income	Number of colleges listed	Whether completed FAFSA	Error for forecast of student valuation (e)	Error for forecast of college valuation (ξ)
Yes	Poverty Dum	Yes	Yes	2.46	0.38
Yes	No	Yes	Yes	2.65	0.42
Yes	Yes	No	Yes	1.55	0.00
Yes	Poverty Dum	No	No	2.94	0.38
Yes	No	No	No	3.33	0.42

The table reports the standard deviation of the forecast errors in each of the five counterfactuals. Note that I assume that neither the number of colleges listed on the FAFSA nor whether the student completed the FAFSA directly affect the college's valuation for the student. Thus, for ξ the standard deviation in counterfactual 3 is zero and rows 4 and 5 are identical to rows 1 and 2. No sample weights were used.

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Postsecondary Student Aid Study, 2008, and Beginning Postsecondary Students, 2004.

Table 8: Information Loss in the Counterfactuals

the uncertainty introduced by restricting colleges' information, some students, up to 13 percent, are priced out of the elite college market and end up inefficiently attending nonelite colleges. Second, among those students who remain at elite colleges, price falls by an average of \$684 in counterfactual 1, \$748 in counterfactual 2, \$540 in counterfactual 3, \$860 in counterfactual 4, and \$986 in counterfactual 5. In other words, restricting FAFSA information lowers transaction prices for students who remain at elite colleges by 4.1–7.5 percent, depending on the counterfactual.[45]

6.2.2 Welfare

As was just discussed, colleges inefficiently price many students out of the elite market when they cannot use the FAFSA. College payoffs fall and colleges are worse off under each counterfactual than in baseline. But making a welfare calculation for the students requires us to take a stand on how they value the way in which colleges spend their tuition revenues. At least two possibilities suggest themselves: first, colleges themselves consume their marginal dollar of surplus (the standard assumption); second, colleges spend their marginal dollar of surplus on something that students greatly value, perhaps by supplying a valuable public good.[46]

The first case is more standard and is the one assumed in Table 9. If colleges consume their marginal dollar of surplus, then a reduction in transaction prices would simply transfer surplus from colleges to students. The estimates indicate that average student surplus would rise by \$622 in counterfactual 1, \$673 in counterfactual 2, \$481 in counterfactual 3, \$755 in counterfactual 4, and \$827 in counterfactual 5. However, total surplus per student would also fall due to the misallocation of students to the nonelite sector: between \$78 and \$234 depending on the counterfactual. Although as many as 13 percent of students would be misallocated, total surplus

[45]Because students who are priced out of the elite market tend to be paying lower prices in baseline, average transaction prices do not fall by as much.

[46]In section F.7, I explore an extension of the model that explicitly incorporates college spending.

	Baseline	Counterfactual 1	2	3	4	5
Panel A. Levels						
Consumer (student) surplus per student	**$5,023**	**$5,645**	**$5,697**	**$5,504**	**$5,778**	**$5,851**
	($4168, $5033)	($4730, $5646)	($4758, $5691)	($4549, $5452)	($4849, $5745)	($4938, $5832)
Total surplus per student	**$15,417**	**$15,339**	**$15,322**	**$15,317**	**$15,279**	**$15,183**
	($13770, $15553)	($13688, $15449)	($13664, $15428)	($13676, $15457)	($13632, $15403)	($13532, $15285)
Of those who remain at elite colleges:						
Mean student share of surplus	**30.1%**	**37.5%**	**37.7%**	**36.7%**	**38.3%**	**38.4%**
	(27.7%, 30.4%)	(35.3%, 38.1%)	(35.5%, 38.3%)	(33.5%, 36.9%)	(35.9%, 38.4%)	(36.3%, 38.6%)
Mean transaction price	**$13,158**	**$13,066**	**$13,024**	**$13,305**	**$13,040**	**$13,088**
	($12818, $13530)	($12757, $13541)	($12738, $13544)	($12980, $13790)	($12733, $13530)	($12694, $13513)
Panel B. Changes Relative to Baseline						
Consumer (student) surplus per student	**$0**	**$622**	**$673**	**$481**	**$755**	**$827**
	(---)	($443, $743)	($485, $802)	($261, $538)	($566, $840)	($664, $941)
Total surplus per student	**$0**	**-$78**	**-$95**	**-$100**	**-$138**	**-$234**
	(---)	($-161, $-45)	($-215, $-57)	($-174, $-30)	($-247, $-86)	($-371, $-163)
Percent of students who inefficiently choose a non-elite college	**0.0%**	**8.0%**	**8.5%**	**9.0%**	**10.4%**	**12.7%**
	(---)	(7.3%, 13%)	(7.8%, 14.3%)	(6.6%, 13.5%)	(9.6%, 15.9%)	(12%, 18.5%)
Of those who remain at elite colleges:						
Mean change in student share of surplus	**0.0%**	**5.8%**	**5.8%**	**4.7%**	**6.1%**	**5.9%**
	(---)	(5.1%, 6.8%)	(5.1%, 6.9%)	(3.3%, 5.4%)	(5.1%, 6.9%)	(5.1%, 7%)
Mean change in transaction price	**$0**	**-$684**	**-$748**	**-$540**	**-$860**	**-$986**
	(---)	($-869, $-491)	($-962, $-542)	($-631, $-291)	($-1027, $-648)	($-1208, $-799)
Percent of students with price drop	**0.0%**	**71.6%**	**71.3%**	**80.1%**	**75.6%**	**68.2%**
	(---)	(67.5%, 74.9%)	(67.1%, 74.1%)	(69.8%, 85.6%)	(69.1%, 78.9%)	(63.6%, 72.5%)
Within-college variance in price (in millions)	**38.81**	**-15.0%**	**-16.2%**	**-9.6%**	**-16.6%**	**-22.2%**
		(-19.5%, -5.4%)	(-20%, -5.6%)	(-12.9%, 0.9%)	(-20.4%, -5.4%)	(-27.4%, -11%)
FAFSA Information Available						
Parent income	Yes	Poverty Dummy	No	Yes	Poverty Dummy	No
Number of schools listed on FAFSA	Yes	Yes	Yes	No	No	No
Whether completed FAFSA	Yes	Yes	Yes	Yes	No	No

Column 1 contains baseline estimates while columns 2-6 contain estimates for the five counterfactuals. Point estimates are in bold. 95% percent confidence intervals in parentheses were calculated using 1,000 bootstrap replications. Dollar amounts are expressed in terms of dollars per year. The percentages in rows 7 and 9 are percentage points. The percentages in row 11 are percent changes in within-college price variance relative to a base of 38,811,673. No sample weights were used.

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Postsecondary Student Aid Study, 2008, and Beginning Postsecondary Students, 2004.

Table 9: Counterfactual Estimates

per student would fall by less than 2 percent because the students who would be misallocated tend to have low match surpluses to begin with. Thus, if colleges consume their marginal dollar of surplus, restricting FAFSA information primarily transfers surplus from colleges to students with a small amount of that surplus being lost in student misallocation.

Now consider the case where colleges' marginal dollar of surplus benefits society as a whole, including students. In this case, the welfare implications become ambiguous and nearly impossible to quantify. Colleges may spend their marginal dollar of surplus on an extremely valuable public good so that reducing college surplus will dramatically reduce social welfare and perhaps even make the students themselves worse off in the process, despite the lower prices they are paying. But in order for students to be harmed by the reduction in college surplus, they would need to value the marginal dollar of college surplus at more than $0.89 for counterfactual 1, $0.88 for counterfactual 2, $0.83 for counterfactual 3, $0.85 for counterfactual 4, and $0.78 for counterfactual 5.

6.2.3 Reduced Form Pricing Patterns

When colleges are no longer permitted to use FAFSA information in their pricing, they may attempt to use other student characteristics to proxy for the lost FAFSA information. Table 10 reports the results of a reduced-form regression of transaction price on student covariates and college fixed effects. This specification is reminiscent of the reduced-form specification from section 4.1.[47] Then I run the same regression for each counterfactual, substituting in an estimate of each student's counterfactual transaction price. Table 10 provides a glimpse into how observed pricing patterns would change under the three counterfactuals. As would be expected, prices become less correlated with income when income information is restricted. However, the coefficient does not drop all the way to zero, because other student covariates, such as test scores and race, behave as proxies for income. This helps to explain why the coefficient on ACT score becomes smaller in magnitude when income information is restricted, while the coefficients for racial minorities (black and Hispanic) become larger.[48] Figure 5 plots the fitted values of transaction price against parent adjusted gross income. For each of the three counterfactuals, it plots the estimated income gradient from the baseline regression in Table 10 and the estimated income gradient in the counterfactual, holding other covariates fixed. Just as in Table 10, the income gradient flattens when income information is restricted, but not when the number of colleges listed on the FAFSA is restricted.

Dependent Variable: Transaction Price	**Baseline**	**Counterfactual** 1	2	3	4	5
Parent AGI (in $10,000's)	220.1 (27.86) ***	63.8 (25.19) *	44.1 (24.98)	212.0 (26.99) ***	60.9 (24.85) *	49.4 (24.69) *
ACT score	-196.6 (47.54) ***	-151.9 (48.44) **	-142.6 (48.10) **	-214.2 (49.49) ***	-146.5 (49.80) **	-147.8 (48.99) **
High school GPA	-1087.1 (365.7) **	-1316.7 (367.5) ***	-1270.5 (366.8) ***	-1160.5 (371.6) **	-1249.1 (373.3) ***	-1425.4 (370.7) ***
Earned AP credit in high school	-193.8 (404.8)	-19.3 (378.3)	-42.7 (375.8)	-192.3 (376.6)	-93.7 (379.9)	-261.0 (376.0)
Completed FAFSA	-1110.1 (491.1) *	-1785.7 (455.3) ***	-1908.2 (454.0) ***	-778.9 (458.7)	-939.6 (462.6) *	1110.6 (455.0) *
Num of add'l colleges listed on FAFSA	-527.9 (112.6) ***	-395.6 (114.3) ***	-396.1 (113.7) ***	-548.8 (112.4) ***	-555.9 (114.2) ***	-609.7 (115.3) ***
Age as of 12/31/07	77.8 (283.8)	112.9 (251.3)	112.4 (251.5)	219.1 (256.8)	126.0 (255.9)	325.4 (256.4)
Female	-553.4 (314.5)	-552.7 (306.8)	-576.3 (307.0)	-629.9 (313.5) *	-578.6 (311.5)	-559.3 (310.9)
Black	-2168.8 (636.6) ***	-3048.8 (643.6) ***	-3240.4 (639.0) ***	-2710.1 (652.8) ***	-3198.8 (654.1) ***	-3425.0 (652.7) ***
Hispanic	-1826.3 (638.8) **	-2351.3 (623.2) ***	-2477.9 (626.7) ***	-1799.0 (618.2) **	-2084.4 (622.0) ***	-2287.3 (626.1) ***
Asian	-110.6 (822.9)	34.2 (717.0)	3.3 (720.1)	415.9 (708.5)	96.4 (714.9)	-5.8 (711.0)
Other / Multiple	-865.1 (810.3)	-1503.2 (844.4)	-1719.1 (828.3) *	-1358.2 (841.7)	-1298.6 (820.1)	-1669.7 (822.7) *
Out-of-State, Public	8722.5 (837.8) ***	8804.1 (683.1) ***	8715.0 (676.7) ***	8939.6 (698.0) ***	8576.4 (678.8) ***	7987.8 (628.7) ***
College fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2210	2030	2020	2010	1980	1930
R-squared	0.744	0.772	0.772	0.778	0.772	0.769
FAFSA Information Available						
Parent income	Yes	Poverty Dummy	No	Yes	Poverty Dummy	No
Number of schools listed on FAFSA	Yes	Yes	Yes	No	No	No
Whether completed FAFSA	Yes	Yes	Yes	Yes	No	No

All six regressions include students from the elite sample. The dependent variable in column 1 is transaction price (sticker price minus discount). In columns 2-6, the dependent variable is the counterfactual price the student would have been charged based on the structural estimates. In each counterfactual, students who would be priced out of the elite sector are omitted from the regression. Robust standard errors are in parentheses. Sampling weights were used in all six regressions (NPSAS variable WTA000).

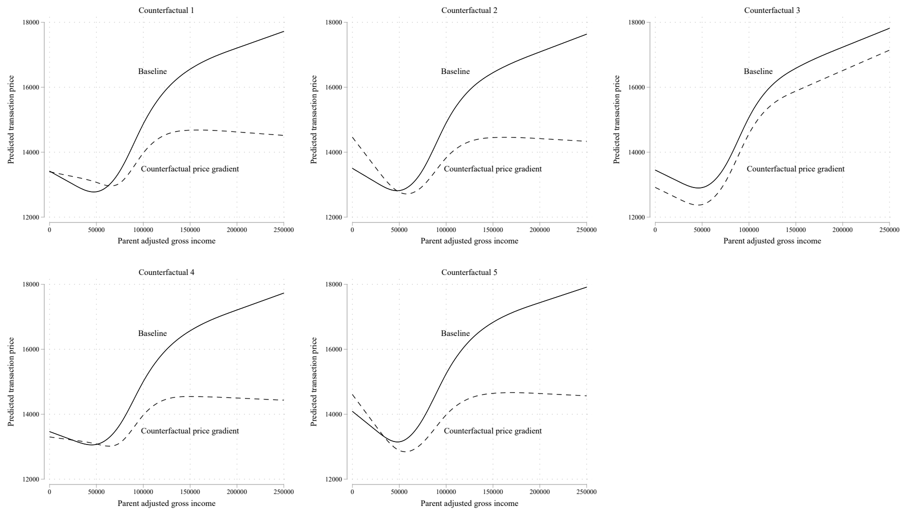
* p<0.05, ** p<.01, *** p<.001

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Postsecondary Student Aid Study, 2008, and Beginning Postsecondary Students, 2004.

Table 10: Comparing Baseline and Counterfactual Pricing Patterns at Elite Colleges

[47]In contrast with the reduced-form specification from section 4.1, the dependent variable in Table 10 is the transaction price rather than the tuition discount.

[48]The coefficient on high school GPA becomes more negative because it turns that, although high school GPA is positively correlated with income, it is negatively correlated with income *holding constant ACT score.*



The fitted values plotted here come from five pairs of regressions. In each pair, the fitted values represent a white, female student with all other covariates set to their sample means. Each pair of regressions only include students who were not priced out of the elite college market in the corresponding counterfactual. Parent adjusted gross income was included as a cubic spline.

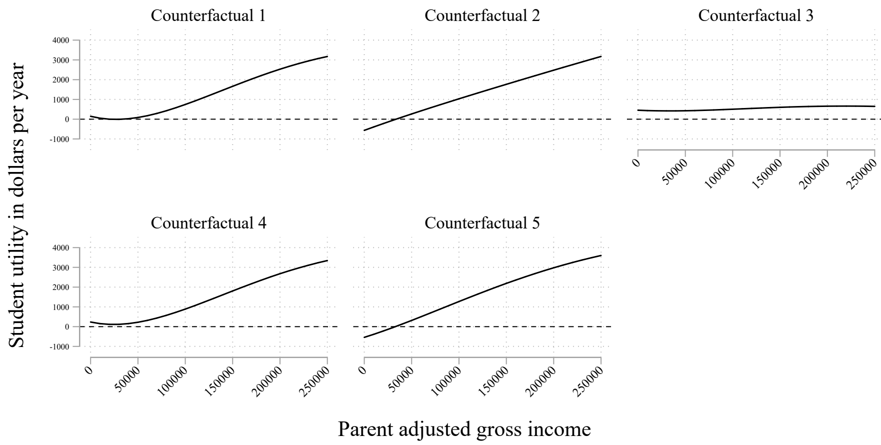
SOURCE: U.S. Department of Education, National Center for Education Statistics, National Postsecondary Student Aid Study, 2008, and Beginning Postsecondary Students, 20

Figure 5: Change in Observed, Within-College Income Gradient

6.2.4 Heterogeneity and Distributional Consequences

We would expect restricting FAFSA information to affect students differently depending on their income. In Tables 11 and 12, I look at how the five counterfactuals differentially affect low-, middle-, and high-income students. Figure 6 plots the fitted values from regressing each student's change in utility, relative to baseline, on a polynomial in parent income. Surprisingly, restricting income information alone (counterfactuals 1 and 2) benefits most students with little, if any, harm to low-income students, on average. This comes from the fact that in these counterfactuals colleges are still able to distinguish students above and below the poverty line. However, the lowest income students are harmed when income information is restricted entirely (counterfactuals 2 and 5). Still, the income gradient doesn't turn negative until well below $50,000 in parent income, indicating that many middle income students also benefit from restricting income information. Strikingly, restricting the number of colleges listed on the FAFSA (counterfactual 3) helps students in a way that is more income neutral.

To what degree does price discrimination redistribute resources from higher income students to lower income students? Table 11 demonstrates how colleges use the FAFSA to extract surplus from some students while redistributing some of it to other students. When elite colleges are permitted to use the FAFSA, they raise prices on some students. If a student does not respond by leaving the elite sector, then her college has successfully extracted some of her surplus. Her college may also use the FAFSA to lower prices for other students, thereby transferring some of her extracted surplus to her fellow students. Colleges use the FAFSA to transfer utility in two ways: by charging less to current students and by attracting new students with lower prices. For example, students in the bottom third of the income distribution receive, on average, $200



The fitted values plotted here come from a separate regression for each counterfactual. In all five regressions, the dependent variable is the estimated change in utility relative to baseline and the independent variable is parent adjusted gross income (included as a fourth-order polynomial). No other covariates are included.

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Postsecondary Student Aid Study, 2008, and Beginning Postsecondary Students, 2004.

Figure 6: Change In Student Utility When FAFSA Info Restricted

more per year in utility than they would if FAFSA information were completely restricted. This occurs because 16.8 percent of them would be priced out of the elite market if the FAFSA were restricted and because those who would remain would pay, on average, \$164 more per year. In contrast, those in the middle third receive \$571 less utility and those in the upper third receive \$2,109 less utility per year than they otherwise would if the FAFSA were completely restricted. But colleges do not transfer all of the extracted surplus to other students. In the final row of Table 11, I calculate the ratio between the total surplus that colleges transfer to students and the total surplus they extract by virtue of the FAFSA. If colleges were price discriminating in a purely redistributionist way, this ratio would be 100 percent; all of the extracted surplus would be transferred to other students. Putting it differently, if we think of price discrimination as a tax, how much of the "tax revenue" is being transferred to other students and how much is accruing to the colleges? The last line of Table 11 shows that, depending on the counterfactual, 12.2–22.2 percent of the surplus colleges extract from students is being transferred to other students. For instance, when comparing our baseline results to counterfactual 5, colleges transfer 22.2 percent of the extracted surplus to other students and keep the remaining 77.8 percent.

We have just seen that, depending on the counterfactual, colleges transfer between one-eighth and one-fifth of the surplus they extract from students to other students. But to what extent is that redistribution moving surplus from high-income students to low-income students? As illustrated in Table 12, although lower-income students may be worse off *on average* when FAFSA information is restricted, many are not. Across all five counterfactuals, one-third or more of low-income students are better off than in baseline. Moreover, while 59.3 percent and 86.3 percent of middle- and high-income students benefit from restricting all FAFSA information, this also

		Counterfactual				
		1	2	3	4	5
Bottom third of parent income	**Utility change**	-$13	-$217	$425	$115	-$200
	Price change (if stay at elite college)	$4	$221	-$482	-$156	$164
	Percent leaving elite sector	8.8%	10.7%	9.5%	12.2%	16.8%
Middle third of parent income	**Utility change**	$262	$480	$461	$397	$571
	Price change (if stay at elite college)	-$296	-$540	-$529	-$474	-$698
	Percent leaving elite sector	8.7%	9.4%	10.6%	12.5%	14.1%
Top third of parent income	**Utility change**	$1,615	$1,755	$556	$1,752	$2,109
	Price change (if stay at elite college)	-$1,732	-$1,860	-$607	-$1,882	-$2,280
	Percent leaving elite sector	6.4%	5.4%	6.9%	6.5%	7.1%
Percent of "extracted utility" transferred to other students		18.5%	19.4%	12.2%	14.2%	22.2%
FAFSA Information Available						
Parent income		Poverty Dum	No	Yes	Poverty Dum	No
Number of schools listed on FAFSA		Yes	Yes	No	No	No
Whether completed FAFSA		Yes	Yes	Yes	No	No

When colleges can no longer use the FAFSA to price discriminate, some students see their prices fall, relative to baseline, some see their prices rise, and the rest are priced out of the elite market. The three super rows in the table were calculated separately for each of the three income terciles. The first subrowwith each super row reports the change in mean utility. The second subrow reports the mean change in transaction price for those students who are not priced out of the market. The third subrow reports the percent of students who are priced out of the elite market. The table also reports the fraction of utility that colleges are able to extract via price discrimination that is transfered to other students. This measures the degree to which colleges use FAFSA information to price discriminate in a way that redistributes from some students to others versus simply increasing college payoffs. No sample weights were used.

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Postsecondary Student Aid Study, 2008, and Beginning Postsecondary Students, 2004.

Table 11: Effect of Restricting FAFSA Info on Student Welfare and Price by Income Group

	Counterfactual				
	1	2	3	4	5
Bottom third of parent income	45.1%	33.8%	65.9%	48.5%	33.0%
Middle third of parent income	61.9%	70.6%	72.3%	65.8%	59.3%
Top third of parent income	90.8%	91.3%	80.3%	88.7%	86.3%
FAFSA Information Available					
Parent income	Poverty Dum	No	Yes	Poverty Dum	No
Number of schools listed on FAFSA	Yes	Yes	No	No	No
Whether completed FAFSA	Yes	Yes	Yes	No	No

When colleges can no longer use the FAFSA to price discriminate, some students benefit from lower prices while others are hurt by either higher prices or being priced out of the elite market. Each cell reports the percent of students who benefit, relative to baseline, in each counterfactual in the corresponding tercile of the distribution of parent adjusted gross income. No sample weights were used.

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Postsecondary Student Aid Study, 2008, and Beginning Postsecondary Students, 2004.

Table 12: Percent of Students Who Benefit From Restricting FAFSA Info by Income Group

implies that 40.7 percent and 13.7 percent of middle- and high-income students actually benefit from allowing colleges to use the FAFSA. Table 12 demonstrates that, although colleges use the FAFSA to price discriminate in a way that redistributes from higher- to lower-income students on average, they do so with less precision than one might hope for. Importantly, this imprecision comes, not from a lack of information about student income, but from the incentives and competitive forces that colleges face.

7 Conclusion

Pricing in the higher education market is complex, and colleges commonly use tuition discounts to price discriminate. Economic theory tells us that a seller must have some information about a buyer's willingness to pay in order to price discriminate. Colleges are fortunate because they have access to the FAFSA, which provides detailed and reliable information about their students' finances. In order to quantify the importance of the FAFSA information in enabling price discrimination, I build and estimate a structural model of college pricing. I recast the pricing problem as a first-price auction in utility bids and show that the model is identified from data on student-level transaction prices.

Consistent with the predictions of the model, I find that more desirable students (proxied for by ACT score and high school GPA) do pay less than their peers, while those with a higher willingness to pay (proxied for by income) pay more. Those who list more colleges on the FAFSA also pay less. These patterns emerge primarily among private and very selective public colleges. I label these "elite" colleges and focus on them when estimating the structural model.

I estimate the model using student-level data on transaction prices and student characteristics. I simulate five counterfactuals wherein colleges are restricted from using some or all of the FAFSA information. I find that restricting FAFSA information affects the average transaction price through two channels. First, due to the uncertainty introduced by restricting colleges' information, some students, up to 13 percent, are priced out of the elite college market and end up inefficiently attending nonelite colleges. Second, among those students who remain at elite colleges, price falls by between \$540 and \$986 on average, thereby raising student surplus by up to \$827 per student. These estimates highlight an important policy tradeoff—restricting colleges' use of FAFSA information raises student welfare but also leads to a misallocation of students. Depending on the counterfactual, these inefficient matches lower total surplus by between 0.5–1.5 percent. I also find evidence that colleges would use other student characteristics to proxy, albeit imperfectly, for the lost FAFSA information.

These average effects mask the fact that the effect on students differs by income. On average, low-income students lose \$200 per year when all FAFSA information is restricted, while high-income students gain \$2,109. Although colleges use the FAFSA to transfer surplus from high-income to low-income students, this redistribution is both incomplete and imprecise. Of the student surplus they extract by virtue of the FAFSA information, colleges only transfer 22.2

percent to other students. Moreover, one-third of low-income students are actually harmed while one in seven high-income students benefit from allowing colleges to use FAFSA information in their pricing.

Acknowledgments. I would like to thank Derek Neal, Brent Hickman, Steve Levitt, and Chad Syverson for their helpful input throughout the project. I am also grateful for comments from Kris Hult, Trevor Gallen, Xan Vangsathorn, Jonathan Hall, Devin Pope, Ali Hortaçsu, Stephen Raudenbush, Lars Lefgren, Brennan Platt, George-Levi Gayle, Stephen Ryan, Barton Hamilton, Bruce Peterson, and workshop participants at the University of Chicago, Brigham Young University, University of Wisconsin, Bates White Economic Consulting, W.E. Upjohn Institute for Employment Research, University of Iowa, Tepper School of Business, Washington University in St. Louis, and the Stanford Institute for Theoretical Economics 2017 summer workshop. Kevin Brown and NORC at the University of Chicago provided indispensable help in obtaining the restricted-use data. Brad Hershbein and the W.E. Upjohn Institute were also extremely helpful in procuring data, and Benjamin Jones provided valuable editing assistance.

Funding. I gratefully acknowledge financial support from the University of Chicago Predoctoral Training Grant in Education, funded by the Institute of Education Sciences (grant number R305 B090025), the W.E. Upjohn Institute Postdoctoral Fellowship, and a grant from the Weidenbaum Center on the Economy, Government, and Public Policy.

References

Athey, Susan and Haile, Philip A. Identification of Standard Auction Models. *Econometrica*, 70 (6):2107–2140, 2002.

Avery, Christopher and Hoxby, Caroline Minter. Do and should financial aid packages affect students' college choices? *College choices: The economics of where to go, when to go, and how to pay for it*, pages 239–302, 2004.

Baum, Sandy; Elliott, Diane Cardenas, and Ma, Jennifer. Trends in Student Aid. Technical report, The College Board, 2014.

Bergemann, Dirk; Brooks, Benjamin, and Morris, Stephen. The Limits of Price Discrimination. *Cowles Foundation For Research In Economics*, pages 1–48, 2013.

Bettinger, Eric P.; Terry Long, Bridget; Oreopoulos, Philip, and Sanbonmatsu, Lisa. The Role of Application Assistance and Information in College Decisions: Results from the H&R Block FAFSA Experiment. *The Quarterly Journal of Economics*, pages 1205–1242, 2012.

Cunningham, Alisa F. Changes in patterns of prices and financial aid. *U.S. Department of Education Statistics*, 2005.

Dynarski, S. The behavioral and distributional implications of aid for college. *American Economic Review*, 92(2):279–285, 2002.

Dynarski, Susan M. and Scott-Clayton, Judith E. The Cost of Complexity in Federal Student Aid: Lessons from Optimal Tax Theory and Behavioral Economics. *National Tax Journal*, 59(2): 319–356, 2006.

Epple, Dennis; Romano, Richard, and Sieg, Holger. Admission, Tuition, and Financial Aid Policies in the Market for Higher Education. *Econometrica*, 74(4):885–928, 2006.

Fillmore, Ian. A Semi-Parametric Estimator of Conditional Distributions. 2017.

Fu, Chao. Equilibrium Tuition, Applications, Admissions, and Enrollment in the College Market. *Journal of Political Economy*, 122(2):225–281, 2014.

Grant, Tim. Families maximize their eligibility for financial aid . *Pittsburgh Post-Gazette*, pages 1–4, Oct 2006.

Guerre, Emmanuel; Perrigne, Isabelle, and Vuong, Quang. Optimal Nonparametric Estimation of First-Price Auctions. *Econometrica*, 68(3):525–574, 2000.

Hubbard, Timothy P.; Li, Tong, and Paarsch, Harry J. Semiparametric estimation in models of first-price, sealed-bid auctions with affiliation. *Journal of Econometrics*, 168(1):4–16, 2012.

Krishna, Vijay. *Auction Theory*. Academic Press/Elsevier, 2 edition, 2010. ISBN 9780123745071.

Lawson, R and Zerkle, Ann. Price discrimination in college tuition: an empirical case study. *Journal of Economics and Finance Education*, 5:1–7, 2006.

Li, Tong; Perrigne, Isabelle, and Vuong, Quang. Structural Estimation of the Affiliated Private Value Auction Model. *The RAND Journal of Economics*, 33(2):171–193, 2002.

Long, Bridget Terry. How have college decisions changed over time? An application of the conditional logistic choice model. *Journal of Econometrics*, 121(1-2):271–296, 2004.

Peña, Pablo A. Pricing in the Not-for-Profit Sector: Can Wealth Growth at American Colleges Explain Chronic Tuition Increases? *Journal of Human Capital*, 4(3):242–273, 2010.

Ramsay, J O. Estimating Smooth Monotone Functions. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 60(2):365–375, 1998.

Roberts, James W. Unobserved heterogeneity and reserve prices in auctions. *RAND Journal of Economics*, 44(4):712–732, 2013.

Rothschild, Michael and White, Lawrence J. The Analytics of the Pricing of Higher Education and Other Services in Which the Customers Are Inputs. *Journal of Political Economy*, 103(3): 573–586, 1995.

Tiffany, Frederick G and Ankrom, Jeff A. The competitive use of price discrimination by colleges. *Eastern Economic Journal*, pages 99–110, 1998.

van der Klaauw, Wilbert. Estimating the Effect of Financial Aid Offers on College Enrollment: A Regression-Discontinuity Approach. *International Economic Review*, 43(4):1249–1287, 2002.

Varian, H R. Price discrimination. *Handbook of industrial organization*, 1:597–654, 1989.

Weston, Liz. Four college financial aid maneuvers that can backfire, 2014.

Williams, Richard. Generalized ordered logit/partial proportional odds models for ordinal dependent variables. *The Stata Journal*, 6(1):58–82, 2006.