The Impact of Grants on Schools and Students Evidence from the Cal Grant *

Jonathan Louis Gu†

Latest Version

October 28, 2019

Abstract

This article estimates the causal impact of additional government funding on school choice. I construct a structural equilibrium model of the application-admission-scholarship-enrollment game between students and schools. I show how to estimate the causal impact in the presence of omitted variable bias and measurement error. A jump in Cal Grant aid at a GPA threshold serves as a shift in net tuition that is independent from omitted variables. I find that the vast majority of California Cal Grant recipients would have enrolled in-state even without the Cal Grant. I predict that removing the Cal Grant would decrease in in-state enrollment by one-tenth, and the majority of this decrease leads to non-enrollment.

Keywords: Education, Student Aid, Causality, Discrete Choice, Regression Discontinuity, Measurement Error, Missing Data, Selective Sampling

*Thank you Edward E. Leamer, Moshe Buchinsky, Sarah Reber, and Till von Wachter for your constructive criticism and support. Thank you Mengshan Cui for being my wife and also for your support.

†E-mail address: jonathangu@ucla.edu

Contents

1 Introduction

Postsecondary education is one of the main levers of promoting country-wide growth (Hanushek and Woessmann, 2015). States have long offered subsidized education at public institutions, and recently some states have made most in-state options altogether free (Cities of Promise, 2019). State-based grants, federal grants, and college tuition have all doubled over the last twenty years (a 40% increase in real terms Pingel (2017)). The Cal Grant is the largest source of financial aid in California.[1] This article uses the Cal Grant to estimate the causal impact of grants and scholarships on college choice.

My paper estimates the efficacy of the Cal Grant at achieving its goals. Like nearly all state-based grant aid, the Cal Grant is provided for in-state students to attend in-state schools. These state-based grants have three possible goals: (1) increasing overall enrollment, (2) encouraging more in-state enrollment, and (3) helping students fund their higher education. These three goals split recipients into three categories: (1) access-switchers who are induced to enroll in-state instead of declining to enroll anywhere, (2) out-of-state switchers who are induced to enroll in-state instead of enrolling out-of-state, and (3) always-takers who were originally going to attend an in-state school already.

I find that the Cal Grant is the best at helping California students fund their postsecondary education because 93% of Cal Grant recipients would have enrolled in-state even without the Cal Grant. I also find that removing the Cal Grant would decrease in-state enrollment from 67% to 62%, with out-of-state enrollment concurrently increasing from 10% to 12%, and non-enrollment increasing from 23% to 26%.

Most research does not find a significant impact of governmental aid on overall enrollment – Goal (1) (Rubin, 2011; Bettinger, 2004; Denning et al., 2019), and evidence suggests that state-based aid can convince students to attend less selective in-state schools instead of more selective out-of-state schools – Goal (2) (Avery and M. Hoxby, 2003; Cohodes and Goodman, 2014). The Cal Grant does a good job of helping students fund their postsecondary education by reducing dropout rates – Goal (3) (Bettinger et al., 2019). Mountjoy (2019) highlights the potential for targeted student-aid to divert students from their original choices by showing that many recipients of free community college were diverted from attended four-year colleges.

In addition to predicting the distribution of enrollment if Cal Grant were removed, my model of the equilibrium between schools and students can also predict the impact of other educational government policies. For example, I find that making community

[1]On average the Cal Grant covers 37% of University of California Tuition and Fees, is received by 17% of graduating high school seniors, and accounts for 13% of all postsecondary funding by the state. Figure 1 shows the relative importance for the Cal Grant program both in 2004 and 2018.

colleges free would increase community college enrollment from 29% to 32%, but the majority of this decrease comes from students that would have enrolled in a different four-year school without the funding.

I contribute to the literature on the college market by specifically modeling net tuition – the online advertised tuition (sticker price) minus any governmental grants and school scholarships. Students apply to schools based on the net tuition offers they expect to receive, and they enroll based on the net tuition offers they actually receive. Schools make their admissions and scholarship policies in anticipation of student applications and enrollments. I estimate the causal impact of grants and net tuition on this application-admission-scholarship-enrollment equilibrium between students and schools.

Although the college market has been modeled before as a structural equilibrium (Fu, 2014; Epple et al., 2006), student-school level net tuition has not been studied in this setting. Instead, previous structural estimations of the equilibrium between schools and students use the sticker price as inputs for the cost of schooling. However, most students don't pay the full sticker price of a university to enroll, and nearly a quarter of students attended college without paying any tuition or fees (Figure 2). Any analysis of the monetary cost of schooling needs to examine the net tuition instead.

The Educational Longitudinal Study of 2002 (ELS2002) is the best dataset for studying the impact of grants and scholarships because it has government grants, school scholarships, entire application sets, admissions outcomes, and enrollments for each student. The grant data reveals detailed student characteristics related to the amount of government funding received. I model school policies with admissions and school-scholarships, and I model student decisions with applications and enrollments. This dataset gives me the best chance at examining the impact of net tuition on the interaction between schools and students.

Section 2 discusses the ELS2002 in further detail and previews the key data issues. Section 3 presents the model that parameterizes student and school behavior. Section 4 discusses the strategies and assumptions that I need to overcome the key issues. Section 5 shows how to write the likelihood function and presents a modified simulated maximum likelihood estimation method. Section 6 shows the estimated results. Section 7 simulates what would happen if I removed the Cal Grant, and examines the Cal Grant's effectiveness at accomplishing the three goals.

2 Data

The Educational Longitudinal Study of 2002 (ELS2002) is centered around a survey of 16,197 sophomores from 751 different high schools nationwide. In addition to the survey, it also includes loan data from the National Student Loan Data System and official transcripts from each high school and postsecondary institution attended by the students. I also merge school characteristics from the College Scorecard.[2]

The ELS2002 is a panel survey of one cohort. The surveyors (NCES) constructed the dataset by randomly sampling from all the high schools in the nation (weighting by high school enrollment). 751 out of the 1,200 high schools agreed to participate. The surveyors then randomly selected 20-30 students from each high school. I compare the ELS2002 with Census data from the American Community Survey in Table 1 to check how well the ELS2002 represents students both nationwide and in California. I find that students from wealthier families are slightly underrepresented in the ELS2002. This bias in the wealth of ELS2002 respondents doesn't matter much because wealthy students don't qualify for the Pell Grant or the Cal Grant. I also compare the ELS2002 to the high school population in 2017 to find that Hispanic students are underrepresented compared to today – the share of students that identify as Hispanic increased by 50% since 2002.

The students in the ELS2002 are first surveyed as sophomores in high school in 2002, then three times more – in 2004, 2006, and 2011. I focus on California because my main identification strategy relies on the California Cal Grant. To estimate my model, I use students from the ELS2002 that both responded in 2006 and have their high school transcripts available. I need students to respond in 2006 (the second follow-up survey) in order to observe their applications, admissions, enrollment, and net tuition. I need each student's high school transcript because I calculate the student's Cal Grant eligibility from detailed class-by-class performance. The relative proportions of race and income are not changed when I only take the respondents (Table 2). We see that the sample represents geographic regions in California quite well in Figure 3.

It is vital to include two-year schools (community and technical colleges) in the analysis of postsecondary college attainment because only 523 students enrolled at a four-year institution, while 581 students enrolled at a two-year community college. We can see from Figure 4 that many students who attend a community college at first eventually attain a bachelor's degree. We can also see that only 80% of the students enrolled in four-year schools have attained bachelor degrees by age 25.

[2]The College Scorecard provides the online advertised tuition (sticker price) of each school, number of undergrads, average SAT scores, average faculty salary of professors, whether the school is a public institution, and whether the school is four-year bachelors granting institution.

Although the ELS2002 is a vast resource, the most important features of the ELS2002 are that I observe each student's entire application set, admissions outcomes, and enrollment decision. The median student applies to only 2 or 3 schools. Table 3 shows that students from wealthier families tend to apply to more schools, and these schools tend to be more expensive, more likely to offer bachelor degrees, and have greater average SAT scores. Wealthier students also have a greater high school GPA.

Not only does this data reveal how students choose between schools during the enrollment process, but it also shows how students choose their application sets.[3] Since the Cal Grant incentives students to enroll in-state, my dataset should contain students that seriously consider enrolling out-of-state. Figure 5 shows that although only 13.6% of enrollments were out-of-state, 19.4% of application portfolios included at least one out-of-state school. Figure 6 shows that students value the "location" of the school the most. However, they do not qualify whether "location" refers to the distance from home or school environment. I partially capture this location preference with distance-from-home and in-state as school attributes, but "location" undoubtedly also refers to unobserved preferences. We also see that cost ranks just as highly as the program and reputation of a school.

Now we come to the first data issue that I confront in my analysis: omitted variable bias. Just as students are more likely to apply to schools that would admit them, schools are more likely to admit students who are more likely to enroll. There might be something I don't observe about the exchange between students and schools that signals their preferences for each other. For example, I don't see the application essay or any recruitment from college sports teams. Without accounting for this omitted variable bias, I might incorrectly attribute some of the observed relationship between admissions and enrollment to a causal interpretation. The same omitted variables issue exists for school scholarships. Subsection 4.1 discusses how I deal with omitted variable bias.

I only observe a noisy measure of net tuition for the enrolled school. For the student's enrolled school, the ELS2002 records the amount of tuition/fees paid for by grants/scholarships (these survey responses are displayed in Figure 2). The students answered the question "For your first term at your first postsecondary institution, what proportion of tuition and fees were paid for by grants and scholarships?" with a multiple choice selection comprised of: "All","At least half but not all", "Less than half", and "None". I translate these answers to percentages: "100%","66%","33%", and "0%" respectively. I calculate the net

[3]The two main competitors to the ELS2002 are the National Longitudinal Survey of Youth 1997 and the High school Longitudinal Study of 2009. Neither collect the entire application set of students, nor do they collect information about the net tuition.

tuition by multiplying the sticker-price by this percentage. The data shows that students with greater GPA and less income pay less net tuition in Figure 7.

My second data issue is the mismeasurement of net tuition. The lack of datasets with accurate grant and scholarship data in conjunction with application and enrollment data has prevented previous authors from estimating the impact of net tuition on the matches between schools and students. Subsection 4.3 discusses how to estimate the causal impact of net tuition in the presence of measurement error.

The third data issue is common across many papers that examine choices – even though students know the net tuition of all the schools to which they were admitted, only the net tuition of the enrolled school is in the data set. This issue is mitigated for me because I do observe each student's option set when they are deciding between schools for enrollment. Furthermore, for each school that admitted the student, I also see separate indicators of whether the student received a grant, scholarship, or tuition waiver. I confront the selected observation of net tuition for the enrolled school, and admissions for the applied schools in Subsection 4.4.

The two main sources of government student aid in California are the California Cal Grant and the Federal Pell Grant.[4] As shown in Figure 1, on average, the Cal Grant covers 37% of University of California Tuition and Fees, is received by 17% of graduating high school seniors, and accounts for 13% of all postsecondary funding by the state. The government also provides subsidized loans to students. PLUS loans (backed by parents) had no upper limit, and Stafford and Perkins loans (backed by the student) had annual maximums that totaled $9,500.

The California Cal Grant is central to my strategy for identifying the causal impact of grants on the college market, and the causal impact of net tuition on enrollment. The ELS2002 allows me to model governmental student aid in great detail. I observe FAFSA applications for students that applied for federal student aid, and complete high school transcripts for every student. The FAFSA applications and the high school transcripts are necessary for determining each student's Cal Grant eligibility. I discuss how to use a merit-based threshold in the amount of Cal Grant aid to identify the causal impacts in Subsection 4.2. More than half of students with parental income less than $50,000 receive some student aid from the government (Figure 8), and nearly three-quarters of these students receive some student aid.

Note that students only apply for financial aid with a vague idea of how much funding they will receive. However, schools make admissions and scholarship decisions with full knowledge of each student's government student aid. Students receive the admissions

[4] The Pell Grant is further discussed in Appendix A.

decisions, and all the financial assistance – both government and school funded – in one admissions packet. Figure 9 shows that every school admission packet includes a net tuition calculation that subtracts governmental grants and school scholarships from the sticker price of schools.

As discussed in (Bettinger et al., 2012; David Deming, 2009), students are resistant to filing FAFSA and applying to too many schools. Only 45% of students with parental income less than $50,000 filed for federal student aid (FAFSA), and the average student applied to only three schools (Figure 10). I should take into account the non-monetary cost student applications and FAFSA filing.

Many low-income students didn't file FAFSA. The chief reason for not filing was because students thought they could pay without aid (Figure 11). The second most important reason (even among low-income students) was that the students didn't deem themselves eligible. The ELS2002 survey did not directly obtain US citizenship status from their participants, but they did survey the parents about the birthplace of their children. Although only 79% of parents responded to the survey, 81% of respondents indicated that their child was born in the United States.

3 Model

This section details how I model the student application and enrollment decision. I show a timeline of the decisions and describe how parameters interact with student-school attributes to lead to the final match between students and schools.

The governmental grant policy g_{ij} is fixed and known by both students and schools. Students already know whether they qualify for the Cal Grant A or the Cal Grant B program when they are applying and should be able to forecast how much governmental aid they would gain by filing a FAFSA application. As described in Subsection 3.1, I associate a non-monetary cost to obtaining all the information required to file the FAFSA application, because some student's don't file FAFSA even they would have benefited from filing.

At the start of the application process, the schools J set their admissions and net tuition policies. These policies depend on student and school attributes, governmental aid, and school preferences: (η, η_a). Students don't know school preferences (η, η_a) when they are applying, so they must forecast the expected admissions and net tuitions for each school.

Students I know observable school and student attributes and their preferences for each school ε_{ij}. Given the school policies, students decide on their application set, O_i, and whether they file FAFSA, d_{fi}. In order to make this decision, students need to choose the combination (O_i, d_{fi}) which maximizes their expected utility under the uncertainty inherent in school preferences (η, η_a).

After students apply to each school and file FAFSA, schools reveal actual admissions and net tuitions. Now each student makes her enrollment decision j_i^* from the application set: O_i^*.[5]

3.1 Student Decision

Here I specify how each student i values school j relative to her outside option.

Assumption 1. *Each student* i *values school* j *relative to her outside option as below:*
$V(\chi_{ij}, c_{ij}, \varepsilon_{ij}) = \chi_{ij}\alpha + c_{ij}\beta + \varepsilon_{ij}$

V represents how the student values each school relative to the outside option $\varnothing$: $V(\varnothing) = 0$. χ_{ij} is the net tuition, and ε_{ij} represents student preferences for each school. c_{ij} represents student-school level characteristics: for the student: GPA, income, and for the school: sticker tuition, distance, mean SAT score, mean faculty salary, and whether the school is in California, is a public institution, and whether it is a four-year school.

In Assumption 1, α tells me how net tuition impacts student valuation and choice probabilities. Estimating α in an asymptotically unbiased manner is central for making causal statements about how changes in net tuition impacts changes in prices.

I specify an indirect utility function because I am interested in the estimation of choice probabilities. In Appendix B, I break Assumption 1 into three smaller assumptions by showing a lifetime maximization problem which can microfound an indirect utility of this form. The biggest underlying assumption is that students can borrow enough to attend any school. The most straightforward argument for these loose borrowing constraints comes from the presence of the PLUS loan, which was loans from the government to the schools and had an upper bound equal to the cost of attendance (COA) of the school. Recall that the COA is a school reported figure that is supposed to sum to tuition and fees plus anticipated housing and supply costs.

Figure 12 shows that 198 out of 910 enrolled California students took loans to attend the first year of college. The average loan taken for the first year was less than \$8,000. 40% of student loan packages were at least partially funded with PLUS loans.[6]

[5]The option set O_i^* only includes the schools that admitted the student. Writing O_i^* is without loss of generality because removing an item from the option set is the same as raising its price to infinity.

[6]Brown et al. (2011) shows that it may be essential to model parents and students separately for school funding. Since Stafford and Perkins loans also have requirements for Adjusted Gross Income to be less than a certain threshold, some students from middle-income families may fall into a valley where they don't receive Pell or Cal Grants, and they also don't qualify for Stafford or Perkins loans. If parents also

The student enrolls in the school that maximizes her indirect utility. For brevity I assume the outside option $\varnothing \in O_i^*$, where $V(\varnothing) = 0$.

$$\max_{j \in O_i^*} V(x_{ij}, \varepsilon_{ij}) \tag{1}$$

During the application step, each student compares all possible application sets and chooses the set that provides the highest expected utility. There is a cost to applying to each additional school and filing FAFSA. The cost is necessary to explain why students don't apply to every school, and the FAFSA penalty is necessary to justify why every student doesn't file FAFSA.

Assumption 2. *Students evaluate the cost of filing FAFSA* d_f *and applying to application set* O *as follows:*
$C_A(|O|, d_f) \equiv \beta_A + \beta_{Ao}|O| + \beta_{Ad} d_f$

To feasibly model the decision for which application set to choose, I construct a consideration set $\mathcal{J}_i$ for each student. The consideration set is the union of the schools that the student applied to with the nearest 2-year and 4-year schools, and the most popular 6 in-state and 8 out-of-state schools. The most popular schools are displayed in Figure 13. To keep the computation feasible, I only considered 40 random combinations of schools that are within two of the size of the student's actual application set.

The student takes expectations over school preferences η to find the value of filing FAFSA (d_f) and applying to set $O \subset J$:

$$W(O, d_f, c_i, \vec{\varepsilon}_i) \equiv E_{\vec{\eta}, \vec{\eta}_a}\left(\max_{j \in \{\emptyset, O\}} V(\chi(\eta_{ij}, \eta_{aij}), c_{ij}, \varepsilon_{ij}) \right) - C_A(|O|, d_f)$$

The optimal application and FAFSA decision solves:

$$\max_{(O, d_f) \in \mathbb{P}(\mathcal{J}_i) \times \{0,1\}} W(O, d_f, c_i, \vec{\varepsilon}_i) \tag{2}$$

Recall that net tuition and admissions are functions of school preferences (η, η_a). The school decision is discussed in the next section.

don't contribute their Expected Family Contribution (which isn't a legal obligation), then these students cannot afford full tuition at many schools.

3.2 School Policies

The school's maximization problem can take many different forms. Rothschild and White (1995); Epple et al. (2006); Fu (2014) all agree that it would make sense for the education production to depend on both student ability and monetary inputs. Appendix C shows how to model a school's objective function as a maximization problem subject to capacity, revenue, and diversity constraints. The key takeaways from modeling the true objective function of the school are: (1) a binding capacity constraint means some students rejected,[7] (2) a binding revenue constraint means some students have to pay nonzero net tuition, and (3) a binding diversity constraints means some races and genders are admitted more, or receive better net tuition offers even if the actual objective function of the schools don't include race or gender.

Instead of specifying the entire maximization problem of schools, I will assume that the symmetric admission and net tuition policies of schools take the following form:

Assumption 3. *In the symmetric subgame perfect equilibrium, the school policies for admissions ($e = \mathbb{1}(s > 0)$), and net tuition (χ) take the following form:*

$$\begin{aligned} s &= z\pi_a + g\gamma_{ag} + c\gamma_a + \eta_a \\ \chi &= z\pi + g\gamma_{tg} + c\gamma_t + \eta \\ &= \textit{Tuition} - \underbrace{\Big(z + g\Big)}_{\textit{Gov. Grant}} - \underbrace{\Big((-c\gamma_t) - z(\pi + 1) - g(\gamma_{tg} + 1) - \eta\Big)}_{\textit{School Scholarship}} \end{aligned}$$

Where the student is admitted $e = 1$ if $(s > 0)$, χ is the net tuition, z is the increase in grant aid from being above the $CalGPA$ threshold, and g are the other government grants. I discuss the $CalGPA$ threshold further in Subsection 4.2. This linear approximation of the subgame perfect strategies which captures how net tuition and admissions are changed by an increase in cal grant aid at a threshold (z). Schools have full knowledge of government grants before they decide to offer any of their scholarships. Therefore government grants can crowd out school scholarships – π can have a magnitude that is less than 1.

I use these two specifications to model how students predict potential net tuition offers and admission probabilities for potential school matches. In Assumption 6 below I allow student preferences (ε) to correlate with school preferences (η_a, η). Students can know

[7] Many schools in my dataset have open admissions, which means that their capacity constraint isn't binding. When is school is designated as having open admissions according to the college scorecard, then I forgo estimating their admissions policy.

something about potential offers from schools beyond what I observe.

I implicitly assume that students' beliefs about the general equilibrium game played between students and schools agree with the actual outcomes observed in the data because these specifications are formed from the actual outcomes. Buchinsky and Leslie (2010) shows that expectations of labor market outcomes formed long before outcome realization may not agree with the information available at the time.[8] In my application, the student isn't forecasting so far into the future to guess potential admissions and scholarship outcomes. Note that I use school characteristics from the previous year for all school characteristics except the sticker price and COA – the student forms expectations based on information available at the time of application.

4 Identification Strategy

Now that I have parameterized student and school decisions above in Section 3, I can discuss how I confront three data issues: (1) omitted variable bias, (2) measurement error, and (3) nonresponse. I also discuss how a threshold in Cal Grant aid eligibility can serve as a source of randomization in net tuition.

4.1 Omitted Variable Bias

I can't estimate a causal effect from two students with different net tuition offers. Imagine two students, Beth-1 and Beth-2, with the same observed characteristics who are choosing between UCLA or not enrolling at all. Let's say Beth-1 is offered a $5,000 scholarship and enrolls, but Beth-2 isn't offered a scholarship and didn't enroll.

I cannot use this data as evidence that the scholarship (decreased net tuition) caused Beth-1 to enroll because the scholarship could have been due to an omitted variable that concurrently caused UCLA to offer the scholarship, and Beth-1 to enroll at UCLA. A possible omitted variable is if Beth-1 was recruited to the UCLA soccer team, and this recruitment isn't observed in my data. There are many other possible reasons: perhaps Beth-1 did well in an interview, or perhaps Beth-1 knew of an obscure UCLA scholarship.

The biggest issue is the matching problem in determining the impact of net tuition on college choice. If there is an omitted variable that affects both net tuition and college choice – such as an application essay – then I may be incorrectly attributing a causal interpretation to some of the relationship between net tuition and college choice. For

[8] A better approximation of expectation formation would be to use the outcomes observed by students from a grade above this cohort or use aggregate labor market statistics from the Current Population Survey.

example, if Beth-1 and Beth-2 have the same observed characteristics, and only Beth-1 receives a $5,000 scholarship and enrolls, then I cannot say that the $5,000 caused Beth-1 to enroll because Beth-1 could have written an amazing college application essay that successfully conveyed her strong preference for the school, and in return the school gave Beth-1 a $5,000 award for "best" application essay.

Government grants impact student choices through net tuition. As you recall, net tuition is the online advertised tuition minus government grants and school scholarships. To estimate the impact of grants on college choice, I must estimate two effects: the impact of grants on net tuition and the impact of net tuition on college choice.

Since observed characteristics completely determine government grants, I do not have a situation where Beth-1 and Beth-2 have the same observables, but Beth-1 gets more government grants. I must find a source of variation in government grants that is arguably random. This is where the $CalGPA$ threshold of the Cal Grant comes in.

4.2 Cal Grant Threshold

A student is eligible to obtain much more financial aid from the Cal Grant program if her $CalGPA$ is greater than 3. The $CalGPA$ is calculated using the unweighted GPA from academic courses taken during the sophomore and junior years of high school. In 2004, the Cal Grant A program yielded a maximum of $8,322 for private schools and $6,141 for public schools. Only students who enroll in-state can receive the Cal Grant. The Cal Grant A provided more funding than the Cal Grant B, which yielded a maximum of $1,551 and was for students with $CalGPA$ between 2 and 3.

The Cost of Attendance (COA) determines the maximum amount of government student aid a student can receive to attend each school. Each school sends its COA to the government, and it can differ for in-state out-of-state residents. The COA includes the sticker price, on-campus room and board, and allowances for supplies and dependent care.[9]

Students must file the FAFSA to obtain any governmental aid.[10] When a student files the FAFSA, she provides her parents' and her tax returns. The government uses this information to calculate each student's Adjusted Gross Income (AGI) and Expected Family Contribution (EFC). The amount of money each student can receive from the Cal Grant program depends on each student's financial need for each school. Financial need is calculated as each school's COA minus the student's EFC. The student's parents must also

[9]The cost of attendance at UCLA in 2019 was $35,791 (for in-state residents) while annual tuition was $11,442 and fees were $4808. I recover the costs of attendance from FAFSA applications.

[10]The Pell Grant is discussed in Appendix A.

have Adjusted Gross Income (AGI) less than \$67,600 for the student to be financially eligible.[11] The actual amount awarded to each student from the Cal Grant program is the lesser of the maximum award and the financial need.

Recall from Assumption 3 that the increase in grant aid due to the threshold (z) impacts admissions and net tuition in a homogenous, linear manner: (π_a, π).

If Cal Grant aid were a binary treatment and assigned at random, then we could estimate the average impact of increasing Cal Grant aid on net tuition by subtracting the mean of the treated group from the mean of the untreated group. My situation here differs from the ideal case in three ways: (1) Cal Grant aid eligibility isn't randomly assigned; (2) the amount of Cal Grant aid each student is eligible for is a continuous variable; and (3) only the students that have filed FAFSA receive cal grant aid. I will discuss issues (2) and (3) first.

(2) In the idealized experiment with a randomized continuous treatment, I can either non-parametrically model the impact by binning the data and then recording the average net tuition for each level of Cal Grant aid, or I can assume the functional form of the impact of Cal Grant aid on net tuition and then match the data to the model. I assume that the amount of Cal Grant Aid received linearly affects the net tuition (Assumption 3) and that this effect is homogeneous across student-school pairs.

(3) In my setting, the amount of Cal Grant aid received is only partially determined by whether the student's CalGPA is greater than 3. Since each student must also file the FAFSA application to receive aid, I have a situation of imperfect compliance. My assumption of a constant treatment effect is already much stronger than a monotonicity assumption that would be required in a situation with heterogeneous treatment effects (as discussed in Imbens and Angrist (1994))

(1) For the Cal Grant A eligibility to be similar to a random assignment, I make a regression discontinuity type assumption (Lee and Lemieux, 2010) that students don't precisely control their CalGPA in the region $[2.8, 3.2]$:

Assumption 4. *Conditional on observed controls and when CalGPA in* $[2.8, 3.2]$*:*
CalGPA *is independent from preferences.*
$\mathsf{CalGPA} \perp\!\!\!\perp (\eta, \eta_a, \varepsilon) | c_{fafsa}, d_f$

Where c_{fafsa} are controls related to the student's FAFSA application, and $d_f = \mathbb{1}_{\text{filed FAFSA}}$ is whether the student filed FAFSA. We can examine whether Cal Grant A eligibility is as good as random in the neighborhood of the CalGPA threshold by checking if characteristics are balanced across this threshold. Figure 14 provides evidence that that the CalGPA

[11] I do not use this need-based threshold because every student that has AGI just near \$67,600 would have received negligible amounts of aid anyway.

threshold is not associate with any trend in characteristics that are related to the outcome of net tuition except for GPA. We see that the 3.0 threshold in CalGPA is not associated with any changes in the online advertised tuition, the average SAT score, or the adjusted gross income of the students in my sample. Furthermore, the urbanicity of students does not impact his location relative to the $CalGPA$ threshold.

The actual Cal Grant Aid amount received is determined by whether the student is above the threshold ($\mathbb{1}_{CalGPA>3}$). If the student were above the threshold, she would receive Cal Grant A: $q_A(c_{fafsa})$, and if she were below the threshold, she would receive Cal Grant B: $q_B(c_{fafsa})$. As discussed above, the two structural equations of aid amounts are constant with respect to the grade:

$$q_A(c_{fafsa}) = \mathbb{1}(same-state)max(0, min(COA-EFC, 8322\mathbb{1}(public)+6141(1-\mathbb{1}(public)))) \tag{3}$$

$$q_B(c_{fafsa}) = \mathbb{1}(same-state)max(0, min(COA-EFC, 1551)) \tag{4}$$

Let z_e be the additional Cal Grant funding a student is eligible for:

$$z_e = q_A(c_{fafsa}) - q_B(c_{fafsa}) \tag{5}$$

Note that Cal Grant eligibility (z_e) is entirely determined by observed characteristics c_{fafsa} and the $CalGPA$. Now we can define the actual additional funding received as the interaction between the additional funding the student is eligible for multiplied by the student's decision to file FAFSA:

$$z = z_e d_f \tag{6}$$

Then we also have the key independence of z:

Lemma 1. $z \perp\!\!\!\perp (\eta, \eta_a, \varepsilon) | c_{fafsa}, d_f$

The independence condition Lemma 1 is the central point of this section. This condition allows me to use z as a surrogate for shifts in cal grant funding that is uncorrelated with school preferences (omitted variables). In turn, I will also use z as a surrogate to observe shifts in net tuition that is uncorrelated with student preferences. Given this identification condition, I can estimate the impact of Cal Grant aid on net tuition π, and subsequently, the impact of net tuition on college choice α.

The amount of government aid received by a student is completely determined by observed characteristics of the student and the school: (EFC, AGI, COA, public, same-state,

CalGPA). Since observables completely determine governmental aid, two students with exactly identical observed characteristics would receive the same amount of government aid, and I am left with no variation in the net tuition.

The increase in Cal Grant aid at the CalGPA threshold can help solve this issue if we believe that the increase in aid is independent from any unobserved reasons for the school to give increased scholarship (η), or for Beth-1 to enroll at UCLA ε.

Let's say Beth-1 and Beth-2 both tried their hardest in school, and are observationally identical except that Beth-1 has $\mathsf{CalGPA} = 3.01$ and Beth-2 has $\mathsf{CalGPA} = 2.99$. Then the Cal Grant program gives an additional \$6,500 to Beth-1 because her CalGPA is greater than 3. I could claim the additional $\$6,500$ in Cal Grant aid is not due to any unobserved reason as long as I believe Beth-1's and Beth-2's location relative to the threshold is random.

4.3 Measurement Error

As discussed in Section 2, the net tuition that I observe is imputed from a categorical survey response. If the student responded that either "all" or "none" of her tuition and fees were paid by grants or scholarships, then there is no measurement error. However, if a student responded that either "more than half but not all" or "some but less than half" of their tuition and fees paid by grants and scholarships, then I impute the net tuition by multiplying the actual tuition and fees of the school by 33% and 66% respectively. The measurement error is the difference between my imputed net tuition value and actual net tuition (unobserved to me).

The increased Cal Grant aid from locating above the CalGPA threshold can help with the measurement error. Continuing with the example of Beth-1 and Beth-2 as above. Imagine I observe fifty Beth-1s with $\mathsf{CalGPA} = 3.01$ and fifty Beth-2s with $\mathsf{CalGPA} = 2.99$. The Beth-1s have an additional $\$6,500$ in Cal Grant aid as before.

As long as Beth-1s' and Beth-2s' position relative to the CalGPA threshold is unrelated to the measurement error, then the Beth-1s and Beth-2s have the same average measurement error. Then I can estimate the impact of the threshold $\$6,500$ on net tuition by subtracting average imputed net tuition from the two groups. Note that I can estimate the impact as long as the average measurement error is the same across the two groups – the measurement error need not be mean zero.

I formally define the difference between the noisy observed net tuition x and the actual net tuition χ as the measurement error m.

$$x = \chi + m \tag{7}$$

If the measurement error is independent from the increase in Cal Grant aid at the threshold z, then the average impact of z on the noisy net tuition x will be the same as the average impact of z on the actual net tuition χ:

Assumption 5. *The measurement error is independent of the threshold surrogate z.*
$m \perp\!\!\!\perp z$

4.4 Selected Sample

I only observe the survey response for the amount of tuition/fees paid for by grants/scholarships for the student's enrolled school.[12] Likewise, I only observe the admissions outcomes for the schools in the student's application set. If I do not account for the sample selection, then I will make biased estimates of the impact of Cal Grant aid, and any imputed net tuitions or admissions probabilities will be biased as well.

To establish the intuition behind correcting for a selected sample, imagine that you are Beth's parent and that you would like to know her average weekly spending. You know that each week Beth either buys ice cream or doesn't and on the weeks that she buys ice cream, she spends an extra \$10. Finally, you know that Beth only reports her spending when she doesn't buy ice cream.

Let's say the Beth told you her spending half of the time, and her reported spending averaged \$5. You could naively guess that her weekly spending also averaged \$5. But you know that Beth averaged \$15 on weeks when she didn't report, so you successfully estimate Beth's overall spending averaged \$10.

The functional forms of school policies Assumption 3 and student valuations Assumption 1 together with a distributional assumption allow me to account for observing a selected sample. My distributional assumption is that the unobserved school preferences (η, η_a), the unobserved student preferences (ε), and the measurement error (m)[13] are jointly distributed multivariate normal:

[12]For each school that the student was admitted to, I do observe some information regarding the net tuition: I see separate indicators of whether the student received a grant, scholarship, tuition waiver, or work-study.

[13]Since the measurement error is imputed from a bounded categorical variable, it actually cannot be distributed normal. However, keeping the normal assumption simplifies the mathematics and allows me to express the impact of having larger measurement error on my estimates. The bounds on the measurement error are useful for stating the degree of confidence in my results. Leamer (1987) shows how knowing the size of the measurement error translates to regions of maximum likelihood estimates.

Assumption 6. *The unobserved preferences are i.i.d multivariate normal across student-school pairs* (i, j)*:*

$$\begin{pmatrix} \varepsilon \\ \frac{\eta}{\sigma_a} \\ \eta_a \\ \frac{m-\mu_m}{\sigma_m} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho & \rho_a & 0 \\ \rho & 1 & \rho_s & 0 \\ \rho_a & \rho_s & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \right)$$

This distributional assumption parameterizes the potential correlation between student preferences ε and school preferences (η, η_a). I would overestimate the negative impact of net tuition on college choice if $\rho < 0$. Likewise, we incorrectly estimate the impact of increased admissions probability on college if $\rho_a \neq 0$.

Of the three preferences, I only let the net tuition preference have nontrivial variance σ_a because the other two preferences are specified to match choice probabilities. If I doubled the standard deviation of η_a, ε, I could always also double the coefficients π_a, α to get the same choice probabilities.

The mean μ_m and variance σ_m^2 of the measurement error don't impact the value of my final estimates of (α, π, π_a) because I have already assumed the measurement error is independent from the increased Cal Grant aid at the threshold (z) with Assumption 5.

A larger variance (σ_m^2) will lead to greater standard errors in our estimates of key model parameters (α, π, π_a). Although I have already assumed m to be normal, since the noisy measure of net tuition x is actually imputed from a categorical variable that's bounded below by 0 and above by the full tuition and fees, we can have worst-case bounds on the size of the variance of the measurement error σ_m^2.

5 Estimation

To estimate this model, I specify the likelihood of observing the data and break the estimation into two stages. The first stage estimates the school-side parameters: the impact of grant aid on net tuition (π) and admissions (π_a), and the correlation between the admissions and net tuition decision (ρ_s). The second stage estimates the student-side parameters: the impact of net tuition on college choice (α), and the correlations between student and school preferences (ρ, ρ_a).[14]

For each student I observe the application set and FAFSA filing decision (O^*, d_f^*), the admissions outcome for all applied schools: $[e_j^*]_{j \in O^*}$, the enrollment choice j^*, and a noisy measure of net tuition for the enrolled school: x_{j^*}.

[14] This section omits discussion of control variables (c), and also omits student subscripts (i).

I express the overall likelihood as the multiplication of a school-side probability and a student-side probability:

$$\begin{aligned}
&Pr_{stu,sch}(\text{ apply } (O^*, d_f^*) \cap \text{ } j^* \text{ decide } x_{j^*} \cap [\text{j admits } e_j^*]_{j \in O^*} \cap \text{ enroll } j^*) \\
=&Pr_{sch}(j^* \text{ decide } x_{j^*}) Pr_{sch}([\text{j admits } e_j^*]_{j \in O^*} | j^* \text{ decide } x_{j^*}) \times \qquad (8) \\
&Pr_{stu}(\text{ apply } (O^*, d_f^*) \cap i \text{ enroll } j^* | j^* \text{ decide } x_{j^*} \cap [\text{j admits } e_j^*]_{j \in O^*}) \qquad (9)
\end{aligned}$$

Upon taking logs, we see that the school-side (Equation 8) is separated from the student side (Equation 9). I discuss each estimation in detail in the following subsections.

5.1 School-side parameters

I observe school admissions $[e_j^*]_{j \in O^*}$ and net tuition decisions x_{j^*}. Recall that admissions and net tuition follow the following structural equations:

- Admission: $e_j = \mathbb{1}(z_j \pi_a + \eta_{a,j} > 0)$
- Net tuition: $\chi_j = z_j \pi_j + \eta_j$
- Measurement Error: $x_j = \chi_j + m_j$

I would like to estimate the impact of cal grant aid on net tuitions (π) and admissions (π_a). I also need to discover the correlation between the school preferences: $\rho_s = corr(\eta_a, \eta)$.

The likelihood for the school-side parameters breaks into three parts:

$$\begin{aligned}
&Pr_{sch}(j^* \text{ decide } x_{j^*}) Pr_{sch}([\text{j admits } e_j^*]_{j \in O^*} | j^* \text{ decide } x_{j^*}) \qquad (10) \\
=& \prod_{j \in \{O^* \setminus j^*\}} Pr(z_j \pi_a + \eta_{a,j} > 0)^{(e_j^*)} Pr(z_j \pi_a + \eta_{a,j} < 0)^{(1-e_j^*)} \times \qquad (11) \\
&Pr(x_{j^*} = z_j \pi_{j^*} + \eta_j + m_j) \times \qquad (12) \\
&Pr(z_{j^*} \pi_a + \eta_{a,j^*} > 0 | x_j = z_{j^*} \pi_{j^*} + \eta_{j^*} + m_j) \qquad (13)
\end{aligned}$$

Now I can take full advantage of multivariate normality (Assumption 6). Upon taking logs Equation 11 is a probit specification, Equation 12 is an ordinary least squares. Equation 13 is a type-2 Tobit specification which identifies the correlation between school admissions and net tuition preferences (ρ_s). The conditional distribution is easy to express once we remember that η_a conditioned on $(\eta + m)$ is also normally distributed.

The likelihood function is not concave because I am trying the estimate variances (σ_η) and correlations (ρ_s). When a function isn't concave, any maximum obtained from an estimation procedure may only be a local maximum. However, Olsen (1982) notes that the likelihood is concave conditional on these two parameters. Therefore the estimation strategy is clear:

1. Make a grid $[\widetilde{\sigma_\eta}, \widetilde{\rho_s}]_{1,\ldots,K_g}$.

2. Maximize likelihood (Equation 10) to obtain $(\tilde{\pi}, \widetilde{\pi_a})$ for each pair $(\widetilde{\sigma_\eta}, \widetilde{\rho_s})_{k_g}$.

3. My estimate $(\hat{\pi}, \widehat{\pi_a}, \widehat{\rho_s}, \widehat{\sigma_\eta})$ is the set $(\tilde{\pi}, \widetilde{\pi_a}, \widetilde{\sigma_\eta}, \widetilde{\rho_s})$ with the highest likelihood.

5.2 Student-side parameters

To estimate the impact of net tuition on student choice (α), I modify the standard simulated maximum likelihood method to allow student preferences (ε) to correlate with school preferences (η, η_a) according to ρ and ρ_a respectively. Without estimating these correlations, my estimate of α will be asymptotically biased.

At this stage of the estimation, I already have estimates for the impact of net tuition on admissions $(\widetilde{\pi_a})$ and net tuition $(\tilde{\pi})$. I have also estimated the variance of the net tuition preference $(\widetilde{\sigma_\eta})$, and the correlation between the two school preferences $(\widetilde{\rho_s})$.

Before the estimation process, I simulate the preferences:

$$\begin{bmatrix} \omega_\varepsilon \\ \frac{\eta}{\widetilde{\sigma_\eta}} \\ \eta_a \\ m \end{bmatrix} \sim \mathrm{Normal}(\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}), \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & \widetilde{\rho_s} & 0 \\ 0 & \widetilde{\rho_s} & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix})$$

Where I will use ω_ε to generate the student preferences (ε).

The student-side data consists of the application set (O^*), FAFSA filing decision (d_f^*), and enrollment choice (j^*). Recall that the student-side likelihood is conditioned on the school decisions:

$$\mathrm{Pr}_{stu}(\text{apply } (O^*, d_f^*) \cap \text{ enroll } j^* | j^* \text{ decide } x_{j^*} \cap [j \text{ admits } e_j^*]_{j \in O^*}) \tag{14}$$

Let $O^{**} \equiv \{(j \in O^*) \cap (e_j^* = 1)\}$ denote the observed admitted set. The probability of a specific enrollment choice is:

$$\begin{aligned}\Pr(\text{ enroll } j^*) &= \Pr\Big(V(\chi_{j^*}, \varepsilon_{j^*}) \geqslant \max_{j \in O^{**}} V(\chi_j, \varepsilon_j)\Big) \\ &= \Pr\Big(\chi_{j^*}\alpha + \varepsilon_{j^*} \geqslant \max_{j \in O^{**}} (\chi_j \alpha + \varepsilon_j)\Big) \\ &= \Pr\Big(z_{j^*}\pi\alpha + (\eta_{j^*}\alpha + \varepsilon_{j^*}) \geqslant \max_{j \in O^{**}} (z_j \pi\alpha + (\eta_j \alpha + \varepsilon_j))\Big) \end{aligned} \tag{15}$$

The probability of observing a specific application set is:

$$\begin{aligned}\Pr(\text{ apply } (O^*, d_f^*)) &= \Pr\Big(W(O^*, d_f^*, V, \vec{\varepsilon}) \geqslant \max_{(O,d_f) \in \mathbb{P}(\mathcal{J}) \times \{0,1\}} (W(O, d_f, V, \vec{\varepsilon}))\Big) \\ = \Pr\Big(&E_{\vec{\eta},\vec{\eta}_a}\big(\max_{j \in \{\emptyset, O^*(\eta_{aj})\}} V(\chi(\eta_j, \eta_{aj}), \varepsilon_j)\big) - C_A(|O^*|, d_f^*) \geqslant \\ &\max_{(O,d_f) \in \mathbb{P}(\mathcal{J}) \times \{0,1\}} \big(E_{\vec{\eta},\vec{\eta}_a}\big(\max_{j \in \{\emptyset, O(\eta_{aj})\}} V(\chi(\eta_j, \eta_{aj}), \varepsilon_j)\big) - C_A(|O|, d_f)\big)\Big) \\ = \Pr\Big(&E_{\vec{\eta},\vec{\eta}_a}\big(\max_{j \in \{\emptyset, O^*(\eta_{aj})\}} z_j\pi\alpha + (\eta_j\alpha + \varepsilon_j)\big) - C_A(|O^*|, d_f^*) \geqslant \\ &\max_{(O,d_f) \in \mathbb{P}(\mathcal{J}) \times \{0,1\}} \big(E_{\vec{\eta},\vec{\eta}_a}\big(\max_{j \in \{\emptyset, O(\eta_{aj})\}} z_j\pi\alpha + (\eta_j\alpha + \varepsilon_j)\big) - C_A(|O|, d_f)\big)\Big) \end{aligned} \tag{16}$$

Equations (15) and (16) show how I plug-in the increased Cal Grant aid at the threshold ($z\pi$) for the net tuition (χ) instead of using a control function approach as suggested by recent literature (Blundell and Powell, 2004; Chernozhukov et al., 2019). Please refer to Appendix D for more information on why the control function approach is incorrect when the first stage dependent variable is measured with error.[15] The key takeaway is that I must be careful to maximize my likelihood with respect to the new error distribution ($\eta\alpha + \varepsilon$) instead of (ε).

The results from a maximum likelihood optimization procedure may only be a local maximum because the likelihood function isn't concave. However, the likelihood is concave conditional on the correlations. Therefore I define a grid $[(\tilde{\rho}, \widetilde{\rho_a})]_{1,\dots,K_g}$ and optimize the likelihood function for each pair of correlations.

For each pair of correlations: $(\tilde{\rho}, \widetilde{\rho_a})_{k_g}$, I can use the simulated preferences $(\omega_\varepsilon, \eta, \eta_a)$ to generate the student preferences ε with the correct correlations.[16]

[15] In the presence of measurement error of net tuition, the observed net tuition and the leftover error is still correlated even with the presence of a "control residual".

[16] The correlations between student preferences and school preferences $(\tilde{\rho}, \widetilde{\rho_a})_{k_g}$ define coefficients ψ, ψ_a which allow me to generate the student preferences according to: $\varepsilon = \frac{\omega_\varepsilon + \eta\psi + \eta_a\psi_a}{\sqrt{1+\psi^2+\psi_a^2}}$.

I must draw from the conditional distribution defined by the student likelihood (Equation 14). Let $[\tilde{\varepsilon}, \tilde{\eta}, \widetilde{\eta_a}, \tilde{m}]_{j,k}$ be the kth simulated draw that satisfies the following conditions:

$$(\eta_{j^*} + m_{j^*} = x_{j^*} - z_{j^*}) \cap (\eta_{aj^*} \geqslant -z_{j^*}\pi_a) \text{ for enrolled } j^* \tag{17}$$

$$(\eta_{ak} \geqslant -z_k\pi_a) \text{ if applied and admitted} \tag{18}$$

$$(\eta_{ak} < -z_k\pi_a) \text{ if applied and rejected} \tag{19}$$

As shown in the expectation in Equation 16, students must guess the school preferences during their application step. These school decisions are normally distributed conditional on student preferences:

$$\left(\begin{bmatrix} \frac{\eta}{\sigma_\eta} \\ \eta_a \end{bmatrix} \middle| \varepsilon = e \right) \sim N\left(\begin{bmatrix} \rho \\ \rho_a \end{bmatrix} e, \begin{bmatrix} 1-\rho^2 & \rho_s - \rho\rho_a \\ \rho_s - \rho\rho_a & 1-\rho^2 \end{bmatrix} \right) \tag{20}$$

For each simulated draw of student preference $\tilde{\varepsilon}_{jk}$, I also need to simulate potential school decisions: $[\widetilde{\eta}_\varepsilon, \widetilde{\eta_{a\varepsilon}}]_{j,k,k_\varepsilon}$ from the conditional distribution defined in Equation 20. These new draws let me approximate the expectation in the application decision:

$$\begin{aligned} W(O, d_f, \vec{\varepsilon}) =& E_{\vec{\eta}, \vec{\eta}_a}\Big(\max_{j \in \{\emptyset, O(\vec{\eta_a})\}} z_j\pi\alpha + (\eta_j\alpha + \varepsilon_j)\Big) - C_A(|O|, d_f) \\ \approx& \widetilde{W}(O, d_f, [\varepsilon, [\widetilde{\eta}_\varepsilon, \widetilde{\eta_{a\varepsilon}}]_{k_\varepsilon \in K_\varepsilon}]_{j\in J}) \\ \equiv& \frac{1}{|K_\varepsilon|} \sum_{k_\varepsilon \in K_\varepsilon} \left(\max_{j \in \{\emptyset, O([\widetilde{\eta_{a\varepsilon k_\varepsilon}}]_{j\in J})\}} z_j\pi\alpha + (\widetilde{\eta}_{\varepsilon j, k_\varepsilon}\alpha + \tilde{\varepsilon}_j) \right) - C_A(|O|, d_f, x) \end{aligned}$$

Where $O([\widetilde{\eta_{a\varepsilon k_\varepsilon}}]_{j\in J})$ denotes simulated admissions outcomes and $\vec{\varepsilon} \equiv [\varepsilon]_{j\in J}$ is an abbreviation for the vector of all school preferences. Finally, I can express the probability that each student applies to (O^*, d_f^*) and enrolls at j^* as a simulated maximum likelihood:

$$\Pr_{stu}(\text{apply to } (O^*, d_f^*) \cap \text{ enroll at } j^* | j^* \text{ decide } x_{j^*} \cap [k \text{ admits } e_j^*]_{j \in O^*}) = \tag{21}$$

$$\int_{\vec{r}} \mathbf{1}\Big(\big(W(O^*, d_f^*, \vec{r}) \geqslant \max_{(O, d_f) \in \mathbb{P}(\mathcal{J}) \times \{0,1\}} W(O, d_f, \vec{r})\big) \cap \big(V_z(z_{j^*}\hat{\pi}, \eta_{j^*}, r_{j^*}) \geqslant \max_{j \in O^{**}} V_z(z_j \hat{\pi}, \eta_j, r_j)\big)\Big) dF_{\vec{\varepsilon}|(17,18,19)}(\vec{r}) \tag{22}$$

$$\approx \frac{1}{|K|} \sum_{k \in K} \left(\frac{\exp(\widetilde{W}(O^*, d_f^*, \left[[\tilde{\varepsilon}, [\widetilde{\eta}_\varepsilon, \widetilde{\eta_{a\varepsilon}}]_{k_\varepsilon \in K_\varepsilon}]_{j \in J}\right]_k)/\tau_1)}{\sum_{(O, d_f)} \exp(\widetilde{W}(O, d_f, \left[[\tilde{\varepsilon}, [\widetilde{\eta}_\varepsilon, \widetilde{\eta_{a\varepsilon}}]_{k_\varepsilon \in K_\varepsilon}]_{j \in J}\right]_k)/\tau_1)} \cdot \frac{\exp([V_z(z_{j^*}\hat{\pi}, \tilde{\eta}_{j^*,k}, \tilde{\varepsilon}_{j^*,k})]/\tau_2)}{\sum_{j \subset O^{**}} \exp([V_z(z_j \hat{\pi}, \tilde{\eta}_{j,k}, \tilde{\varepsilon}_{j,k})]/\tau_2)} \right) \tag{23}$$

Where $V_z(z\hat{\pi}, \eta, \varepsilon) \equiv z\hat{\pi}\alpha + (\eta\alpha + \varepsilon)$ is the student's indirect utility with the increased Cal Grant aid $(z\hat{\pi})$ plugged in for net tuition (χ).[17] When τ_1 and τ_2 approach zero, Equation 23 approaches the integrated frequency in Equation 22.

In summary, the estimation proceeds as below:

1. Make a grid of correlations between student and school preferences $[(\tilde{\rho}, \widetilde{\rho_a})]_{1,\ldots,K_g}$
2. For each pair of correlations $(\tilde{\rho}, \widetilde{\rho_a})_{k_g}$ perform simulated maximum likelihood as described above to estimate $\hat{\alpha}_{k_g}$.
3. Take the set $(\hat{\rho}, \widehat{\rho_a}, \hat{\alpha})_{k_g}$ with the greatest likelihood.

6 Results

I find that each additional dollar of Cal Grant aid leads to an 81 cent reduction in net tuition and that the additional aid doesn't impact the probability of admissions. Subsection 6.2 shows that the net tuition significantly affects college choice and that the use of a threshold identification strategy leads to a smaller estimated impact.

6.1 School Policies

Table 5 shows results from the regression discontinuity specification. The first two columns predict net tuition paid by the student for each student-school enrollment match. The

[17] The term $\left[[\tilde{\varepsilon}, [\widetilde{\eta}_\varepsilon, \widetilde{\eta_{a\varepsilon}}]_{k_\varepsilon \in K_\varepsilon}]_{j \in J}\right]_k$ documents the all the simulated errors. We have K sets of student preferences. Each student preference also have K_ε sets of school preferences (for estimating the application decision).

next two columns show the marginal effects of a probit model of whether a student was admitted to the school. For students with CalGPA between the bandwidth $[2.8, 3.2]$, we have 191 student-school enrollments and 743 student-school applications.

The first coefficient shows each additional dollar of state-based aid received will reduce the net tuition paid by the student by 81 cents. We can see that the intent-to-treat effect of simply having CalGPA greater than 3.0 isn't as strong. Therefore we should take these specification results as evidence that Cal Grant aid reduces the amount of net tuition received. Our results agree with Turner (2017); Long (2004), which finds that each additional dollar of grant aid reduces net tuition by 20%-30%.

We don't see any impact of additional received aid on the probability of admission. The honors weighted GPA matters much more than the CalGPA for the sake of school admissions - a 1 point increase in weighted GPA increases admission chances by 25%. We also see that in-state schools, public schools, and schools with greater average SAT scores all display lower probabilities of admission.

Although the coefficient of increased Cal Grant aid due to the threshold is the only one we can interpret as a causal result, the data suggests that in-state public schools are much cheaper than in-state private schools, and that students would only enroll out-of-state if they receive a discount to do so.

6.2 Model Results

Table 6 describes the results from the simulated maximum likelihood estimation process. I use 1,000 bootstraps to arrive at the standard deviation estimates shown. For the instrumented method, 98% of the bootstrapped net tuition coefficients were negative, and without using the instrument, 100% of the bootstrapped net tuition coefficients were negative. Therefore we have consistently strong evidence that great net-tuition negatively impacts the chance that student applies to or enroll at a school.

In Figure 15, we see that 92% of the bootstrapped estimates of the impact of net tuition on college with the discontinuity based identification were less than the same estimate without the discontinuity based identification. I bootstrapped the entire estimation process 1,000 times to obtain this distribution of coefficients on the net tuition. I estimated the impact of net tuition on the college choice with and without the threshold on each bootstrapped dataset. To estimate the impact of net tuition without the discontinuity based identification, I formed a simpler likelihood that did not plug in the increase in Cal Grant aid at the threshold $(z\pi)$ for net tuition (χ).

Although my attempt to estimate the cost of filing FAFSA and the cost of applying to an additional school did not yield believable results, the ratio of FAFSA Penalty over Additional School Penalty averaged 0.24 and had a bootstrapped standard deviation of a tight 0.12. I can believably claim that the non-monetary cost to filing an additional FAFSA application is about $\frac{1}{4}$ the cost of applying to an additional school.

These coefficients allow me to estimate the importance of net tuition relative to the other observed attributes (such as school SAT score, or average faculty salary). After adjusting the coefficients by the standard deviation of their respective independent variables, I find that net tuition accounts for 51%[18] of the observed variation in the indirect utility function.

7 Policy Changes

This section discusses the efficacy of the Cal Grant by simulating what would've happened in the absence of Cal Grant funding. I go on to analyze the potential impact of making all community colleges "free" by subsidizing $1,000 per community college enrollee.

This article does not ask for new equilibria. A subgame perfect Nash equilibrium is justified by a set of strategies (best response functions) by students and schools. I hold the current strategies constant and estimate how net tuition impacts applications and enrollment for the student. I also hold current strategies constant and estimate how Cal Grant aid impacts admissions and school scholarships for the schools. My predictions ask: At the current strategies, what would happen if I removed the Cal Grant A program? This isn't the same as asking: What would be the new equilibrium be if I removed the Cal Grant A program?

7.1 Removing the Cal Grant

The model estimates of how net tuition changes in response to grant aid (π) and how applications and enrollment responds to changes in net tuition (α) together tell us the impact of changes in Cal Grant aid. Removing the Cal Grant changes the net tuition of every in-state school that a student could apply to (as long as she qualifies for Cal Grant aid).[19] I use the model to generate probabilities of application and enrollment for every potential school. Figure 17a shows the predicted choice probabilities of all Cal Grant

[18] A 95% confidence interval using bootstrapped coefficients yields: [34%,66%].

[19] Recall that every school has a different cost of attendance, so removing the Cal Grant differentially changes the net tuition of every school.

recipients who had CalGPA between 3 and 3.2. All students in this group enrolled in-state, and the model predicts that the average student had an 83% chance to enroll in-state with the Cal Grant.

Figure 17b shows predicted enrollment chances for all students who were both near the CalGPA threshold and were intended to be treated: they had CalGPA between 3.0 and 3.2 and were financially qualified for aid. This group makes up 7% of graduating high school seniors in 2004 – about 35,000 students (30% of seniors had CalGPA greater than 3 overall and financially qualified for aid – 150,000 students.). Removing the Cal Grant reduced the average student's probability to enroll in-state enrollment from 67% to 62%, with more than half the decrease going to no enrollment, and less than half the decrease going to out-of-state enrollment. Table 7 uses the bootstrapped model coefficients to show 90% confidence bounds on the enrollment probabilities as shown in Figure 17. The model predicted enrollment chances could vary by plus or minus 2%.

Recall that we can evaluate the effectiveness of the Cal Grant by comparing its distributional impact to the three goals: (1) increasing overall postsecondary enrollment, (2) encouraging students to enroll in-state, and (3) helping students fund their higher education. One must keep the costs in mind while evaluating the benefits: In 2004, the Cal Grant A program paid out $5,600 per student.[20] The Cal Grant A was awarded to about 6,250 first time enrollees students who had CalGPA between 3 and 3.2. The state paid a total of $35 million for these 6,250 students. The model estimates that the chance to enroll in-state increased from 78% to 83% for Cal Grant recipients. This translates to 380 more students enrolling in-state. If California didn't care at all about helping students fund their college education and was solely focused on increasing in-state enrollment, then the state spent $90,000 per additional in-state enrollee.

My model shows that 93% of Cal Grant A funding to ($260 million) was spent on students that were already going to enroll in-state anyway. Recall from Figure 16 that there may be some deadweight loss from the point of view of the student when the Cal Grant A funding induces her to switch schools. The intuition is that if a student values her current out-of-state option $500 more than her best in-state option, then even though a $600 grant would cause her to switch to enroll in-state, she would only perceive herself to gain $100. My model calculates that only 3% of the increased funding from the Cal Grant A is lost due to students switching in-state. Therefore the Cal Grant A program is quite efficient at helping students fund their postsecondary education.

[20] The entire Cal Grant A, B, and C program paid out $700 million to 194,000 recipients. The Cal Grant A paid $280 million to 50,000 students. I estimate that one-half of Cal Grant A recipients were first-year enrollees and that one-quarter of first time recipients had CalGPA between 3 and 3.2 (6,250 students). These ratios are calibrated from the ELS2002 and the NASSGAP.

Figure 18 shows predicted enrollment chances by race and sex for all students who were both near the CalGPA threshold and were intended to be treated by the Cal Grant A program: they had CalGPA between 3.0 and 3.2 and were financially qualified for aid. It is essential to keep the target group in mind. All of these students had CalGPA greater than 3, which means they averaged better than a B on academic courses in high school (during their sophomore and junior year). Note that I did not include race or sex in estimating the student utility function. Therefore differences in the distributional impact are based on the parental income and grades of the student.

Females were more likely to go to college than Males, and Asian and African American/Black students were more likely to attend college than Hispanic and White students. This means that Females, Asians, and African Americans are the ones who are most likely to choose to enroll in-state even without the Cal Grant A funding. Therefore these three groups have the most to gain from the Cal Grant A program.

From Subsection 7.1 we already know that the Cal Grant A isn't the correct policy to induce greater in-state enrollment. However, it may be interesting to note that Asian and White students are the most likely to respond to increased Cal Grant funding by reducing their chances not to enroll at all.

7.2 Free Community College

A recent 2017 California bill made California community colleges "free" by providing $1,000 in grant aid to all first and second year community college enrollees. Figure 19 simulates a similar policy in 2017 and shows that the enrollment in community colleges increases by 3%, with the majority of the decrease coming from students who would have enrolled in a different four-year school. It isn't surprising to note that making community college free benefits students who would have attended community college even without the incentive the most.

This highlights the potential dangers of only subsidizing community colleges – students may be diverted from attending a four-year college. This may lead to an overall decrease of bachelor degree attainment in the state. However, it still remains to be established whether these diverted students actually attain less wages in the future. Although the number of students not-enrolling decreased from 23% to 22%, policymakers should take care to calculate whether this small decrease in overall enrollment attains the proposed goals of making community colleges free.

8 Conclusion

This article estimates the causal impact of additional state-based aid on the match between students and schools. I use a jump in Cal Grant aid at a GPA threshold as a surrogate for shifts in net tuition that is independent of omitted variables. I construct a model of the application-admissions-scholarship-enrollment matching game between schools and students and show how to use a modified simulated maximum likelihood to obtain consistent estimates of the causal impact of the Cal Grant on college choice. Since government grants impact college choice through net tuition, I also get consistent estimates of the causal impact of net tuition on college choice.

Readers should be careful when extrapolating from my policy experiments because they do not search for new equilibrium and instead hold fixed the policies of schools and the value functions of the students. The estimates in this article are based on California students in 2004 who had a B average in high school and parental income less than $67,600. Finally, a larger dataset with more students and exact net tuition data would allow me to consider different effects.

The Educational Longitudinal Study of 2002 is a panel dataset that also observes the outcomes of the respondents at age 25. Future work will build on the econometric strategies presented here to examine long term outcomes: dropout rates, degree attainment, migration, and wage earnings.

I evaluate the Cal Grant by using the estimated causal impacts to simulate removing the Cal Grant. I find that only 6% of Cal Grant recipients were induced to switch from a different option to enroll in-state. Since 93% of Cal Grant A funding was received by students that were already going to enroll in-state anyway, the Cal Grant only lost 3% of its value to deadweight loss from switching. I find that the Cal Grant helps pay for schooling, but it doesn't encourage more in-state enrollment.

Once I estimate the impact of government aid on net tuition and the impact of net tuition on college choice, I can simulate the distributional effects of a variety of government policies. I find that "free community college" would increase community college enrollment by 3%, with two-thirds of this increase coming from students who were diverted from enrolling in a four-year institution.

Policymakers have been quite interested in reducing the costs of postsecondary education. However, there has been scant evidence on the impact of student-school level net tuition on the equilibrium matches between schools and students. This article presents a strategy for modeling the college enrollment game and identifying the impact of grants and scholarships on college choice.

References

Avery, C. and C. M. Hoxby (2003, 03). Do and should financial aid packages affect students' college choices?

Bettinger, E. (2004). How financial aid affects persistence. In *College choices: The economics of where to go, when to go, and how to pay for it*, pp. 207–238. University of Chicago Press.

Bettinger, E., O. Gurantz, L. Kawano, B. Sacerdote, and M. Stevens (2019, February). The long-run impacts of financial aid: Evidence from california's cal grant. *American Economic Journal: Economic Policy 11*(1), 64–94.

Bettinger, E. P., B. T. Long, P. Oreopoulos, and L. Sanbonmatsu (2012). The role of application assistance and information in college decisions: Results from the h&r block fafsa experiment. *The Quarterly Journal of Economics 127*(3), 1205–1242.

Blundell, R. W. and J. L. Powell (2004). Endogeneity in semiparametric binary response models. *The Review of Economic Studies 71*(3), 655–679.

Brown, M., J. Karl Scholz, and A. Seshadri (2011). A new test of borrowing constraints for education. *The Review of Economic Studies 79*(2), 511–538.

Buchinsky, M. and P. Leslie (2010). Educational attainment and the changing u.s. wage structure: Dynamic implications on young individuals' choices. *Journal of Labor Economics 28*(3), 675–692.

Chernozhukov, V., I. Fernández-Val, and W. K. Newey (2019). Nonseparable multinomial choice models in cross-section and panel data. *Journal of Econometrics 211*(1), 104–116.

Cities of Promise (2019 (accessed 10/21/2019)). States with promise programs. https://citiesofpromise.com/promise-programs-listed-by-state/.

Cohodes, S. R. and J. S. Goodman (2014). Merit aid, college quality, and college completion: Massachusetts' adams scholarship as an in-kind subsidy. *American Economic Journal: Applied Economics 6*(4), 251–85.

David Deming, S. D. (2009, September). Into college, out of poverty? policies to increase the postsecondary attainment of the poor. NBER Working Paper 15387.

Denning, J. T., B. M. Marx, and L. J. Turner (2019, July). Propelled: The effects of grants on graduation, earnings, and welfare. *American Economic Journal: Applied Economics 11*(3), 193–224.

Epple, D., R. Romano, and H. Sieg (2006). Admission, tuition, and financial aid policies in the market for higher education. *Econometrica 74*(4), 885–928.

Fu, C. (2014). Equilibrium tuition, applications, admissions and enrollment in the college market. *Journal of Political Economy 122*(2), 225–281.

Hanushek, E. A. and L. Woessmann (2015). *The Knowledge Capital of Nations*.

Imbens, G. W. and J. D. Angrist (1994). Identification and estimation of local average treatment effects. *Econometrica 62*(2), 467–475.

Leamer, E. E. (1987). Errors in variables in linear systems. *Econometrica 55*(4), 893–999.

Lee, D. S. and T. Lemieux (2010). Regression discontinuity designs in economics. *Journal of economic literature 48*(2), 281–355.

Long, B. T. (2004). How do financial aid policies affect colleges? the institutional impact of the georgia hope scholarship. *Journal of Human Resources 39*(4), 1045–1066.

Mountjoy, J. (2019). Community colleges and upward mobility. *Available at SSRN 3373801*.

Olsen, R. J. (1982). Distributional tests for selectivity bias and a more robust likelihood estimator. *International Economic Review*, 223–240.

Pingel, S. (2017). Need-based state financial aid. policy snapshot. *Education Commission of the States*.

Rothschild, M. and L. J. White (1995). The analytics of the pricing of higher education and other services in which the customers are inputs. *Journal of Political Economy 103*(3), 573–586.

Rubin, R. B. (2011). The pell and the poor: A regression-discontinuity analysis of on-time college enrollment. *Research in Higher Education 52*(7), 675–692.

Turner, L. J. (2017). The economic incidence of federal student grant aid.