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

Ian Fillmore*†

November 8, 2014

Abstract

In the United States, a form called the Free Application for Federal Student Aid, or FAFSA, is used to determine eligibility for federal aid. The FAFSA collects extensive financial information, checks it for accuracy against several government databases, including the IRS, and then shares the information with colleges. I demonstrate that sharing the FAFSA with colleges enables them to engage in substantial price discrimination with widespread repercussions for the cost of a college education as well as the equilibrium sorting of students into colleges. I build a structural model of college pricing and price discrimination, and show that the model is identified from student-level data on prices and student characteristics. Reduced form estimates are consistent with several predictions of the model. According to my structural estimates, on average elite colleges capture 70% of the student-college match surplus through their student-specific prices. Withholding FAFSA information would lower prices for middle- and high-income students while raising them for low-income students. On average prices would fall by $825, and the within-college price variance would also drop by 17%. By using the FAFSA to price discriminate, elite colleges effectively levy a 1.9% tax on adjusted gross income coupled with a $709 lump sum rebate. However, with less information to use when price discriminating, they would inefficiently price up to 12% of students out of the elite market. My findings highlight a policy tradeoff between increasing the welfare of middle- and high-income students at the expense of total efficiency, college welfare, and the welfare of low-income students.

*Contact Information: ianfillmore@uchicago.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 and Brigham Young University. Kevin Brown and NORC at the University of Chicago provided indispensable help in obtaining the restricted-use data. 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).
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 up to six colleges where the student is considering attending. The government converts this FAFSA information into an “expected family contribution” (EFC) which is used to determine eligibility for federal aid. A student’s EFC depends on many factors and represents the government’s assessment of what that student’s family should be expected to contribute toward her education. The government applies a fixed formula to the student’s EFC to determine her eligibility for federal aid, but it does not directly dispense any of the aid itself. Rather the government enlists colleges as partners in distributing federal dollars to college students.

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. For instance, if a college has a posted sticker price of $20,000 per year, but it offers the student a $15,000 discount, the relevant transaction price is not $20,000 but $5,000. Each college offers a similar “price quote” to the student, and she chooses the college that makes her the most attractive offer with price being one component of that offer.

Why would colleges care about the FAFSA? Intuitively, 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. It is high-quality because the government imposes penalties, in the form of fines or jail time, for misreporting. Perhaps even more importantly

---

1 See Tables 353 and 355 of the 2011 Digest of Education Statistics.
2 Most high school seniors complete the FAFSA at the same time as their college applications. A primary reason for doing this is so they can compare price offers when choosing which college to attend.
3 Research by van der Klaauw (2002) demonstrates that these discounts are indeed effective at attracting students.
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 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. 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.

Economic theory tells us that a seller must have some information about a buyer’s willingness-to-pay in order price discriminate. Do colleges use the FAFSA to price discriminate? If so, what would happen if we withheld some or all of the FAFSA information from colleges? Using student-level data from the 2008 wave of the National Postsecondary Student Aid Study, I provide reduced form evidence that colleges make extensive use of the FAFSA when offering discounts. Then I estimate a structural model to simulate counterfactuals wherein some or all of the FAFSA is withheld from 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 “wins”—the student enrolls—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 if it increases its chances of attracting a particularly desirable student. Within this model, a college may have market power for two reasons: i) if the college knows it faces few competitors for a given student then it does not need to make as generous an offer, and ii) 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 auction literature. 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. The model highlights how the preferences of students and colleges both determine prices, but still allows me to remain fairly agnostic about precisely what students and colleges value. The model predicts that colleges will charge students less if they are more valuable to colleges (for instance, those with higher test scores), all else equal. But colleges will charge students more if those students place a high value on attending the college, all else equal. 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.

In the reduced form analysis I test these predictions by regressing tuition discounts on student

---

4It appears to be public knowledge that many colleges verify all of their FAFSA forms Grant 2006; Weston 2014.
characteristics. I include college fixed effects to isolate variation among students at the same college. Higher-income students tend to receive smaller discounts (pay higher prices), and this effect is driven predominantly by private and very selective public colleges. Higher ability students—as measured by test scores and high school GPA—tend to receive larger discounts. Students who complete the FAFSA receive substantially larger discounts, which is not surprising since most colleges require a student’s FAFSA before considering a discount. 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, and this effect appears to be driven predominantly by private colleges. Moreover, if we control for the number of colleges the student actually applied to, the coefficient on the number listed on the FAFSA remains unchanged while the coefficient on the number of actual applications is effectively zero. The number of colleges listed on the FAFSA is unrelated to a student’s discount if she completes the FAFSA after the college application 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 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 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 find that elite colleges successfully capture an average of 70% of the total match surplus through their student-specific prices, leaving the remaining 30% to the students. Students who list more colleges on the FAFSA reap the benefit of intensified competition; those who list six colleges receive an average of 42% of the surplus while those who list only one college receive an average of 18%. On average, students at elite colleges value attending their current college $18,380 more than their outside option of attending a non-elite college.

Using my structural estimates, I simulate a policy change whereby some or all of the FAFSA information is withheld from colleges. For simplicity, 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. I simulate three counterfactuals: a) the government withholds income information only, b) the government withholds the signal about competitors only, and c) the government withholds income information, the signal about competitors, and whether the student even completed the FAFSA at all. In all three counterfactuals, I assume that colleges can always see basic demographic characteristics such as age, gender and race as well as measures of student quality like ACT score and high school GPA, and that colleges use these characteristics to proxy for the lost FAFSA information to the extent possible. In the counterfactuals I hold the application strategy of students fixed. I find that all three policy changes would improve student welfare by lowering prices on average by as much as $825. However, with less information to use when price
discriminating, colleges would inefficiently price up to 12% of students out of the elite market. With less information, colleges 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 non-elite college. These inefficient matches lower total welfare, even though student welfare as a whole rises. Moreover, although all three policy changes raise student welfare, they each have different distributional consequences.

If we withheld income information only, 72% of students would experience a drop in price, prices would fall by an average of $743 per year, and 8.4% of students currently attending elite colleges would be priced out of the market and opt to attend a non-elite college. Prices would fall for middle- and high-income students. Low-income students, those with adjusted gross income below $32,300 would actually see prices rise on average.

If we withheld the signal about number of competitors, 67% of students would experience a drop in price, prices would fall by an average of $382, and 11.3% of students would switch from elite to less-selective colleges. Compared with the counterfactual of withholding income information, withholding the number of colleges listed on the FAFSA would result in smaller gains for students and larger losses in efficiency. However, in contrast with withholding income information, this counterfactual does not disproportionately hurt low-income students.

Finally, if we withheld all the FAFSA information, 62% of students would experience a drop in price, prices would fall by an average of $825, and 11.9% of students would switch from elite to less-selective colleges. Withholding all FAFSA information results in the largest gains for students and the largest losses in efficiency, although the efficiency losses are not substantially more than those that occur from withholding the number of colleges listed on the FAFSA. Just as when we withhold income information only, prices fall for middle-and high-income students. Students with adjusted gross income below $36,670 would on average see their prices rise.

Each of these counterfactuals is instructive in different ways. Of the three, withholding the signal about competitors represents the least ambitious policy proposal. Withholding the list of competitors would be a relatively straightforward modification that would lower prices by an average of 3% while also pricing roughly one in ten students out of the elite college market. This counterfactual demonstrates the value to colleges of obtaining information about their competition. Withholding income information illustrates the degree to which colleges rely on the income information on the FAFSA to price discriminate. In other words, this counterfactual demonstrates the consequences of giving colleges access to detailed and reliable income information. By using income information to price discriminate, colleges effectively levy a 1.6% tax on adjusted gross income coupled with a $524 lump sum rebate. Withholding all FAFSA information completely cuts colleges out of the information loop and demonstrates the consequences of choosing to enlist colleges as partners in administering federal student aid. By using the entire FAFSA to price discriminate, colleges effectively levy a 1.9% tax on adjusted gross income coupled with a $709 lump sum rebate. Because I hold the application
strategies of students fixed, my estimates likely understate the full drop in price that would result in each counterfactual because students will tend to apply to more colleges when the marginal benefit of applying rises. A rough calculation based on my estimates implies that the benefits of applying to an additional college rise by $50 to $250 depending on the counterfactual. If students are induced to apply to more colleges, then these additional applications would in turn serve to intensify competition, further lowering prices. My findings highlight that withholding FAFSA information presents a policy tradeoff between increasing the welfare of middle- and high-income students (by lowering their prices) while lowering total efficiency (by misallocating some students), college welfare (by lowering tuition revenue), and the welfare of low-income students (by raising their prices).

The paper proceeds as follows. Section 2 reviews the previous literature on both government aid and discounts and discusses how this paper both fits into and differs from previous research. Section 3 presents a structural model of college pricing and price discrimination and discusses the intuition and predictions provided by the model. Section 4 tests these predictions using student-level data on tuition discounts. Section 5 proves that the structural model is non-parametrically identified, outlines an empirical strategy, and presents the baseline estimates. Section 6 models the counterfactuals and presents the counterfactual estimates. Section 7 offers concluding thoughts and directions for further work.

2 Review of the Literature

A large literature estimates how government student aid affects student outcomes. Several papers estimate the effect of financial aid on the choice to attend college (Dynarski 2000, 2003; Kane and National Bureau of Economic Research 2003; Manski and Wise 1983; Seftor and Turner 2002). Others look at the effect of financial aid on longer run outcomes such as whether students persist in college, the number of courses they complete, and whether they graduate (Bettinger 2004; Castleman and Long 2013; Dynarski 2003; Goldrick-Rab et al. 2012).

A related set of papers have estimated the effect of tuition discounts on a student’s choice of college (van der Klaauw 2002; Terry Long 2004). Overall, these papers have found that discounts are effective at persuading students to alter their college choices.

Several economists have interpreted tuition discounts as price discrimination (Dynarski 2002; Lawson and Zerkle 2006; Tiffany and Ankrom 1998). It is true that colleges charge different students different prices for essentially the same good. However, a seller may charge different buyers different prices for a different reason—some buyers are more costly to serve than others (Varian 1989; Rothschild and White 1995) make essentially the same point in the context of the college market. Some of the within-college price variation could be because some students are more desirable and hence are effectively less “costly” to enroll. Thus, when I structurally estimate the amount of price discrimination
among colleges, I will also account for factors such as test scores that affect how attractive students are to colleges.

Economic theory tells us that a seller must have information about the buyer’s willingness to pay before it can price discriminate. Typically sellers must either rely on crude proxies for willingness-to-pay, such as age, 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. Theory predicts that, if such information were withheld, 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. \textit{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 result of withholding FAFSA information from colleges remains an empirical question.

A few papers have also looked at the consequences of tuition discounts for the college market as a whole. \textit{Fu} (2014) estimates a very different model from the one in this paper 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. \textit{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 withholding some or all of the FAFSA information from colleges. I allow colleges to see other student characteristics (demographics, test scores, etc.) and to use those characteristics to proxy for the lost FAFSA information. Colleges are still permitted to price discriminate, but they must do so without the benefit of some or all of a student’s FAFSA.

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 \textit{[Bennett 1987]}. He posited that increases in federal financial aid programs did not actually help students and instead merely resulted in higher tuition levels. Empirical support for the Bennett Hypothesis has been mixed with some studies rejecting it \textit{[McPherson and Schapiro 1991]} and others finding modest support \textit{[Long 2003; Cellini and Goldin 2012]}. In contrast to this literature, I am not estimating the effect of financial aid \textit{per se} on prices or price discrimination. Rather, I estimate the effect of sharing FAFSA information with colleges on price discrimination while holding constant the generosity of

\footnote{They use student-level data from the 1995-1996 wave of NPSAS in one portion of their estimation procedure. However, their primary dataset for most of the estimation is at the college level.}
federal aid programs.

3 Modeling the College Market As a First-Price Auction

3.1 The Model

The reduced form evidence from the previous section suggests that some types of colleges possess enough market power to price discriminate while others do not. Therefore, I divide colleges into elite and non-elite colleges. I define elite colleges to be private and very selective public four-year colleges. Non-elite colleges comprise the remainder—less selective public colleges—and I assume that they operate in a competitive fringe. 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 submitting an application. The college chooses to participate by admitting the student and making a take-it-or-leave-it offer. Student $i$ evaluates her offers using the utility function $u_{ij} = v_{ij} - p_{ij}$ where $v_{ij}$ represents her valuation, in dollars, of attending the college and $p_{ij}$ is the price college $j$ offers her. I normalize the utility of a student’s outside option, attending a non-elite college, to be zero. If none of the student’s offers provide her with positive utility, then she can always enroll in a non-elite college. One strength of my model is that I am able to remain agnostic about the preferences of students. $v_{ij}$ depends on college $j$’s characteristics as well as how the student values those characteristics, but I do not place any structure on the nature of those preferences. Rather, I infer $v$ from equilibrium behavior.

The college “wins” the auction—the student enrolls—if it makes the best offer as judged by the student. Denote the space of college payoffs for enrolling a student by $\Pi$. College $j$’s payoff 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. $X_i$ denotes characteristics of student $i$ that are observed to both the college and the econometrician while $z_j$ is observed to the college only. $-w_{ij}$ is $j$’s willingness-to-receive—the lowest price that the college would be willing to offer student $i$. It varies across students with a more desirable student having a higher valuation and hence a lower willingness-to-receive. $w_{ij}$ 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 the college values having high ability students on campus, or maybe it anticipates 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.
In this model, colleges may have market power for two reasons: i) 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 ii) 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 does not know the number of bidders in the auction \( n_j \), but it does receive a noisy signal \( \tilde{n}_j \) and it knows the distribution of bidders conditional on the value of the noisy signal \( \rho(n|\tilde{n}) \). The assumption that colleges perfectly observe \( v_{ij} \) is important. This assumption allows me to back out student preferences from the behavior of the colleges while remaining completely agnostic about precisely what students value in a college or in the college experience. For example, I do not need to take a stand on which college characteristics students value, nor do I need to parse out how much of the college experience is investment and how much is consumption. What does it mean to assume that the college perfectly observe \( v_{ij} \)? In essence, this assumption says that the college perfectly knows how it compares with non-elite colleges in the competitive fringe, so that all of the college’s uncertainty about a student’s willingness-to-pay is driven by uncertainty about the \( v \)'s and \( w \)'s of its competitors.

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 model 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

\[
\{ (v_{ij} + w_{ij}) - (v_{ij} - p_{ij}) \} \mathbb{P}[u_{ij} \geq u_{i\ell} \forall \ell \neq j|\tilde{n}_i]
\]

\[
= (s_{ij} - u_{ij}) \mathbb{P}[u_{ij} \geq \beta(s_{ij}|X_i) \forall \ell \neq j|X_i, \tilde{n}_i]
\]

\[
= (s_{ij} - u_{ij}) \mathbb{P}[\beta^{-1}(u_{ij}|X_i) \geq s_{ij} \forall \ell \neq j|X_i, \tilde{n}_i]
\]

\[
= (s_{ij} - u_{ij}) \left\{ \sum_{n=1}^{N} F_{S|X_i}^{-1} \left( \beta^{-1}(u_{ij}|X_i) \rho(n|\tilde{n}_i) \right) \right\} \tag{1}
\]

where \( F_{S|X_i} \) is the distribution of match surpluses, conditional on student covariates \( X_i \), with support \( S = [\underline{s}, \bar{s}] \), 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 \( w_{ij} \). In the language of auctions, 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

---

\(^6\)After conditioning on \( X_i \), idiosyncratic variation in \( w_{ij} \) only occurs through \( z_i \).
are now in terms of utility (expressed in dollars), and the number of bidders is uncertain. The reserve price, now a reservation utility, is known to all the bidders. Because the student always has the option of attending a non-elite 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

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

(2)

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 additional match surplus. 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 $F_{S|X_i}$.

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 $\beta'$. Intuitively, 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 due to 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. If $v_{ij}$ rises, then the student values the college more. The college still increases its utility bid, but this time it actually raises its price offer. The intuition is that if the student values the college one dollar more, then the college will extract some of that dollar by raising its price but by less than a dollar, so that its utility bid also rises. In essence, 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 $\tilde{n}$ 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, we would not expect prices to be related to the actual number of bidders.

---

7The uncertainty in the number of bidders makes the algebra a little messy, but does not pose a problem as long as I can separately identify and estimate $\rho(n|\tilde{n})$. 
4 Reduced-Form Estimates

In this section I provide a brief description of the data and test some of the 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 dataset 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. 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. NPSAS provides a comprehensive picture of a student’s expenses and financial aid. Unlike many other datasets, 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 amounts 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 postsecondary 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)
- \( \leq 30 \) 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 9 or more full-time-equivalent months
- Was not on an athletic scholarship
- Did not receive a tuition waiver due to a parent’s employment at the college

I exclude athletes because their tuition discounts are determined in a very different way than the general population. I also exclude students with tuition waivers because those waivers are not really
discounts; rather, they represent a non-wage benefit to the student’s parent who works at the college. The sample is restricted to U.S. residents because they are not eligible to complete the FAFSA.8

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 freshmen sample. Finally, I restrict the freshmen 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 freshmen sample. Tables 1a and 1b contain descriptive statistics for both the freshmen and elite samples. Women constitute 54% of the freshmen 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.

Table 2 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 include in \( X_i \) a dummy for whether the student is an out-of-state student at a public college. As predicted by the model, discounts tend to be larger for students with higher test scores and higher high school GPA’s. But tuition discounts are not purely a function of student ability. They fall with parent income and rise with the number of colleges a student lists on her FAFSA. A $10,000 increase in parent adjusted gross income is associated with a $124 decrease in the student’s tuition discount. On the other hand, students who list one more college on the FAFSA tend to receive $373 more. Discounts are also larger for black and Hispanic students than for white and Asian students.

In Table 3, I focus on the discount-income gradient from Table 2 by interacting parent income with college type and selectivity. Table 3 indicates that the coefficient on income in Table 2 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 non-selective colleges—precisely the colleges least likely to possess enough market power 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 power.

Tables 4 and 5 focus on the effect on discounts of listing more colleges on the FAFSA. Column one of Table 4 replicates the regression from Table 2. 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 by very selective private colleges. Among students at the same very selective private college, those who listed one more college on the FAFSA tended to enjoy a $910 larger tuition discount, all else equal. The rest of the elite colleges, less-selective four-year private colleges and very selective public colleges, also

---

8In the fall of 2010, non-resident aliens accounted for 2.2% of undergraduate enrollment (see Table 237 of the 2011 Digest of Education Statistics).
have positive and economically significant coefficients although they are not statistically significant due to a lack of power. The model predicts that listing more colleges on the FAFSA will be related to larger discounts, but that the number of colleges actually applied to will not (holding constant the number listed on the FAFSA). In columns three and four of Table 4 I exploit a previous wave of the data that explicitly asks freshmen how many colleges they applied to. When the number of applications is included as a control, the coefficient on number of colleges listed remains significant and hardly changes but the coefficient on number of applications is essentially zero. 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 applications season is over, listing more colleges on the FAFSA does not appear to be related to the students’ discounts. Taken together, these results show that colleges use information from the FAFSA about potential competitors when offering discounts. It appears that elite colleges, especially very selective private colleges, incorporate the information about potential competitors into their discounts and offer larger discounts when they believe they face stiffer competition for a student.

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 except for one problem: the bids are 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) 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 bidding distribution). My approach is similar, except I must deal with the fact that the bids themselves are not observed.

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

\[
\pi'(s|X_i) = 1 - \beta'(s|X_i) = 1 - (s - \beta(s|X_i)) \frac{\sum_{n=1}^{\pi}(n-1)F^n_{\pi|X_i}(s)f_{s|X_i}(s)\rho(n|\bar{n}_i)}{\sum_{n=1}^{\pi}F^n_{\pi|X_i}(s)\rho(n|\bar{n}_i)}
\]

\[
\sum_{n=1}^{\pi}(n-1)F^n_{\pi|X_i}(s)f_{s|X_i}(s)\rho(n|\bar{n}_i) - \pi(s|X_i)
\]

\[
\Rightarrow \pi(s|X_i) = \frac{\sum_{n=1}^{\pi}F^n_{\pi|X_i}(s)\rho(n|\bar{n}_i)}{\sum_{n=1}^{\pi}(n-1)F^n_{\pi|X_i}(s)f_{s|X_i}(s)\rho(n|\bar{n}_i)} - \pi(s|X_i)
\]

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

\[
\pi'(s|X_i) = \left( 1 + \pi(s|X_i) \right) \left( \frac{\sum_{n=1}^{\pi}(n-1)F^n_{\pi|X_i}(\pi(s|X_i))f_{\pi|X_i}(\pi(s|X_i))\rho(n|\bar{n}_i)}{\sum_{n=1}^{\pi}F^n_{\pi|X_i}(\pi(s|X_i))\rho(n|\bar{n}_i)} \right)^{-1} 0 \leq s
\]

\[
\pi(0|X_i) = 0
\]

Finally, 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 equation (4). So we can write

\[
\psi'(\pi|X_i) = 1 + \pi \frac{\sum_{n=1}^{\pi}(n-1)F^n_{\pi|X_i}(\pi)\rho(n|\bar{n}_i)}{\sum_{n=1}^{\pi}F^n_{\pi|X_i}(\pi)\rho(n|\bar{n}_i)} 0 \leq \pi
\]

\[
\psi(0|X_i) = 0
\]

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

In section 5.1 I prove that the model can be identified from equilibrium transaction prices, but the intuition behind the proofs is not complicated. From equation (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, \( F_{\pi|X} \) is just a shifted version of \( F_{p|X} \) 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 the college ever charges 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 \( \rho(n|\bar{n}) \) can then be combined with equation (6) to identify all the primitives of the model.
5.1 Identification

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

*Proof.* Define the function $y(X,j) \equiv \inf\{\text{supp}(p|X,j)\}$ 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 $\text{supp}(p|X,j)$ 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 $\rho(n|\tilde{n})$.

*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

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

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 $\rho(n|\tilde{n})$.

---

9The college would never be willing to charge a student less than $-w_{ij}$.

10To be more precise, I actually identify the truncated distribution of $F_{S|X_i}$ conditional on $S \geq 0$ which is the standard result in the presence of a binding reserve price.
Proof. Combining $F_{\pi|X}$ and $\rho(n|\bar{n})$ 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 Data and Empirical Strategy

I estimate the model using student-level data from the 2007-2008 wave of the National Postsecondary Student Aid Study (NPSAS). NPSAS contains information on various student characteristics as well as detailed information about prices and discounts. For the structural estimation I focus on freshmen at elite colleges—private and very selective four-year colleges. Tables 1a and 1b 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, Perrigne, and Vuong (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 non-parametrically identified, the identification proof is not constructive because it does not immediately suggest an estimator that could be used in a finite dataset. 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

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

where $\alpha_1(X,j) > 0$, $\alpha_2(X,j) > 0$, and $\bar{p}(X,j)$ 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 $\bar{p}$, I estimate several quantiles of the
price distribution $F_{p|X,j}$ using the quantile regression\textsuperscript{11}

$$F_{p|X,j}^{-1}(t) = X\gamma^t + a_j^t \quad t = 0.05, 0.10, \ldots, 0.40$$ \hspace{1cm} (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 by estimating the parameters $a_1(X,j) > 0$, $a_2(X,j) > 0$, and $p(X,j)$ via nonlinear least squares, subject to the constraint that $p(X_i,j) \leq p_{ij}\textsuperscript{12}$. My estimate of willingness-to-receive is $-\hat{w}_{ij} = \hat{p}(X,j)$. Armed with an estimate $\hat{w}_{ij}$, 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\textsuperscript{13}

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 non-elite 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 [1] contains a histogram of estimated college payoffs for the elite sample. The distribution skews right. College payoffs are typically less than $20,000 although they can range as high as $50,000. Recall

\textsuperscript{11}I include all observations from the full sample, including upperclassmen, from the full sample 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.

\textsuperscript{12}This constraint binds for around 3% of the students in my sample.

\textsuperscript{13}I briefly outline this estimation procedure in the appendix. The estimator is discussed in detail in my working paper [Fillmore 2014].
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.

The first three rows of column one of Table 6 contain a summary of my structural estimates. Total surplus per student, net of attending a non-elite college, averages $16,377 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 non-elite college. On average, students receive $5,222 of consumer surplus while the remainder accrues to the colleges. Note that these estimates do not imply that students value attending college at a mere $5,222 per year. Rather, this number represents the student surplus above and beyond what the student would have received if she had attended a non-elite college. Adding the average student surplus of $5,222 with the average transaction price of $13,158 implies that on average students at elite colleges are willing to pay $18,380 to attend their current colleges rather than a non-elite college.

On average students receive 30% of the total match surplus, and colleges are able to extract an average of 70% of the match surplus through their individualized prices. 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 42.3% of the match surplus while students who list only one college receive just 18%.

Because colleges are extracting a large amount of the match surplus, students may have less incentive to apply to additional colleges. Searching over 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 most of it through its prices. Although I do not directly model students’ application decisions, my estimates can speak to the incentives students face when choosing how many colleges to apply to. In Table 9 I calculate the expected return from applying to an additional college. In this calculation I assume that students apply to colleges randomly. Thus, for each student I calculate the probability that a draw from the match surplus distribution $F_{S|X}$ would exceed the student’s current best match $s_{ij}$. Then I multiply this probability by the expected match surplus given that the draw did exceed $s_{ij}$. This calculation indicates that the expected return to applying to an additional college is $739. However if, as seems likely, students tend to apply to colleges that are a better match first, then the estimates in Table 9 will overstate the returns of an additional application.

6 The Counterfactuals

For simplicity, I assume that the FAFSA conveys three pieces of information: I) parent adjusted gross income, II) a noisy signal about the number of competitors the college faces, and III) the fact that the student chose to complete the FAFSA. I simulate three counterfactuals:
1. Colleges lose access to income information only.

2. Colleges lose access to the noisy signal of the number of competitors only.

3. Colleges lose access to income information, the signal of the number of competitors, and whether the student completed the FAFSA.

I assume that colleges can always see 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 observe in the counterfactual. Colleges always make the best use of the information they have, so they can use the student covariates they see to proxy for those they don’t.

### 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.\footnote{Note that $e_i$ is independent of the observable student covariates $\tilde{X}_i$.} $e_i$ has distribution function $F_e$ with density $f_e$ strictly positive over the support $[e, \bar{e}]$. Importantly, $e_i$ is not college specific. The second shock, $\xi_{ij}$, is a student-college 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.

Define $\tilde{u}_{ij} = \tilde{v}_{ij} - p_{ij} = u_{ij} - e_i$ 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, $\tilde{\pi}_{ij} = \tilde{w}_{j}(\tilde{X}_i) + p_{ij} = \pi_{ij} - \xi_{ij}$ 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...
\( \bar{s}_{ij} = \bar{\pi}_{ij} + \bar{u}_{ij} \). The college maximizes its expected payoff

\[
\max_{\bar{\pi}_{ij}} \mathbb{E}_{\tau|\bar{\pi}} [\bar{\pi}_{ij} \mathbb{P}[j \text{ wins}]]
\]

\[
\rightarrow \max_{\bar{\pi}_{ij}} \mathbb{E}_{\tau|\bar{\pi}} [\bar{\pi}_{ij} \mathbb{P}[u_{ij} > 0 \cap \bar{u}_{ij} > \bar{u}_{ij} \ \ell \neq j]]
\]

\[
\rightarrow \max_{\bar{\pi}_{ij}} \mathbb{E}_{\tau|\bar{\pi}} [\bar{\pi}_{ij} \mathbb{P}[\bar{u}_{ij} + e_i > 0 \cap \bar{u}_{ij} + e_i > \bar{u}_{ij} + e_i \ \ell \neq j]]
\]

\[
\rightarrow \max_{\bar{\pi}_{ij}} \bar{\pi}_{ij} \mathbb{P}[e_i > -\bar{u}_{ij}] \mathbb{P}[\bar{u}_{ij} > \bar{u}_{ij} \ \ell \neq j]
\]

\[
\rightarrow \max_{\bar{u}_{ij}} (\bar{s}_{ij} - \bar{u}_{ij}) \mathbb{P}[e_i > -\bar{u}_{ij}] \mathbb{P}[\bar{u}_{ij} > \bar{\beta}(\bar{s}_{ij}|\bar{X}_i) \ \ell \neq j]
\]

\[
\rightarrow \max_{\bar{u}_{ij}} (\bar{s}_{ij} - \bar{u}_{ij})(1 - F_{\ell}(-\bar{u}_{ij})) \sum_{n=1}^{\pi} F_{n|\bar{X}_i}(\bar{\beta}^{-1}(\bar{u}_{ij}|\bar{X}_i)) \rho(n|\bar{n}_i)
\]

The first order condition for the college yields the ODE

\[
\bar{\beta}'(\bar{s}|\bar{X}_i) = \frac{(\bar{s} - \bar{\beta}(\bar{s}|\bar{X}_i)) \sum_{n=1}^{\pi} F_{n|\bar{X}_i}(\bar{\beta}(\bar{s}|\bar{X}_i)) \rho(n|\bar{n}_i)}{1 - (\bar{s} - \bar{\beta}(\bar{s}|\bar{X}_i)) \sum_{n=1}^{\pi} F_{n|\bar{X}_i}(\bar{\beta}(\bar{s}|\bar{X}_i))}
\]

(10)

with the initial condition

\[
\bar{\beta}(-\bar{e}|\bar{X}_i) = -\bar{e}.
\]

The initial condition now says that the lowest value bidder is the college with an observed match surplus \( \bar{s}_{ij} \) 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 \( \bar{s}_{ij} = -\bar{e} \). As \( \bar{s}_{ij} \) 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 \( \bar{\beta} = -\bar{e} \), 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: i) \( F_{\bar{S}|\bar{X}_i} \) will differ, perhaps substantially, from \( F_{S|X_i} \); ii) in counterfactuals 2 and 3, colleges will lose access to the noisy signal \( \bar{n}_i \) which will affect the distribution \( \rho(n|\bar{n}_i) \); iii) colleges now face uncertainty over their position relative to the competitive fringe; iv) each college’s observed match surplus \( \bar{s}_{ij} \) 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 mistakenly 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. 

\[ \text{In counterfactual 2, } \bar{n}_i \text{ becomes a binary indicator for whether the student completed the FAFSA. In counterfactual 3, } \rho(n|\bar{n}_i) \text{ becomes the marginal distribution } \rho(n). \]
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 mistakenly 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 college’s two forecast errors \( e_i \) and \( \xi_i \) for \( v_{ij} \) and \( w_{ij} \) 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.\(^{16}\)

Table 6 compares the baseline structural estimates with three counterfactuals: 1) colleges lose access to income information only, 2) colleges lose access to the number of colleges listed on the FAFSA only, and 3) colleges lose access to income information, the number of colleges listed, and even whether the student completed the FAFSA at all. In all three counterfactuals, most students pay lower prices (receive larger discounts). Prices drop by an average of $743 in counterfactual 1, $382 in counterfactual 2, and $825 in counterfactual 3. The percentage of students who experience a price drop amounts to 71.5%, 66.6% and 61.6% respectively. When we withhold income information the within-college variance in prices drops by 10.9% in counterfactual one and 16.7% in counterfactual two. Due to the uncertainty introduced by restricting colleges’ information, some students, up to 11.9%, are misallocated and end up inefficiently attending non-elite colleges. While this misallocation lowers total surplus, student surplus is higher in all three counterfactuals due to the lower prices students are paying.

Table 8 looks at the counterfactual results in a different way. I run 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.\(^{17}\) Then I run the same regression for each counterfactual, substituting in my estimate of each student’s counterfactual transaction price. Table 8 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 withheld from colleges. 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 withheld while the coefficients for racial minorities (black and Hispanic) become larger.

In Table 10 I look at how the three counterfactuals differentially affect low-and high-income stu-

---

\(^{16}\)See the appendix for more details.

\(^{17}\)In contrast with the reduced form specification from section 4, the dependent variable in Table 8 is the transaction price rather than the tuition discount.
dents. I regress a student’s price change (relative to baseline) on her parent adjusted gross income. Withholding income information lowers prices for high-income students but raises them for low-income students. Looking at it the other way, sharing income information with colleges effectively levies a 1.6% tax on adjusted gross income combined with a $524 lump sum rebate. The full effect of sharing the FAFSA with colleges is to effectively levy a 1.9% tax rate on adjusted gross income combined with a $709 lump sum rebate. In contrast, sharing the number of colleges listed on the FAFSA neither helps nor hurts low-income students compared with high-income students.

In all three counterfactuals, students receive a larger share of the match surplus which will strengthen their incentives to apply to more colleges. In Table 9 I calculate the expected return from applying to an additional college (again assuming that the student applies to colleges randomly). Relative to baseline, the return to applying rises by between $50 and $250 depending on the counterfactual. This suggests that my estimates might be understating the full effects of withholding FAFSA information from colleges. If students respond to these incentives by applying to more colleges, then the increased competition will further lower prices.

7 Conclusion

Pricing in the higher education market is complex. Tuition discounts are a common form of price discrimination for colleges. 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 equal, colleges will charge students less if they are more attractive 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. I include college fixed effects to isolate variation among students in the same school. At elite colleges, I find that more attractive students (based on ACT score and high school GPA) do pay less than their peers, while higher income students pay more. Those who list more colleges on the FAFSA also pay less. These patterns, which are consistent with the predictions of the model, show up primarily among private and very selective public colleges. I label these elite colleges and focus on them when estimating the structural model.

I estimate the model using student-level data on transaction prices and student characteristics. I simulate three counterfactuals: 1) withhold income information only, 2) withhold the number of colleges listed on the FAFSA only, and 3) withhold income information, the number of colleges listed,
and whether the student completed the FAFSA. My estimation results highlight an important policy tradeoff—withholding student information from colleges lowers average prices but also leads to a misallocation of students. The intuition is that colleges will not be able to tailor their prices as precisely if the government withholds some or all of the FAFSA. As a result, elite colleges will sometimes mistakenly over-price a student and cause her to (inefficiently) attend a non-elite college. Depending on the counterfactual, I estimate that between 8.4% and 11.9% of students who currently attend elite colleges would end up attending non-elite colleges if we withheld some or all of the FAFSA.

However, my estimates indicate that the welfare gain to students is larger than the loss to efficiency. Thus, students are better off on average, and colleges are worse off, when FAFSA information is withheld. This can be seen in Table 6 by noticing that student prices fall on average, so revenue per student must also fall. My estimates suggest that by providing FAFSA information to colleges, the federal government effectively harms the average student by $696 per year while benefiting the average elite college by $897 per year per student. However, the effect on students differs by income. Sharing the FAFSA with colleges enables them to price discriminate in a way that amounts to a 1.9% tax on adjusted gross income coupled with a $709 lump sum rebate, so that students with parent adjusted gross income below $36,670 are helped, while those with higher incomes are harmed. My findings highlight that withholding FAFSA information presents a policy tradeoff between increasing the welfare of middle- and high-income students (by lowering their prices) while lowering total efficiency (by misallocating some students), college welfare (by lowering tuition revenue), and the welfare of low-income students (by raising their prices).

The model as presented captures the basic elements of competition in the U.S. college market. However, there are at least two extensions to the model that one might consider. First, the model requires me to make the strong assumption that a college perfectly learns a student’s valuation for that college, relative to non-elite colleges in the competitive fringe, during the application process. This assumption implies that all of a college’s uncertainty about a student’s willingness-to-pay is driven by uncertainty about the student’s match quality with the college’s competitors. One could imagine trying to relax this assumption by assuming instead that colleges only observe a noisy signal about a student’s valuation. Doing so will likely require parametric assumptions about the distribution of that noise as well as a way to identify this distribution from the data. Second, I assume a symmetric conditionally independent private values framework. Clearly, some colleges (i.e. private colleges) are stronger bidders than others (i.e. public colleges), but a symmetric information structure is still appropriate if bidders do not know the identities of their competitors. One might worry that colleges do have some information about the types of colleges they are competing against. If so, then this would suggest an asymmetric auction framework. I plan to incorporate both of these additional features into the model in future work.
References


8 Figures

![Histogram of Estimated College Payoffs](image)

Figure 1: Histogram of Estimated College Payoffs (restricted to sample used in structural estimation)

26
Appendices

A Data Appendix

The college selectivity variable used throughout the paper was 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 selectivity categories based on a few characteristics. Two-year colleges are put into their own category as were 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. (Cunningham 2005)

Table E-1, reproduced here for reference, gives 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.”
types of selectivity measures such as Peterson’s Selectivity Ranking.29 The selectivity variable appeared to assign institutions to categories in ways that would be expected (table E-1).

<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>Northern Kentucky University</td>
<td></td>
</tr>
<tr>
<td>SUNY-Binghamton</td>
<td>Ohio State University</td>
<td>University of Northern Alabama</td>
<td>Texas Southern University</td>
<td></td>
</tr>
<tr>
<td>University of Virginia</td>
<td>University of Oregon</td>
<td>Winston-Salem State University</td>
<td>University of Toledo</td>
<td></td>
</tr>
<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></td>
<td></td>
</tr>
<tr>
<td>Williams College</td>
<td>University of San Francisco</td>
<td>Albertus Magnus College</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.

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 $X$. I 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 $p = .05, .10, .15, \ldots, .95$. For each observation $i$, calculate the fitted value from each quantile regression $\tilde{y}^{p}_{i}$. This is an estimate of the $p$th quantile of the random variable $Y$. For each observation $i$, collect the points $\{(\tilde{y}^{p}_{i}, p)\}_{p}$ and append the point $(\min y, 0)$ to the beginning and the point $(\max y, 1)$ to the end. I 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). I restrict 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. I perform this procedure for each observation $i$ in the dataset. This gives me estimates of $F_{Y|X_{i}}$ and $f_{Y|X_{i}}$ for each observation.

The method outlined here is semi-parametric. 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 we specify the regressions. 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 simple linear specification, the method still allows the vector of coefficients $\beta^{p}$ to differ for each value of $p$. In my application, I adopt a linear specification.
C Simulating the Counterfactuals

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

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

2. Repeat step 1 for $\tilde{w}_{ij}$ 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 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 non-elite college and receives zero utility. Otherwise, the student remains at her college and pays a price equal to $v_{ij} - (\tilde{u}_{ij} + e_i)$.

D Modeling College Spending by Incorporating Dynamics

When colleges price discriminate they are generally 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 as well as 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 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.\(^{18}\) 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

\[
Q' = (1 - \delta)Q + p_zZ \tag{11}
\]

\[
E' = (1 + r) \left( E - Z + \sum_{i=1}^{I} (p_{ij} + S_{ij})q_{ij}(p_{ij}; Q) + \sum_{i=1}^{I} d_{ij}a_{ij} \right) \tag{12}
\]

\( \delta \) 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 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.

\(^{18}\)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.
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

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

Note that this first order condition for the college has the same form as in the static model\textsuperscript{19}. 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. Students who are expected to make larger future donations have a lower willingness-to-receive (the college is willing to charge them less). At colleges with a higher marginal value of quality, willingness-to-receive will be less sensitive to $w_{ij}$. Intuitively, if a college desperately needs money to invest in quality improvements, then it can’t afford to be picky about the quality 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

\textsuperscript{19} $a_{ij}$ is analogous to $P[i$ chooses $j]$ and the first order conditions from the static model in the paper and the dynamic model presented here have the same form $(p - c)q' = q$. 

31
are spent entirely on college quality, and the marginal value (to the college) of quality shows up in the
denominator of \( \frac{w_{ij}}{p_{j}^{p}V_{Q}} \). 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 \( i \).

E Matching Models

Matching models represent a popular framework for thinking about the college market. In 1962,
Gale and Shapley \cite{Gale1962} proposed their famous deferred-acceptance algorithm for understanding how students get matched with colleges. In 1982, Kelso and Crawford \cite{Kelso1982} 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 \cite{Boyd2013}, husbands and wives \cite{Choo2006,Hitsch2010}, and companies and venture capitalists \cite{Sorensen2007} to name a few examples. Fu \cite{Fu2014} 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,\cite{Athey2001} 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

\(^{20}\)A college may be matched with multiple students.
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 on some un-modeled 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 auctions framework.
### Table 1a. Cell Counts

<table>
<thead>
<tr>
<th></th>
<th>Freshmen Sample</th>
<th>Elite Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Colleges</td>
<td>Students</td>
</tr>
<tr>
<td>All Colleges</td>
<td>1,110</td>
<td>6,780</td>
</tr>
<tr>
<td>Very Selective</td>
<td>180</td>
<td>1,120</td>
</tr>
<tr>
<td>Moderately Selective</td>
<td>440</td>
<td>2,570</td>
</tr>
<tr>
<td>Not Selective</td>
<td>160</td>
<td>920</td>
</tr>
<tr>
<td>Two-Year</td>
<td>330</td>
<td>2,170</td>
</tr>
<tr>
<td>Private</td>
<td>430</td>
<td>1,830</td>
</tr>
<tr>
<td>Very Selective</td>
<td>120</td>
<td>580</td>
</tr>
<tr>
<td>Moderately Selective</td>
<td>240</td>
<td>960</td>
</tr>
<tr>
<td>Not Selective</td>
<td>60</td>
<td>220</td>
</tr>
<tr>
<td>Two-Year</td>
<td>10</td>
<td>70</td>
</tr>
<tr>
<td>Public</td>
<td>690</td>
<td>4,960</td>
</tr>
<tr>
<td>Very Selective</td>
<td>70</td>
<td>540</td>
</tr>
<tr>
<td>Moderately Selective</td>
<td>190</td>
<td>1,610</td>
</tr>
<tr>
<td>Not Selective</td>
<td>100</td>
<td>710</td>
</tr>
<tr>
<td>Two-Year</td>
<td>320</td>
<td>2,100</td>
</tr>
<tr>
<td>Public, In-State</td>
<td>---</td>
<td>4,580</td>
</tr>
<tr>
<td>Very Selective</td>
<td>---</td>
<td>490</td>
</tr>
<tr>
<td>Moderately Selective</td>
<td>---</td>
<td>1,440</td>
</tr>
<tr>
<td>Not Selective</td>
<td>---</td>
<td>650</td>
</tr>
<tr>
<td>Two-Year</td>
<td>---</td>
<td>2,010</td>
</tr>
<tr>
<td>Public, Out-Of-State</td>
<td>---</td>
<td>370</td>
</tr>
<tr>
<td>Very Selective</td>
<td>---</td>
<td>50</td>
</tr>
<tr>
<td>Moderately Selective</td>
<td>---</td>
<td>170</td>
</tr>
<tr>
<td>Not Selective</td>
<td>---</td>
<td>60</td>
</tr>
<tr>
<td>Two-Year</td>
<td>---</td>
<td>90</td>
</tr>
</tbody>
</table>

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>
<thead>
<tr>
<th></th>
<th>Freshmen Sample</th>
<th></th>
<th>Elite Sample</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Std Dev</td>
<td>Min</td>
<td>Max</td>
</tr>
<tr>
<td>Tuition Discount</td>
<td>$2,914</td>
<td>($5,722)</td>
<td>$0</td>
<td>$37,000</td>
</tr>
<tr>
<td>Student Received Discount</td>
<td>0.38</td>
<td>(0.49)</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Sticker Price</td>
<td>$10,168</td>
<td>($9,737)</td>
<td>$275</td>
<td>$39,289</td>
</tr>
<tr>
<td>Parent Adjusted Gross Income</td>
<td>$65,400</td>
<td>($55,502)</td>
<td>$0</td>
<td>$496,347</td>
</tr>
<tr>
<td>ACT Score</td>
<td>21.2</td>
<td>(4.7)</td>
<td>11</td>
<td>36</td>
</tr>
<tr>
<td>High School GPA</td>
<td>3.42</td>
<td>(0.57)</td>
<td>0.90</td>
<td>4.00</td>
</tr>
<tr>
<td>Earned AP Credit</td>
<td>0.22</td>
<td>(0.42)</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Parents With College Degree</td>
<td>0.68</td>
<td>(0.80)</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Completed FAFSA</td>
<td>0.82</td>
<td>(0.38)</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Additional Colleges Listed on FAFSA</td>
<td>1.2</td>
<td>(1.7)</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>Age as of 12/31/07</td>
<td>18.6</td>
<td>(0.9)</td>
<td>17</td>
<td>23</td>
</tr>
<tr>
<td>Female</td>
<td>0.54</td>
<td>(0.50)</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>White</td>
<td>0.65</td>
<td>(0.48)</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Black</td>
<td>0.13</td>
<td>(0.34)</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Hispanic</td>
<td>0.12</td>
<td>(0.33)</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Asian</td>
<td>0.05</td>
<td>(0.22)</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Other</td>
<td>0.04</td>
<td>(0.20)</td>
<td>0</td>
<td>1</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 6 college on the FAFSA, so the number of additional colleges listed runs from 0 to 5. The elite sample is a subset of the freshmen sample and consists of freshmen at private and very selective public four-year colleges. No sample weights were used.
### Table 2. Within-College Discounting Patterns

**Dependent Variable: Tuition Discount**

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

**College fixed effects** | Yes

**Observations** | 5640
**R-squared** | 0.670

The regression includes students from the freshmen sample. The omitted race category is "white." For students who completed the FAFSA, "number of additional colleges listed" ranges from zero to 5 (students can list up to six colleges on the FAFSA). For those who did not complete the FAFSA, "number of additional colleges listed" is set to zero and the dummy "completed FAFSA" is included. Robust standard errors are in parentheses. Sampling weights were used (WTA000).

* p<0.05, ** p<.01, *** p<.001
### Table 3. Interacting Income With College Type

<table>
<thead>
<tr>
<th>Parent AGI (in $10,000's)</th>
<th>Very Selective</th>
<th>Moderate Selective</th>
<th>Not Selective</th>
<th>Two-Year</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Public</td>
<td>Private</td>
<td>Public</td>
<td>Private</td>
</tr>
<tr>
<td>Public</td>
<td>-89.8 (22.2) ***</td>
<td>-37.9 (14.7) **</td>
<td>28.0 (26.1)</td>
<td>23.2 (12.0)</td>
</tr>
<tr>
<td>Private</td>
<td>-473.0 (62.4) ***</td>
<td>-87.1 (40.6) *</td>
<td>-185.2 (86.0) *</td>
<td>145.5 (139.5)</td>
</tr>
</tbody>
</table>

College fixed effects Yes
Observations 5640
R-squared 0.693

See notes to Table 2. The regression specification here is identical to that in Table 2 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 (WTA000).

* p<0.05, ** p<0.01, *** p<0.001
### Table 4. The Number of Colleges Listed on a Student’s FAFSA

<table>
<thead>
<tr>
<th>Number of additional colleges listed on FAFSA</th>
<th>Freshmen Sample</th>
<th>Freshmen Sample</th>
<th>BPS 03-04</th>
<th>BPS 03-04</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Dependent Variable: Tuition Discount</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number listed on FAFSA</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Very Selective</td>
<td>Public</td>
<td>131.8 (90.02)</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>
</tr>
<tr>
<td>Moderately Selective</td>
<td>Public</td>
<td>56.0 (62.38)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Private</td>
<td>249.3 (155.8)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not Selective</td>
<td>Public</td>
<td>44.9 (104.6)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Private</td>
<td>264.5 (307.7)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Two-Year</td>
<td>Public</td>
<td>-0.5 (36.80)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Private</td>
<td>-463.8 (204.5) *</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of colleges applied to</td>
<td>6.9 (26.79)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>College fixed effects</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Observations</td>
<td>5640</td>
<td>5640</td>
<td>5290</td>
<td>5290</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.670</td>
<td>0.714</td>
<td>0.639</td>
<td>0.639</td>
</tr>
</tbody>
</table>

See notes to Table 2. The regressions reported in columns 1 and 2 are identical to the regression in Table 2 except that in column 2 the number of additional colleges listed on the FAFSA has been interacted with college selectivity. I set number of 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 (WTA000) were used in all regressions.

* p<0.05, ** p<.01, *** p<.001
Table 5. Colleges Listed and Month FAFSA Completed

| Number of additional colleges listed on FAFSA | January       | 530.6  (144.2) *** |
|                                            | February      | 290.7  (81.7) *** |
|                                            | March         | 11.3   (91.5)    |
|                                            | April or later| 77.4   (95.1)    |

College fixed effects: Yes

Observations: 5630
R-squared: 0.673

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

* p<0.05, ** p<.01, *** p<.001
### Table 6. 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>$5,222</td>
<td>$5,891</td>
<td>$5,541</td>
<td>$5,918</td>
</tr>
<tr>
<td></td>
<td>($4493, $5262)</td>
<td>($5046, $6018)</td>
<td>($4772, $5610)</td>
<td>($5173, $6023)</td>
</tr>
<tr>
<td>Total Surplus Per Student</td>
<td>$16,377</td>
<td>$16,293</td>
<td>$16,208</td>
<td>$16,176</td>
</tr>
<tr>
<td></td>
<td>($15081, $16975)</td>
<td>($15023, $16889)</td>
<td>($14985, $16851)</td>
<td>($14949, $16767)</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>30.3%</td>
<td>37.4%</td>
<td>35.3%</td>
<td>36.9%</td>
</tr>
<tr>
<td></td>
<td>(28.1%, 30.2%)</td>
<td>(34.6%, 37.7%)</td>
<td>(32.8%, 35.7%)</td>
<td>(34.8%, 37.3%)</td>
</tr>
<tr>
<td>Mean Transaction Price</td>
<td>$13,158</td>
<td>$13,135</td>
<td>$13,650</td>
<td>$13,291</td>
</tr>
<tr>
<td></td>
<td>($12771, $13530)</td>
<td>($12738, $13579)</td>
<td>($13236, $14100)</td>
<td>($12931, $13727)</td>
</tr>
</tbody>
</table>

### Changes Relative to Baseline

<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>Consumer (Student) Surplus Per Student</td>
<td>$0</td>
<td>$670</td>
<td>$319</td>
<td>$696</td>
</tr>
<tr>
<td></td>
<td>(---)</td>
<td>($442, $799)</td>
<td>($198, $432)</td>
<td>($546, $860)</td>
</tr>
<tr>
<td>Total Surplus Per Student</td>
<td>$0</td>
<td>-$84</td>
<td>-$168</td>
<td>-$201</td>
</tr>
<tr>
<td></td>
<td>(---)</td>
<td>($-156, $-41)</td>
<td>($-246, $-67)</td>
<td>($-299, $-131)</td>
</tr>
<tr>
<td>Percent of students who switch to a non-elite college</td>
<td>0.0%</td>
<td>8.4%</td>
<td>11.3%</td>
<td>11.9%</td>
</tr>
<tr>
<td></td>
<td>(---)</td>
<td>(6.8%, 12.5%)</td>
<td>(0%, 0%)</td>
<td>(0%, 0%)</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>3.3%</td>
<td>4.9%</td>
</tr>
<tr>
<td></td>
<td>(---)</td>
<td>(5.2%, 6.8%)</td>
<td>(2.7%, 4.6%)</td>
<td>(4.2%, 6.2%)</td>
</tr>
<tr>
<td>Mean Change in Transaction Price</td>
<td>$0</td>
<td>-$743</td>
<td>-$382</td>
<td>-$825</td>
</tr>
<tr>
<td></td>
<td>(---)</td>
<td>($-931, $-482)</td>
<td>($-550, $-227)</td>
<td>($-1074, $-634)</td>
</tr>
<tr>
<td>Percent of Students With Price Drop</td>
<td>0.0%</td>
<td>71.5%</td>
<td>66.6%</td>
<td>61.6%</td>
</tr>
<tr>
<td></td>
<td>(---)</td>
<td>(68.7%, 75.7%)</td>
<td>(62.3%, 75.8%)</td>
<td>(58.8%, 68.4%)</td>
</tr>
<tr>
<td>Within-College Variance in Price</td>
<td>$38,812</td>
<td>-10.9%</td>
<td>-1.6%</td>
<td>-16.7%</td>
</tr>
<tr>
<td></td>
<td>($27911, $34061)</td>
<td>(-17.5%, -3.1%)</td>
<td>(-9.4%, 8.2%)</td>
<td>(-24.6%, 8.5%)</td>
</tr>
</tbody>
</table>

Column 1 contains baseline estimates while columns 2-4 contain estimates for the three counterfactuals. Point estimates are in bold. 95% percent confidence intervals in parentheses were calculated using 200 bootstrap replications. The percentage changes in rows 6 and 7 are in percentage points. The percentage changes in row 10 are percent changes in within-college price variance relative to a base of $38,812. No sample weights were used.
Table 7. 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>23.8%</td>
</tr>
<tr>
<td>1</td>
<td>18.0%</td>
</tr>
<tr>
<td>2</td>
<td>33.5%</td>
</tr>
<tr>
<td>3</td>
<td>38.0%</td>
</tr>
<tr>
<td>4</td>
<td>39.4%</td>
</tr>
<tr>
<td>5</td>
<td>41.6%</td>
</tr>
<tr>
<td>6</td>
<td>42.3%</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 8. Comparing Baseline and Counterfactual Pricing Patterns

<table>
<thead>
<tr>
<th>Dependent Variable: Transaction Price</th>
<th>Baseline</th>
<th>Parent Income Withheld</th>
<th>Colleges Listed Withheld</th>
<th>Entire FAFSA Withheld</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parent AGI (in $10,000's)</td>
<td>220.1</td>
<td>(27.9) ***</td>
<td>249.7 (30.7) ***</td>
<td>66.7 (29.0) *</td>
</tr>
<tr>
<td>ACT score</td>
<td>-196.6</td>
<td>(47.5) ***</td>
<td>-136.0 (54.1) *</td>
<td>-105.8 (51.3) *</td>
</tr>
<tr>
<td>High school GPA</td>
<td>-1087.1</td>
<td>(365.7) **</td>
<td>-1269.3 (403.1) **</td>
<td>-1378.8 (379.4) ***</td>
</tr>
<tr>
<td>Earned AP credit in high school</td>
<td>-193.8</td>
<td>(404.8)</td>
<td>89.4 (425.1)</td>
<td>125.7 (406.1)</td>
</tr>
<tr>
<td>Completed FAFSA</td>
<td>-1638.1</td>
<td>(428.5) ***</td>
<td>-2721.0 (475.9) ***</td>
<td>-651.9 (446.4)</td>
</tr>
<tr>
<td>Number of additional colleges listed on FAFSA</td>
<td>-527.9 (112.6) ***</td>
<td>13.2 (121.9)</td>
<td>-22.1 (121.2)</td>
<td></td>
</tr>
<tr>
<td>Age as of 12/31/07</td>
<td>77.8</td>
<td>(283.8)</td>
<td>59.8 (273.3)</td>
<td>213.8 (256.4)</td>
</tr>
<tr>
<td>Female</td>
<td>-553.4</td>
<td>(314.5)</td>
<td>-479.9 (346.8)</td>
<td>-501.2 (328.7)</td>
</tr>
<tr>
<td>Black</td>
<td>-2168.8</td>
<td>(636.6) ***</td>
<td>-2927.4 (662.9) ***</td>
<td>-2442.2 (674.0) ***</td>
</tr>
<tr>
<td>Hispanic</td>
<td>-1826.3</td>
<td>(638.8) ***</td>
<td>-1492.4 (702.5) *</td>
<td>-1982.1 (685.5) **</td>
</tr>
<tr>
<td>Asian</td>
<td>-110.6</td>
<td>(822.9)</td>
<td>202.5 (752.5)</td>
<td>-44.2 (725.3)</td>
</tr>
<tr>
<td>Other/Multiple</td>
<td>-865.1</td>
<td>(810.3)</td>
<td>-857.9 (924.7)</td>
<td>-1073.6 (915.6)</td>
</tr>
<tr>
<td>Out-of-State, Public</td>
<td>8722.5</td>
<td>(837.8) ***</td>
<td>8287.9 (690.3) ***</td>
<td>7294.9 (648.0) ***</td>
</tr>
</tbody>
</table>

All four regressions include students from the elite sample. The dependent variable in column 1 is transaction price (sticker minus discount). In columns 2-4, 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 for all regressions (WTA000).

* p<0.05, ** p<.01, *** p<.001
Table 9. Average Return From Applying to An Additional College

<table>
<thead>
<tr>
<th></th>
<th>FAFSA Information Withheld</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Baseline</td>
</tr>
<tr>
<td></td>
<td>$739</td>
</tr>
</tbody>
</table>

Each cell reports the average return from applying to an additional college for students in that cell. Returns represent the expected utility gain if the new college beats out 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 will overstate the returns to applying to an additional college. No sample weights were used.
### Table 10. Distributional Effects of Price Changes

<table>
<thead>
<tr>
<th>Dependent Variable: Change in Price</th>
<th>Parent Income Withheld</th>
<th>Colleges Listed Withheld</th>
<th>Entire FAFSA Withheld</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parent AGI (in $10,000's)</td>
<td>-162.1 (2.60) ***</td>
<td>-8.9 (5.19)</td>
<td>-193.4 (6.44) ***</td>
</tr>
<tr>
<td>Constant</td>
<td>523.6 (26.36) ***</td>
<td>-312.9 (52.85) ***</td>
<td>709.2 (66.06) ***</td>
</tr>
<tr>
<td>Observations</td>
<td>2030</td>
<td>1960</td>
<td>1950</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.657</td>
<td>0.001</td>
<td>0.317</td>
</tr>
</tbody>
</table>

All three regressions include students from the elite sample. 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 (WTA000).

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