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

Ian Fillmore*†

Abstract

In the United States, the federal government grants colleges access to a student’s Free Application for Federal Student Aid, or FAFSA, which enables them to engage in substantial price discrimination. I build and estimate a model of competitive college pricing and simulate counterfactuals wherein colleges are restricted from using some or all of the FAFSA information. I find that although allowing colleges to use FAFSA information does increase efficiency somewhat and lower prices for some, usually low-income, students, its main effect is to boost tuition revenue, primarily at the expense of middle- and high-income students.

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

---

*Washington University in St. Louis, St. Louis, MO, U.S.A., ianfillmore@wustl.edu
†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, 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, and Washington University in St. Louis. 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. I also 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) and from the W.E. Upjohn Institute Postdoctoral Fellowship.
1 Introduction

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. 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 up to six colleges the student is considering attending. The government applies a fixed formula to this FAFSA information 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 specified 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 needs a way to distribute aid to millions of students across the country, and enlisting the help of colleges seems like an obvious solution to this logistic headache. 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. As I will demonstrate, 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, and many colleges require students to complete the FAFSA before 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

---

1See Tables 353 and 355 of the 2011 Digest of Education Statistics.
As a result, the transaction price—sticker price minus institutional discount—varies tremendously across students at the same college.

There are several notions of price in higher education, but in this paper I will focus on what I call the transaction price, which is the relevant price from the perspectives of both the student and the college. The transaction price is simply the college’s sticker price minus all institutional discounts (technically, institutional grants) she receives from her college. To illustrate, suppose a college has a posted sticker price of $20,000 per year but it offers the student a $15,000 discount. Then the transaction price is the difference, $5,000, which is the amount transferred from the student to the college. Each college offers a similar “price quote” to the student, and she weighs each transaction price against the other non-price characteristics of each college when deciding where to attend. One might object that, if the student also receives outside financial aid, such as a $2,000 Pell grant from the government, then her price is really $3,000 rather than $5,000. However, although 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. Moreover, Pell grants can also be used to pay for consumption, so if a college were to offer the student 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 to spend out of her available resources (including outside financial aid) as well as how much revenue the college will earn from that student.

Why do colleges care about the FAFSA? Economic theory tells us that a seller must have some information about a buyer’s willingness to pay in order to price discriminate, and the FAFSA 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 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 the IRS, in a process called verification. If a student’s FAFSA

---

2. Most high school seniors complete the FAFSA at the same time as their college applications. A major reason for doing this is so they can compare price offers when choosing which college to attend.
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.\footnote{It appears to be public knowledge that many colleges verify all of their FAFSA forms (Grant 2006; Weston 2014).} Effectively, the FAFSA grants colleges generous access to the IRS and other government databases and allows them to use that information to learn about a student’s willingness to pay.

Do colleges use the information on the FAFSA to price discriminate? If so, what would happen if we restricted the colleges’ ability to use some or all of the FAFSA information in their pricing? To answer these questions, I build a model of competitive 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. A college enrolls the student if it makes the best offer, as judged by the student. Students care about both price and other college characteristics, so a student may be willing to pay more to attend a particularly attractive college. 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. Within the model, a college 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, and 2) if a student places a high value on attending a particular college, then that college has some room to extract surplus from the student by offering a higher price (lower discount). I show how to reformulate the model in terms of a first-price auction in utility bids with conditionally independent private values and endogenous entry, which allows me to leverage
both theory and empirical methods from the auctions literature. The model predicts that, all else being equal, colleges will charge more desirable students less (for instance, if they have higher test scores). But colleges will charge students more if those students place a high value on attending the college. If colleges believe they face more competitors for the student, they will also charge the student less, as they bid more aggressively to try to attract her. 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 test the model’s qualitative predictions by regressing tuition discounts on student characteristics and including college fixed effects to isolate variation among students at the same college. I use income as a proxy for willingness to pay and find that higher-income students tend to receive smaller discounts (pay higher prices) than their classmates, and this effect is driven predominantly by private and very selective public colleges. Higher-ability students—proxied for by test scores and high school grade-point average (GPA)—tend to receive larger discounts (pay lower prices) than their classmates. Colleges do not necessarily know how many competitors they face, but if the student completes the FAFSA, they do see the list of colleges to which she had her information sent, which serves as a noisy signal about the true number of competitors. Students who list more colleges receive larger discounts. Moreover, when the number of applications is included in the regression, the coefficient on number of colleges listed hardly changes, but the coefficient on number of applications is essentially zero. The number of colleges listed on the FAFSA is also unrelated to a student’s discount if she completes the FAFSA after the college admissions season (January through March). In summary, I find that pricing patterns at elite colleges, defined as private and very selective public colleges, are consistent with the qualitative predictions of the model. Prices at elite colleges tend to be higher when students have a higher willingness to pay (higher income), lower when students are of higher quality (higher test scores and high school GPA), and lower when colleges believe they face more competitors (more colleges listed on the FAFSA).

I estimate my structural model using data on freshmen at elite colleges and

\[4\text{See Appendix A.1.2 for additional detail regarding the college classifications used in the paper.}\]
find that colleges successfully capture an average of 70 percent of the total match surplus through their student-specific prices, leaving the remaining 30 percent to the students. On average, students at elite colleges value attending their current college $18,035 more than their outside option of attending a nonelite college, but their average consumer surplus falls to $4,877 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 41 percent of the surplus, while those who list only one college receive an average of only 18 percent.

Using my structural estimates, I simulate a policy change wherein colleges are no longer permitted to use some or all of the FAFSA information in their pricing. For simplicity’s sake, I assume that the FAFSA conveys three pieces of information: parent adjusted gross income, a noisy signal about the number of competitors the college faces, and the fact that the student chose to complete the FAFSA at all. I simulate three counterfactuals: 1) the government restricts income information only, 2) the government restricts the signal about competitors only, and 3) the government restricts income information, the signal about competitors, and even whether the student completed the FAFSA. In all three counterfactuals, I hold the application strategy of students fixed and assume that colleges can always use basic demographic characteristics (age, gender, and race) as well as measures of student quality (ACT scores and high school GPA) to proxy for the lost FAFSA information.

I find that all three policy changes would affect overall prices, the allocation of students to colleges, welfare, and observed pricing patterns. Moreover, I also estimate heterogeneous effects that vary by student income. I find that average prices would fall by as much as $826 per student per year thereby lowering colleges’ tuition revenues.\footnote{The welfare consequences of 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.} However, with less information to use when price discriminating, colleges would inefficiently price up to 12.5 percent of students out of the elite market because they would be unable to tailor their prices as precisely and on occasion would inefficiently charge a student more than she is willing to pay. In response, this student, who should have been matched with an elite college, instead opts to attend a nonelite college. These inefficient matches lower total surplus, although this lost surplus is less than one-third the size of the lowered prices to students. I also find evidence that colleges would use other student characteristics to proxy,
albeit imperfectly, for the lost FAFSA information. Although all three policies lower prices on average, they each have different distributional consequences. In counterfactuals 1 and 3, colleges use the FAFSA to effectively levy an income “tax” coupled with a lump sum rebate. Thus, the lowest-income students benefit from colleges using income information, but that benefit gradually disappears as income rises so that middle- and high-income students pay more than they would if the information were restricted. Although this yields a degree of redistribution, less than half of the “tax revenue” that colleges raise by virtue of the FAFSA information is actually distributed to other students in the form of lower prices. For instance, in counterfactual 3, 65 percent of the tax revenue accrues to colleges in the form of higher tuition revenue, and only 35 percent is transferred to other students. Additionally, this redistribution varies considerably, even among students with the same income: in counterfactual 3 where all FAFSA information is restricted, 46 percent of low-income students actually pay more as a result of price discrimination while 11 percent of high-income students pay less. In contrast to counterfactuals 1 and 3, restricting information about a college’s competitors (counterfactual 2) lowers average prices in an income neutral way. Taken as a whole, the results indicate that although allowing colleges to use FAFSA information does increase efficiency somewhat and lower prices for some students, its main effect is to boost tuition revenue, primarily at the expense of middle- and high-income students.

The paper proceeds as follows. Section 2 reviews the previous literature and discusses how this paper both fits into 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.

2 Review of the Literature

My paper is the first to examine the importance of FAFSA information for college pricing. It is most closely related to two papers that have looked at the de-
terminants 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. Epple et al. (2006) use primarily college-level data 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 policy 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.

My paper is also related to a literature testing the Bennett Hypothesis. In 1987, William Bennett, then U.S. Secretary of Education, proposed what has become known as the Bennett Hypothesis (Bennett 1987). He posited that increases in federal financial aid programs do not actually help students and instead merely result in higher tuition levels. Empirical support for the Bennett Hypothesis has been mixed, with some studies rejecting it (McPherson and Schapiro 1991) and others finding modest support (Cellini and Goldin 2012; Epple et al. 2013; Long 2003). In contrast to this literature, I do not estimate the effect of financial aid per se on prices or price discrimination. Rather, while holding constant the generosity of federal aid programs, I estimate the consequences of allowing colleges to use FAFSA information in their pricing.

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

---

6 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.
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.

Economic theory tells us that a seller must have information about a buyer’s willingness to pay before it can price discriminate. Typically sellers must either rely on crude proxies for willingness to pay or come up with a clever way of inducing buyers to (partially) reveal their willingness to pay. Fortunately for colleges, they receive free access to a student’s FAFSA, which provides detailed and reliable financial information as well as a signal about the competitors the college will be facing. If such information were restricted, colleges would not be able to price discriminate as precisely, because they would have to rely on lower-quality information. However, theory can say little in general about the consequences for welfare. 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 withheld. Hence, the consequence of restricting colleges’ use of FAFSA information remains an empirical question.

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

3.1 Institutional Detail

In order to think about modeling the U.S. college market, we must first understand the real world mechanics of how students are matched with colleges and how prices are determined. To this end, let’s quickly walk through the timeline of a student who is applying for college. During her junior year of high school (and possibly even earlier) the 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. Colleges also consider a student’s high school grade point average (GPA) as an indicator of academic motivation and achievement. In the fall of her senior year, the student applies to one or more colleges. These applications are typically due in November or December. 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 also 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, information regarding any federal financial aid she might qualify for, and an offer of a discount. 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 non-selective 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. Students want to attend a great college (as judged by them) at the lowest possible price, and colleges want to attract great 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 tuition revenue if the student enrolls 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.

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

---

7 Colleges are 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.
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 in-
centivize her to choose one college over another. The same is true for most private
grants and scholarships—they are typically portable and do not affect the student’s preferences over colleges.

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.

In the model, a student invites a college to participate in her auction by submit-
ting 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 \( v_{ij} \) repre-
sents her valuation, in dollars, of attending the college and \( p_{ij} \) is the price college
j offers her. 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 \( v \) from equilibrium behavior.

Elite colleges compete for students on price. The college “wins” the auction—
the student enrolls—if it makes the best offer, \( u \), as judged by the student. Colleges
care about maximizing both the quality of their students as well as tuition revenue.
Let \( \Pi \) denote the space of college payoffs for enrolling a student. College j’s pay-
off from enrolling student i is \( \pi_{ij} = w_{ij} + p_{ij} \), where \( w_{ij} = z_j + \omega(X_i) \) represents
college j’s valuation, in dollars, of enrolling the student. The vector \( X_i \) denotes

\[ \text{Because federal and private grants are portable from college to college, they cancel out and do not affect the student’s decision. This also means that colleges will be unable to extract federal aid from students by, for instance, charging students more when they receive a larger Pell grant from the government.} \]
characteristics of student \( i \) that are observed to both the college and the econometrician, while \( z_j \) is observed to the college only. \( w_{ij} \) varies across students, with a more desirable student having a higher valuation, thus lowering the college’s willingness to receive for that student. \( w_{ij} \) 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 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 some students may have a higher \( w_{ij} \) than 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 the data. Lastly, note that since \( w_{ij} \) is college \( j \)’s valuation, \(-w_{ij}\) is \( j \)’s willingness to receive—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.

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 a noisy signal, \( \tilde{n}_i \), and it knows the probability of \( n \) bidders conditional on the value of the noisy signal, \( \rho(n|\tilde{n}) \).

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. What does it mean to assume that the college perfectly observes \( v_{ij} \)? 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 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

\[ ^9 \text{As we will see later, } w_{ij} \text{ is negative for most students.} \]
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.

College \( j \) makes a take-it-or-leave-it price offer, \( p_{ij} \), to maximize

\[
\pi_{ij} \mathbb{P}[j \text{ wins}] = (w_{ij} + p_{ij}) \mathbb{P}[u_{ij} \geq u_{i\ell} \forall \ell \neq j | \tilde{n}_i].
\]

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 \( j \), and rewrite the college’s objective function as

\[
\begin{align*}
\{ (v_{ij} + w_{ij}) - (v_{ij} - p_{ij}) \} \mathbb{P}[u_{ij} \geq u_{i\ell} \forall \ell \neq j | \tilde{n}_i] \\
= (s_{ij} - u_{ij}) \mathbb{P}[u_{ij} \geq \beta(s_{i\ell}|X_i) \forall \ell \neq j | X_i, \tilde{n}_i] \\
= (s_{ij} - u_{ij}) \mathbb{P}[\beta^{-1}(u_{ij}|X_i) \geq s_{i\ell} \forall \ell \neq j | X_i, \tilde{n}_i] \\
= (s_{ij} - u_{ij}) \left\{ \sum_{n=1}^{\pi} F_{S|X_i}^{n-1} \left( \beta^{-1}(u_{ij}|X_i) \right) \rho(n|\tilde{n}_i) \right\},
\end{align*}
\]

where \( F_{S|X_i} \) is the distribution of match surpluses, conditional on student covariates \( X_i \), with support \( S = [s, \bar{s}] \), and \( \beta(s|X_i) \) is the equilibrium bid function conditional on covariates. As is standard in the auction literature, I assume that the density \( f_{S|X_i} \) is strictly positive over the entire support. I have assumed that the \( s_{ij} \) are drawn independently from the distribution \( F_{S|X_i} \), conditional on the covariates \( X_i \). 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 \( w_{ij} \).\footnote{After conditioning on \( X_i \), idiosyncratic variation in \( w_{ij} \) only occurs through \( z_j \).}

In the language of the auctions literature, I have assumed a conditionally independent private values environment.

This setup is very similar to a canonical first-price auction with independent and
symmetric private values. It differs from the usual model in that the private values have been replaced by \( s_{ij} \), the bids are now in terms of student utility (expressed in dollars), and the number of bidders is uncertain.\(^{11}\) 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, no college will bother to enter unless \( s_{ij} \geq 0 \). Intuitively, the college offers a portion of \( s_{ij} \) to the student and keeps the rest for itself; 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. Taking the first order condition for the utility bid \( u_{ij} \) yields the ODE

\[
\beta'(s|X_i) = (s - \beta(s|X_i)) \frac{\sum_{n=1}^{\pi}(n - 1)F_{S|X_i}^{n-1}(s)f_{S|X_i}(s)\rho(n|\tilde{n}_i)}{\sum_{n=1}^{\pi} F_{S|X_i}^n(s)\rho(n|\tilde{n}_i)},
\]

with the initial condition \( \beta(0|X_i) = 0 \). If \( s_{ij} = 0 \), then the college can only just match the student’s outside option. When \( s_{ij} > 0 \), the college offers the student some, but not all, of the positive match surplus.\(^{12}\)

### 3.3 Qualitative Predictions of the Model

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 \( \beta' \). In other words, when the value of the match rises, the college offers a fraction \( \beta' \) 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

---

\(^{11}\)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 \( \rho(n|\tilde{n}) \).

\(^{12}\)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 \( F_{S|X_i} \).
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 a noisy signal $\bar{n}$ about the actual number of bidders, $n$. Colleges will respond to a higher signal by making more aggressive bids and thus offering lower prices. However, holding constant the noisy signal and other student characteristics, price 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.

The higher education market is extremely diverse, and the NPSAS sampling scheme reflects that diversity by sampling students at a wide variety of post-
secondary institutions, ranging from cosmetology programs to Ivy League universities. I restrict myself to “traditional” college students, which I define as meeting the following criteria:

- Degree-seeking undergraduate with no prior bachelor’s degree
- U.S. resident (not foreign)
- Less than 31 years old
- Attended a public or private not-for-profit college in the 50 states (plus D.C.) during the 2007–2008 school year
- Attended only one college during the 2007–2008 school year
- Enrolled for nine or more full-time-equivalent months
- Was not on an athletic scholarship
- Did not receive a tuition waiver because a parent’s employment at the college

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 those waivers are not really discounts; rather, they represent a nonwage benefit to the student’s parent who works at the college. The sample is restricted to U.S. residents because foreign students are not eligible to complete the FAFSA.¹³

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. Tables 1 and 2 contain 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 gross income of just over $65,000.

¹³In the fall of 2010, nonresident aliens accounted for 2.2 percent of undergraduate enrollment (see Table 237 of the 2011 Digest of Education Statistics).
The elite sample consists of freshmen at private and very selective public four-year colleges. Each cell contains the raw count of the number of colleges and students in the sample. Per NCES requirements, counts have been rounded to the nearest ten.

Table 1: Cell Counts
### Descriptive Statistics

<table>
<thead>
<tr>
<th></th>
<th>Freshman sample</th>
<th>Elite sample</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Std dev</td>
</tr>
<tr>
<td>Tuition discount</td>
<td>$2,914</td>
<td>($5,722)</td>
</tr>
<tr>
<td>Student received discount</td>
<td>0.38</td>
<td>(0.49)</td>
</tr>
<tr>
<td>Sticker price</td>
<td>$10,168</td>
<td>($9,737)</td>
</tr>
<tr>
<td>Parent adjusted gross income</td>
<td>$65,400</td>
<td>($55,502)</td>
</tr>
<tr>
<td>ACT score</td>
<td>21.2</td>
<td>(4.7)</td>
</tr>
<tr>
<td>High school GPA</td>
<td>3.42</td>
<td>(0.57)</td>
</tr>
<tr>
<td>Earned AP credit</td>
<td>0.22</td>
<td>(0.42)</td>
</tr>
<tr>
<td>Parents with college degree</td>
<td>0.68</td>
<td>(0.80)</td>
</tr>
<tr>
<td>Completed FAFSA</td>
<td>0.82</td>
<td>(0.38)</td>
</tr>
<tr>
<td>Additional colleges listed on FAFSA</td>
<td>1.2</td>
<td>(1.7)</td>
</tr>
<tr>
<td>Age as of 12/31/07</td>
<td>18.6</td>
<td>(0.9)</td>
</tr>
<tr>
<td>Female</td>
<td>0.54</td>
<td>(0.50)</td>
</tr>
<tr>
<td>White</td>
<td>0.65</td>
<td>(0.48)</td>
</tr>
<tr>
<td>Black</td>
<td>0.13</td>
<td>(0.34)</td>
</tr>
<tr>
<td>Hispanic</td>
<td>0.12</td>
<td>(0.33)</td>
</tr>
<tr>
<td>Asian</td>
<td>0.05</td>
<td>(0.22)</td>
</tr>
<tr>
<td>Other</td>
<td>0.04</td>
<td>(0.20)</td>
</tr>
</tbody>
</table>

Sticker prices come from NPSAS variable TUITION2. Tuition discounts come from NPSAS variable INGRTAMA. "Additional colleges listed on FAFSA" is only calculated for those students who completed the FAFSA. Students can list up to six colleges on the FAFSA, so the number of additional colleges listed runs from zero to five. The elite sample is a subset of the freshman sample and consists of freshmen at private and very selective public four-year colleges. No sample weights were used.

Table 2: Descriptive Statistics
4.2 Testing Qualitative Predictions of the Model

The model makes four qualitative predictions.

1. All else equal, elite colleges will offer a higher price to students whom it believes are willing to pay more.

2. All else equal, elite colleges will offer a lower price to students whom it views as more desirable.

3. All else equal, elite colleges will offer a lower price to students when it believes it is facing stiffer competition.

4. 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 3 reports the estimates from the following regression

\[
\text{Tuition Discount}_{ij} = X_i \beta + a_j + e_{ij},
\]

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. 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. 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.

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 a related study, Beginning Postsecondary Students (BPS), that focuses on first-time freshmen in NPSAS does ask about appli-
Within-College Discounting Patterns

Dependent variable: tuition discount

<table>
<thead>
<tr>
<th></th>
<th>Estimate (Standard Error)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parent AGI (in $10,000s)</td>
<td>-123.5 (15.22) ***</td>
</tr>
<tr>
<td>Number of additional colleges listed on FAFSA</td>
<td>372.7 (56.64) ***</td>
</tr>
<tr>
<td>Completed FAFSA</td>
<td>850.0 (164.3) ***</td>
</tr>
<tr>
<td>ACT score</td>
<td>133.4 (18.43) ***</td>
</tr>
<tr>
<td>High school GPA</td>
<td>394.1 (101.1) ***</td>
</tr>
<tr>
<td>Earned AP credit in high school</td>
<td>277.7 (174.2)</td>
</tr>
<tr>
<td>Number of parents with college degree</td>
<td>93.8 (90.59)</td>
</tr>
<tr>
<td>Age as of 12/31/07</td>
<td>-37.5 (58.41)</td>
</tr>
<tr>
<td>Female</td>
<td>113.3 (115.6)</td>
</tr>
<tr>
<td>Black</td>
<td>580.2 (232.0) *</td>
</tr>
<tr>
<td>Hispanic</td>
<td>581.0 (245.8) *</td>
</tr>
<tr>
<td>Asian</td>
<td>44.1 (391.9)</td>
</tr>
<tr>
<td>Other / multiple</td>
<td>549.7 (360.7)</td>
</tr>
<tr>
<td>Out-of-state, public</td>
<td>1135.5 (295.4) ***</td>
</tr>
</tbody>
</table>

College fixed effects: Yes
Observations: 5640
R-squared: 0.670

The regression includes students from the freshman sample. The omitted race category is "white." For students who completed the FAFSA, "number of additional colleges listed" ranges from zero to five (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. Robust standard errors are in parentheses. Sampling weights were used (NPSAS variable WTA000).

* p<0.05, ** p<0.01, *** p<0.001

Table 3: Within-College Discounting Patterns

---

20
Figures 1 and 2 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 1 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 2 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.

Tables 4 and 5 focus on the relationship between tuition discounts and listing more colleges on the FAFSA. Column one of Table 4 replicates the regression from Table 3. In column two, I interact number of colleges listed with college type and selectivity, which shows that the coefficient in column one is driven predominantly

---

14See Appendix section A.2 for more information about the BPS data.
Figure 2: Histogram of Applications Given Schools Listed on FAFSA
by very selective private colleges. 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. The rest of the elite colleges, consisting of less-selective four-year private colleges and very selective public colleges, also have positive and economically significant coefficients, although they are not statistically significant.

The fourth prediction of the model says that, holding constant the number of colleges listed on the FAFSA, the number of colleges actually applied to will be unrelated to a student’s tuition discount. In columns three and four of Table 4, I use BPS data from the 2003-2004 school year that explicitly asks freshmen how many colleges they applied to. When the number of applications is included in the regression, the coefficient on number of colleges listed remains significant and hardly changes, but the coefficient on number of applications is essentially zero.

If the relationship between tuition discounts and listing more colleges on the FAFSA is really due to colleges responding to increased competition, then the relationship should disappear for students who complete the FAFSA after admissions season is over. In Table 5, I interact the number of colleges a student lists on her FAFSA with the month in which she completed the FAFSA. Table 5 shows that the positive coefficient in column one of Table 4 is driven entirely by students who complete the FAFSA during the admissions season (January through March). For those students who complete the FAFSA later in the year, after the admissions season is over, listing more colleges on the FAFSA does not appear to be related to a student’s discount.

An additional implication of the model is that some colleges (elite colleges) will use discounts to attract students while others (nonelite colleges) will not. In Table 6, I focus on the discount-income gradient from Table 3 by interacting parent income with college type and selectivity. Table 6 indicates that the coefficient on income in Table 3 is driven largely by private and very selective public four-year colleges. At very selective private colleges, a $10,000 increase in parent adjusted gross income is associated with a $473 drop in tuition discount. Tuition discounts are unrelated to income at two-year and nonselective colleges—precisely the colleges least likely to possess the market power necessary to engage in price discrimination. On the other hand, tuition discounts are negatively correlated with income at private and very selective public colleges—those colleges most likely to have the necessary market
### The Number of Colleges Listed on a Student’s FAFSA

<table>
<thead>
<tr>
<th>Number of additional colleges listed on FAFSA</th>
<th>Dependent variable: tuition discount</th>
<th>Freshman sample</th>
<th>Freshman sample</th>
<th>BPS 03-04</th>
<th>BPS 03-04</th>
</tr>
</thead>
<tbody>
<tr>
<td>Very selective</td>
<td>Public</td>
<td>131.8 (90.02)</td>
<td></td>
<td>249.3 (36.60) ***</td>
<td>245.5 (38.91) ***</td>
</tr>
<tr>
<td></td>
<td>Private</td>
<td>909.8 (239.9) ***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Moderately selective</td>
<td>Public</td>
<td>56.0 (62.38)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Private</td>
<td>249.3 (155.8)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not selective</td>
<td>Public</td>
<td>44.9 (104.6)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Private</td>
<td>264.5 (307.7)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Two-year</td>
<td>Public</td>
<td>-0.5 (36.80)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Private</td>
<td>-463.8 (204.5) *</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Note:**

- Number listed on FAFSA: 6.9 (26.79)
- College fixed effects: Yes
- Observations: 5640
- $R^2$: 0.670

See note to Table 3. The regressions reported in columns 1 and 2 are identical to the regression in Table 3 except that in column 2 “Number of additional colleges listed on the FAFSA” has been interacted with college selectivity. I set “Number of additional colleges listed on the FAFSA” to zero for students who did not complete the FAFSA and always include a dummy for whether the student completed the FAFSA. In column 2 when I interact “Number of colleges listed” with college type, I also interact the dummy for completing the FAFSA with college type. 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). Robust standard errors are in parentheses. Sampling weights (NPSAS variable WTA000) were used in all regressions.

* p<0.05, ** p<0.01, *** p<0.001

Table 4: The Number of Colleges Listed on a Student’s FAFSA
### Colleges Listed and Month FAFSA Completed

**Dependent variable: tuition discount**

<table>
<thead>
<tr>
<th>Number of additional colleges listed on FAFSA</th>
<th>January</th>
<th>February</th>
<th>March</th>
<th>April or later</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>530.6 (144.2) ***</td>
<td>290.7 (81.7) ***</td>
<td>11.3 (91.5)</td>
<td>77.4 (95.1)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>College fixed effects</th>
<th>Observations</th>
<th>$R$-squared</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>5630</td>
<td>0.673</td>
</tr>
</tbody>
</table>

See note to Table 3. The regression specification here is identical to that in Table 3 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<0.01, *** p<0.001

Table 5: Colleges Listed and Month FAFSA Completed

Power. Figure 3 further emphasizes this point. It plots the fitted values of tuition discounts against parent adjusted gross income, holding other covariates fixed. Again, we can see that the income gradient in Table 3 is driven almost entirely by private and very selective public colleges.

### 5 Structural Identification and Estimation

Standard empirical auction methods combine economic theory with observed bids to estimate the structural primitives of the auction model. We could proceed to 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 actually observed in the data. Rather, I observe the tuition offer, $p$, which is only one component of the utility bid $\beta(s) = v - p$. At first glance, this appears to be a serious problem. But I will show how we can still identify the model using data on student characteristics and transaction prices. The identification strategy is similar in spirit to that used by [Guerre et al. (2000)](Guerre+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
## Interacting Income With College Type

*Dependent variable: tuition discount*

<table>
<thead>
<tr>
<th>College fixed effects</th>
<th>Yes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observations</td>
<td>5640</td>
</tr>
<tr>
<td><em>R</em>-squared</td>
<td>0.693</td>
</tr>
</tbody>
</table>

See note to Table 3. The regression specification here is identical to that in Table 3 except that "Parent adjusted gross income" has been interacted with college type. The remaining covariates were included but not reported here. Robust standard errors are in parentheses. Sampling weights were used (NPSAS variable WTA000).

* p<0.05, ** p<0.01, *** p<0.001

Table 6: Interacting Income With College Type
The fitted values plotted here come from three separate regressions. In all three regressions, the fitted values represent a white female student with all other covariates set to their sample means. Parent adjusted gross income was included as a fourth-order polynomial.

Figure 3: Predicted Discounts, Holding Other Covariates Fixed
approach is similar, except I must first deal with the fact that the bids themselves are not observed.

The identification strategy is as follows. Define the 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 \( \pi(s|X_i) \) is monotone in \( s \) with \( \pi'(s|X_i) \in (0, 1) \). Denote the distribution of college payoffs, \( \pi \), conditional on \( X_i \) by \( F_{\pi|X_i} \). Now the derivative of the payoff function \( \pi(s|X_i) \) is

\[
\pi'(s|X_i) = 1 - \beta'(s|X_i) = 1 - (s - \beta(s|X_i)) \frac{\sum_{n=1}^{\pi}(n - 1) F_{S|X_i}^{n-1}(s) f_{S|X_i}(s) \rho(n|\tilde{m}_i)}{\sum_{n=1}^{\pi} F_{S|X_i}^{n}(s) \rho(n|\tilde{m}_i)}
\]

\[
\Rightarrow \quad \pi'(s|X_i) = \frac{\sum_{n=1}^{\pi} F_{S|X_i}^{n}(s) \rho(n|\tilde{m}_i)}{\sum_{n=1}^{\pi} (n - 1) F_{S|X_i}^{n-1}(s) f_{S|X_i}(s) \rho(n|\tilde{m}_i) - \pi(s|X_i)}
\]

\[
\Rightarrow \quad \pi(s|X_i) = \frac{\sum_{n=1}^{\pi} F_{S|X_i}^{n}(s) \rho(n|\tilde{m}_i)}{\sum_{n=1}^{\pi} (n - 1) F_{S|X_i}^{n-1}(s) f_{S|X_i}(s) \rho(n|\tilde{m}_i)} \left( \frac{1}{\pi'(s|X_i)} - 1 \right). \tag{3}
\]

I have now rewritten \( \pi(s|X_i) \) in terms of the distribution of college payoffs rather than the surplus distribution by using the fact that \( F_{S|X_i}(s) = F_{\pi|X_i}(\pi(s|X_i)) \) and therefore \( f_{\pi|X_i}(\pi(s|X_i)) = \frac{f_{S|X_i}(s)}{\pi'(s|X_i)} \). Solving (3) for \( \pi'(s|X_i) \) gives

\[
\pi'(s|X_i) = \left( 1 + \pi(s|X_i) \frac{\sum_{n=1}^{\pi}(n - 1) F_{\pi|X_i}^{n-1}(s) f_{\pi|X_i}(s) \rho(n|\tilde{m}_i)}{\sum_{n=1}^{\pi} F_{\pi|X_i}^{n}(s) \rho(n|\tilde{m}_i)} \right)^{-1} \quad 0 \leq s \tag{4}
\]

\[
\pi(0|X_i) = 0. \tag{5}
\]

Notice that since \( \pi : S \rightarrow \Pi \) is monotone its inverse \( \psi : \Pi \rightarrow S \) exists and has a derivative that is simply the reciprocal of \( \pi' \). So we can write

\[
\psi'(\pi|X_i) = 1 + \pi \frac{\sum_{n=1}^{\pi}(n - 1) F_{\pi|X_i}^{n-1}(s) f_{\pi|X_i}(s) \rho(n|\tilde{m}_i)}{\sum_{n=1}^{\pi} F_{\pi|X_i}^{n}(s) \rho(n|\tilde{m}_i)} \quad 0 \leq \pi \tag{6}
\]

\[
\psi(0|X_i) = 0. \tag{7}
\]

Finally, Equation (6) can be solved by simply integrating from 0 to \( \pi \). \( \psi(\pi|X) \) is the equilibrium inverse payoff function; it maps from the space of college payoffs, \( \Pi \), to the space of match surpluses, \( S \). \( \psi(\pi|X) \) depends only on the equilibrium distribution of college payoffs, \( F_{\pi|X} \), and the distribution of potential bidders, \( \rho(n|\tilde{m}) \). Equation (6) provides the key to identifying the model.

In section 5.1 I prove that the model can be identified from equilibrium transac-
tion prices, but the intuition behind the proofs is not complicated. From Equation (6) 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, $F_{\pi|X}$, is just a shifted version of the distribution of transaction prices, $F_{p|X}$, 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 covariates can be used to estimate the distribution of college payoffs $F_{\pi|X}$. $F_{\pi|X}$, $f_{\pi|X}$, and $\rho(n|\tilde{n})$ can then be combined with Equation (6) to identify the primitives of the model.

5.1 Identification

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

Proof. Define the function $y(X,j) \equiv \inf\{\supp(p|X,j)\}$ 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 $G_{\pi|X}$. Note that $\pi|X,j$ is just a shifted version of the transaction price.

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 $\supp(p|X,j)$.

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 (5) that $\pi(0|X) = 0$; that is, the college receives zero payoff from matches with zero surplus. Since 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 $\rho(n|\tilde{n})$.

\[\text{15} \text{The college would never be willing to charge a student less than } -w_{ij}.\]
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, 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

$$G_{\pi|X}(z) = \sum_{n=1}^{\pi} \rho(n|\tilde{n})F_{\pi|X}(z).$$

Define the transformation $T(\alpha) \equiv \sum_{n=1}^{\pi} \rho(n|\tilde{n})\alpha^n$ 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 $\rho(n|\tilde{n})$.

Proof. Combining $F_{\pi|X}$ and $\rho(n|\tilde{n})$ with Equations (6) and (7) 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

$$F_{\pi|X}(z) = F_{S|X}(\psi(\pi|X)|X)$$

$$f_{\pi|X}(z) = f_{S|X}(\psi(\pi|X)|X)\psi'(\pi|X)$$

■

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). The NPSAS contains information on various student characteristics as well as detailed information about prices

---

16 To be more precise, I actually identify the truncated distribution of $F_{S|X}$ conditional on $S \geq 0$, which is the standard result in the presence of a binding reserve price.
and discounts. For the structural estimation I focus on freshmen at elite colleges—private and very selective four-year colleges. Tables 1 and 2 provide cell counts and summary statistics of key variables. To estimate the model, 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, $\pi_{ij}$. In the second step, I estimate the distribution of college payoffs $F_{\pi|X}$ and combine this with the equilibrium conditions in Equations (6) and (7) to recover $F_{S|X}$, the equilibrium bid function $\beta(s|X)$, and an estimate of the match surplus, $\hat{s}_{ij}$.

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

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

(8)

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. 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 local approximation of the left tail. Whichever perspective we take, in order to estimate $\alpha_1$, $\alpha_2$, and $\underline{p}$, I estimate several quantiles of the price distribution $F_{p|X,j}$ using the quantile regression17

$$F_{p|X,j}^{-1}(t) = X\gamma^t + a_j^t \quad t = .05, .10, \ldots, .40,$$

(9)

where the $a_j^t$ are college fixed effects. 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 (8) to these points.

17I include all observations from the full sample, including upperclassmen, in these quantile regressions. This is the only place 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.
by estimating the parameters $\alpha_1(X, j) > 0$, $\alpha_2(X, j) > 0$, and $p(X, j)$ via nonlinear least squares, subject to the constraint that $p(X_{ij}, j) \leq p_{ij}$. Armed with an estimate $\hat{w}_{ij} = -\hat{p}(X, j)$, I simply estimate the college payoff as $\hat{\pi}_{ij} = \hat{w}_{ij} + p_{ij}$.

The second step in the estimation process begins by estimating the distribution of college payoffs. Students in the auction 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 $F_{\pi|X_i}$ which is the distribution of college payoffs conditional on student covariates. In order to do so in a flexible yet feasible way, I employ the following procedure to estimate the primitives of the model:

1. Estimate the cdf $\hat{G}_{\pi|X_i}$ and pdf $\hat{g}_{\pi|X_i}$ of estimated college payoffs for each student.\(^{19}\)

2. Calculate $\hat{F}_{\pi|X_i}$, the parent distribution of $\hat{G}_{\pi|X_i}$, for each student. Also calculate the density $\hat{f}_{\pi|X_i}$.

3. For each student, use $\hat{F}_{\pi|X_i}$ and $\hat{f}_{\pi|X_i}$ from step 2 along with Equations (6) and (7) to solve for the inverse markup function $\psi(\pi|X_i)$.

4. For each student, calculate $\hat{s}_{ij} = \psi(\hat{\pi}_{ij}|X_i)$, which is the estimated match surplus from matching student $i$ with the college she ended up attending. Calculate the implied value the student places on college $j$ relative to a nonelite college $\hat{v}_{ij} = \hat{s}_{ij} - \hat{w}_{ij}$. Also calculate $F_{S|X_i}$, the distribution of match surpluses conditional on student covariates.

### 5.3 Baseline Structural Estimates

Figure 4 contains a histogram of estimated college payoffs for the elite sample. The distribution skews right, and college payoffs are typically less than $20,000, although they can range as high as $50,000. Recall that these payoffs represent the difference between a student’s transaction price and the lowest price her college would have been willing to accept from her.

---

\(^{18}\)This constraint binds for about 3 percent of the students in my sample.

\(^{19}\)I briefly outline this estimation procedure in the appendix. The estimator is discussed in detail in my working paper (Fillmore 2014).
Figure 4: Histogram of Estimated College Payoffs (restricted to sample used in structural estimation)
On average, colleges are able to extract an average of 70 percent of the match surplus through their individualized prices while students receive the remaining 30 percent. However, the student’s share of the surplus depends on how much competition the college believes it faces. Table 7 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 percent of the match surplus, while students who list only one college receive just 18 percent.

### Student Share of Match Surplus

<table>
<thead>
<tr>
<th>Number of schools listed on FAFSA</th>
<th>Average student share of match surplus (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>No FAFSA</td>
<td>23.7</td>
</tr>
<tr>
<td>1</td>
<td>18.0</td>
</tr>
<tr>
<td>2</td>
<td>32.9</td>
</tr>
<tr>
<td>3</td>
<td>36.6</td>
</tr>
<tr>
<td>4</td>
<td>39.1</td>
</tr>
<tr>
<td>5</td>
<td>40.7</td>
</tr>
<tr>
<td>6</td>
<td>41.3</td>
</tr>
</tbody>
</table>

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.

Table 7: 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 $15,413 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 $4,877 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 $4,877 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 $4,877 with the average transaction price of $13,158 implies that on average students at elite colleges are willing to pay $18,035 to attend their current colleges rather than a nonelite college (their outside option).

Because colleges are extracting a large amount of the match surplus, students may have less incentive to apply to additional colleges. Thus, the expected value
of applying to college $j'$ if the student’s current best option is $j$ can be written as

$$\mathbb{E}[\text{Value of Applying to } j'|j] = \mathbb{P}[s_{ij'} > s_{ij}] \times \mathbb{E}[\beta(s_{ij'}) - \beta(s_{ij})|s_{ij'} > s_{ij}].$$

The first term, $\mathbb{P}[s_{ij'} > s_{ij}]$, represents the probability that $j'$ beats $j$, while the second term, $\mathbb{E}[\beta(s_{ij'}) - \beta(s_{ij})|s_{ij'} > s_{ij}]$, represents the expected marginal gain in the event that $j'$ beats $j$. Applying to one additional college only pays off if the college beats out the student’s current best match. And if it does, the student will only receive a fraction of the additional match surplus because the new college will extract much of it through price discrimination. Although I do not directly model students’ application decisions, my estimates can speak to the incentives students face when choosing how many applications to send out. In Table 14, I calculate the expected return from applying to an additional college. In this calculation I assume the extreme case that students apply to colleges randomly. Thus, for each student I assume $\mathbb{P}[s_{ij'} > s_{ij}] = 1 - F_{S|X_i}(s_{ij})$. Then I multiply this probability by the conditional expected utility bid increase (conditional on $s_{ij'}$ exceeding $s_{ij}$). This calculation indicates that the expected return to applying to an additional college at random is $693.20$

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 my sample of dependent freshmen, parent adjusted gross income alone can explain 73 percent of the variation in a student’s Expected Family Contribution.\footnote{\label{fn:1}A student’s Expected Family Contribution is the federal government’s assessment of how much the student and her family should be expected to contribute toward her education. This number amounts to a sufficient statistic for determining a student’s federal aid eligibility.} 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

\footnote{However, if, as seems likely, students tend to apply to colleges that are a better match first, then $\mathbb{P}[s_{ij'} > s_{ij}] < 1 - F_{S|X_i}(s_{ij})$ and the estimates in Table 14 will overstate the returns of an additional application.}
college faces (the number of colleges listed on the FAFSA), and 3) the fact that the student chose to complete the FAFSA at all. I simulate three counterfactuals:

1. Colleges cannot use income information only.

2. Colleges cannot use the noisy signal of the number of competitors only.

3. Colleges cannot use income information, the signal of the number of competitors, or even whether the student completed the FAFSA.

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. $\tilde{X}_i$ denotes the limited set of student covariates that the college can use in the counterfactual. Colleges always make the best use of the information they have, so they can use these student covariates to proxy for those that they cannot use.

### 6.1 Modeling the Counterfactuals

In all three counterfactuals, I model the loss of information as a pair of shocks to the college’s beliefs about $v_{ij}$ and $w_{ij}$. The first shock, $e_i$, is a shock to each college’s assessment of the value that student $i$ places on the college. That is,

$$v_{ij} = \tilde{v}_{ij} + e_i,$$

where $v_{ij}$ is the true match surplus, $\tilde{v}_{ij}$ is a forecast of $v_{ij}$ given $\tilde{X}_i$, and $e_i$ represents a mean zero forecast error resulting from the lost information. $e_i$ 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. The second shock, $\xi_i$, is a student-specific shock to $w_{ij}$:

$$w_{ij} = z_j + \tilde{\omega}(\tilde{X}_i) + \xi_i.$$

Again, $\tilde{\omega}_{ij}$ is a forecast of $w_{ij}$, given the limited information set $\tilde{X}_i$, and $\xi_i$ is the associated (mean zero) forecast error.

---

22 Note that $e_i$ is independent of the observable student covariates $\tilde{X}_i$. 
Define $\tilde{u}_{ij} \equiv \tilde{v}_{ij} - p_{ij} = u_{ij} + e_i$ to be the utility offer that college $j$ thinks it is making to student $i$. In reality, $j$’s utility offer is $u_{ij} = \tilde{u}_{ij} + e_i$. Similarly, $\tilde{\pi}_{ij} = \tilde{w}_j(\tilde{X}_i) + p_{ij} = \pi_{ij} - \xi_i$ is the payoff that college $j$ thinks it will get if it enrolls the student. Finally, define the college’s forecast of the match surplus $\tilde{s}_{ij} = \tilde{\pi}_{ij} + \tilde{u}_{ij}$.

The college makes a tuition offer to maximize its expected payoff:

$$\max_{\tilde{p}_{ij}} \mathbb{E}_{\pi|\tilde{\pi}}[\pi_{ij} P[\tilde{u}_{ij} + e_i > 0 \cap \tilde{u}_{ij} + e_i > \tilde{u}_{i\ell} + e_i \mid \ell \neq j]]$$

$$\rightarrow \max_{\tilde{p}_{ij}} \tilde{\pi}_{ij} P[e_i > -\tilde{u}_{ij}] \mathbb{P} [\tilde{u}_{ij} > \tilde{u}_{i\ell} \mid \ell \neq j]$$

$$\rightarrow \max_{\tilde{u}_{ij}} (\tilde{s}_{ij} - \tilde{u}_{ij})(1 - F_e(-\tilde{u}_{ij})) \sum_{n=1}^{\tilde{\pi}} F_{\tilde{S}|\tilde{X}_i}^{-1}(\tilde{\beta}^{-1}(\tilde{u}_{ij}|\tilde{X}_i)) \rho(n|\tilde{n}_i)$$

The first order condition for the college yields the ODE

$$\tilde{\beta}'(\tilde{s}|\tilde{X}_i) = \frac{\sum_{n=1}^{\pi} (n-1) F_{\tilde{S}|\tilde{X}_i}^{-1}(\tilde{s}) f_{\tilde{S}|\tilde{X}_i}(\tilde{s}) \rho(n|\tilde{n}_i)}{\sum_{n=1}^{\pi} F_{\tilde{S}|\tilde{X}_i}^{-1}(\tilde{s}) \rho(n|\tilde{n}_i)}$$

(10)

with the initial condition

$$\tilde{\beta}(\tilde{s}|\tilde{X}_i) = -\tilde{e}.$$

The initial condition now says that the lowest-value bidder is the college with an observed match surplus so low that even the most favorable draw of $e$ possible would only just make student $i$ indifferent between college $j$ and the competitive fringe. A college in this situation will bid its observed match surplus, $\tilde{s}_{ij} = -\tilde{e}$. As $\tilde{s}_{ij}$ rises, the college must consider competition from other bidders as well as competition from the competitive fringe. Notice that the numerator in Equation (10) is analogous to Equation (2). The denominator in Equation (10) reflects college $j$’s uncertainty over whether it will beat the competitive fringe. However, once $\tilde{\beta} = -\tilde{e}$, the denominator equals one thereafter, because college $j$ is now making an offer large enough that competition from the fringe is no longer a concern.

Counterfactual bids (and prices) will differ from those in baseline for four reasons: 1) $F_{\tilde{S}|X_i}$ will differ, perhaps substantially, from $F_{S|X_i}$; 2) in counterfactuals 2 and 3, colleges will lose access to the noisy signal $\tilde{n}_i$ which will affect the distri-
bution $\rho(n|\tilde{n}_i)$, 3) colleges now face uncertainty over their position relative to the competitive fringe; 4) each college’s observed match surplus, $\tilde{s}_{ij}$, will differ from the true (baseline) match surplus, $s_{ij}$, by the realization of the information shock $e_i + \xi_i$. 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 $\xi_i$ do not vary by college, the relative rankings of colleges are unaffected. However, the relative ranking of the competitive fringe could be affected. On occasion, the winning bidder will not beat the competitive fringe because $e_i$ turned out to be unexpectedly low (negative). The student appears to have a high willingness to pay when she actually does not, so the colleges make higher price offers and lose the student to the fringe. Because the 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 $\xi_i$, for $\nu_{ij}$ and $\omega_{ij}$, respectively. I also estimate the distribution of the college’s forecast of match surplus $F_{\tilde{s}|\tilde{X}_i}$. Combining $F_{\tilde{s}|\tilde{X}_i}$, $F_{e}$, and Equation (10) allows me to calculate the counterfactual equilibrium bid function and simulate counterfactual outcomes. Table 8 provides a sense of the magnitude of these forecast errors, as well as how much information about parent income and the number of colleges listed on the FAFSA a college can recover from using the remaining student covariates as proxies. The results in Table 8 indicate that most of the college’s forecast error is coming from its uncertainty about student preferences, $e_i$, rather than from uncertainty about its own valuation for the student, $\xi_i$.

6.2.1 Prices and Misallocation

Table 9 compares the baseline structural estimates with three counterfactuals: 1) colleges cannot use income information only, 2) colleges cannot use the number of colleges listed on the FAFSA only, and 3) colleges cannot use income information, the number of colleges listed, and even whether the student completed the FAFSA at all. In all three counterfactuals, prices fall by an average of $730 in counterfactual 23In counterfactual 2, $\tilde{n}_i$ becomes a binary indicator for whether the student completed the FAFSA. In counterfactual 3, $\rho(n|\tilde{n}_i)$ becomes the marginal distribution $\rho(n)$.

24See the appendix for more details.
### Information Loss in the Counterfactuals

**Standard deviation of forecast errors**

<table>
<thead>
<tr>
<th>FAFSA information restricted</th>
<th>Number of colleges listed</th>
<th>All FAFSA info</th>
</tr>
</thead>
<tbody>
<tr>
<td>Std dev of $e$</td>
<td>$2,311$</td>
<td>$1,470$</td>
</tr>
<tr>
<td>Std dev of $\xi$</td>
<td>$683$</td>
<td>$0$</td>
</tr>
</tbody>
</table>

**$R^2$ of regressing FAFSA variables on remaining student covariates**

<table>
<thead>
<tr>
<th>FAFSA variable</th>
<th>Parent income</th>
<th>Number of colleges listed</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R$-squared</td>
<td>0.129</td>
<td>0.319</td>
</tr>
</tbody>
</table>

The top panel reports the standard deviation of the forecast errors in each of the three 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, the standard deviation of $\xi$ in column 2 is zero and is the same in column 3 as in column 1. The bottom panel reports the $R^2$ from regressing parent income and number of colleges listed on the remaining student covariates. This provides a sense of the degree to which other student covariates are able to serve as proxies for the FAFSA variables. No sample weights were used.

Table 8: Information Loss in the Counterfactuals
1, $359 in counterfactual 2, and $826 in counterfactual 3. In other words, restricting FAFSA information lowers elite colleges’ tuition revenues per student by up to 6 percent. The percentage of students who experience a price drop amounts to 70 percent, 68 percent, and 62 percent, respectively. Not only do prices fall, but their within-college variance also falls by 13 percent in counterfactual 1 and 19 percent in counterfactual 3. Due to the uncertainty introduced by restricting colleges’ information, some students, up to 12.5 percent, are misallocated and end up inefficiently attending nonelite colleges.

6.2.2 Welfare

What are the welfare consequences of restricting FAFSA information? As was just discussed, restricting FAFSA information lowers elite colleges’ tuition revenues. So making welfare comparisons requires us to take a stand on the social value of the marginal dollar of college spending. At least two possibilities suggest themselves: first, colleges themselves consume their marginal dollar of spending (the standard assumption); second, colleges’ marginal dollar of spending benefits society as a whole (perhaps because they supply a valuable public good).

The first case is more standard and is the one assumed in Table 9. If colleges consume their marginal dollar of spending, then a reduction in transaction prices would simply transfer surplus from colleges to students. The estimates indicate that average student surplus would rise by $655 in counterfactual 1, $318 in counterfactual 2, and $694 in counterfactual 3. However, total surplus per student would also fall somewhat due to the misallocation of students to the nonelite sector: $89 in counterfactual 1, $91 in counterfactual 2, and $198 in counterfactual 3. Although as many as 12.5 percent of students would be misallocated, total surplus per student would fall by less than 1.5 percent. This is because the students who would be misallocated tend to have low match surpluses anyway. Thus, if colleges consume their marginal dollar, restricting FAFSA information primarily transfers surplus from colleges to students with a relatively small loss in total surplus.

Now consider the case where colleges’ marginal dollar of spending benefits society as a whole. In this case, the welfare implications become ambiguous and nearly impossible to quantify. It is logically possible that colleges spend their marginal dollar on an extremely valuable public good, and that reducing tuition revenue will dramatically reduce social welfare and perhaps even make the stu-
## Counterfactual Estimates

<table>
<thead>
<tr>
<th></th>
<th>Baseline</th>
<th>Parent income</th>
<th>Number of colleges listed</th>
<th>All FAFSA info</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Levels</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Consumer (student) surplus per student</td>
<td>$4,877</td>
<td>$5,532</td>
<td>$5,195</td>
<td>$5,571</td>
</tr>
<tr>
<td></td>
<td>($4136, $4896)</td>
<td>($4691, $5563)</td>
<td>($4358, $5228)</td>
<td>($4795, $5599)</td>
</tr>
<tr>
<td>Total surplus per student</td>
<td>$15,413</td>
<td>$15,324</td>
<td>$15,322</td>
<td>$15,215</td>
</tr>
<tr>
<td></td>
<td>($13962, $15734)</td>
<td>($13877, $15654)</td>
<td>($13881, $15630)</td>
<td>($13770, $15532)</td>
</tr>
<tr>
<td>Of those who remain at elite colleges:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean student share of surplus</td>
<td>30.0%</td>
<td>37.2%</td>
<td>35.2%</td>
<td>36.9%</td>
</tr>
<tr>
<td></td>
<td>(27.9%, 30%)</td>
<td>(34.9%, 37.6%)</td>
<td>(32.3%, 35.7%)</td>
<td>(35%, 37.1%)</td>
</tr>
<tr>
<td>Mean transaction price</td>
<td>$13,158</td>
<td>$13,166</td>
<td>$13,497</td>
<td>$13,313</td>
</tr>
<tr>
<td></td>
<td>($12775, $13545)</td>
<td>($12785, $13597)</td>
<td>($13147, $14001)</td>
<td>($12896, $13719)</td>
</tr>
<tr>
<td><strong>Changes relative to baseline</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Consumer (student) surplus per student</td>
<td>$0</td>
<td>$655</td>
<td>$318</td>
<td>$694</td>
</tr>
<tr>
<td></td>
<td>(---)</td>
<td>($452, $777)</td>
<td>($133, $437)</td>
<td>($530, $844)</td>
</tr>
<tr>
<td>Total surplus per student</td>
<td>$0</td>
<td>-$89</td>
<td>-$91</td>
<td>-$198</td>
</tr>
<tr>
<td></td>
<td>(---)</td>
<td>($-149, $-51)</td>
<td>($-185, $-40)</td>
<td>($-283, $-129)</td>
</tr>
<tr>
<td>Percent of students who switch to a nonelite college</td>
<td>0.0%</td>
<td>8.9%</td>
<td>8.8%</td>
<td>12.5%</td>
</tr>
<tr>
<td></td>
<td>(---)</td>
<td>(7.4%, 12.5%)</td>
<td>(6.8%, 13.8%)</td>
<td>(11.2%, 16.5%)</td>
</tr>
<tr>
<td>Of those who remain at elite colleges:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean change in student share of surplus</td>
<td>0.0%</td>
<td>6.0%</td>
<td>3.9%</td>
<td>5.1%</td>
</tr>
<tr>
<td></td>
<td>(---)</td>
<td>(5.2%, 6.9%)</td>
<td>(2.5%, 4.9%)</td>
<td>(4.4%, 6.3%)</td>
</tr>
<tr>
<td>Mean change in transaction price</td>
<td>$0</td>
<td>-$730</td>
<td>-$359</td>
<td>-$826</td>
</tr>
<tr>
<td></td>
<td>(---)</td>
<td>($-900, $-500)</td>
<td>($-527, $-149)</td>
<td>($-1060, $-629)</td>
</tr>
<tr>
<td>Percent of students with price drop</td>
<td>0.0%</td>
<td>70.3%</td>
<td>67.8%</td>
<td>62.3%</td>
</tr>
<tr>
<td></td>
<td>(---)</td>
<td>(68%, 75%)</td>
<td>(60.4%, 75.6%)</td>
<td>(58.7%, 67.7%)</td>
</tr>
<tr>
<td>Within-college variance in price</td>
<td>38,812</td>
<td>-13.2%</td>
<td>-4.7%</td>
<td>-19.1%</td>
</tr>
<tr>
<td></td>
<td>(---)</td>
<td>(-17.5%, -4.8%)</td>
<td>(-11.7%, 3.1%)</td>
<td>(-26.4%, -11.2%)</td>
</tr>
</tbody>
</table>

Column 1 contains baseline estimates, while columns 2 through 4 contain estimates for the three counterfactuals. Point estimates are in bold. 95% percent confidence intervals in parentheses were calculated using 1,000 bootstrap replications. Dollar amounts are expressed in dollars per year. The percentage changes in rows 7 and 8 are in percentage points, while those in row 11 are percentage changes relative to a base of 38,812. No sample weights were used.

Table 9: Counterfactual Estimates
dents themselves worse off in the process.

The welfare implications of restricting FAFSA information hinge on the marginal social value of college spending, which is beyond the scope of this paper. If one believes that the marginal dollar of college spending has large social benefits, then restricting FAFSA information will significantly reduce total surplus and may even make students themselves worse off. On the other hand, if one believes that the marginal dollar of college spending has little or no social value and is instead consumed by the college, then the welfare analysis in Table 9 is correct—restricting FAFSA information primarily transfers surplus from colleges to students with a relatively small loss in total surplus.

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. 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. Figure 5 plots the fitted values 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.

25 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.
Comparing Baseline and Counterfactual Pricing Patterns

<table>
<thead>
<tr>
<th>Dependent variable: transaction price</th>
<th>Baseline</th>
<th>Parent income restricted</th>
<th>Colleges listed restricted</th>
<th>Entire FAFSA restricted</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parent AGI (in $10,000s)</td>
<td>220.1</td>
<td>80.6** (28.07) **</td>
<td>239.0 (28.74) ***</td>
<td>90.2 (28.29) **</td>
</tr>
<tr>
<td>ACT score</td>
<td>-196.6</td>
<td>-144.1** (49.40) **</td>
<td>-153.2 (51.27) **</td>
<td>-123.7 (50.46) *</td>
</tr>
<tr>
<td>High school GPA</td>
<td>-1087.1</td>
<td>-1193.9** (377.0) **</td>
<td>-1328.4 (394.5) ***</td>
<td>-1309.3 (384.3) ***</td>
</tr>
<tr>
<td>Earned AP credit in high school</td>
<td>-193.8</td>
<td>89.8 (392.0)</td>
<td>52.8 (406.5)</td>
<td>127.0 (397.6)</td>
</tr>
<tr>
<td>Completed FAFSA</td>
<td>-1638.1</td>
<td>-2383.6*** (436.3) ***</td>
<td>-2218.4 (447.6) ***</td>
<td>-433.7 (434.1)</td>
</tr>
<tr>
<td>Number of colleges listed on FAFSA</td>
<td>-527.9</td>
<td>-305.9** (113.6) **</td>
<td>-80.9 (118.3)</td>
<td>-96.2 (118.4)</td>
</tr>
<tr>
<td>Age as of 12/31/07</td>
<td>77.8</td>
<td>62.6 (251.6)</td>
<td>64.1 (253.4)</td>
<td>129.3 (255.6)</td>
</tr>
<tr>
<td>Female</td>
<td>-553.4</td>
<td>-541.1 (317.6)</td>
<td>-499.7 (330.3)</td>
<td>-561.6 (323.1)</td>
</tr>
<tr>
<td>Black</td>
<td>-2168.8</td>
<td>-2747.6*** (651.7) ***</td>
<td>-1964.5 (695.6) **</td>
<td>-2435.0 (675.0) ***</td>
</tr>
<tr>
<td>Hispanic</td>
<td>-1826.3</td>
<td>-2362.4*** (649.2) ***</td>
<td>-1318.5 (668.2) *</td>
<td>-1787 (658.1) **</td>
</tr>
<tr>
<td>Asian</td>
<td>-110.6</td>
<td>-112.5 (688.7)</td>
<td>476.6 (728.6)</td>
<td>89.7 (696.0)</td>
</tr>
<tr>
<td>Other / multiple</td>
<td>-865.1</td>
<td>-1203.1 (871.3)</td>
<td>-677.9 (880.6)</td>
<td>-814.7 (879.3)</td>
</tr>
<tr>
<td>Out-of-state, public</td>
<td>8722.5</td>
<td>8245.4*** (678.2) ***</td>
<td>8514.6 (702.0) ***</td>
<td>7369.8 (634.2) ***</td>
</tr>
</tbody>
</table>

College fixed effects | Yes | Yes | Yes | Yes

Observations | 2210 | 2010 | 2020 | 1930

R-squared | 0.744 | 0.754 | 0.755 | 0.742

All four regressions include students from the elite sample. The dependent variable in column 1 is transaction price (sticker minus discount). In columns 2 through 4, the dependent variable is transaction price plus the estimated price change for the student based on the structural estimates. In each counterfactual, students who are priced out of the elite sector are omitted from the regression. Robust standard errors are in parentheses. Sampling weights were used in all four regressions (NPSAS variable WTA000).

* p<0.05, ** p<0.01, *** p<0.001

Table 10: Comparing Baseline and Counterfactual Pricing Patterns
The fitted values plotted here come from three 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 includes students who were not priced out of the elite college market in the corresponding counterfactual. Parent adjusted gross income was included as a fourth-order polynomial.

Figure 5: Change in Observed 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, 12, and 13, I look at how the three counterfactuals differentially affect low-, middle-, and high-income students. In Table 11, I regress a student’s price change (relative to baseline) on her parent adjusted gross income. Restricting income information lowers prices for high-income students but raises them for low-income students. Looking at it the other way, colleges use the income information from the FAFSA to effectively levy a 1.6 percent tax on parent adjusted gross income combined with a $525 lump sum rebate. This implies that students with income of $32,700 are unaffected on average, those with higher incomes pay more, and those with lower incomes pay less than they would if colleges could not use the income information on the FAFSA. Colleges use the entire FAFSA to effectively levy a 2 percent tax rate on adjusted gross income combined with a $723 lump sum rebate, implying that those with incomes above $36,900 pay more while those below pay less than they would if colleges could not use the FAFSA at all. In contrast, colleges use the number of colleges listed on the FAFSA to price discriminate in a way that is income neutral. Figure 6 plots the fitted values from regressions like those in Table 11 but with a fourth-order polynomial in parent income. Qualitatively, the findings are unchanged; colleges use the income information from the FAFSA to effectively levy an income tax coupled with a rebate, but they use the number of colleges listed on the FAFSA to price discriminate in an income neutral way.

To what degree does price discrimination redistribute resources from higher income students to lower income students? Table 12 demonstrates how colleges use the FAFSA to charge middle- and higher-income students more and lower-income students less. For example, students in the bottom third of the income distribution pay on average $428 less per year than they would if FAFSA information were completely restricted. In contrast, those in the middle third pay $538 more and those in the upper third pay $2,312 more per year than they otherwise would. To further quantify the degree of redistribution, I calculate the total tuition revenue that colleges forgo from those, mostly lower-income, students who pay less in baseline and divide by the total tuition revenue that colleges raise from those, mostly higher-income, students who pay more in baseline. If colleges were price discriminating in a purely redistributionist way, this ratio would be 100 percent. Putting
it differently, if we think of price discrimination as a “tax,” how much of the “tax revenue” is being transferred to other students in the form of lower prices and how much is showing up as increased tuition revenues? The last line of Table 12 shows that less than half of the tax revenue is being transferred to other students. For instance, when looking at the FAFSA as a whole, colleges transfer 35 percent of the tax revenue to other students and keep the remaining 65 percent as additional tuition revenue.

### Table 11: Distributional Effects of Price Changes

<table>
<thead>
<tr>
<th>Dependent variable: Price change relative to baseline</th>
<th>Parent income restricted</th>
<th>Colleges listed restricted</th>
<th>Entire FAFSA restricted</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parent AGI (in $10,000s)</td>
<td>-160.7 (2.56) ***</td>
<td>-6.1 (4.72)</td>
<td>-195.9 (6.18) ***</td>
</tr>
<tr>
<td>Constant</td>
<td>525.3 (25.90) ***</td>
<td>-310.8 (48.29) ***</td>
<td>722.6 (63.05) ***</td>
</tr>
<tr>
<td>Observations</td>
<td>2010</td>
<td>2020</td>
<td>1930</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.662</td>
<td>0.001</td>
<td>0.342</td>
</tr>
</tbody>
</table>

All three regressions include students from the elite sample. The dependent variable is the change in the student’s transaction price relative to baseline. In each counterfactual, students who are priced out of the elite sector are omitted from the regression. Robust standard errors are in parentheses. Sampling weights were used for all regressions (NPSAS variable WTA000).

*p<0.05, ** p<0.01, *** p<0.001

We have just seen that only 35 percent of the tuition revenue colleges rise from price discrimination is actually transferred to other students through lower prices. But how precise is that redistribution? As illustrated in Table 13, although lower-income students benefit from price discrimination on average, many do not. At best, only a slight majority of students in the bottom third of the income distribution enjoy lower prices because of price discrimination. Moreover, some middle- and higher-income students actually benefit from price discrimination. Table 13 illustrates 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.

### 6.2.5 Potential Effects on Students’ Application Behavior

When simulating the counterfactuals, I hold students’ application behavior constant. However, the results indicates that students would receive a larger share of
### Price Change Relative to Baseline by Income Group

<table>
<thead>
<tr>
<th>Parent income</th>
<th>Number of colleges listed</th>
<th>All FAFSA info</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bottom third</td>
<td>$212</td>
<td>-$242</td>
</tr>
<tr>
<td>Middle third</td>
<td>-$493</td>
<td>-$361</td>
</tr>
<tr>
<td>Top third</td>
<td>-$1,908</td>
<td>-$471</td>
</tr>
</tbody>
</table>

Percent of "tax revenue" transferred to other students: 17.0% 45.8% 35.4%

When colleges can no longer use the FAFSA to price discriminate, some students see their prices rise, relative to baseline, while others see their prices fall. Each cell in the first three rows reports the average change in price for students in the corresponding tercile of the distribution of parent adjusted gross income. The final row reports the change in price for those who see their prices rise, divided by the change in price for those who see their prices fall. This measures the degree to which colleges use FAFSA information to price discriminate in a way that redistributes money from some students to others, versus simply boosting tuition revenues.

Table 12: Price Change Relative to Baseline by Income Group

### Students Seeing Price Rise Relative to Baseline by Income Group

<table>
<thead>
<tr>
<th>Parent income</th>
<th>Number of colleges listed</th>
<th>All FAFSA info</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bottom third</td>
<td>58.6%</td>
<td>34.1%</td>
</tr>
<tr>
<td>Middle third</td>
<td>20.4%</td>
<td>28.4%</td>
</tr>
<tr>
<td>Top third</td>
<td>2.3%</td>
<td>25.6%</td>
</tr>
</tbody>
</table>

When colleges can no longer use the FAFSA to price discriminate, some students see their prices rise, relative to baseline, while others see their prices fall. Each cell reports the percentage of students who would see their prices rise in the corresponding tercile of the distribution of parent adjusted gross income.

Table 13: Percent of Students Seeing Price Rise Relative to Baseline by Income Group
The fitted values plotted here come from three separate regressions. In all three regressions, the dependent variable is the estimated change in price relative to baseline and the independent variable is parent adjusted gross income (included as a fourth-order polynomial). No other covariates were included.

Figure 6: Change In Price When FAFSA Info Restricted
the match surplus which would strengthen their incentive to apply to more colleges. In Table 14, I recalculate the expected return from applying to an additional college (again assuming that the student applies to colleges randomly). Because colleges are bidding more aggressively, relative to baseline, the return to applying rises by $155, $34, and $255 in the three counterfactuals. If students respond to these incentives by applying to more colleges, then the increased competition will further lower prices as colleges are forced to bid more aggressively for students. This suggests that the estimates might be understating the full effects of restricting colleges’ use of FAFSA information.

### Average Return From Applying to An Additional College

<table>
<thead>
<tr>
<th>FAFSA information restricted</th>
<th>Number of colleges listed</th>
<th>All FAFSA info</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline Parent income</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$693</td>
<td>$848</td>
<td>$727</td>
</tr>
<tr>
<td>$848</td>
<td></td>
<td>$948</td>
</tr>
</tbody>
</table>

Each cell reports the average return from applying to an additional college under each counterfactual. Returns represent the expected utility gain (in dollars) if the new college beats the student's current best option, multiplied by the probability that this occurs. The returns were calculated under the assumptions that a) the number of colleges listed on the FAFSA remains fixed and b) students apply to colleges randomly. If, as seems likely, students tend to apply to colleges that are a better match first, these estimates provide an upper bound on the average returns to applying to an additional college. No sample weights were used.

Table 14: Average Return From Applying to an Additional College

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.
The model provides several predictions about college prices. All else being equal, colleges will charge students less if they are more desirable to the college and more if they have a higher willingness to pay. Colleges will also charge a student less when they believe they face more competitors. I test these predictions by regressing a student’s tuition discount on several student characteristics and include college fixed effects to isolate variation among students in the same school. At elite colleges, 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, which are consistent with the predictions of the model, emerge primarily among private and very selective public colleges. I label these as “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 three counterfactuals: 1) restrict income information only, 2) restrict the number of colleges listed on the FAFSA only, and 3) restrict income information, the number of colleges listed, and whether the student completed the FAFSA. My estimation results highlight an important policy tradeoff—restricting colleges’ use of FAFSA information lowers average prices but also leads to a mis-allocation of students. The intuition is that colleges will not be able to tailor their prices as precisely if the government restricts some or all of the FAFSA. As a result, elite colleges will sometimes overcharge a student and cause her to (inefficiently) attend a nonelite college. Depending on the counterfactual, I estimate that between 8.8 percent and 12.5 percent of students who currently attend elite colleges would end up attending nonelite colleges if we restricted some or all of the FAFSA. These inefficient matches lower total surplus, although this lost surplus is less than one-third the size of the lowered prices to students. 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. Sharing the FAFSA with colleges enables them to price discriminate in a way that amounts to a 2 percent tax on adjusted gross income coupled with a $723 lump sum rebate, so that students with parent adjusted gross income below $36,900 are helped, while those with higher incomes are harmed. Although this yields a de-
gree of redistribution, only 35 percent of the “tax revenue” that colleges raise is transferred to other students in the form of lower prices. The remaining 65 percent accrues to colleges in the form of higher tuition revenues. Taken as a whole, the results indicate that although allowing colleges to use FAFSA information does increase efficiency somewhat and does lower prices for some students, its main effect is to boost tuition revenue, primarily at the expense of middle- and high-income students.
References


Appendices

Appendix A  Data Sources and Variable Creation

A.1  National Postsecondary Student Aid Study 2007–2008 (NPSAS)

The National Postsecondary Student Aid Study (NPSAS) is a nationally representative cross-section of college students in the United States. The survey collects data from students on many aspects of the college experience, with a particular focus on understanding how students pay for college. A new wave of NPSAS is collected every three to four years. I use the 2008 wave which surveyed students during the 2007–2008 school year. 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.

A.1.1  Raw Data

NPSAS is a restricted-use dataset obtained from the National Center for Education Statistics (NCES). I obtained access to the data through a data use agreement which prohibits publishing the data. The data comes in the form of several flat text files along with an electronic codebook. The electronic codebook contains the data documentation and allows the user to select desired variables. Once the user has selected the desired variables, the electronic codebook will produce an SPSS script that will read in and format the data. From SPSS, the data can be saved in several formats including Stata (.dta) format.

A.1.2  Variable Creation

Not all students attend college full-time for the full year. I define the number of full-time equivalent months to be equal to the number of full-time months plus half of the number of half-time months plus one-quarter of the number of less-than-half-time months.

The college selectivity variable used throughout the paper is based on a classification developed by the National Center for Education Statistics. The methodology is described in Appendix E of Cunningham (2005). Colleges are assigned to selec-
tivity categories based on a few characteristics. Two-year colleges are put into their own category as are open admission four-year colleges.

For non-open admission institutions, an index was created from two variables: 1) the centile distribution of the percentage of students who were admitted (of those who applied); and 2) the centile distribution of the midpoint between the 25th and 75th percentile SAT/ACT combined scores reported by each institution (ACT scores were converted into SAT equivalents). The two variables were given equal weight for those non-open admission institutions that had data for both, and the combined centile variable was divided into selectivity categories—very selective, moderately selective, and minimally selective—based on breaks in the distribution. Institutions that did not have test score data (about 10 percent of non-open admission institutions) were assigned to the selectivity categories using a combination of percent admitted and whether they required test scores; institutions that did not require test scores were assigned to the “minimally selective” category, while the remainder were assigned according to the range of centiles of “percent admitted” in which they fell. (page E-1)

Table E-1 from Cunningham (2005), reproduced here for reference, provides examples of the types of colleges that fall into each selectivity category. I collapse the “Minimally Selective” and “Open Admission” categories into a single category I label “Not Selective.”

I use the NPSAS variable TUITION2 as my measure of gross tuition or sticker price. Discounts come from the institutional grants variable INGRTAMA. Transaction prices are simply the difference between the sticker price and discount. I adjust all prices and discounts to be in terms of 9 full-time equivalent months (those with less than 9 months were dropped, see below).

The number of additional colleges listed on the FAFSA comes from the NPSAS variables C08100, C08102, C08104, C08106, C08108, C08110. For each student, I added up the number of non-missing entries in these variables and subtracted one. The result was a number between 0 and 5. The variable was set to zero for students who did not complete the FAFSA. The variable was set to missing for those who did complete the FAFSA but had no colleges listed (as a result, these students were excluded from the analysis).
types of selectivity measures such as Peterson's Selectivity Ranking. The selectivity variable appeared to assign institutions to categories in ways that would be expected (table E-1).

The Peterson's Ranking was available for 1,093 of the 1,569 4-year institutions in the study universe. All of the tests had significant Chi Squares and Pearson's r values.

<table>
<thead>
<tr>
<th>Public institutions</th>
<th>Very Selective</th>
<th>Moderately selective</th>
<th>Minimally selective</th>
<th>Open admission</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cornell University</td>
<td>Ball State University</td>
<td>Black Hills State University</td>
<td>University of Kentucky</td>
<td>University</td>
</tr>
<tr>
<td>SUNY-Binghamton</td>
<td>Ohio State University</td>
<td>University of Northern Alabama</td>
<td>Texas Southern University</td>
<td>University of Toledo</td>
</tr>
<tr>
<td>University of Virginia</td>
<td>University of Oregon</td>
<td>University of Winston-Salem State University</td>
<td>University of Toledo</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Private not-for-profit institutions</th>
<th>Very Selective</th>
<th>Moderately selective</th>
<th>Minimally selective</th>
<th>Open admission</th>
</tr>
</thead>
<tbody>
<tr>
<td>Duke University</td>
<td>DePaul University</td>
<td>Cabrini College</td>
<td>University of Rio Grande</td>
<td></td>
</tr>
<tr>
<td>Princeton University</td>
<td>Mary Baldwin College</td>
<td>Wayland Baptist University</td>
<td>Pikeville College</td>
<td></td>
</tr>
<tr>
<td>University of Virginia</td>
<td>University of San Francisco</td>
<td>Albertus Magnus College</td>
<td>Rochester College</td>
<td></td>
</tr>
<tr>
<td>Williams College</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table E-1. Selected 4-year institutions in the study universe, by institutional selectivity

NOTE: Selected institutions are in no particular order.

Table 15: Selected 4-year Institutions, by Institutional Selectivity

High school GPA was reported in ranges in the NPSAS variable HSGPA. Therefore I assigned each student a high school GPA equal to the minimum of the range which she reported.

I collapsed the NPSAS variable RACE into five categories “white,” “black”, “Hispanic”, “Asian”, and “other/multiple.”

A.1.3 Sample Selection

I applied the following sample selection criteria

- Undergraduates only
- Exclude students on athletic scholarship
- Exclude foreign students and those from Puerto Rico
- Exclude students who attended multiple colleges during the 2007–2008 school year
- Exclude students not enrolled in a degree program
• Exclude students who have already earned a bachelor’s degree
• Exclude students who received tuition waivers because of their parent’s employment at the college
• Exclude students who had imputed tuition data from IPEDS
• Exclude students who completed a FAFSA but were missing FAFSA information
• Exclude students over age 30
• Exclude students who attended less than 9 full-time equivalent months
• Exclude colleges outside the 50 states plus Washington, D.C.
• Exclude for-profit colleges
• Exclude less-than-2-year colleges
• Exclude specialized colleges (but keep engineering and business schools)
• Exclude “special use” two-year colleges

In addition, there appears to be some misreporting of tuition prices. It seems that some colleges reported net tuition when they were supposed to report gross tuition. I drop students at colleges where the average transaction price (reported gross tuition minus institutional grants) is not positive.

For most of the reduced form analysis, I further restrict the sample to dependent freshmen. In some cases, I also restrict the analysis to students at elite or nonelite colleges.

A.2 Beginning Postsecondary Students 2003-2009 (BPS)

Beginning Postsecondary Students 2003–2009 is a panel dataset that follows first-time freshmen that were sampled as part of the 2003 wave of NPSAS. Because BPS is derived from NPSAS, it contains much of the same information. However, because BPS is focused on freshmen, it also specifically asks about the number of colleges applied to (BPS variable APPS04) which I top-coded at 10. The raw data, variable creation, and sample selection for BPS were identical to that for NPSAS.
Appendix B  Estimating Conditional Distributions

I employ the following method, described in detail in Fillmore (2014), to estimate conditional distributions. Suppose we are interested in estimating $F_{Y|X}$ and $f_{Y|X}$, the cdf and pdf of the random variable $Y$ conditional on a set of covariates $X$, where $Y$ has support $[y, \bar{y}]$. Begin by estimating a series of quantile regressions of the form

$$F_{Y|X_i}^{-1}(p) = X_i \beta^p$$

for a grid of values for $p$. For instance, the grid could be $p = 0.05, 0.10, 0.15, \ldots, 0.95$. For each observation $i$, calculate the fitted value from each quantile regression $\hat{y}_i^p$. This is an estimate of the $p$th quantile of the random variable $Y|X = X_i$. For each observation $i$, collect the points $\{(\hat{y}_i^p, p)\}$ and append the point $(\min y, 0)$ to the beginning and the point $(\max y, 1)$ to the end. Then take these points—in our example 21 of them—and fit a smooth monotone curve through them using the techniques described in Ramsay (1998), while restricting the curve to pass through the first and last point, which guarantees that it is a proper cdf. Its derivative is then an estimate of the pdf. Perform this procedure for each observation $i$ in the data set. This produces estimates of $F_{Y|X_i}$ and $f_{Y|X_i}$ for each observation.

The method outlined here is semiparametric. The second step, fitting the smooth monotone curve, is flexible enough to fit any monotone twice differentiable distribution function (Ramsay 1998). The first step, fitting the quantile regressions, can also be quite flexible, depending on how one specifies the regressions. Thus, the flexibility of the method as a whole is really driven by the flexibility of the quantile regression specification. However, note that even if we choose a linear specification, as I do in the paper, the method still allows the vector of coefficients $\beta^p$ to differ for each value of $p$.

Appendix C  Simulating the Counterfactuals

In order to simulate the counterfactuals, I must estimate the forecast error terms $e_i$ and $\xi_i$. I use the following procedure:

1. Regress $\hat{\sigma}_{ij}$ on student covariates $X_i$ and store the fitted values $\hat{\sigma}_{ij}^1$. Then regress $\hat{\sigma}_{ij}$ on $\bar{X_i}$, the reduced set of student covariates, and store the fitted values $\hat{\sigma}_{ij}^2$. Calculate $e_i = \hat{\sigma}_{ij}^1 - \hat{\sigma}_{ij}^2$. 

59
2. Repeat step 1 for $\hat{w}_{ij}$ to estimate $\xi_i = \hat{w}_{ij}^1 - \hat{w}_{ij}^2$.

3. Calculate $\tilde{s}_{ij} = \tilde{s}_{ij} - e_i - \xi_i$.

4. Estimate $F_e$ and $f_e$ from the $e_i$ using the same method described in section [B].

5. Estimate $G_{\tilde{s}|\tilde{X}_i}$ and $g_{\tilde{s}|\tilde{X}_i}$ from the $\tilde{s}_{ij}$ using the method from section [B]. Remember that this is the distribution of winning (counterfactual) match surpluses.

6. Solve for $F_{\tilde{s}|\tilde{X}_i}$ and $f_{\tilde{s}|\tilde{X}_i}$, the parent distribution of (counterfactual) match surpluses.

7. Solve for the counterfactual equilibrium bidding function, $\tilde{\beta}(\cdot|\tilde{X}_i)$, using Equation (10).

8. Calculate the observed (to the college) winning bid, $\tilde{u}_{ij} = \tilde{\beta}(\tilde{s}_{ij}|\tilde{X}_i)$.

9. Calculate the true winning bid, $\tilde{u}_{ij} + e_i$. If $\tilde{u}_{ij} + e_i < 0$, then the student switches to a nonelite college and receives zero utility. Otherwise, the student remains at her college and pays a price equal to $v_{ij} - (\tilde{u}_{ij} + e_i)$.

**Appendix D  Modeling College Spending by Incorporating Dynamics**

When colleges price discriminate, they are able to earn more revenue than if they charged a uniform price. If the colleges invest this additional revenue into improving educational quality, then price discrimination may directly improve the quality of education for all students (although as Peña (2010) points out, the colleges may capture some of that increased consumer surplus by raising prices). In this section, I generalize the model to explicitly incorporate both tuition revenues and alumni giving. I show that adding these features to the model does not fundamentally alter the college’s first order condition.

The model as presented in the paper is silent about how colleges use their tuition revenues. Modeling college spending on quality is tricky. For instance, if I assume that colleges use today’s revenues on today’s quality, then the problem becomes quite complicated, because a given student’s demand for college $j$ will depend on her beliefs about the classmates she will have—wealthier classmates...
translate into more tuition revenue and thus a better education. But if I adopt an alternative, and perhaps more realistic, timing assumption, then the model remains tractable. I assume that colleges must purchase all of their quality inputs (faculty, facilities, etc.) one period ahead. This timing assumption means that the quality of college \( j \) will not depend on today’s tuition revenue. Rather, the college’s quality will depend on tuition revenue from last period. Thus, a college’s quality is fixed and common knowledge when student \( i \) is considering whether to attend.

There are a finite number \( I \) of student types indexed by \( i \). Each type contains a mass of students. Students live for two periods. In the first period, they decide which college to attend. In the second they make a donation to their alma mater. Let \( d_{ij} \) be the average donation made by type \( i \) students to college \( j \).

Colleges are infinitely lived. College \( j \) enters the period with a stock of quality \( Q \), an endowment \( E \), and a set of alumni. Let \( a_{ij} \) denote the mass of type \( i \) alumni from college \( j \). The demand for college \( j \) among type \( i \) students is given by \( q_{ij}(p_{ij}; Q) \). Demand depends on both the price offered by the college as well as the college’s quality. The college makes price offers to each student type and collects its tuition revenue from the students that enroll. In addition, the college receives a payoff of \( w_{ij} \) for each type \( i \) student that enrolls. \( w_{ij} \) represents the contribution that type \( i \) students make on campus minus the direct costs of enrolling the student and the opportunity costs of diverting resources away from other activities such as research. \( w_{ij} \) could be positive or negative, depending on the student and college.

The college’s quality and endowment evolve according to the laws of motion

\[
Q' = (1 - \delta)Q + p_zZ \\
E' = (1 + r) \left( E - Z + \sum_{i=1}^{I} (p_{ij} + S_{ij})q_{ij}(p_{ij}; Q) + \sum_{i=1}^{I} d_{ij}w_{ij} \right).
\]

\( \delta \) is the depreciation rate on the stock of college quality. If all quality inputs must be repurchased every year, then \( \delta \) will be one, but if some quality inputs behave more like capital, then \( \delta \) will be less than one. \( Z \) represents the college’s investment in quality, and \( p_z \) is the price of those investments. \( r \) is the interest rate earned on the

\( ^{26} \)Types \( i \) and \( i' \) need not have the same mass of students. Furthermore, the mass of all student types combined need not be one, although it could easily be normalized to one.
endowment. $S_{ij}$ represents a per-student subsidy that the college receives, and is probably most relevant for public colleges if state governments tie appropriations to enrollment levels—especially in-state enrollment.

In order to keep things simple, I will assume that $Z$ is chosen by an external group called the Board of Trustees, and the college takes $Z$ as given when it makes its price offers. This assumption is relatively innocuous because I will be focusing on the steady state of the model, where $Z$ would not differ even if it were chosen by the college.

As mentioned above, the college receives a (possibly negative) payoff $w_{ij}$ for each student it enrolls. It also receives payoff $g(Q)$ from having quality $Q$. The function $g$ is increasing and concave. The Bellman equation for college $j$ is

$$V(Q, E, a_{1j}, \ldots, a_{ij}) = \max_{\{p_{ij}\}} \sum_{i=1}^{I} w_{ij} q_{ij}(p_{ij}; Q) + g(Q) + \beta V(Q', E', q_{1j}(p_{1j}; Q), \ldots, q_{ij}(p_{ij}; Q)),$$

subject to the laws of motion (11) and (12). In the steady state, $Q' = Q$ and $E' = E$ so that

$$Z_{ss}^* = \frac{r}{1 + r} E_{ss}^* + \sum_{i=1}^{I} (p_{ij} + S_{ij}) q_{ij}(p_{ij}; Q_{ss}^*) + \sum_{i=1}^{I} d_{ij} a_{ij}.$$

That is, in the steady state, the college invests all of its tuition revenue and alumni donations, along with any interest income, into college quality. The first order condition with respect to $p_{ij}$ for a college in the steady state is

$$0 = w_{ij} q_{ij}'(p_{ij}; Q) + p_{z} \beta V_{Q} \times \left( (p_{ij} + S_{ij}) q_{ij}(p_{ij}; Q) + q_{ij}(p_{ij}; Q) \right) + \beta V_{a_{ij}} q_{ij}'(p_{ij}; Q),$$

which can be simplified to

$$p_{ij} = \frac{-w_{ij} - \beta V_{a_{ij}}}{p_{z} \beta V_{Q}} - S_{ij} + \frac{q_{ij}(p_{ij}; Q)}{q_{ij}'(p_{ij}; Q)}.$$

Substituting in the Euler condition $V_{a_{ij}} = p_{z} \beta V_{Q} d_{ij}$ and using the fact that we are in the steady state (so that $V_{Q}$ is the same every period), we get

$$p_{ij} = \frac{-w_{ij}}{p_{z} \beta V_{Q} c_{ij}} - \beta d_{ij} - S_{ij} + \frac{q_{ij}(p_{ij}; Q)}{q_{ij}'(p_{ij}; Q)}.$$
Note that this first order condition for the college has the same form as in the static model. The only difference is in the components of the willingness-to-receive term $c_{ij}$. The college simply factors the net present value of a student’s alumni giving and the benefits of her tuition revenue into its willingness to receive. The college has a lower willingness to receive—it is willing to charge less—for students who are expected to make larger future donations or who represent a larger per-student subsidy to the college. At colleges with a higher marginal value of quality, willingness to receive will be less sensitive to the college’s payoff $w_{ij}$. Intuitively, if a college desperately needs money to invest in quality improvements, then it can’t afford to be picky about the type of students it enrolls and will be willing to charge a low price to any student (as long as the total dollar revenue from the student remains nonnegative).

In this section, I have extended the model to incorporate tuition revenues as well as alumni giving. Both of these features affect the college’s pricing decision by altering the willingness-to-receive term $c_{ij}$. The net present value of future donations directly lowers $c_{ij}$. In the steady state, tuition revenues are spent entirely on college quality, and the marginal value (to the college) of quality shows up in the denominator of $\frac{w_{ij}}{p_i z_\beta V Q}$. The takeaway lesson here is that alumni giving and tuition revenues both alter the components of $c_{ij}$ but do not alter the way it is estimated. In any case, $c_{ij}$ can always be interpreted as the lowest price college $j$ is willing to accept from students of type $i$.

**Appendix E Matching Models**

Matching models represent a popular framework for thinking about the college market. More than a half-century ago, Gale and Shapley (1962) proposed their famous deferred-acceptance algorithm for understanding how students get matched with colleges. Twenty years later, Kelso and Crawford (1982) extended their model to allow for transfers between students and colleges. The model of Kelso and Crawford is quite general and has been used in a variety of contexts. A whole literature has emerged, estimating matching models from data on matches between teachers and jobs (Boyd et al. 2013), husbands and wives (Choo and Siow 2006; Hitsch et al.)

$q_{ij}$ is analogous to $P[i \text{ chooses } j]$, and the first order conditions from the static model in the paper and the dynamic model presented here have the same form, $(p - c)q' = q$.  

63
and companies and venture capitalists (Sørensen 2007), to name a few examples. Fu (2014) uses a matching framework to analyze the U.S. college market. Why then do I break with precedent and adopt an auctions framework?

The first answer is that auctions and matching models are not completely unrelated. Kelso and Crawford even frame their model in terms of a set of simultaneous auctions. So the difference is not as stark as it first appears. One can think of an auction as a mechanism for matching objects and bidders, just as the deferred-acceptance algorithm is a mechanism for matching students and colleges.

Yet despite their similarity, there are some important, if subtle, differences between auction and matching models. When estimating an auction model, the baseline assumption is that the data were generated by a Bayes-Nash equilibrium—given the bidding strategies of all the other bidders, bidder $j$ bid optimally. In contrast, when estimating a matching model, the baseline assumption is that the data were generated by a stable matching. Think of a matching as a rule assigning each student to one college (or no college at all). There are many possible matchings, and a particular matching is stable if it is impossible for a student and a college who are currently unmatched to abandon their assigned matches, match with each other, and be better off. Of course it is possible, even likely, that in a stable matching a student will prefer a college other than the one to which she is assigned. But in that case the college will be unwilling to reciprocate.

For a wide variety of auction formats, a monotone Bayes-Nash equilibrium exists (Athey 2001), and the structural primitives of the model are identified by combining bid data with restrictions implied by that equilibrium. In contrast, the empirical matching literature combines data on matches (and possibly transfers if they are observed) with the restrictions implied by stability to partially identify the structural parameters of the model—rather than a point estimate, researchers obtain a set of equally plausible parameters. The hope is that this set of parameters is small enough to still be useful.

Auction models directly model how prices (bids) are formed, while matching models think of the price as a transfer between a student and her college. Matching models do not directly predict what prices will be, nor do they explain the prices that do arise. Rather, matching models predict a range of prices that satisfy the incentive constraints implied by stability. The actual realized price depends

\[28\] A college may be matched with multiple students.
on some unmodeled feature (such as bargaining between the student and her college). In short, matching models are focused on understanding who matches with whom—any transfers that happen to occur are of secondary importance. Since the objective of this paper is to understand prices themselves, I choose to adopt an auction framework.