Title
Topics in Education and Labor

Author
Gu, Jonathan Louis

Publication Date
2020

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Topics in Education and Labor

A dissertation submitted in partial satisfaction

of the requirements for the degree

Doctor of Philosophy in Economics

by

Jonathan Louis Gu

2020

ABSTRACT OF THE DISSERTATION

Topics in Education and Labor

by

Jonathan Louis Gu

Doctor of Philosophy in Economics

University of California, Los Angeles, 2020

Professor Moshe Buchinsky, Chair

This thesis consists of three chapters and covers topics in education and the labor market. The first chapter 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 by taking advantage of a jump in Cal Grant aid at a GPA threshold. 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.

The second chapter estimates the heterogeneous returns to education by place-of-birth for males born between 1930 and 1940. I do not assume the estimate is distributed according to its asymptotic distribution, therefore sidestepping the issues concerning weak instruments. Instead, we form point estimates and credible intervals directly from the posterior distribution of our two-stage estimate. When using weak priors, we obtain the same point estimates as the standard IV-2SLS methods, but we have much larger 95% credible intervals. When using priors determined from cross-validation, we are 95% sure that the returns to education are positive for only four out of nine regions, whereas the standard IV-2SLS asymptotic distribution would yield "significant" results for all nine geographic regions.

The Pasadena Minimum Wage ordinance (Ordinance #7278) passed on March 14, 2016 adopts a minimum wage schedule that is above the state minimum wage through the end of

June 2019. The third chapter studies the impact of the Pasadena minimum wage on earnings, employment, and the number of establishments. We distinguish the effect of the California minimum wage increases from the Pasadena increment because the City of Pasadena can only affect its local increment. Using data from the individual zipcodes within and around Pasadena, we find evidence of a positive impact of California/Pasadena minimum wages on the earnings of restaurant workers and of other low wage industries. Our model implies that a state-wide minimum wage increase of 10% would increase the average quarterly earnings per worker in limited-service restaurants by 8% and in full-service restaurants by 5%. We find that the Pasadena local minimum wage has a negative impact on limited-service restaurants employment and document that a minimum wage increase would decrease the number of firms in low-income industries, such as hair, nail, and skin care services.

The dissertation of Jonathan Louis Gu is approved.

Sarah Reber

Till von Wachter

Edward E. Leamer

Moshe Buchinsky, Committee Chair

University of California, Los Angeles

2020

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

VITA

2011	B.A. (Economics and Mathmatics), University of California, Berkeley.
2016	M.A. (Economics), UCLA, Los Angeles, California.
2015–2020	Teaching Assistant, Economics Department, UCLA.

CHAPTER 1

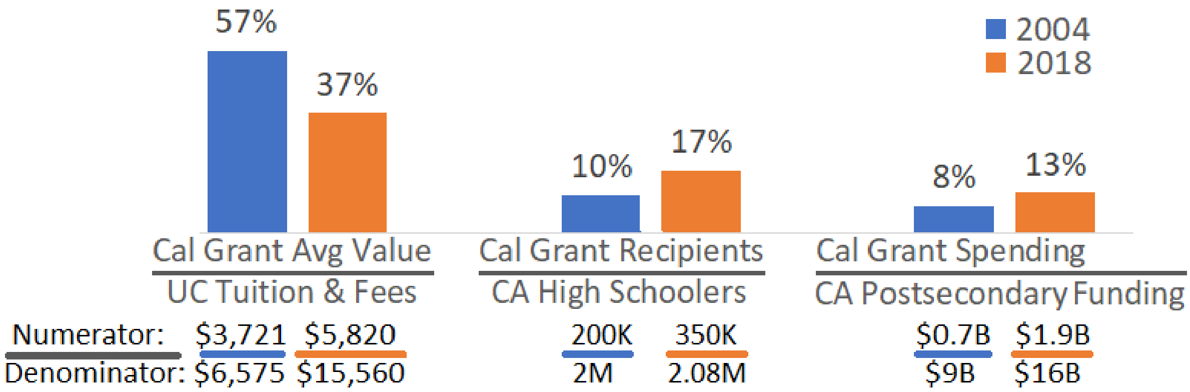
The Impact of Grants on Schools and Students

1.1 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.

[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 1.1 shows the relative importance for the Cal Grant program both in 2004 and 2018.

Figure 1.1: Relative Importance of Cal Grant



This figure shows the relative importance of the Cal Grant in 2004. The first set of bars shows that the Cal Grant had an average value of \$3,721 in 2004, which was 57% of University of California Tuition and Fees. We see that 10% of California high-schoolers were projected to receive a Cal Grant in 2004, and this has recently risen to 17%. We can also see that Cal Grant spending has increased from 8% to 13% of all of California postsecondary funding.
The data is obtained from the NASSGAP survey of state grants, NCES estimates primary and secondary school enrollment, and California Budget spending reports.

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%.

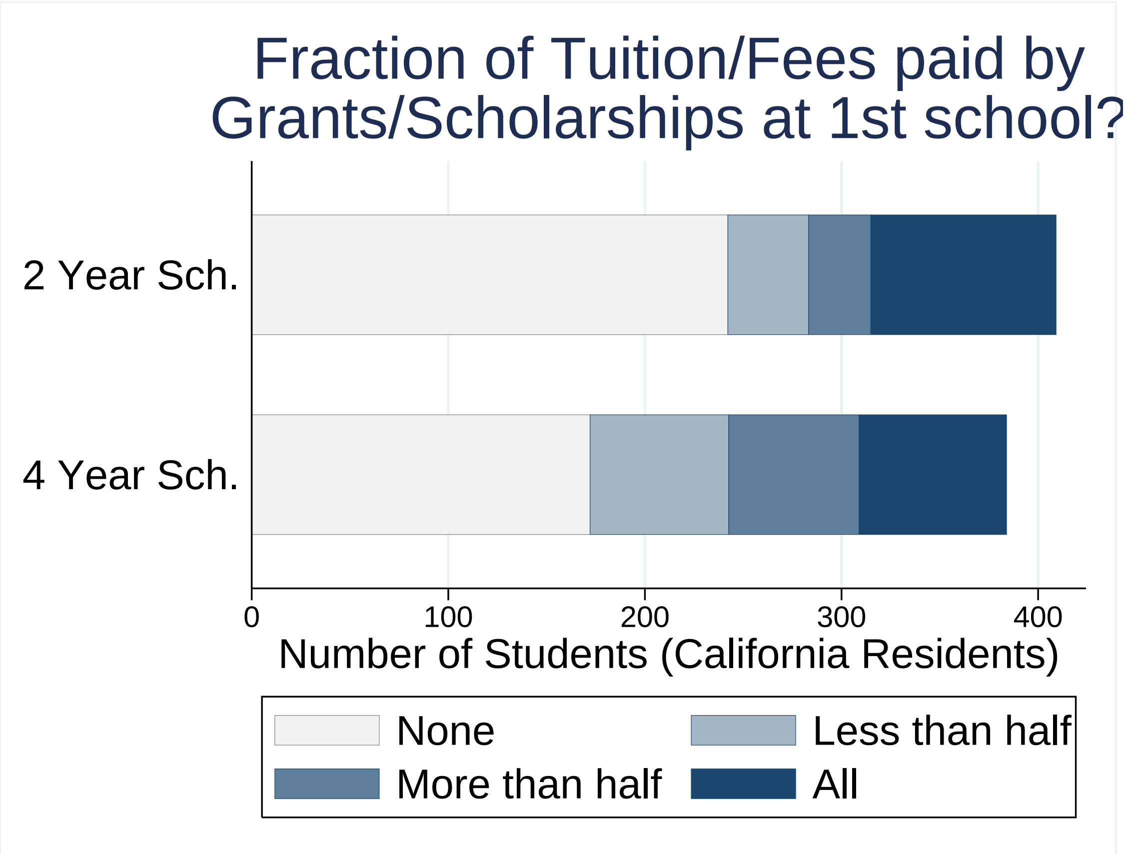
Most research does not find a significant impact of governmental aid on overall enrollment – Goal (1) (Rubin, 2011; Bettinger, 2004; Denning et al., 2019), and evidence suggests that state-based aid can convince students to attend less selective in-state schools instead of more selective out-of-state 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 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 1.2). Any analysis of the monetary cost of schooling needs to examine the net tuition instead.

Figure 1.2: Net Tuition Calculation



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.
Each student that enrolled reported what fraction of tuition and fees she paid. 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 tuition by multiplying the sticker-price by this percentage.

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 2.2 discusses the ELS2002 in further detail and previews the key data issues. section 1.3 presents the model that parameterizes student and school behavior. section 1.4 discusses the strategies and assumptions that I need to overcome the key issues. section 2.5 shows how to write the likelihood function and presents a modified simulated maximum likelihood estimation method. section 2.6 shows the estimated results. section 1.7 simulates what would happen if I removed the Cal Grant, and examines the Cal Grant's effectiveness at accomplishing the three goals.

1.2 Data

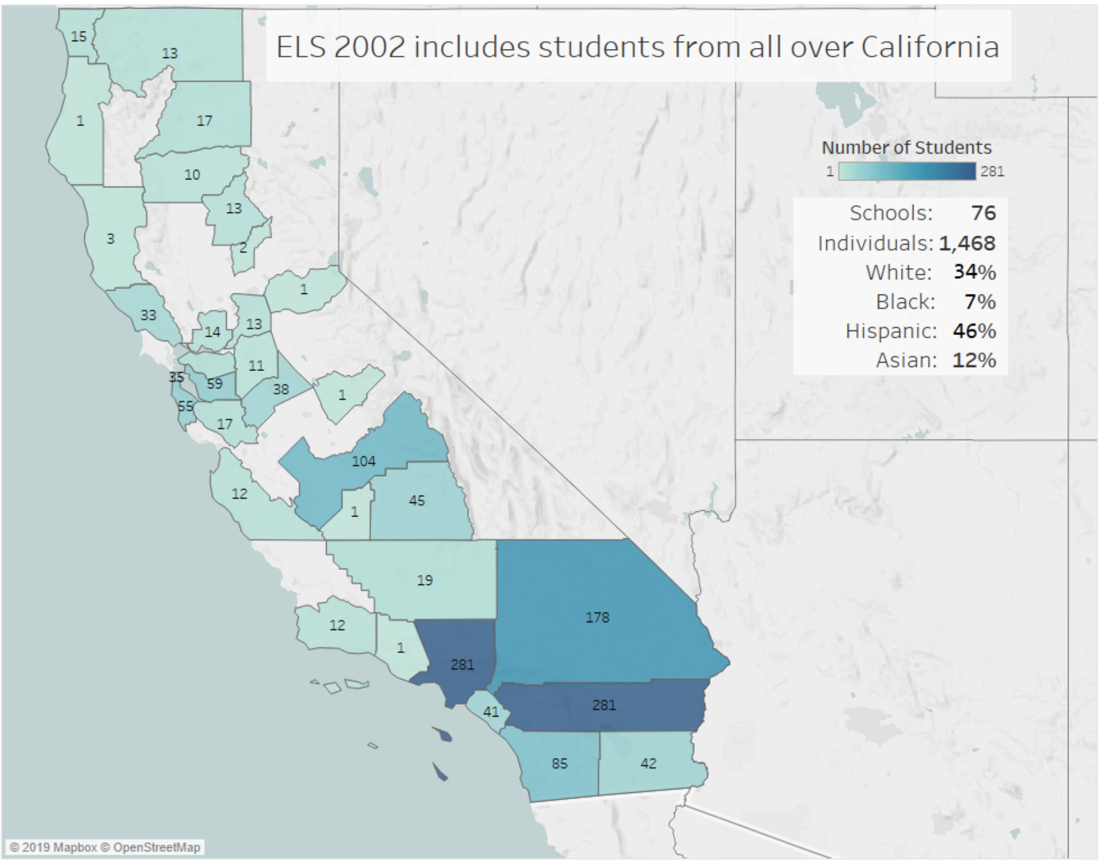
The Educational Longitudinal Study of 2002 (ELS2002) is centered around a survey of 16,197 sophomores from 751 different high schools nationwide. In addition to the survey, it also includes loan data from the National Student Loan Data System and official transcripts from each high school and postsecondary 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 1.1 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.

[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.

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 1.2). We see that the sample represents geographic regions in California quite well in Figure 1.3.

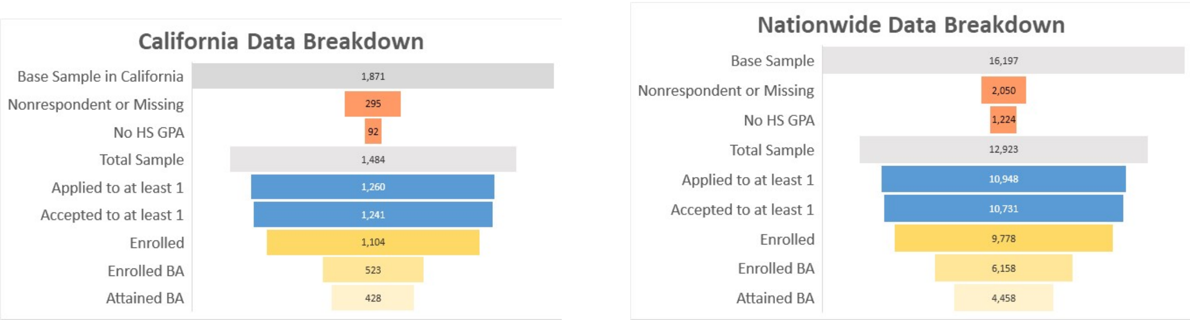
Figure 1.3: Focus on California Students



The ELS2002 contains a good geographical representation of the counties in California.

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 1.4 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.

Figure 1.4: Student Progression Towards Educational Attainment



This figure visualizes the progression of the students as they advance to their eventual degree attainment. The sample is restricted to include students that have an observed high school GPA and are survey respondents in the second and third waves. The final California sample consists of 1,484 students. It is vital to include two-year schools (community and technical colleges) in the analysis of postsecondary college attainment because only 523 of the 1,104 students enrollees enrolled at a four-year institution. We can also see that only 80% of the students enrolled in four-year schools have attained bachelor degrees by age 25.

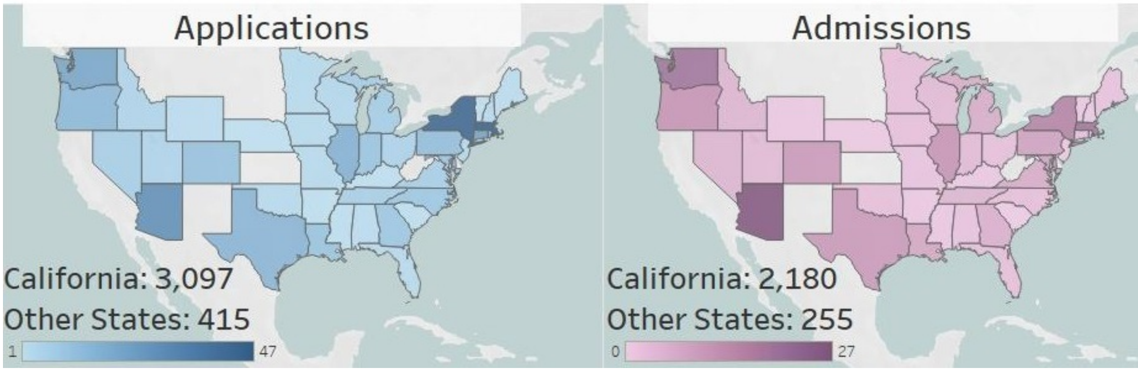
Although the ELS2002 is a vast resource, the most important features of the ELS2002 are that I observe each student's entire application set, admissions outcomes, and enrollment decision. The median student applies to only 2 or 3 schools. Table 1.3 shows that students from wealthier families tend to 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

[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.

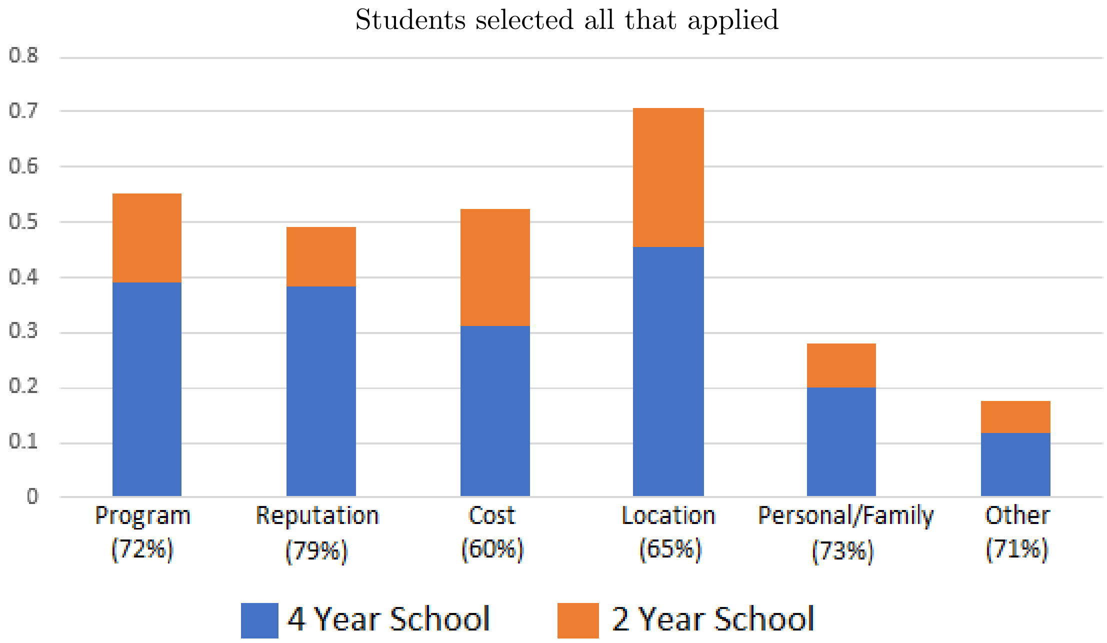
incentives students to enroll in-state, my dataset should contain students that seriously consider enrolling out-of-state. Figure 1.5 shows that although only 13.6% of enrollments were out-of-state, 19.4% of application portfolios included at least one out-of-state school. Figure 1.6 shows that students 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.

Figure 1.5: California Residents Apply and are Admitted Nationwide



Since the Cal Grant incentives students to enroll in-state, my dataset should contain students that seriously consider enrolling out-of-state. The two maps above show that students do indeed seriously consider out-of-state schools to attend. 71% of the sample applied to at least one school. 12% applied out of state. 9% of the sample were admitted to at least one out-state school.

Figure 1.6: Reason Attended Postsecondary School

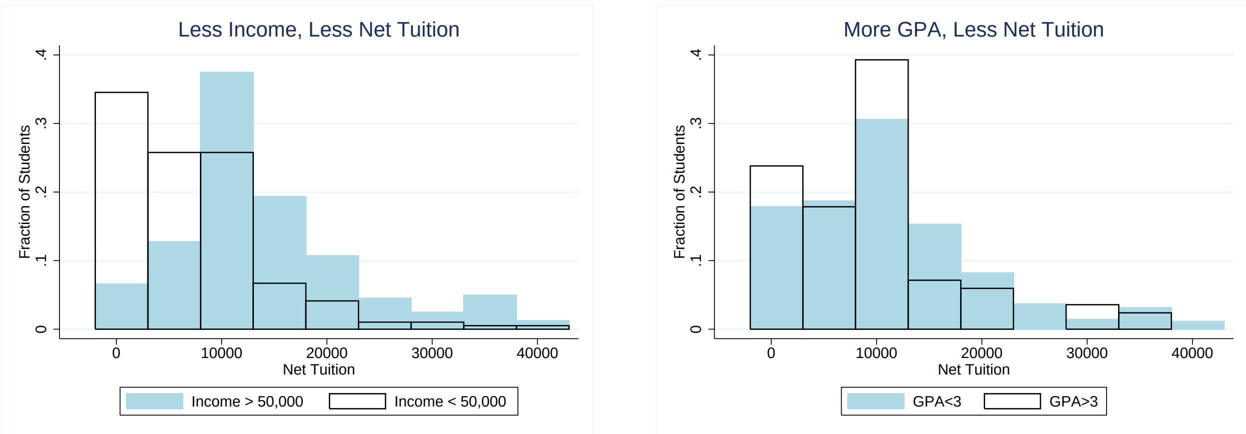


From student survey responses we see that students value the location of the school the most. However, they do not qualify whether location refers to the distance from home or school environment. I can partially capture this location preference with distance-from-home, and in-state as school attributes, but there "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 1.4.1 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 1.2). 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 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 1.7.

Figure 1.7: Better Grades and Less Income ↔ Less Net Tuition



(a) (b)

Students with greater GPA and less income pay less net tuition. Notice that the white boxes are bunched towards zero more, these white boxes represent students with better grades and lower parental income.

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 1.4.3 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.

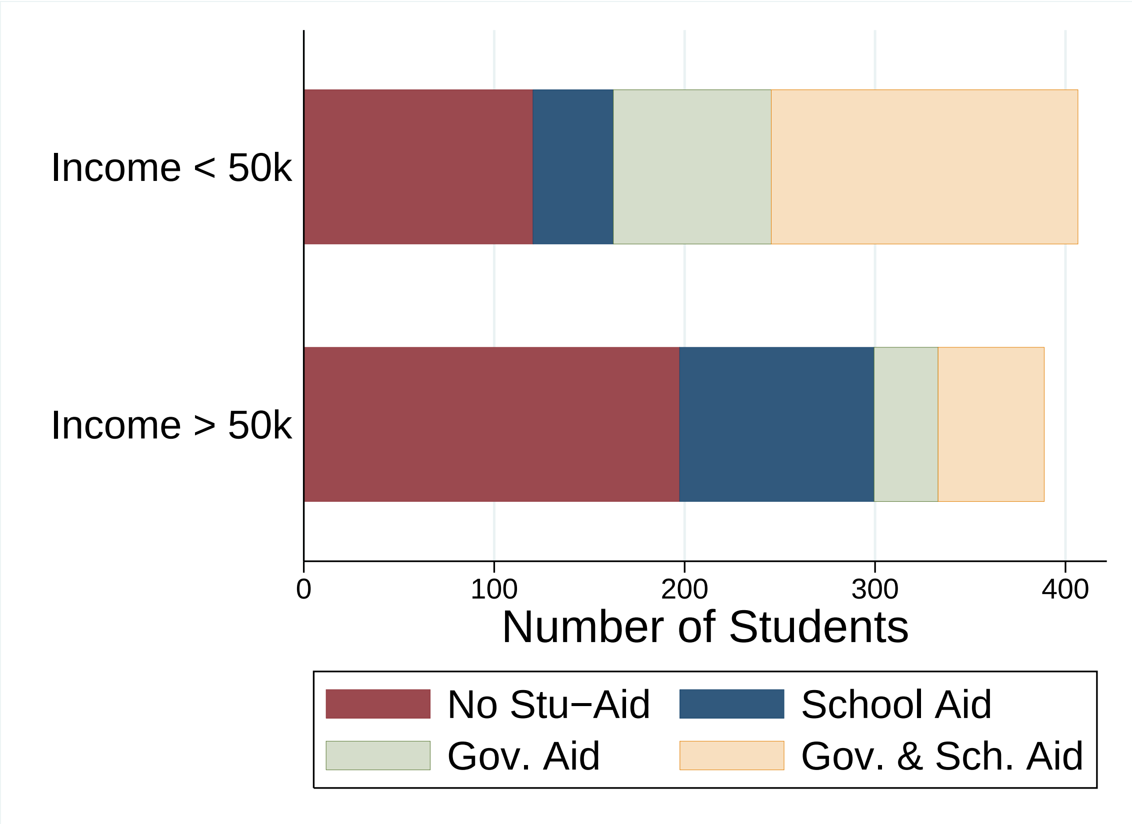
Furthermore, for each school that admitted the student, I also see separate indicators of whether the student received a grant, scholarship, or tuition waiver. I confront the selected observation of net tuition for the enrolled school, and admissions for the applied schools in subsection 1.4.4.

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 1.1, 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 1.4.2. More than half of students with parental income less than $50,000 receive some student aid from the government (Figure 1.8), and nearly three-quarters of these students receive some student aid.

[4]The Pell Grant is further discussed in section 2.8.

Figure 1.8: Most Students Receive Aid



This figure shows that most students receive student aid. More than half of students with parental income less than $50,000 receive some student aid from the government. This figure highlights how important it is to account for net tuition in the college market, instead of the sticker price.

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 decisions, and all the financial assistance – both government and school funded – in one admissions packet. Figure 1.9 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 1.10). I should take into account the non-monetary cost student applications and FAFSA filing.

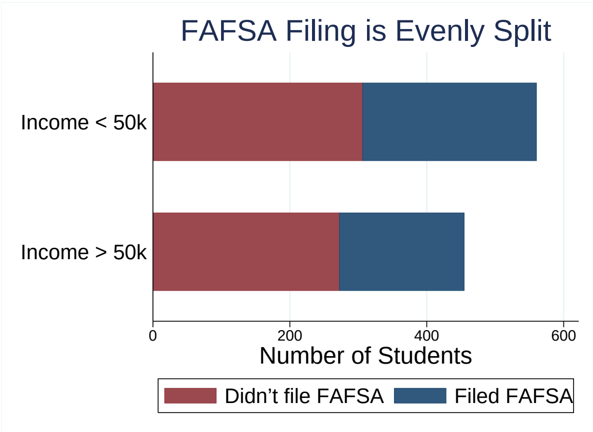
Figure 1.9: Sample Admission Net Tuition Calculation

Award Details (2014 - 2015)

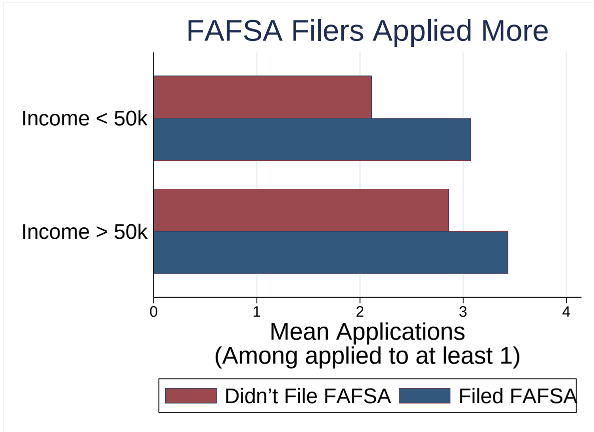
Description	Department	Annual Totals	Status
Gift Aid			
Berkeley Scholarship	Financial Aid and Scholarships	$6,289.00	Confirmed
Cal Grant A Fee Award	Financial Aid and Scholarships	$12,192.00	Confirmed
Federal Pell Grant	Financial Aid and Scholarships	$4,680.00	Confirmed
UC Undergraduate Grant	Financial Aid and Scholarships	$600.00	Confirmed
Other Gift Aid			
SURVEY PRIZE	Financial Aid and Scholarships	$500.00	Confirmed
Total Gift Aid		**$24,261.00**	
Work-study			
Work Study Eligibility	Financial Aid and Scholarships	$3,844.00	Offered
Subsidized Loan			
Federal Subsidized Direct Loan	Financial Aid and Scholarships	$2,966.00	Offered
Unsubsidized Loan			
Federal Unsubsidized Direct Loan	Financial Aid and Scholarships	$1,097.00	Offered
Grand Total		**$32,168.00**	

All postsecondary admission packets include net tuition calculations where they aggregate both school scholarships and governmental grants. This is a sample net tuition calculation from a UC Berkeley admissions letter in 2014.

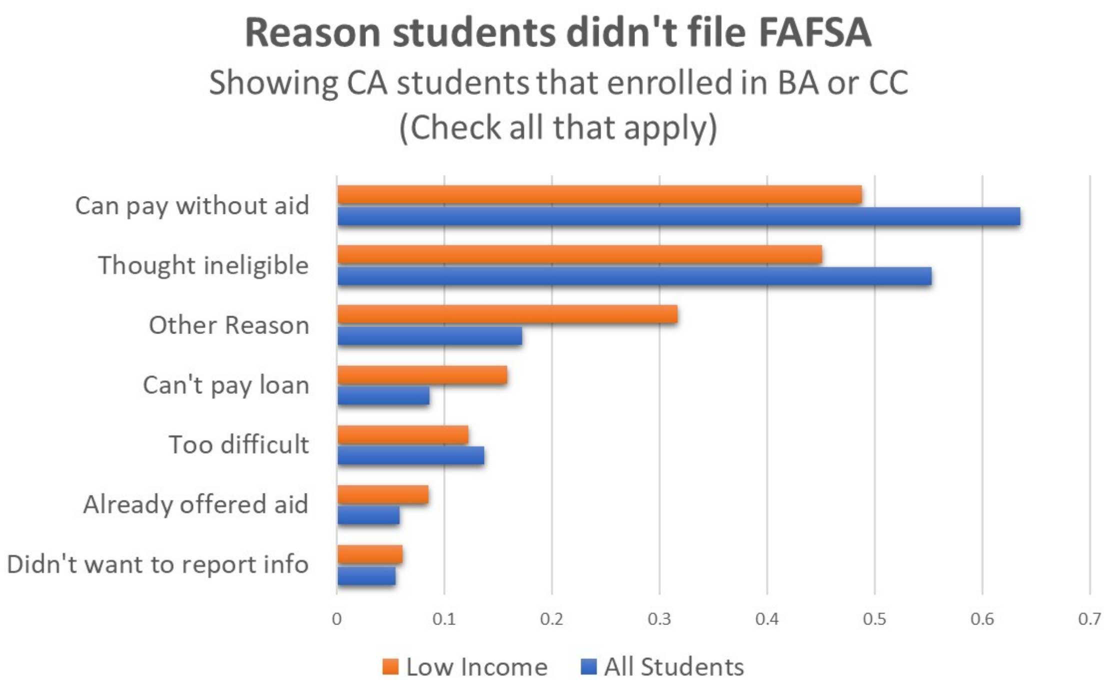
Figure 1.10: FAFSA filing and Applications



(a)



(b)



(c)

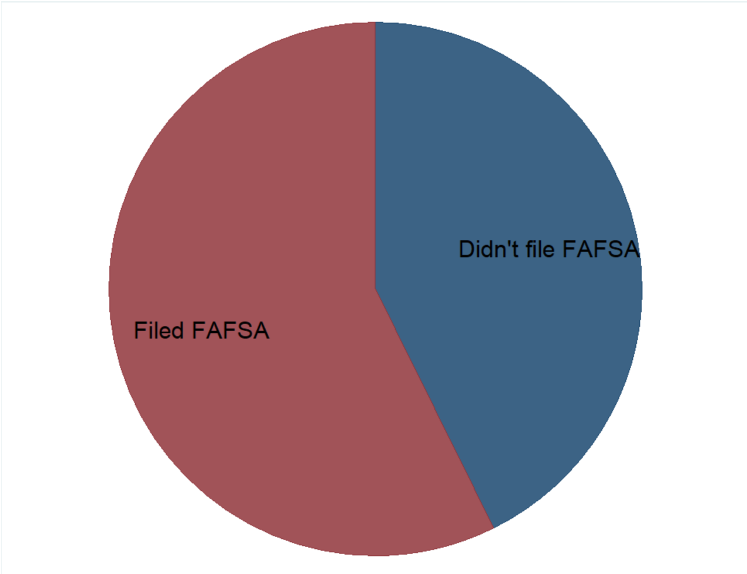
Figure 1.10a shows 45% of students with parental income less than $50k filed FAFSA.

Figure 1.10b shows that FAFSA filers applied to more schools, but overall the average student only applied to three schools.

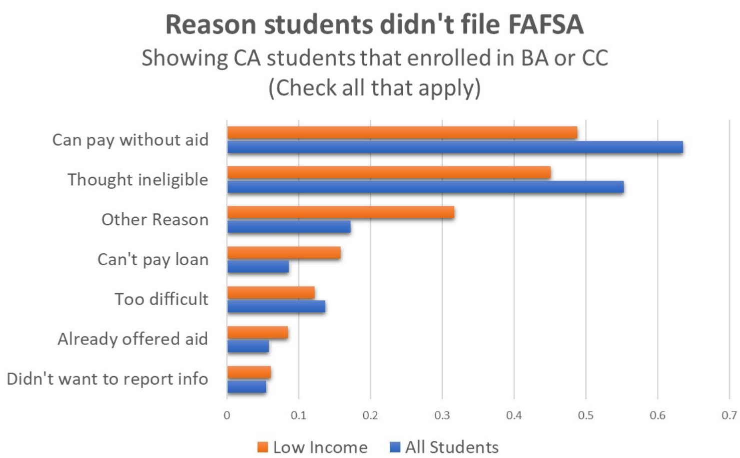
Figure 1.10c shows that the chief reason for not filing was because students thought they could pay without aid. The second most important reason (even among low-income students) was that the students didn't deem themselves eligible. The ELS2002 survey did not directly obtain US citizenship status from 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.

Many low-income students didn't file FAFSA. The chief reason for not filing was because students thought they could pay without aid (Figure 1.11). 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.

Figure 1.11: Many low income students didn't file FAFSA



(a)



(b)

The above two charts show that many low-income students didn't file FAFSA. The chief reason for not filing was because students thought they could pay without aid. 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.

Table 1.1: Demographic Comparison of the ELS2002 sample against the ACS

	CA ELS 2002 (age 16)	CA ACS (14-19)		ACS overall	
	2002	2002	2017	2002	2017
White	34.2%	37.1%	26.8%	67.3%	60.6%
African American/Black	6.9%	7.0%	5.3%	12.0%	12.3%
Hispanic	46.3%	41.7%	50.2%	14.1%	18.0%
Asian	12.0%	10.5%	12.4%	4.4%	5.7%
HH Income: $\leq$ 50k	61.7%	44.3%	30.1%	47.7%	33.5%
HH Income: $\geq$ 50k	38.3%	55.7%	69.9%	52.3%	66.5%
N	1,468	9.5K	30K	1.1M	3.2M
Population Estimate		3.0M	3.1M	285M	325M

Student's from wealthier families are underrepresented in the ELS2002, and the weights provided by the survey seem to exacerbate the problem. For drawing conclusions in current times, we may want to focus more on students that identify as Hispanic because the share of Hispanic students between 14-18 has increased dramatically since 2002.

Table 1.2: ELS2002 Student by Survey Response in 2006 and Transcript Availability

	Responded and have Transcripts	Overall
White	33.4%	34.6%
African American/Black	9.9%	7.9%
Hispanic	46.0%	45.1%
Asian	13.2%	12.4%
HH Income: $\leq$ 50k	59.1%	59.7%
HH Income: $\geq$ 50k	40.9%	40.3%
Overall	79.7%	100%

To estimate my model, I used 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. We should compare demographics between the responded with the overall characteristics to check the representativity of the students I use in my sample. The relative proportions of race and income are not changed when I only take the respondents. Table 1.1 has slightly different percentages because it uses weights tailored for comparison between 2006 respondents.

1.3 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

Table 1.3: Variation in Application Portfolios by Parental Income

Parental Income	N	Proportion applied	HS GPA[1]	Number of Applications[2]	%BA[1]	Sticker[2]
< 10k	546	0.8	2.4	2.0	0.3	2.9k
			(0.8)	[1.0, 3.0]	(0.4)	[1.6k, 7.3k]
10k - 20k	915	0.8	2.5	1.0	0.4	3.6k
			(0.8)	[1.0, 3.0]	(0.4)	[1.8k, 7.7k]
20k - 25k	653	0.8	2.5	2.0	0.4	3.7k
			(0.9)	[1.0, 3.0]	(0.4)	[1.9k, 7.8k]
25k - 35k	1274	0.8	2.6	2.0	0.5	4.2k
			(0.8)	[1.0, 3.0]	(0.4)	[2.2k, 8.6k]
35k - 50k	2082	0.9	2.7	2.0	0.5	4.5k
			(0.8)	[1.0, 3.0]	(0.4)	[2.3k, 9.4k]
50k - 75k	2454	0.9	2.8	2.0	0.6	5.1k
			(0.8)	[1.0, 3.0]	(0.4)	[2.5k, 11.1k]
75k - 100k	1629	0.9	2.9	2.0	0.7	6.4k
			(0.8)	[1.0, 4.0]	(0.4)	[2.9k, 12.1]
100k - 200k	1407	1.0	3.1	3.0	0.8	8.7k
			(0.7)	[2.0, 5.0]	(0.4)	[3.8k, 15.5k]
> 200k	444	1.0	3.2	3.0	0.8	12.8k
			(0.7)	[2.0, 6.0]	(0.3)	[6.3k, 18.8k]

[1]: Mean (standard deviation in parentheses), [2]: Median [interquartile range in parentheses]

It is quite surprising that the median student applies to only 2 or 3 schools. Any model of the student application decision should be able to estimate the cost (both monetary and non-monetary) of applying to additional schools. 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.

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 1.3.1, 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]

1.3.1 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 section 1.10, 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 1.12 shows that 198 out of 910 enrolled California students took loans to attend

[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.

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]

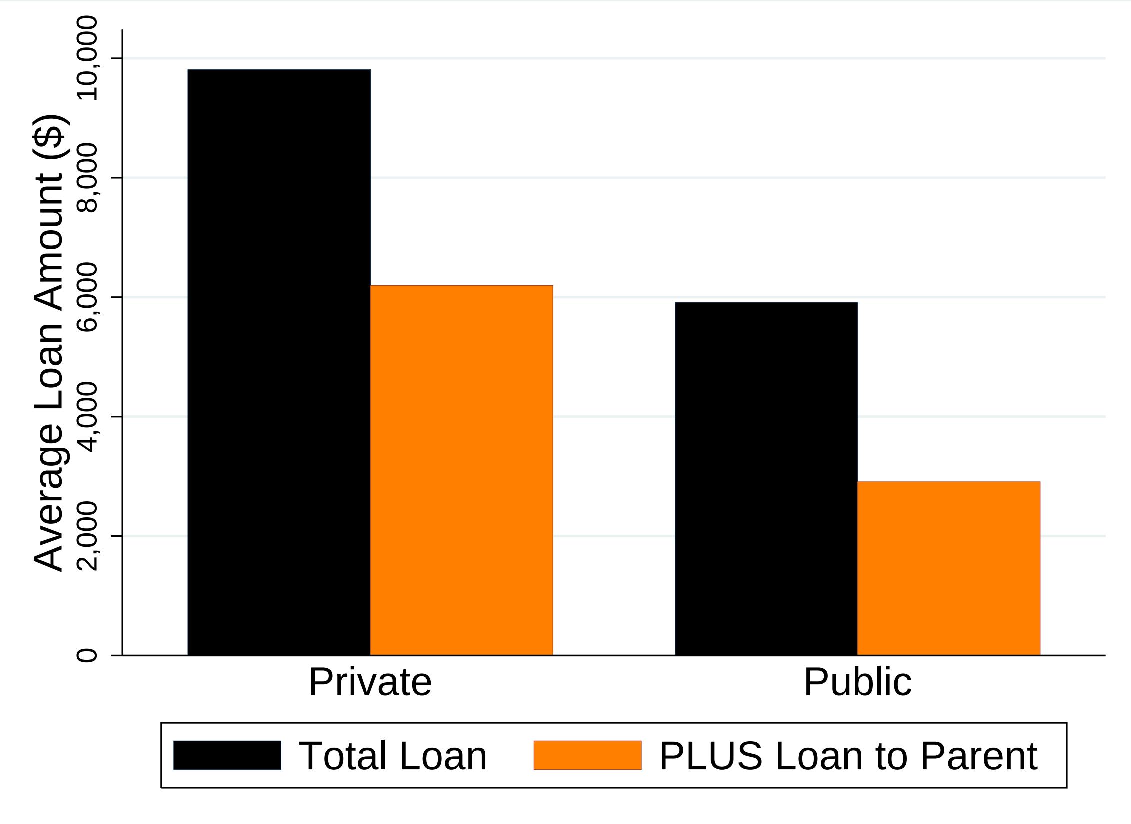
Table 1.4: Student Loan Limits

	Loan Type	Annual Max	Aggregate Max
Stafford (dep.)	Govt.→Student	$5,500	$23,000
Stafford (indep.)	Govt.→Student	$6,000	$23,000
Perkins	School→Student	$4,000	$20,000
Plus	Govt.→Parent	COA - Aid	Unlimited

Shows they have annual maximums of $5,500 and $4,000 respectively. 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.

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

Figure 1.12: Student Loan Amounts



198 out of 910 enrolled California students took loans to attend the first year of college. The average loan was less than $10,000 among those that took loans to enroll in private schools, and the average loan was less than $6,000 to enroll in a public school. 40% of student loan packages were at least partially funded with PLUS loans. Stafford and Perkins loans go directly to the student and Table 1.4 shows they have annual maximums of $5,500 and $4,000 respectively. For reference in 2004, the tuition and fees of all University of California schools totaled $6,575. 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 don't contribute their Expected Family Contribution (which isn't a legal obligation), then these students cannot afford full tuition at many schools.

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

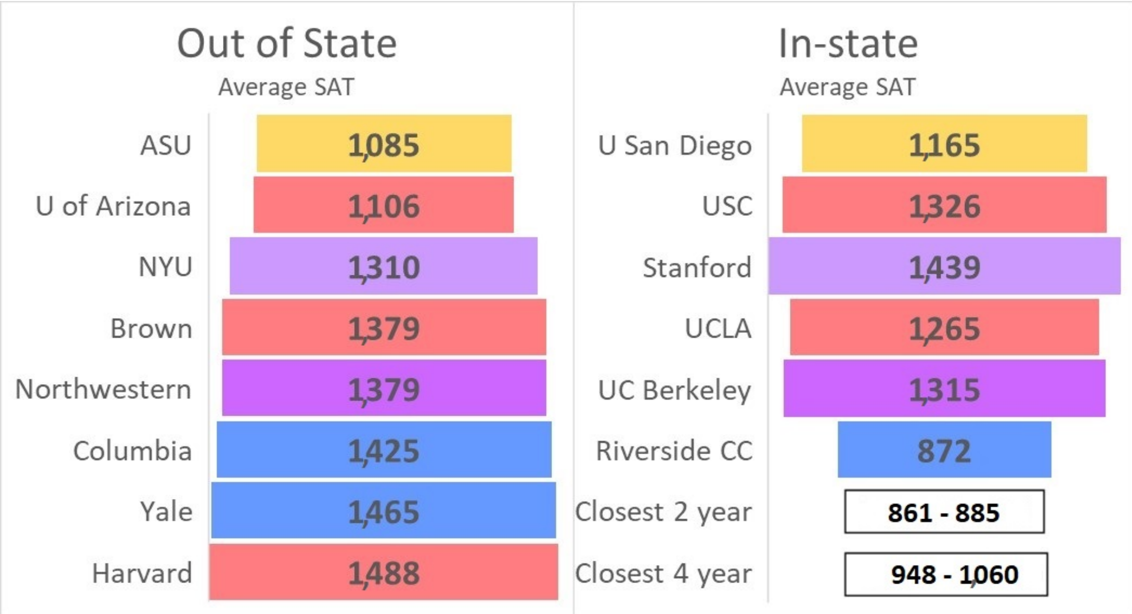
$$\max_{j \in O_i^*} V(x_{ij}, \varepsilon_{ij}) \tag{1.1}$$

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 1.13. 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.

Figure 1.13: Student Application Contemplation Set



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 1.13. 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:

$$\max_{(O,d_f)\in\mathbb{P}(\mathcal{J}_i)\times\{0,1\}} W(O, d_f, c_i, \vec{\varepsilon}_i) \tag{1.2}$$

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

1.3.2 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. section 1.11 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:*

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

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 1.4.2. 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

[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.

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 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.

1.4 Identification Strategy

Now that I have parameterized student and school decisions above in section 1.3, 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.

1.4.1 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

[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.

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 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.

1.4.2 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 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)

[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 section 2.8.

[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.

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.*
$CalGPA \perp\!\!\!\perp (\eta, \eta_a, \varepsilon) | c_{fafsa}, d_f$

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 1.14 provides evidence that that the CalGPA 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:

$$q_A(c_{fafsa}) = \mathbb{1}(same-state)max(0, min(COA-EFC, 8322\mathbb{1}(public)+6141(1-\mathbb{1}(public)))) \tag{1.3}$$

$$q_B(c_{fafsa}) = \mathbb{1}(same-state)max(0, min(COA-EFC, 1551)) \tag{1.4}$$

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

$$z_e = q_A(c_{fafsa}) - q_B(c_{fafsa}) \tag{1.5}$$

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:

$$z = z_e d_f \tag{1.6}$$

Then we also have the key independence of z:

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

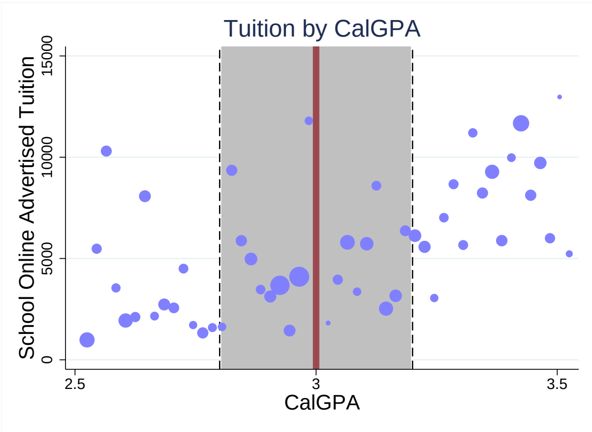
The independence condition Lemma 1 is the central point of this section. This condition allows me to use z as a surrogate for shifts in cal grant funding that is uncorrelated with school preferences (omitted variables). In turn, I will also use z as a surrogate to observe shifts in net tuition that is 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.

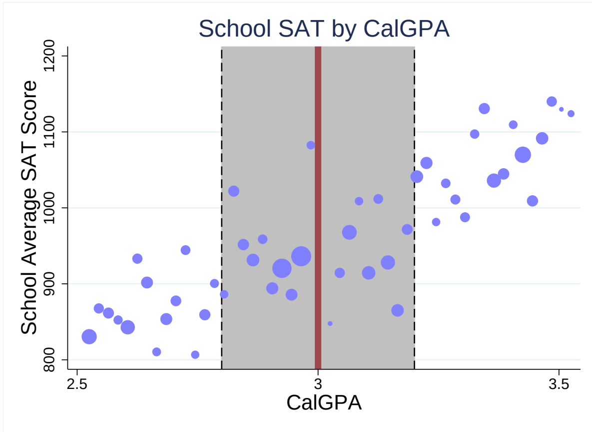
The increase in Cal Grant aid at the $CalGPA$ threshold can help solve this issue if we believe that the increase in aid is independent from any unobserved reasons for the school to give increased scholarship (η), or for Beth-1 to enroll at UCLA ε.

Let's say Beth-1 and Beth-2 both tried their hardest in school, and are observationally identical except that Beth-1 has $CalGPA = 3.01$ and Beth-2 has $CalGPA = 2.99$. 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.

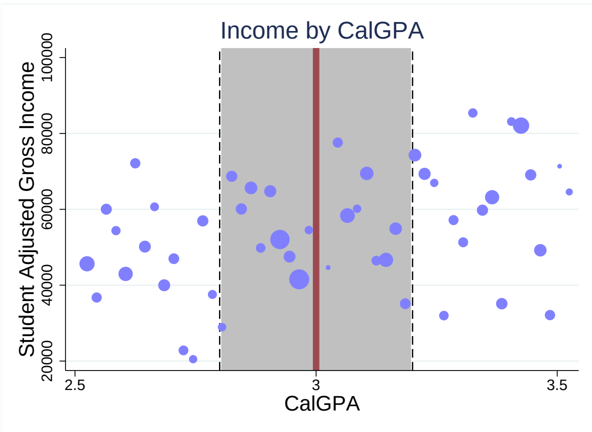
Figure 1.14: Balanced Characteristics



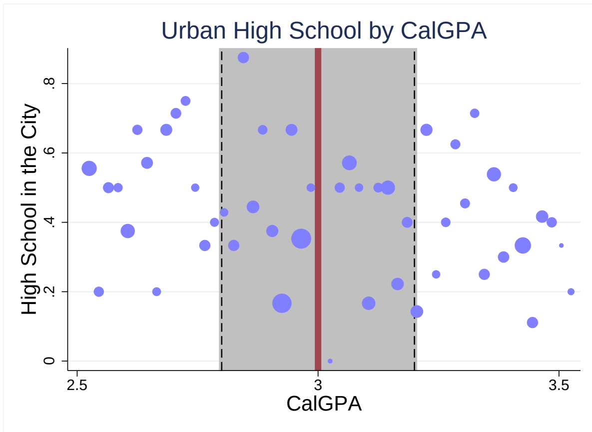
(a)



(b)



(c)



(d)

Students qualify for the Cal Grant A program if they have a $CalGPA$ greater than 3. The $CalGPA$ is calculated from the academic courses taken in the sophomore and junior years of high school. The figures above provide evidence that the $CalGPA$ threshold does not matter for many characteristics that are associated with the outcome of net tuition. Figure 1.14b, Figure 1.14a, Figure 1.14c demonstrate 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. Figure 1.14d shows that the urbanicity of students does not impact his location relative to the $CalGPA$ threshold.

1.4.3 Measurement Error

As discussed in section 2.2, 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 $CalGPA = 3.01$ and fifty Beth-2s with $CalGPA = 2.99$. 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.

$$x = \chi + m \tag{1.7}$$

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.*
$m \perp\!\!\!\perp z$

1.4.4 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:

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

[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.

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

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.

1.5 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] The overall estimation strategy is Simulated Maximum Likelihood, or the Logit-smoothed Accept Reject Simulator, as discussed in (?) and (?), Ben-Akiva and Bolduc (1996) have called it "logit-kernel probit.

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^*}.

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

$$
\begin{aligned}
&Pr_{stu,sch}(\text{ apply } (O^*, d_f^*) \cap \; j^* \text{ decide } x_{j^*} \cap [j \text{ admits } e_j^*]_{j\in O^*} \cap \text{ enroll } j^*) \\
&= Pr_{sch}(j^* \text{ decide } x_{j^*}) Pr_{sch}([j \text{ admits } e_j^*]_{j\in O^*} | j^* \text{ decide } x_{j^*}) \times \qquad (1.8) \\
&\quad Pr_{stu}(\text{ apply } (O^*, d_f^*) \cap i \text{ enroll } j^* | j^* \text{ decide } x_{j^*} \cap [j \text{ admits } e_j^*]_{j\in O^*}) \qquad (1.9)
\end{aligned}
$$

Upon taking logs, we see that the school-side (Equation 1.8) is separated from the student side (Equation 1.9). I discuss each estimation in detail in the following subsections.

1.5.1 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: $\rho_s = corr(\eta_a, \eta)$.

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

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

$$Pr_{sch}(j^* \text{ decide } x_{j^*})Pr_{sch}([j \text{ admits } e_j^*]_{j\in O^*}|j^* \text{ decide } x_{j^*}) \quad (1.10)$$

$$= \prod_{j\in\{O^*\backslash j^*\}} Pr(z_j\pi_a + \eta_{a,j} > 0)^{(e_j^*)} Pr(z_j\pi_a + \eta_{a,j} < 0)^{(1-e_j^*)}\times \quad (1.11)$$

$$Pr(x_{j^*} = z_j\pi_{j^*} + \eta_j + m_j)\times \quad (1.12)$$

$$Pr(z_{j^*}\pi_a + \eta_{a,j^*} > 0|x_j = z_{j^*}\pi_{j^*} + \eta_{j^*} + m_j) \quad (1.13)$$

Now I can take full advantage of multivariate normality (Assumption 6). Upon taking logs Equation 1.11 is a probit specification, Equation 1.12 is an ordinary least squares. Equation 1.13 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 1.10) 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.

1.5.2 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:

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

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:

$$Pr_{stu}(\text{apply } (O^*, d_f^*) \cap \text{ enroll } j^* | j^* \text{ decide } x_{j^*} \cap [j \text{ admits } e_j^*]_{j \in O^*}) \tag{1.14}$$

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

$$\begin{aligned} Pr(\text{ enroll } j^*) &= Pr\left(V(\chi_{j^*}, \varepsilon_{j^*}) \geq \max_{j \in O^{**}} V(\chi_j, \varepsilon_j)\right) \\ &= Pr\left(\chi_{j^*}\alpha + \varepsilon_{j^*} \geq \max_{j \in O^{**}} \left(\chi_j \alpha + \varepsilon_j\right)\right) \\ &= Pr\left(z_{j^*}\pi\alpha + (\eta_{j^*}\alpha + \varepsilon_{j^*}) \geq \max_{j \in O^{**}} \left(z_j \pi\alpha + (\eta_j \alpha + \varepsilon_j)\right)\right) \end{aligned} \tag{1.15}$$

The probability of observing a specific application set is:

$$
\begin{aligned}
Pr(\text{ apply } (O^*, d_f^*)) &= Pr\Bigg(W(O^*, d_f^*, V, \vec{\varepsilon}) \geq \max_{(O,d_f)\in\mathbb{P}(\mathcal{J})\times\{0,1\}} \big(W(O, d_f, V, \vec{\varepsilon})\big)\Bigg) \\
&= Pr\Bigg(E_{\vec{\eta},\vec{\eta}_a}\big(\max_{j\in\{\emptyset,O^*(\eta_{aj})\}} V(\chi(\eta_j,\eta_{aj}),\varepsilon_j)\big) - C_A(|O^*|, d_f^*) \geq \\
&\qquad \max_{(O,d_f)\in\mathbb{P}(\mathcal{J})\times\{0,1\}} \big(E_{\vec{\eta},\vec{\eta}_a}\big(\max_{j\in\{\emptyset,O(\eta_{aj})\}} V(\chi(\eta_j,\eta_{aj}),\varepsilon_j)\big) - C_A(|O|, d_f)\big)\Bigg) \\
&= Pr\Bigg(E_{\vec{\eta},\vec{\eta}_a}\big(\max_{j\in\{\emptyset,O^*(\eta_{aj})\}} z_j\pi\alpha + (\eta_j\alpha + \varepsilon_j)\big) - C_A(|O^*|, d_f^*) \geq \\
&\qquad \max_{(O,d_f)\in\mathbb{P}(\mathcal{J})\times\{0,1\}} \big(E_{\vec{\eta},\vec{\eta}_a}\big(\max_{j\in\{\emptyset,O(\eta_{aj})\}} z_j\pi\alpha + (\eta_j\alpha + \varepsilon_j)\big) - C_A(|O|, d_f)\big)\Bigg) \qquad (1.16)
\end{aligned}
$$

Equations (1.15) and (1.16) 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 section 1.12 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 $[(\tilde{\rho}, \widetilde{\rho_a})]_{1,\ldots,K_g}$ 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]

I must draw from the conditional distribution defined by the student likelihood (Equation 1.14). Let $[\tilde{\varepsilon}, \tilde{\eta}, \widetilde{\eta_a}, \tilde{m}]_{j,k}$ be the kth simulated draw that satisfies the following conditions:

[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: $\varepsilon = \frac{\omega_\varepsilon + \eta\psi + \eta_a\psi_a}{\sqrt{1+\psi^2+\psi_a^2}}$.

$$(\eta_{j^*} + m_{j^*} = x_{j^*} - z_{j^*}) \cap (\eta_{aj^*} \geq -z_{j^*}\pi_a) \text{ for enrolled } j^* \tag{1.17}$$

$$(\eta_{ak} \geq -z_k\pi_a) \text{ if applied and admitted} \tag{1.18}$$

$$(\eta_{ak} < -z_k\pi_a) \text{ if applied and rejected} \tag{1.19}$$

As shown in the expectation in *Equation* 1.16, students must guess the school preferences during their application step. These school decisions are normally distributed conditional on student preferences:

$$\left(\begin{bmatrix} \frac{\eta}{\sigma_\eta} \\ \eta_a \end{bmatrix} \middle| \varepsilon = e \right) \sim N\left(\begin{bmatrix} \rho \\ \rho_a \end{bmatrix} e, \begin{bmatrix} 1-\rho^2 & \rho_s - \rho\rho_a \\ \rho_s - \rho\rho_a & 1-\rho^2 \end{bmatrix} \right) \tag{1.20}$$

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 1.20. 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:

$$Pr_{stu}(\text{apply to } (O^*, d_f^*) \cap \text{ enroll at } j^*|j^* \text{ decide } x_{j^*} \cap [k \text{ admits } e_j^*]_{j \in O^*}) = \tag{1.21}$$

$$\int_{\vec{r}} \mathbf{1}\Big(\big(W(O^*, d_f^*, \vec{r}) \geq \max_{(O,d_f) \in \mathbb{P}(\mathcal{J}) \times \{0,1\}} W(O, d_f, \vec{r})\big) \cap$$

$$\big(V_z(z_{j^*}\hat{\pi}, \eta_{j^*}, r_{j^*}) \geq \max_{j \in O^{**}} V_z(z_j \hat{\pi}, \eta_j, r_j)\big)\Big) dF_{\vec{\varepsilon}|(1.17,1.18,1.19)}(\vec{r}) \tag{1.22}$$

$$\approx \frac{1}{|K|} \sum_{k \in K} \Bigg(\frac{exp(\widetilde{W}(O^*, d_f^*, \big[[\tilde{\varepsilon}, [\widetilde{\eta}_\varepsilon, \widetilde{\eta}_{a\varepsilon}]_{k_\varepsilon \in K_\varepsilon}]_{j \in J}\big]_k)/\tau_1)}{\sum\limits_{(O,d_f)} exp(\widetilde{W}(O, d_f, \big[[\tilde{\varepsilon}, [\widetilde{\eta}_\varepsilon, \widetilde{\eta}_{a\varepsilon}]_{k_\varepsilon \in K_\varepsilon}]_{j \in J}\big]_k)/\tau_1)} \cdot$$

$$\frac{exp([V_z(z_{j^*}\hat{\pi}, \tilde{\eta}_{j^*,k}, \tilde{\varepsilon}_{j^*,k})]/\tau_2)}{\sum\limits_{j \subset O^{**}} exp([V_z(z_j \hat{\pi}, \tilde{\eta}_{j,k}, \tilde{\varepsilon}_{j,k})]/\tau_2)} \Bigg) \tag{1.23}$$

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 1.23 approaches the integrated frequency in Equation 1.22.

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.

1.6 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 1.6.2 shows that the net tuition significantly affects college choice and that the use of a threshold identification strategy leads to a smaller estimated impact.

[17]The term $\big[[\tilde{\varepsilon}, [\widetilde{\eta}_\varepsilon, \widetilde{\eta}_{a\varepsilon}]_{k_\varepsilon \in K_\varepsilon}]_{j \in J}\big]_k$ 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).

1.6.1 School Policies

Table 1.5 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 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.

1.6.2 Model Results

Table 1.6 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

Table 1.5: Net Tuition and Admissions

CalGPA Bandwidth: [2.8,3.2]	Net Tuition (OLS) Coeff.	SE	Admission (Probit) Marginal Effect.	SE
Cal Grant Aid Received	-0.81*	(0.36)	-0.00	(0.000)
$\mathbb{1}_{CalGPA>3}$	3,088.57	(1,769.55)	0.06	(0.066)
$Pell_{FAFSA} \times \mathbb{1}_{FAFSA}$	-0.57	(0.49)	-0.00	(0.000)
$\mathbb{1}_{In\text{-}state}$	11,911.11*	(3,036.27)	-0.22*	(0.096)
$\mathbb{1}_{PublicSch}$	496.99	(4,493.29)	-0.32*	(0.109)
$\mathbb{1}_{In\text{-}state}\mathbb{1}_{PublicSch}$	-9,233.84	(5,294.38)	0.08	(0.152)
$\mathbb{1}_{FAFSA}$	1,848.23	(1,199.62)	0.09*	(0.041)
$CalGPA$	-11,160.23	(7,390.59)	-0.07	(0.281)
Honors GPA	-117.02	(2,329.38)	0.25*	(0.091)
Sch. Sticker	-0.20	(0.44)	-0.00	(0.000)
Sch. COA	0.45	(0.50)	-0.00	(0.000)
Sch. SAT	-11.84	(8.55)	-0.001*	(0.0003)
Grant	-7,082.20*	(2,217.91)		
Grant × Income	0.03*	(0.01)		
Parental Income	0.01	(0.01)	0.00	(0.000)
Log Distance	105.95	(373.43)	-0.01	(0.011)
$CalGrant_{FAFSA}$	-0.19	(0.28)	0.00	(0.000)
$Pell_{FAFSA}$	0.46	(0.35)	0.00	(0.000)
$\mathbb{1}_{BASch}$	2,327.83	(1,470.63)	-0.05	(0.065)
Grant × Sticker	0.04	(0.16)		
Constant	42,090.70	(21,665.20)		
Observations	191		743	
R^2	0.62			

Coefficients more than double their standard deviation have *.

The first two columns predict net tuition paid by the student for each student-school enrollment match. The 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 find 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.

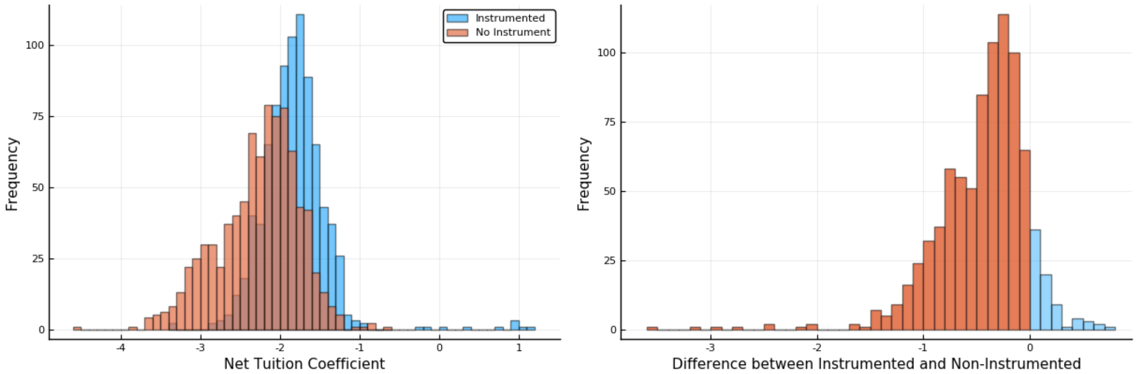
Although the first coefficient is the only one we can interpret as a causal result, the data suggest that in-state public schools are much cheaper than in-state private schools and that students would only enroll out-of-state if they receive a discount to do so.

that student applies to or enroll at a school.

In Figure 1.15, 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

Figure 1.15: The Net Tuition Impact is Less when Estimated with Threshold

(a) Bootstrapped Coefficients on Net Tuition (b) Net Tuition Difference is Negative



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).

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.

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

Table 1.6: Simulated Maximum Likelihood Results

	Instrumented		No Instrument	
	Coefficient	SE	Coefficient	SE
Net Tuition	-1.78*	(0.45)	-2.20*	(0.51)
Student GPA	-0.20	(0.29)	-0.19	(0.29)
Student Income	0.22	(0.28)	0.34	(0.28)
Same State	2.13*	(0.99)	2.36*	(1.08)
Public School	-2.55	(1.70)	-2.68	(1.99)
Bachelors	-1.62*	(0.54)	-1.60*	(0.56)
School SAT	1.08	(0.47)	1.04*	(0.48)
Avg Faculty Salary	-0.47	(0.44)	-0.45	(0.46)
Log Distance	-0.09	(0.23)	-0.10	(0.24)
School Sticker Price	-0.19	(0.87)	-0.01	(0.98)
App Cost Constant	-28.36*	(0.37)	-10.98	(7.94)
App Cost FAFSA Penalty	0.40	(1.61)	0.45	(1.06)
App Cost Additional School Penalty	1.80	(6.96)	1.96	(6.44)
Constant	7.17	(1.85)	7.86	(2.15)

* denotes coefficients more than double their standard deviation.

The estimated impact of net tuition on student applications and enrollment is less than what would have been estimated without this discontinuity based identification. I bootstrapped the entire estimation process 1,000 times to arrive at the standard deviation estimates shown.
Although my attempt to estimate the cost of applying to an additional school and the penalty of filing FAFSA did not yield significant results, the ratio $\frac{FAFSAPenalty}{AdditionalSchoolPenalty}$ averaged 0.24, and had a bootstrapped standard deviation of 0.12, which means that the non-monetary cost to filing an additional FAFSA application is about $\frac{1}{4}$ the cost of applying to an additional school.

1.7 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?

1.7.1 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 1.17a shows the predicted choice probabilities of all Cal Grant 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 1.17b 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 1.7 uses the bootstrapped model coefficients to show 90% confidence bounds on the enrollment probabilities as shown in Figure 3.11. 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

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

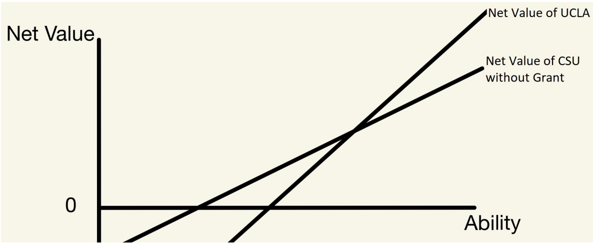
[20] The entire Cal Grant A, B, and C program paid out $700 million to 194,000 recipients. The Cal Grant A paid $280 million to 50,000 students. I estimate that one-half of Cal Grant A recipients were first-year enrollees and that one-quarter of first time recipients had $CalGPA$ between 3 and 3.2 (6,250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

$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.

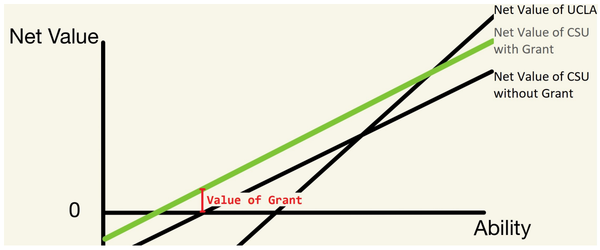
My model shows that 93% of Cal Grant A funding to ($260 million) was spent on students that were already going to enroll in-state anyway. Recall from Figure 1.16 that there may be some deadweight loss from the point of view of the student when the Cal Grant A funding induces her to switch schools. The 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.

Figure 1.16: Thought Experiment Illustrating the Impact of School Specific Grants

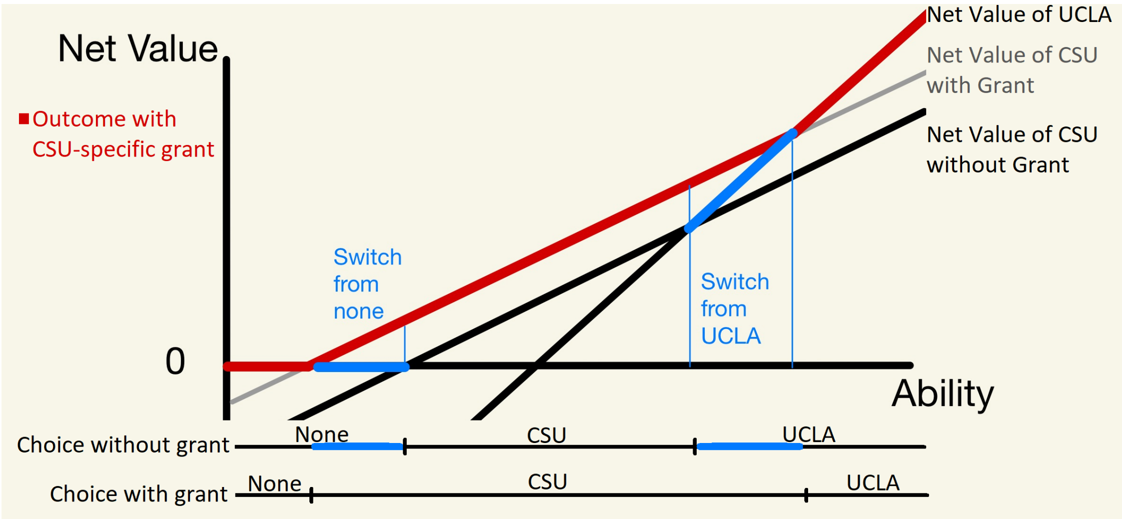
(a) Value varies with ability and cost



(b) Grant to Cal State enrollees



(c) Three groups of switchers



In this thought experiment, the net value is utility − net tuition. Assume for now that utility varies with some ability measure as in Figure 1.16a. Figure 1.16b shows the increase in the net value of Cal State University as a result of the grant. Figure 1.16c illustrates that three groups of recipients. Among the new CSU enrollees with the CSU-specific grant, there are students that weren't going to enroll, students that switched from enrolling at UCLA, and also students that didn't switch at all.

Students that switched from a different option to enroll in the funded-school gain less than the value of the grant. This is a sort of deadweight loss: Although the government gives the same grant to all enrollees of the funded-school, switchers don't gain the full value because they value their original option more than the funded-school without the grant.

Figure 1.18 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 1.7.1 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.

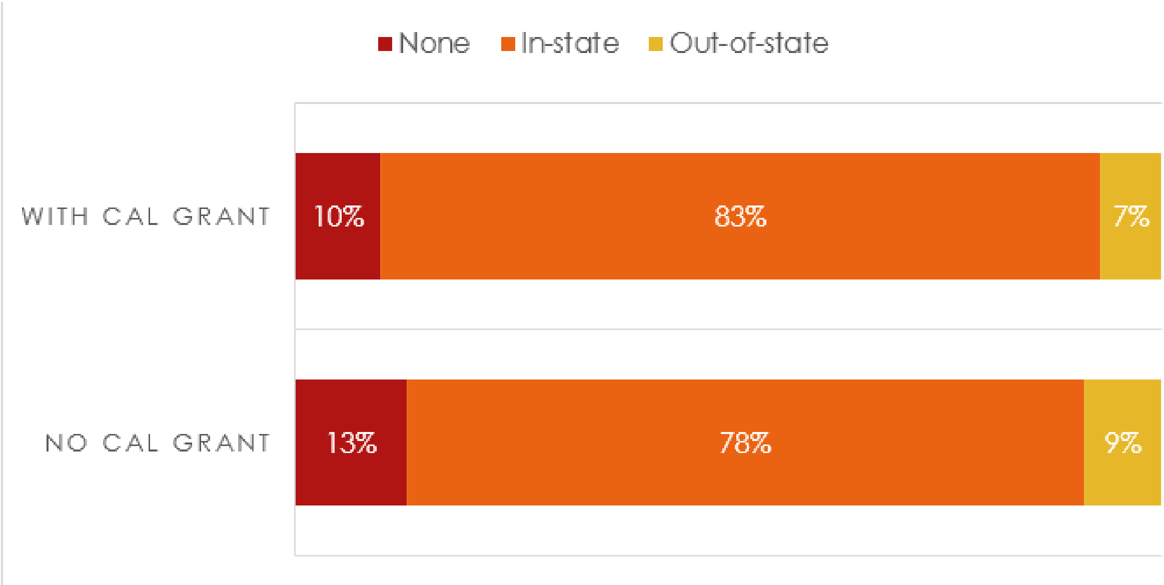
Table 1.7: Confidence in Model Predictions

Treatment On Treated	None	In-state	Out-of-state
With Cal Grant	(7%, 11%)	(81%, 86%)	(5%, 10%)
No Cal Grant	(10%, 14%)	(75%, 82%)	(7%, 12%)
Intent to Treat	**None**	**In-state**	**Out-of-state**
With Cal Grant	(19%, 24%)	(64%, 71%)	(8%, 15%)
No Cal Grant	(21%, 27%)	(59%, 67%)	(9%, 17%)

This table uses the bootstrapped model coefficients to show 90% confidence bounds on the enrollment probabilities as shown in Figure 3.11. We can see that the average enrollment chance can varies by about 2%.

Figure 1.17: Model Predicted Enrollment Chance

(a) Impact of Cal Grant on Cal Grant Recipients



(b) Impact of Cal Grant on Cal Grant Eligible

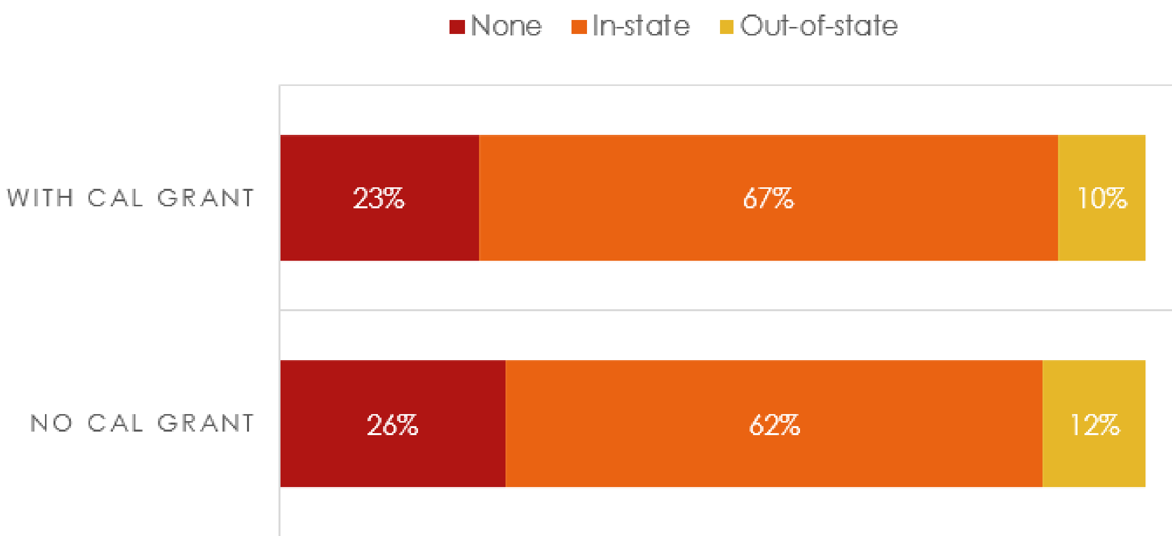
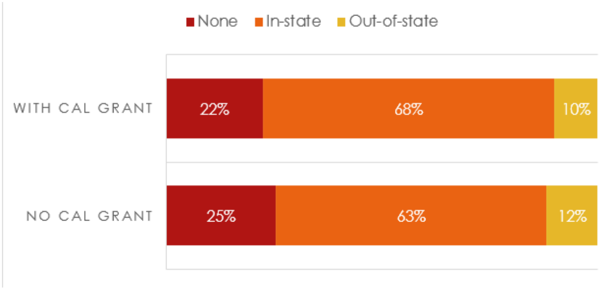


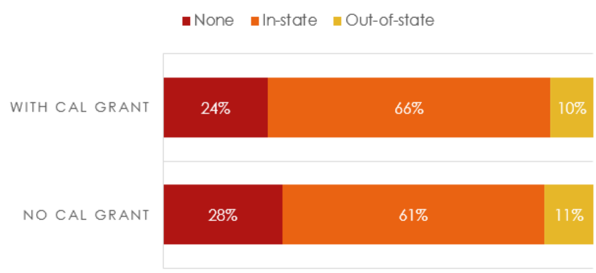
Figure 1.17a shows the predicted choice probabilities of all Cal Grant recipients who had *CalGPA* between 3 and 3.2. All students in this group enrolled in-state, and the model predicts that the average student had an 83% chance to enroll in-state with the Cal Grant.
Figure 1.17b shows predicted enrollment 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 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 1.7 uses the bootstrapped model coefficients to show 90% confidence bounds on the enrollment probabilities as shown in Figure 3.11. The model predicted enrollment chances could vary by plus or minus 2%.

Figure 1.18: Model Predicted Enrollment Chance On Cal Grant Grant Eligible

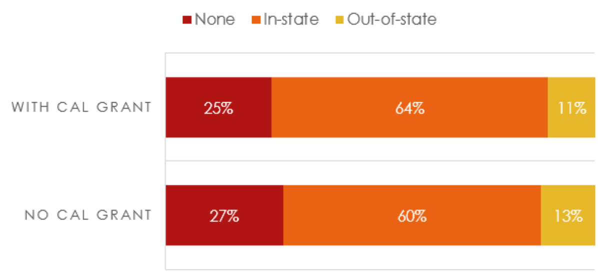
(a) Female



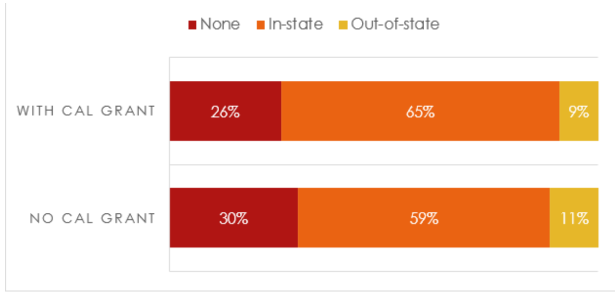
(b) Male



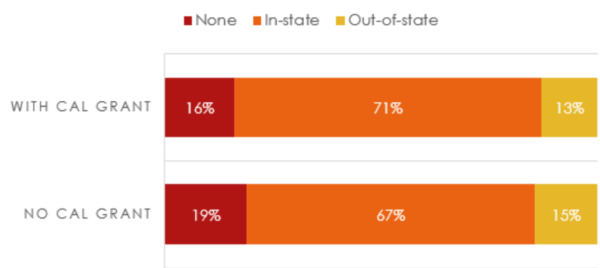
(c) Hispanic



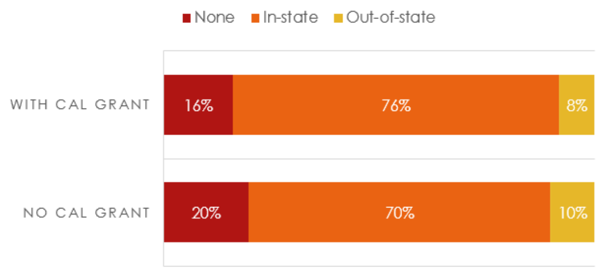
(d) White



(e) African American/Black



(f) Asian



These figures show 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. 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). Since I did not include race or sex in estimating the student utility function. 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.[5] These three groups have the most to gain from the Cal Grant A program.

1.7.2 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 1.19 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.

Figure 1.19: Reason Attended Postsecondary School

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 1.19 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 bachelors degree attainment in the state. However, it 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.

1.8 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.

1.9 Appendix A: Pell Grant

I observe the actual amount of Pell Grant received by each student through the NSLDS database. The Pell Grant yielded a maximum of \$4,050 in 2004. The award chart in Figure 1.20 shows that Pell Grants are calculated deterministically from the COA and EFC. I use this chart to determine the amount of Pell Grant that would have been awarded to students for each non-enrolled school. I also determine the grant award amounts for each student that didn't file FAFSA.[21]

Figure 1.20: Pell Grant Award Amount

Full Time

Federal Pell Grant Program
Regular Payment Schedule for Determining
Full-Time Scheduled Awards in the 2004-2005 Award Period
January 2004

$4,050 Maximum

Expected Family Contribution

Cost of Attendance	0 To 0	1 To 100	101 To 200	201 To 300	301 To 400	401 To 500	501 To 600	601 To 700	701 To 800	801 To 900	901 To 1000	1001 To 1100	1101 To 1200	1201 To 1300	1301 To 1400	1401 To 1500	1501 To 1600	1601 To 1700	1701 To 1800	1801 To 1900	1901 To 2000
2800 - 2899	2850	2800	2700	2600	2500	2400	2300	2200	2100	2000	1900	1800	1700	1600	1500	1400	1300	1200	1100	1000	900
2900 - 2999	2950	2900	2800	2700	2600	2500	2400	2300	2200	2100	2000	1900	1800	1700	1600	1500	1400	1300	1200	1100	1000
3000 - 3099	3050	3000	2900	2800	2700	2600	2500	2400	2300	2200	2100	2000	1900	1800	1700	1600	1500	1400	1300	1200	1100
3100 - 3199	3150	3100	3000	2900	2800	2700	2600	2500	2400	2300	2200	2100	2000	1900	1800	1700	1600	1500	1400	1300	1200
3200 - 3299	3250	3200	3100	3000	2900	2800	2700	2600	2500	2400	2300	2200	2100	2000	1900	1800	1700	1600	1500	1400	1300
3300 - 3399	3350	3300	3200	3100	3000	2900	2800	2700	2600	2500	2400	2300	2200	2100	2000	1900	1800	1700	1600	1500	1400
3400 - 3499	3450	3400	3300	3200	3100	3000	2900	2800	2700	2600	2500	2400	2300	2200	2100	2000	1900	1800	1700	1600	1500
3500 - 3599	3550	3500	3400	3300	3200	3100	3000	2900	2800	2700	2600	2500	2400	2300	2200	2100	2000	1900	1800	1700	1600
3600 - 3699	3650	3600	3500	3400	3300	3200	3100	3000	2900	2800	2700	2600	2500	2400	2300	2200	2100	2000	1900	1800	1700
3700 - 3799	3750	3700	3600	3500	3400	3300	3200	3100	3000	2900	2800	2700	2600	2500	2400	2300	2200	2100	2000	1900	1800
3800 - 3899	3850	3800	3700	3600	3500	3400	3300	3200	3100	3000	2900	2800	2700	2600	2500	2400	2300	2200	2100	2000	1900
3900 - 3999	3950	3900	3800	3700	3600	3500	3400	3300	3200	3100	3000	2900	2800	2700	2600	2500	2400	2300	2200	2100	2000
4000 - 4049	4025	3975	3875	3775	3675	3575	3475	3375	3275	3175	3075	2975	2875	2775	2675	2575	2475	2375	2275	2175	2075
4050 - 99999	4050	4000	3900	3800	3700	3600	3500	3400	3300	3200	3100	3000	2900	2800	2700	2600	2500	2400	2300	2200	2100

Cost of Attendance	2001 To 2100	2101 To 2200	2201 To 2300	2301 To 2400	2401 To 2500	2501 To 2600	2601 To 2700	2701 To 2800	2801 To 2900	2901 To 3000	3001 To 3100	3101 To 3200	3201 To 3300	3301 To 3400	3401 To 3500	3501 To 3600	3601 To 3700	3701 To 3800	3801 To 3850	3851 To 99999
2800 - 2899	800	700	600	500	400	400	400	0	0	0	0	0	0	0	0	0	0	0	0	0
2900 - 2999	900	800	700	600	500	400	400	400	0	0	0	0	0	0	0	0	0	0	0	0
3000 - 3099	1000	900	800	700	600	500	400	400	400	0	0	0	0	0	0	0	0	0	0	0
3100 - 3199	1100	1000	900	800	700	600	500	400	400	400	0	0	0	0	0	0	0	0	0	0
3200 - 3299	1200	1100	1000	900	800	700	600	500	400	400	400	0	0	0	0	0	0	0	0	0
3300 - 3399	1300	1200	1100	1000	900	800	700	600	500	400	400	400	0	0	0	0	0	0	0	0
3400 - 3499	1400	1300	1200	1100	1000	900	800	700	600	500	400	400	400	0	0	0	0	0	0	0
3500 - 3599	1500	1400	1300	1200	1100	1000	900	800	700	600	500	400	400	400	0	0	0	0	0	0
3600 - 3699	1600	1500	1400	1300	1200	1100	1000	900	800	700	600	500	400	400	400	0	0	0	0	0
3700 - 3799	1700	1600	1500	1400	1300	1200	1100	1000	900	800	700	600	500	400	400	400	0	0	0	0
3800 - 3899	1800	1700	1600	1500	1400	1300	1200	1100	1000	900	800	700	600	500	400	400	400	0	0	0
3900 - 3999	1900	1800	1700	1600	1500	1400	1300	1200	1100	1000	900	800	700	600	500	400	400	400	0	0
4000 - 4049	1975	1875	1775	1675	1575	1475	1375	1275	1175	1075	975	875	775	675	575	475	400	400	400	0
4050 - 99999	2000	1900	1800	1700	1600	1500	1400	1300	1200	1100	1000	900	800	700	600	500	400	400	400	0

Pell Grants are calculated deterministically from the COA and EFC. I use this chart to determine the amount of Pell Grant that would have been awarded to students for each school that she didn't enroll in. I also determine the grant award amounts for each student that didn't file FAFSA.

1.10 Appendix B: Student Utility

This section justifies how I can express a student's value of a school as a linear indirect utility function as done by Assumption 1 in the main paper. The result follows once I plug the simpler budget constraint into utility.

Student i chooses between schools j in admitted set O_i^*:

[21] I use consideration sets to narrow down the number of schools that each student considered. This is process is detailed in subsection 1.3.1.

$$\max_{j \in O_i^*} U([cons_{ijt}]_{t=0}^{T}, c_{ij}, \varepsilon_{ij})$$

$$\chi_{ij} + cons_{ij0} + save_{ij0} \leq Wealth_{ij}$$

$$-save_{ij0} \leq MaxBorrow_{ij}$$

$$cons_{ijt} + loanpay_{ijt} + save_{ijt} \leq save_{ijt-1}(1+r) + wage_t(c_{ij}, \varepsilon_{ij}) \forall t \geq 1$$

$$\sum_{t=1}^{T} \frac{loanpay_{it}}{(1+r)^t} = -save_{ij0}$$

Where $[cons_{ijt}]_{t=0}^{T}$ is the vector of lifetime consumption, and $-save_{ij0}$ is the amount of student loans taken to enroll. Given my data, the differences between human capital accumulation $wage_t(c_{ij}, \varepsilon_{ij})$ and student preferences $U([cons_{ijt}]_{t=0}^{T}, c_{ij}, \varepsilon_{ij})$ are not identified. Since the ELS2002 actually does have the student's earnings at age 25, I do potentially have variation in the dataset that allows me to identify a difference between these two functions. I will leave the separate identification of human capital accumulation and preferences for future work. I can write a simpler budget constraint:

$$Consume_{ij} \leq Income(c_{ij}, \varepsilon_{ij}) - \chi_{ij} \tag{1.24}$$

$$Consume_{ij} \equiv \sum_{t=0}^{T} \frac{cons_{ijt}}{(1+r)^t} \tag{1.25}$$

$$Income_{ij} \equiv Wealth_{ij} + \sum_{t=1}^{T} \frac{wage_t(c_{ij}, \varepsilon_{ij})}{(1+r)^t} \tag{1.26}$$

This simpler budget constraint simply states that lifetime consumption needs to be less than lifetime income. Note that throughout this article I have used a simplified discount rate $(1+r)$ to stand in for both the expected financial savings rate, and the future discount rate. In reality these two discount rates are random variables that are most likely different from each other. I will leave the discussion of savings rates and discount rates to future work.

We can discuss the three assumptions that allow me to express a student's enrollment decision as a maximization over a linear indirect utility:

Assumption 7. *(Borrow) Students can borrow enough for any school and savings can be any value.*

- $\chi_{ij} \leq Wealth_{ij} + MaxBorrow_{ij} \forall j \in J, i \in I$
- $save_{ijt}$ *takes any value for* $t \geq 1$.

The biggest assumption behind specifying an indirect utility is that students are able to borrow enough to attend any school. The easiest argument for these loose borrowing constraints comes from the presence of the PLUS loan, which had an upper bound equal to the cost of attendance (COA) of the school. Recall that the cost of attendance is the tuition and fees plus anticipated housing and supply costs.

Assumption 8. *(Consume) Utility cares about the sum of discounted consumption (and increases with it.)*

$$U([cons_{ijt}]_{t=0}^{T}, c_{ij}, \varepsilon_{ij}) = U(Consume_{ij}, c_{ij}, \varepsilon_{ij})$$
$$\frac{\partial U}{\partial Consume_{ij}} > 0$$

Students should value consumption over their lifetime. This is an assumption that lifetime consumption can be boiled down into a single index ($Consume_{ij}$) that matches the way lifetime income can be boiled down into a single net present value. I also assume that utility is always increasing in consumption so that I can plug income directly into the utility function.

Assumption 9. *(Linear) Utility is linear in* $(Income, c_{ij}, \varepsilon)$ *and Income is linear in* (c_{ij}, ε)

When the student utilities and lifetime income are linear functions of the observed characteristics, I can collapse the entire maximization problem to make a linear indirect utility function. We can think of this as a first-order Taylor series expansion of the utility function.

These three assumptions together justify a linear indirect utility function as assumed by Assumption 1 in the main article.

1.11 Appendix C: School Utility

This section discusses the school maximization problem. When deciding admission sand scholarships, the school has four priorities in mind:

- The intelligence and ability of their incoming class.
- The capacity constraint of their school.
- The revenue constraint of their school.
- The income/gender/racial composition of their school.

These priorities define the schools maximization problem. Schools choose admissions policy $f_{e,j}$, and net tuition policy $f_{\chi,j}$:

$$
\begin{aligned}
f_{e,j}(c_{ij}, \eta_{aij}) &= \mathbb{1}(j \text{ admits } i) \\
f_{\chi,j}(c_{ij}, \eta_{ij}) &= \text{ amount } i \text{ pays out of pocket to enroll in } j
\end{aligned}
$$

While optimizing, schools take for granted the admissions policies of other schools, $\overrightarrow{f}_{e,-j}$, net tuition policies of other schools, $\overrightarrow{f}_{\chi,-j}$, and the student chance to apply and enroll, $Pr_{ae,i}$:

$$
Pr_{ae,i}(f_{e,j}, f_{\chi,j}; \overrightarrow{f}_{e-j}, \overrightarrow{f}_{\chi-j}, \overrightarrow{c}_{ij}, \eta_{ij}, \eta_{aij}) = \\
E_{\varepsilon_i}\left(\mathbb{1}(j \in O_i^*(\overrightarrow{f_e}, \overrightarrow{f_\chi})) \cap \mathbb{1}(j = j_i^*(\overrightarrow{f_e}, \overrightarrow{f_\chi}))|\eta_{ij}, \eta_{aij}\right)
$$

Recall that students don't know school preferences $(\overrightarrow{\eta}_i, \overrightarrow{\eta}_a)$ when applying. Then I can define the school payoff as follows:

$$\sum_{i \in I} U_\pi(c_{ij}, \eta_{ij}, \eta_{aij}) f_{e,j}(c_{ij}, \eta_{aij}) Pr_{ae,i}(f_{e,j}, f_{\chi,j}; \overrightarrow{f}_{e-j}, \overrightarrow{f}_{\chi-j}, \overrightarrow{c}_{ij}) \tag{1.27}$$

$$\text{s.t.} \sum_{i \in I} f_{e,j}(c_{ij}, \eta_{aij}) Pr_{ae,i}(f_{e,j}, f_{\chi,j}; \overrightarrow{f}_{e-j}, \overrightarrow{f}_{\chi-j}, \overrightarrow{c}_{ij}) \leq \kappa_c(c_j) \tag{1.28}$$

$$\sum_{i \in I} f_{\chi,j}(c_{ij}, \eta_{aij}) f_{e,j}(c_{ij}, \eta_{aij}) Pr_{ae,i}(f_{e,j}, f_{\chi,j}; \overrightarrow{f}_{e-j}, \overrightarrow{f}_{\chi-j}, \overrightarrow{c}_{ij}) \geq \kappa_\pi(c_j) \tag{1.29}$$

$$\sum_{i \in I} U_{\pi,(k)}(c_{ij}, \eta_{ij}, \eta_{aij}) f_{e,j}(c_{ij}, \eta_{aij}) Pr_{ae,i}(f_{e,j}, f_{\chi,j}; \overrightarrow{f}_{e-j}, \overrightarrow{f}_{\chi-j}, \overrightarrow{c}_{ij}) \geq \kappa_{\pi,k}(c_j) \tag{1.30}$$

Where U_π and $U_{\pi,(k)}$ are the schools preferences for the student with characteristics c_{ij}.

Upon writing the this maximization problem in a Lagrange formulation, we see that the capacity constraint (Equation 1.28) means some students aren't admitted, the revenue constraint (Equation 1.29) means net tuition is positive for some students, and the diversity constraints (Equation 1.30) mean some races/genders/income classes are more preferred.

It is possible to get a handle on the admissions, revenue, and diversity constraints by examining the outcomes of the model that I predicted. If a school had open admissions, then I can only know about the school's revenue constraint. If the school was selective, I could use the expected total enrollment predicted by my model as a stand-in for the capacity constraint. I can also calculate the expected total enrollment for each race/gender/subgroup to examine diversity requirements. Unless all the students were admitted for free, I could use the expected total net tuition received as a stand-in for the revenue constraint.

In the main paper, I assumed that admission policies are fixed and unchanged in response to changes in government policy, student behavior, and the behavior of other schools. Instead, I could conduct the counterfactual: What would have happened without the Cal Grant, but schools still needed to have the same enrollment level. In the counterfactual in my paper above, removing the cal grant would reduce the number of in-state enrollees, but if schools need to have the same enrollment, they may increase admission chances to achieve this target. I can simulate increased admission chances until the number of in-state enrollees is at the same level as before. Even though the number of in-state enrollees is the same, the composition of enrolled students will have changed.

1.12 Appendix D: Measurement Error Can Corrupt Control Function

If net tuition is correlated with any unobserved variables that also impact school valuation, then the observed relationship between net tuition and school valuation does not only contain the causal relationship between the two. This correlation is known as omitted variable bias, or in economic jargon, it means net tuition is endogenous to school valuation.

If I have a source of random variation in net tuition, then I can detect the causal impact of net tuition with a "two-stage least squares" approach. I estimate the impact of the random shift on net tuition, and then I estimate the impact of the random shift of school valuation. I can deduce the impact of net tuition on school valuation from these two estimates. This two-stage approach is commonly known as "two-stage least squares." Recent literature with nonparametric models has advertised the "control function" approach. The "control function" approach estimates the impact of the random shift of net tuition and generates a "first-stage residual," which is supposed to account for the part of net tuition that is correlated with any unobserved determinants of school valuation. Finally, the new second stage would estimate the impact of net tuition on school valuation while including the newly generated "first-stage residual." The key difference between these two approaches is the second stage: "two-stage least squares" estimates the relationship between the random shift and school valuation in the second stage, while "control function" estimates the relationship between net tuition and school valuation while adding a "first-stage residual" as a control variable.

In a simple example below, I will show that when net tuition is measured with error, the "control function" approach doesn't do what it advertises. Even controlling for the "first-stage residual," the observed net tuition is still correlated with unobserved determinants of school valuation.

Let's assume that school valuation is observed. I observe N data points $(y, x, z)_{i=1}^{N}$, which are structurally related as below:

$$y = \chi\alpha + \varepsilon \tag{1.31}$$

$$\chi = z\pi + \eta \tag{1.32}$$

$$x = \chi + m \tag{1.33}$$

Where y is the school valuation, x is the observed net tuition, and z is a random shift of net tuition. Notice from Equation 1.33 that observed net tuition is the actual net tuition measured with error.

The endogeneity of net tuition comes from a potential correlation between η and ε. To make the exposition simpler, let's assume that $\varepsilon = \eta\rho + \omega$.

Finally we have the identifying assumption:

$$z \perp\!\!\!\perp (\eta, \omega) \tag{1.34}$$

I would like to estimate α in an unbiased manner. The "two-stage least squares" approach gives us the estimator: $\hat{\alpha} = \frac{z'y}{z'x}$. Even with the presence of measurement error, we still have asymptotic consistency: $\hat{\alpha} \xrightarrow{p} \alpha$. (Note that the estimate of the first stage coefficient $\hat{\pi} = \frac{z'x}{z'z}$ is consistent: $\hat{\pi} \xrightarrow{p} \pi$.)

The "control function" approach generates a first stage residual $\hat{\eta}_{wrong} = x - z\hat{\pi}$, and then adds this residual as a control variable to obtain the estimates:

$$\begin{bmatrix} \hat{\alpha}_{wrong} \\ \hat{\rho}_{wrong} \end{bmatrix} = \left[\begin{bmatrix} x \\ \hat{\eta}_{wrong} \end{bmatrix}' \begin{bmatrix} x \\ \hat{\eta}_{wrong} \end{bmatrix} \right]^{-1} \begin{bmatrix} x \\ \hat{\eta}_{wrong} \end{bmatrix}' y$$

Once we realize that the observed net tuition (x) has measurement error (m), then we see that the first stage residual actually converges to a different error term: $\hat{\eta}_{wrong} \vec{p} x - z\pi = \eta + m$

.

The true structural equation with the first stage residual included is:

$$y = \underbrace{(\chi + m)}_{x} \alpha + \underbrace{(\eta + m)}_{\hat{\eta}_{cf}} \rho + (\omega - m(\alpha + \rho))$$

Notice that the leftover error term is still correlated with the observed net tuition, even though we controlled for the (incorrect) first-stage residual: $Cov((\chi + m), (\omega - m(\alpha + \rho))) = -(\alpha + \rho)\sigma_m^2 \neq 0$. Therefore the estimate from a control function approach will be asymptotically inconsistent.

I am actually trying to maximize a choice probability. I would like to show that in a choice probability setting one must take care to specify the correct underlying distribution when doing a plug-in estimate (which is a choice-estimate analog of "two-stage least squares").

In a choice probability setting, I only observe $(e, x, z)_i^N$, where $e = \mathbb{1}(y > 0)$ denotes the choice. Here in my example the student enrolls in the school as long as the school valuation is above 0. The maximum likelihood setup looks like:

$$\hat{\pi}, \hat{\alpha}, \hat{\rho} = \arg\max_{\alpha,\pi} \sum_{i=1}^{N} \big\{ log(Pr(z_i \pi + \eta)) + \\ e_i log(Pr(z_i \pi \alpha + \underbrace{\eta_i \alpha + \varepsilon_i}_{\text{new error}} > 0)) + (1 - e_i) log(Pr(z_i \pi \alpha + \underbrace{\eta_i \alpha + \varepsilon_i}_{\text{new error}} < 0)) \big\}$$

Note the new underlying distribution: $\eta_i \alpha + \varepsilon_i$. Therefore any multinomial choice model would have to take care to use the correct underlying distribution.

CHAPTER 2

The Heterogeneous Returns to Education

2.1 Introduction

There is a long history of using instrumental variables to determine the causal effect of schooling on education. Angrist and Krueger (1991) finds that each additional year of schooling can boost earnings by about 8%. The discussion of methods for determining our confidence in these point estimates has spurred a lot of literature. Current empirical applications of instrumental variables face must first pass a binary test of whether the instruments are "strong" enough for us to assume that the likelihood function is quadratic, and therefore the estimate is distributed normal with variance as derived from the delta method. This article is the first to estimate and make statements of confidence about the heterogeneous returns to education by region of birth using the same 1980 Census data used by Angrist and Krueger. We do not suffer from "weak instruments" by avoiding using an asymptotic approximation (Leamer, 2010) of the instrumental variables two-stage least squares (IV-2SLS) estimate.

Instead, we sample from the posterior distribution of our two-stage estimate and form point estimates and measures of confidence directly from the samples. When using weak priors, we obtain the same point estimates as the standard IV-2SLS methods, but we have much larger 95% credible intervals. When using the "best" priors as determined from cross-validation, we find that we are 95% sure that the returns to education are positive for only four out of nine regions, whereas the standard IV-2SLS approach would yield "significant" results for all nine geographic regions.

Bound et al. (1995) uses simulated (fake) instruments to obtain similar results of "significance" to the results obtained by (Angrist and Krueger, 1991). In the case of IV-2SLS, standard definitions of significance is defined by whether zero lies in a 95% confidence interval formed by applying the delta method to the two-stage least squares estimate. The delta method is an appeal to the large sample asymptotics of the IV-2SLS estimate. However, we know that the just-identified IV-2SLS estimate does not have finite moments (Nelson and Startz, 1990; Phillips, 2009) [1], and therefore the appeal to asymptotics in the case of IV-2SLS analysis is diminished.

Chamberlain and Imbens (1996) explains that sampling from the posterior instead of using an asymptotic approximation leads to tight posterior intervals when using the data, and wide posterior intervals when using randomly simulated instruments. Lopes and Polson (2014) and Hoogerheide et al. (2007) discuss a variety of priors to further analyze the causal returns to education in the same setting. This article is the first to add the analysis of heterogeneous treatment effects to the tradition of applying Bayesian credible inference methods towards empirical data.

Due to the increased availability of big data and large scale AB tests in the tech industry, readily available code packages (EconML, 2019) inspired by (Newey and Powell, 2003; Hartford et al., 2017) have surfaced that advertise the ability to determine heterogeneous treatment effects in a nonparametric manner. For the initial impact of the AB test on an initial outcome, the method for discovering causal heterogeneous effects is quite straightforward. However, once we ask for the causal effect of the initial outcome on a secondary outcome, we begin to confront the issue of weak instruments. Staiger and Stock (1994); Stock and Yogo (2002) discuss conditions under which the instrument is "strong enough" to assume the IV-2SLS estimate is normal.

Their approach requires the source of heterogeneity to be non-stochastic, or independent from the omitted factors that determine the treatment. In the case of heterogeneous returns to schooling by region-of-birth, this means that we require the region-of-birth to be independent from the omitted variables that determine the years of schooling. Instead, we implement a strategy that avoids this pitfall at the cost of including estimating additional first-stage equations.

section 2.2 summarizes the data, section 2.3 compares using the posterior distribution

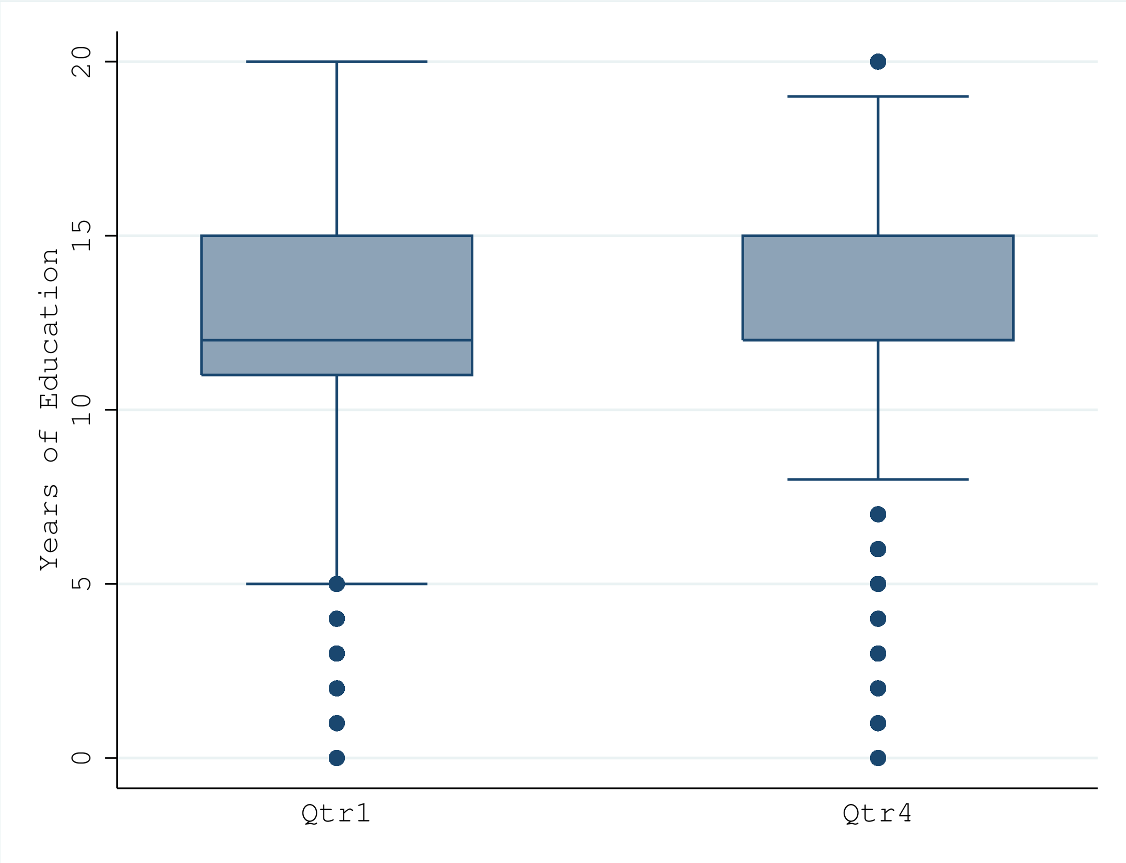
[1] The number of finite moments depends on the degree of over-identification.

with standard instrumental variables research design, section 2.4 sets up a model of heterogeneous treatment effects and shows why we have many first-stage equations, section 2.5 discusses the method of sampling from the posterior, section 2.6 summarizes our results, and section 2.7 concludes.

2.2 Data

Following (Angrist and Krueger, 1991), we use the 5 percent Public Use sample (the A Sample) of the US population as of April 1, 1980. Following Chamberlain and Imbens (1996), we filtered down to native-born African American/Black or Caucasian/white males with birthdays in the first or fourth quarter of a year between 1930 and 1939 (inclusive). The key insight for our identification strategy is that men born in the fourth quarter tend to have about one more year of schooling. Figure 2.1 shows the interquartile ranges of the years of education against quarter-of-birth. For census participants born in the fourth quarter, we see that the first quartile of the level of schooling is equal to the median. This suggests there is a binding lower bound to the number of years of education for census participants born in the fourth quarter.

Figure 2.1: Quarter of Birth Against Years of Education



This figure shows the interquartile ranges of the years of education against quarter-of-birth. For census participants born in the fourth quarter, we see that the first quartile of the level of schooling is equal to the median. This suggests there is a binding lower bound to the number of years of education for census participants born in the fourth quarter.

I limit the sample to men with positive wage and salary earnings, and positive weeks worked in 1979, leaving 202,859 participants in the sample. We compute weekly earnings by dividing annual earnings by weeks worked. Table 2.1 shows that the average man was 45 years old and earned $430 dollars per week in 1979. Eighty four percent of these men were married, and nine percent of these men were African American/Black. The level of schooling is determined by the years of completed schooling, where twelve corresponds to completing high school. The original paper from 1991 conditioned on some concomitant variables such as place of current residence, and marital status, however conditioning on concomitant variables can corrupt the causal interpretation of the data, so I remove these controls. For this article, we weight each census respondent equally, and we leave extensions to weighted calculations for future research.

Table 2.1: Descriptive Statistics

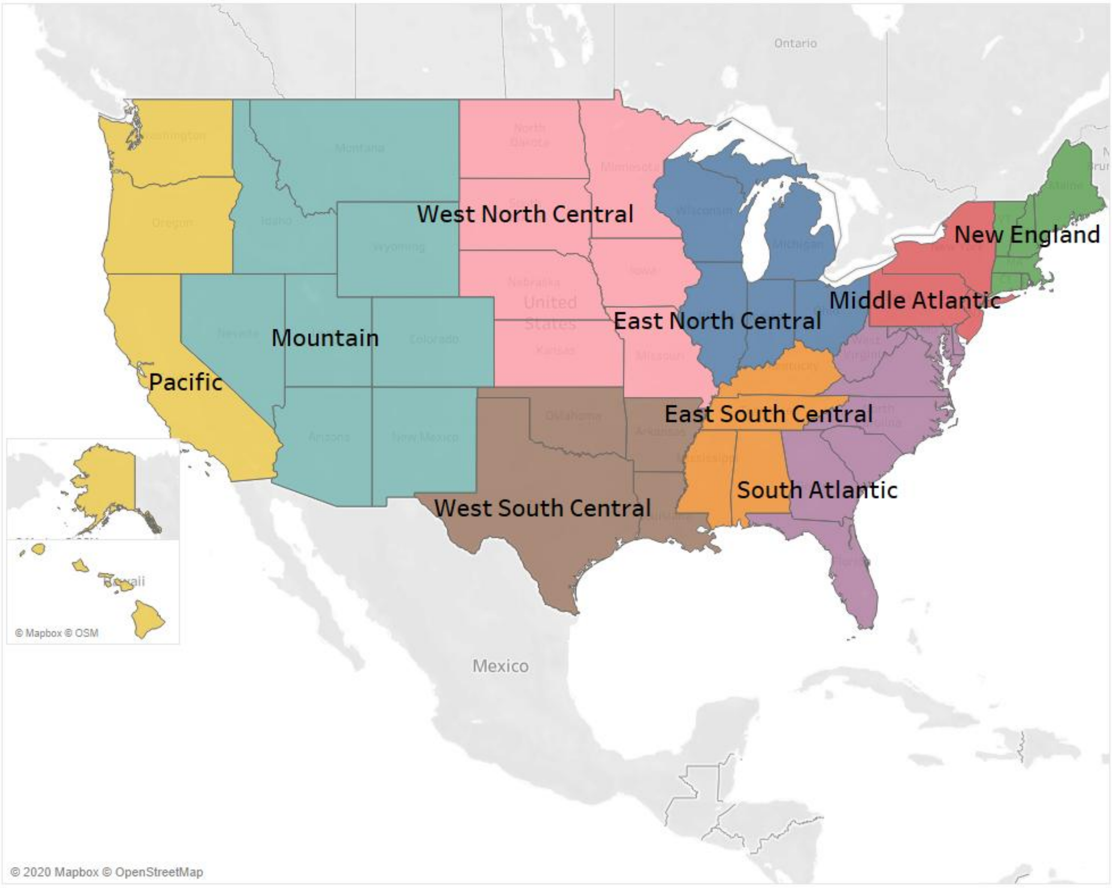
	Mean	SD	10th Pct	90th Pct
Weekly Wages	427.4837	253.9709	179.4828	692.4039
Years of Education	12.68296	3.301034	8	17
Age	44.90577	2.948244	41	49
Married	.8406331	.3660187	0	1
African America	.0946618	.292748	0	0

N = 202,859 individuals in dataset.

Source: 1980 Census (5% Public Use Sample A). We filtered down to native born African American/Black or Caucasian/white males with birthdays in the first or fourth quarter of a year between 1930 and 1939 (inclusive) with positive wage and salary earnings, and positive weeks worked in 1979.

Since some of the states are much smaller than the others, we follow the original paper in using census defined geographic regions as the source of potentially heterogeneous treatment effects. Since the region of birth is determined chronologically before any educational and employment decisions we can worry less about the issue of concomitant variables. Figure 2.2 displays the census regions. Table 2.2 lists out the states that make up each region.

Figure 2.2: Census Regions



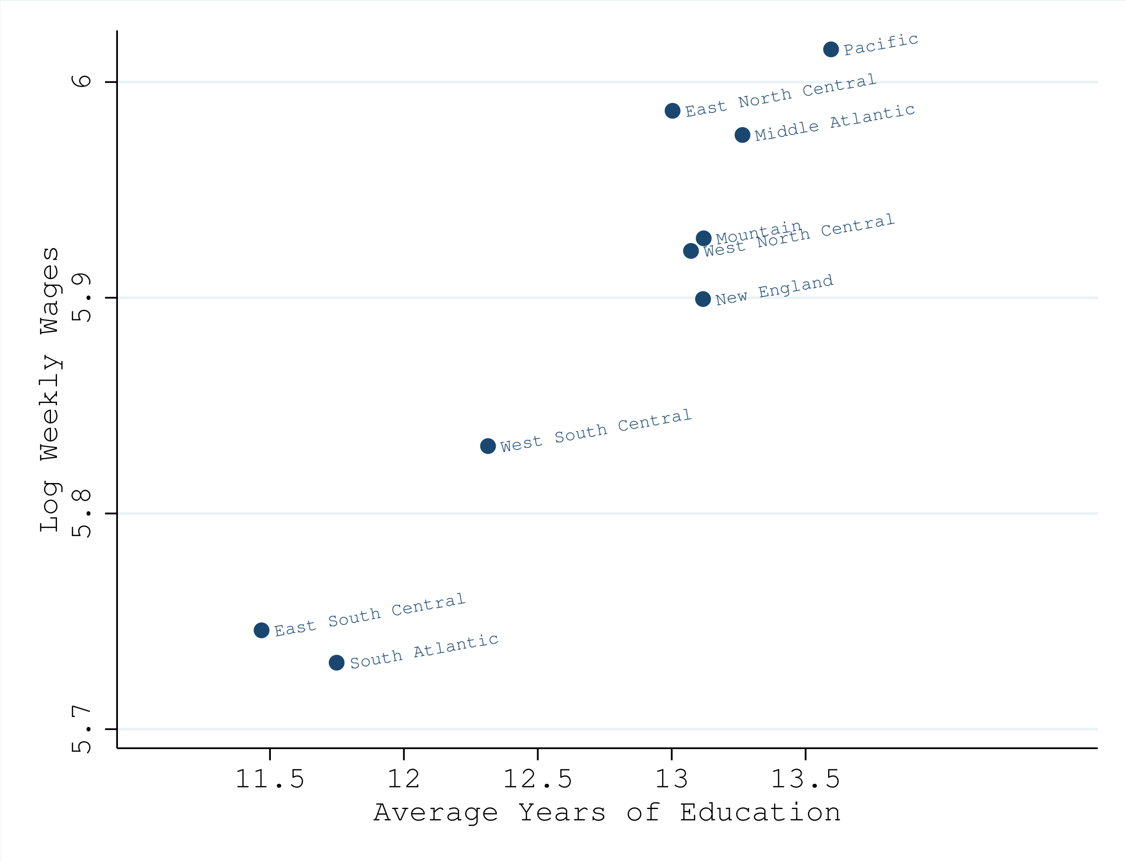
This figure shows the census defined geographic regions. Table 2.2 lists out the states that make up each region.

Table 2.2: Geographic Regions

Census Region	Birth State
East North Central	Illinois, Indiana, Michigan, Ohio, Wisconsin
East South Central	Alabama, Kentucky, Mississippi, Tennessee
Mountain	Arizona, Colorado, Idaho, Montana, Nevada, New Mexico, Utah, Wyoming
South Atlantic	Delaware, District Of Columbia, Florida, Georgia, Maryland, North Carolina, South Carolina, Virginia, West Virginia
West North Central	Iowa, Kansas, Minnesota, Missouri, Nebraska, North Dakota, South Dakota
West South Central	Arkansas, Louisiana, Oklahoma, Texas
Middle Atlantic	New Jersey, New York, Pennsylvania
New England	Connecticut, Maine, Massachusetts, New Hampshire, Rhode Island, Vermont
Pacific	Alaska, California, Hawaii, Oregon, Washington

Figure 2.3 shows the relationship between the average log weekly earnings against the average years of education for each census region. We see a general positive trend, but more importantly, we see that each region has differing average levels of weekly earnings and education. This article checks if the returns to education differs across these geographic regions.

Figure 2.3: Income Against Education by Census Region



This figure shows the relationship between the average log weekly earnings against the average years of education for each census region. We see a general positive trend, but more importantly, we see that each region has differing average levels of weekly earnings and education. This article checks if the returns to education differs across these geographic regions.

The quarter-of-birth is our instrument. We follow the original paper in creating more instruments by interacting the quarter of bith with the census regions and the years of birth. Adding all these instruments would make us worry about adding "weak" instruments, which would negatively affect inference in the IV-2SLS estimation approach because weak instruments make the IV-2SLS more biased and further away from being normally distributed. Due to this issue, we don't assume that the final estimate is distributed normally. Instead we sample from the posterior of the our instrumental variables estimate.

2.3 Comparison of Instrumental Variables Methods

This section compares the standard Instrumental Variables approach to a one one that examines the actual posterior distribution of a Bayesian estimate. To easily compare with previous literature, we focus on the nationwide returns to schooling first (instead of the heterogeneous returns to schooling). The current rule of thumb for for when one "can" use the instrumental variables is that the first stage F-statistic should be large – greater than 20. This section confirms that a confidence interval obtained from using the actual posterior distribution of the IV-2SLS approach agrees with one obtained from using the asymptotic distribution. We also plug in a variety of priors using estimates of the returns to education obtained from other studies and other datasets, by using these other estimates as priors, we essentially pool the data from the two datasets. Of course, this is assuming that the two datasets are independent, and that the returns to education are homogeneous.

2.3.1 Setup of Simple Model

We start of with a simple model of the returns to education. We assume that the returns to education are homogeneous for now:

$$Y_i = S_i\beta_S + X_i\beta_{SX} + \beta_{Sc} + \varepsilon_i \tag{2.1}$$

$$S_i = Z_i\pi_S + X_i\pi_{SX} + \pi_{Sc} + \eta_i \tag{2.2}$$

Where Y_i is weekly earnings in 1979, S_i is the level of schooling, and our instrument, Z_i, is an indicator of whether the student was born in the fourth quarter. The controls X_i include the year of birth, birthplace, and race.

We can also remove the projection of the data onto X. Let $\perp$ refer to the residuals after projecting onto the controls and the constant:

$$Y_i^{\perp} = S_i^{\perp}\beta_S + \varepsilon_i^{\perp} \tag{2.3}$$

$$S_i^{\perp} = Z_i^{\perp}\pi_S + \eta_i^{\perp} \tag{2.4}$$

The simple instrumental variables assumptions are as follows:

Assumption 10. *The instrument is relevant, excluded, and unconfounded:*

$$\textit{Relevance: } \pi_S \neq 0 \tag{2.5}$$

$$\textit{Exclusion: } (Z_i^{\perp} \perp\!\!\!\perp Y_i^{\perp})|S_i \tag{2.6}$$

$$\textit{Unconfoundedness: } Z_i^{\perp} \perp\!\!\!\perp (\varepsilon_i, \eta_i) \tag{2.7}$$

Finally, we assume that the errors are distributed normally:

Assumption 11. *The errors are multivariate normal with mean zero and variance* Σ*:*

$$(\varepsilon_i, \eta_i) \sim N((0,0), \Sigma)$$

The IV2SLS estimate is the value of β_S that maximizes the likelihood of observing our data. Recall that by the Frish-Waugh-Lovell theorem, we can simply use the data after removing the projection onto X to obtain our point estimates. Let the subscript N denote a column of stacked data for each individual:

$$\widetilde{\pi}_{2SLS} = (Z_N^{\perp\prime} Z_N^{\perp})^{-1}(Z_N^{\perp\prime} S_N^{\perp}) \tag{2.8}$$

$$\hat{S}_N^{\perp} = Z_N^{\perp}\widetilde{\pi}_{2SLS} \tag{2.9}$$

$$\widetilde{\beta}_{2SLS} = (\hat{S}_N^{\perp\prime}\hat{S}_N^{\perp})^{-1}(\hat{S}_N^{\perp\prime} Y_N^{\perp}) \tag{2.10}$$

According to the standard research design for IV-2SLS estimates, Staiger and Stock (1994); Stock and Yogo (2002) propose using the first stage F-test for whether π is zero:

$$F_Z \equiv (N-1)\frac{SSE_{null} - SSE_Z}{SSE_Z} \tag{2.11}$$

$$SSE_Z = \hat{\eta}_N^{\perp\prime}\hat{\eta}_N^{\perp} \tag{2.12}$$

$$SSE_{null} = S_N^{\perp\prime}S_N^{\perp} \tag{2.13}$$

$$\hat{\eta}_N^{\perp} = S_N^{\perp} - \hat{S}_N^{\perp} \tag{2.14}$$

The current standard in instrumental variables analysis is to use the asymptotic distribution of the estimate if F_Z is greater 10 (Staiger and Stock, 1994) or 20 (Stock and Yogo, 2002). The asymptotic distribution of the IV-2SLS estimate is normal, centered at the true value, and with variance derived from taking the inverse of the Fisher information matrix (or equivilantly, from using the delta method):

$$V\hat{a}r(\widetilde{\beta}_{2SLS}) = (\widehat{RH}_{2N}'\widehat{RH}_{2N})^{-1}\hat{\eta}_N'\hat{\eta}_N/(N-1) \tag{2.15}$$

$$\widehat{RH2} = [Z\widetilde{\pi}_{2SLS}, X, 1] \tag{2.16}$$

$$\hat{\eta}_N = Y_N - \widehat{RH}_{2N}\widetilde{\beta} \tag{2.17}$$

Where $R\hat{H}2$ denotes the right-hand side of the second stage with the predicted levels of schooling plugged in for the actual level of schooling, and $\widetilde{\beta} \equiv (\widehat{RH2}'\widehat{RH2})^{-1}\widehat{RH2}'Y_N$ is the full set of coefficients in the regression that includes the predicted levels of schooling.

The discussion on weak instruments, and when it is appropriate to use the asymptotic distribution is far from over, see (Young, 2019; Andrews et al., 2019; Bekker, 1994; Hahn and Hausman, 2002) for history of instrumental variables estimates and the variety of methods to determining confidence. The main takeaway is that Equation 2.15 can yield variances that can be many orders of magnitude too small. In the just identified case (one treatment and one instrument), the IV-2SLS estimate doesn't even have finite variance, therefore the variance as derived from the delta method is infinitely too small, and it is inappropriate to use bootstrap to estimate an infinite variance.

Instead of assuming the asymptotic distribution, we can look at the posterior dsitribution of our estimate (Lopes and Polson, 2014; Hoogerheide et al., 2007; Chamberlain and Imbens, 1996). To do so, we put priors on the covariance of the errors Σ and the coeffients β, π:

Prior 1.

$$\Sigma \sim IW(n_0 = 0.01, \Sigma_0 = (0.01) * \boldsymbol{I}_2) \tag{2.18}$$

$$(\beta, \pi) \sim N((\mu_\beta, \mu_\pi), \begin{pmatrix} \lambda_\beta^{-1} & 0 \\ 0 & \lambda_\pi^{-1} \end{pmatrix}) \tag{2.19}$$

Where μ_β and μ_π are vectors that represent our prior point estimates for the values of the coefficients β and π, and λ_β^{-1} and λ_π^{-1} are diagonal matrices that represent the strength of our priors. Throughout this exercise, we keep all prior variances at 1 million, except for the prior on the key coefficient that represents how much schooling affects earnings.

The reduced form set of simultaneous equations is:

$$
\begin{aligned}
Y_i &= (Z_i\pi_S + X_i\pi_{SX} + \pi_{Sc} + \eta_i)\beta_S + X_i\beta_{SX} + \beta_{Sc} + \varepsilon_i && (2.20)\\
&= Z_i\pi_S\beta_S + X_i(\pi_{SX}\beta_S + \beta_{SX}) + \pi_{Sc}\beta_S + \beta_{Sc} + \underbrace{(\eta_i\beta_S + \varepsilon_i)}_{\eta_{y,i}} && (2.21)\\
&= Z_i\pi_S\beta_S + X_i(\pi_{SX}\beta_S + \beta_{SX}) + \pi_{Sc}\beta_S + \beta_{Sc} + \eta_{y,i} && (2.22)\\
S_i &= Z_i\pi_S + X_i\pi_{SX} + \pi_{Sc} + \eta_i && (2.23)\\
\eta_{y,i} &\equiv (\eta_i\beta_S + \varepsilon_i) && (2.24)
\end{aligned}
$$

Therefore the reduced from errors $(\eta_{y,i}, \eta_i)$ have the reduced form covariance matrix:

$$\Omega = \begin{pmatrix} 1 & \beta_S \\ 0 & 1 \end{pmatrix} \Sigma \begin{pmatrix} 1 & 0 \\ \beta_S & 1 \end{pmatrix} \tag{2.25}$$

$$= \begin{pmatrix} \omega_{yy} & \omega_{ys} \\ \omega_{sy} & \omega_{ss} \end{pmatrix} \tag{2.26}$$

2.3.2 Gibbs sampling from Posterior of Simple Model

In order to sample from the posterior distribution of $\widetilde{\beta}_{2SLS}$, we express the conditional distribution of the three key parameters β_S, π_S, Σ on each other:

$$(\Sigma|\beta_S, \pi_S, Y_N, S_N, Z_N, X_N) \sim IW(n_0 + N, \Sigma_0 + N\hat{\Sigma}_N) \tag{2.27}$$

$$(\beta_S|\Sigma, \pi_S, Y_N, S_N, Z_N, X_N) \sim N(b_S, B_S) \tag{2.28}$$

$$(\pi_S|\Sigma, \beta_S, Y_N, S_N, Z_N, X_N) \sim N(p_S, P_S) \tag{2.29}$$

$$\begin{aligned}
\hat{\Sigma}_N &= [\hat{\varepsilon}_N, \hat{\eta}_N]'[\hat{\varepsilon}_N, \hat{\eta}_N] \\
B_S^{-1} &= \lambda_\beta + \widetilde{\widehat{RH}}'_{2N}\widetilde{\widehat{RH}}_{2N} \\
(B_S)^{-1}b_S &= \lambda_\beta\mu_\beta + \widetilde{\widehat{RH}}'_{2N}\widetilde{Y}_N \\
P_S^{-1} &= \lambda_\pi + \widetilde{RH}'_{1N}\widetilde{RH}_{1N} \\
(P_S)^{-1}p &= \lambda_\pi\mu_\pi + \widetilde{RH}'_{1N}\widetilde{S}_N \\
\widetilde{Y}_N &= (Y_N - (\hat{\eta}_N)\omega_{ss}^{-1}\omega_{sy})\omega_{yy|s}^{-0.5} \\
\widetilde{S}_N &= (S_N - (\hat{\eta}_{y,N})\omega_{yy}^{-1}\omega_{ys})\omega_{ss|y}^{-0.5} \\
\widetilde{\widehat{RH}}_{2N} &= \widehat{RH}_{2N}\omega_{yy|s}^{-0.5} \\
\widetilde{RH}_{1N} &= RH_{1N}\omega_{ss|y}^{-0.5} \\
\omega_{yy|s} &= \omega_{yy} - \omega_{ys}\omega_{ss}^{-1}\omega_{sy} \\
\omega_{ss|y} &= \omega_{ss} - \omega_{sy}\omega_{yy}^{-1}\omega_{ys} \\
\hat{\eta}_N &= S_N - RH_{1N}\pi_S \\
\hat{\varepsilon}_N &= Y_N - RH_{2N}\beta_S \\
\hat{\eta}_{y,N} &= Y_N - \widehat{RH}_{2N}\beta_S \\
RH_{1N} &= [Z_N, X_N, 1] \\
RH_{2N} &= [S_N, X_N, 1] \\
\widehat{RH}_{2N} &= [RH_{1N}\pi, X_N, 1]
\end{aligned}$$

Please see section 2.5 for a detailed explanation and intuition behind the calculations here. Theses conditional posteriors are derived by multiplying the liliehood by the prior, and then noticing that the resulting functional form takes the shape of Inverse Wishart or Normal distributions respspectively. In order to draw from the posterior distribution of our instrumental variables estimate, we perform Gibbs sampling by iteravely drawing from the distributions expressed in Equations (2.27 ,2.28, 2.29).

2.3.3 Comparison of Methods

Table 2.3: Comparison of Methods of Inference for Instrumental Variables

Standard Research Design

		First Stage F-Statistic	Point Estimate	95% Confidence Interval Based on Asymptotic Distribution: Lower Bound	95% Confidence Interval Based on Asymptotic Distribution: Upper Bound
[1]	OLS		0.0633	0.0624	0.0641
[2]	IV-2SLS	110.5070	0.0685	0.0356	0.1013

Bayesian Posterior Calculation with Five Different Priors

	Priors Based on ...	Prior Distribution is Normal with ... Mean	Prior Distribution is Normal with ... Standard Deviation	Posterior Distribution: Mean	Posterior Distribution: 95% Credible Interval Lower Bound	Posterior Distribution: 95% Credible Interval Upper Bound
[3]	-	0	1000	0.0684	0.0354	0.1018
[4]	-	0	0.05	0.0612	0.0292	0.0931
[5]	-	0.15	0.05	0.0765	0.0457	0.1080
[6]	*Bettinger et al. (2019)* Data: All CA students eligible for Cal Grant (1998 to 2000). IV: Discontinuity in GPA eligibility for Cal Grant.	0.054	0.019	0.0618	0.0367	0.0864
[7]	*Card (1995b)* Data: NLSY (1966 Cohort) IV: Indicator for living near a 4-year college.	0.097	0.048	0.0715	0.0407	0.1028

Notes:

[1-2]: These point estimates, first stage F-statistic, and 95% confidence intervals are what you would find in a standard instrumental variables research design. The first-stage F-statistic is large enough for us to assume the IV-2SLS estimate is distributed according to its asymptotic distribution.

[3-7]: These point estimates and 95% credible intervals are obtained from sampling from the posterior distribution of the IV-2SLS estimate. The five estimates only vary the prior on the coefficient that determines the impact of schooling on earnings. All other coefficients use a normal prior with mean zero and variance one million.

[3]: Since a prior with such a high variance is relatively uninformative, the point estimate and the 95% credible interval agree with [2].

[4-5]: If we had a relatively strong prior belief that were no returns to education, then this prior combined with the evidence would lead us to reject our prior point estimate at a 95% credible level. Likewise, if we had a relatively strong prior belief that each year of schooling increased earnings by 15%, then the posterior would lead us to reject our prior point estimate at a 95% credible level.

[6-7]: We use the point estimates and standard deviations reported by other papers. Assuming the returns to education are homogeneous, then the posterior point estimates and credible intervals represent pooling evidence from the two datasets. Since the previous papers' point estimates all lie within the 95% posterior credbile interval, the 1980 census data of men born between 1930 and 1940 does not lead us to reject the estimates found in previous literature at a 95% credible level.

Table 2.3 summarizes the variety of methods of determining confidence in the instrumental variables estimates. The first two rows show point estimates, first stage F-statistic, and 95% confidence intervals as one would find in a standard instrumental variables research design. The first-stage F-statistic is large enough for us to assume the IV-2SLS estimate is distributed according to its asymptotic distribution. The OLS reports a much tighter confidence interval, but without an identification strategy the estimate does not have a causal interpretation.

The final four rows shows point estimates and 95% credible intervals obtained from sampling from sampling from the posterior distribution of the IV-2SLS estimate. The five estimates only vary the prior on the coefficient that determines the impact of schooling on earnings. All other coefficients use a normal prior with mean zero and variance one million.

The third row reports nearly identical point estimates and 95% credible intervals to the standard IV-2SLS approach because we use a weak prior for the returns to education. This supports the claim from (Staiger and Stock, 1994; Stock and Yogo, 2002) that a large enough F-statistic allows us to use the asymptotic distribution of the IV-2SLS estimate. When we examine the heterogeneous returns to education in the next section – with multiple treatments – the asymptotic distribution and the posterior distribution of the IV-2SLS estimates will not agree anymore.

The fourth and fifth row show priors that would have their point estimates rejected at a 95% level when combined with the evidence. If we had a relatively strong prior belief that were no returns to education, then this prior combined with the evidence would lead us to reject our prior point estimate at a 95% credible level. Likewise, if we had a relatively strong prior belief that each year of schooling increased earnings by 15%, then the posterior would lead us to reject our prior point estimate at a 95% credible level.

Instead of using a weak prior for the returns to schooling, the last three rows use the point estimates and standard deviations reported by other papers. Assuming the returns to education are homogeneous, then the posterior point estimates and credible intervals represents pooling evidence from the two datasets. Since the previous papers' point estimates all lie within the 95% posterior credbile interval, we can say that the 1980 census data of men born between 1930 and 1940 does not lead us to reject the estimates found in previous literature at a 95% credible level.

The last two rows use the point estimates and standard deviations reported by other papers. Assuming the returns to education are homogeneous, then the posterior point estimates and credible intervals represent pooling evidence from the two datasets. Since the previous papers' point estimates all lie within the 95% posterior credbile interval, the 1980 census data of men born between 1930 and 1940 does not lead us to reject the estimates found in previous literature at a 95% credible level.

2.4 Model for Heterogeneous Returns to Schooling

We would like to estimate the heterogeneous impact of schooling on log weekly earnings. The heterogeneity is with respect to the region of birth [2]. Throughout this paper, data for each individual, i, is in row form, and N in the subscript denotes vertically stacked data for all the participants.

$$Y_i = S_i\beta_S + S_iG_i\beta_{SG} + G_i\beta_G + X_i\beta_X + \beta_c + \varepsilon_i \tag{2.30}$$

Y_i is log weekly earnings, S_i is years of schooling, $\underset{1\times(K_G-1)}{G_i}$ is a vector of indicators for each region of birth, excluding one to prevent over-saturation. We have a $K_G = 9$ geographic regions. The controls X_i include race and indicators for each age between 1930 and 1940. We have ten controls after leaving out one age indicator and one race indicator to prevent over-saturating the model. Let $K_X = 10$ represent the number of additional controls.

The instrument for schooling is an indicator of whether the respondent was born in the fourth quarter. We have the following reduced form relationships between schooling and the instrument:

$$S_i = Z_i\pi_{S1} + G_i\pi_{G1} + X_i\pi_{X1} + \pi_{c,1} + \eta_{i1} \tag{2.31}$$

$$S_iG_{k,i} = Z_i\pi_{Sk} + G_i\pi_{Gk} + X_i\pi_{Xk} + \pi_{c,k} + \eta_{ik} \text{ for } k \in \{2, ..., K_G\} \tag{2.32}$$

[2] For region definitions, see Figure 2.3. We leave the extension to random treatment effects for future research.

Where Z_i is a $1 \times K_Z$ vector of instruments, which includes any interactions between the quarter-of-birth and any other variables. In our analyis, $K_Z = 19$, which represents interacting the quarter-of-birth with region, age, and race indicators. This is the same approach taken in the original paper. Bound et al. (1995) delves into the issue of many and weak instruments to conclude that it is not appropriate to assume the resulting two-stage least squares estimate is normally distributed due to the issue of weak instruments. Chamberlain and Imbens (1996) shows how we can overcome this issue by examining the posterior distribution instead. This article is an extension of the posterior oriented analysis that examines heterogeneous treatment effects.

2.4.1 Instrumental Variables Assumptions

We have the usual identifying assumptions:

Assumption 12. *The instrument is relevant, excluded, and unconfounded:*

$$\textit{Relevance: } (\pi_S'\pi_S) \textit{is is invertible} \tag{2.33}$$

$$\textit{Exclusion: } (Z_i \perp\!\!\!\perp Y_i)|S_i, X_i \tag{2.34}$$

$$\textit{Unconfoundedness: } Z_i \perp\!\!\!\perp (\varepsilon_i, \eta_i)|X_i \tag{2.35}$$

Where $\underset{KZ \times KG}{\pi_S}$ is the matrix of coefficients that predict the treatments from the instruments. Some sufficient conditions for $(\pi_S'\pi_S)$ to be invertible are that $KZ > KG$ (we have more instruments than treatments), and that $\pi_{S,k}$ are all linearly independent (π_S is full column rank). This second sufficient condition means that the set of instruments should impact each treatment differently, otherwise we wouldn't be able to identify the impact two two different treatments from this set of instruments.

Note that Assumptions (2.34,2.35) are conditional on the control variables X_i, which can make the assumptions fore tenable: controlling for the birthyear and the birthplace, the location is indepdendent from omitted variables that determine schooling and earnings.

2.4.2 Augmented First-Stage Reduced Form

We have augmented the initial reduced form relationship (Equation 2.31) with expressions for each interaction between quarter-of-birth and geographic region (Equation 2.32). This is a departure from widely available "nonparametric heterogeneous instrumental variables" implementations in the Python package EconML (2019). To see why, we only plug one equation (Equation 2.31) into Equation 2.30:

$$\begin{aligned} Y_i = &(Z_i\pi_{S1} + G_i\pi_{G1} + X_i\pi_{X1} + \pi_{c,1} + \eta_{i1})\beta_{S1} + \\ &(Z_i\pi_{S1} + G_i\pi_{G1} + X_i\pi_{X1} + \pi_{c,1} + \eta_{i1})G_i\beta_{SG} + \\ &G_i\beta_G + X_i\beta_X + \beta_c + \varepsilon_i \end{aligned} \tag{2.36}$$

$$= Z_i(\pi_{S1}\beta_{S1} + \pi_{S1}G_i\beta_{SG}) + \tag{2.37}$$

$$\begin{aligned} &G_i(\pi_{G1}\beta_{S1} + \pi_{G1}G_i\beta_{SG} + \beta_G) + \\ &X_i(\pi_{X1}\beta_{S1} + \pi_{X1}G_i\beta_{SG} + \beta_X) + \\ &\beta_c + \pi_{c,1}(\beta_{S1}) + \pi_{c,1}G_i\beta_{SG} + \\ &\eta_{i1}\beta_{S1} + \underbrace{\eta_{i1}G_i}_{\text{interaction}}(\beta_{SG}) + \varepsilon_i \end{aligned} \tag{2.38}$$

Due to the interaction between the first stage reduced form errors and the geographic region indicators in Equation 2.38, the two-stage least squares estimate will still be biased unless we assume the source of heterogeneity, G_i, is a fixed, non-stochastic vector or if we assume G_i is independent from the reduced form errors. This is like saying that any omitted factors that determine the the years of schooling for each person are independent from the place of birth!

An alternative is to instrument for the heterogeneous treatments also, which is why we include first stage reduced form representations of all the interactions between the amount of education with geography. To see this, let's first compactify the notation by representing the treatments by letting $\mathbf{S}_i$ represent all the interacted treatment effects: $\underset{1\times K_G}{\mathbf{S}_i} \equiv [S_i \quad S_iG_i]$. Then we can stack the coefficients on the treatments β_S and the first stage coefficients on the instruments π_S:

$$\mathbf{S}_i = Z_i\pi_S + G_i\pi_G + X_i\pi_X + \pi_c + \eta_i \tag{2.39}$$

$$\begin{aligned} Y_i &= \mathbf{S}_i\beta_S + G_i\beta_G + X_i\beta_X + \beta_c + \varepsilon_i \qquad (2.40) \\ &= (Z_i\pi_S + G_i\pi_G + X_i\pi_X + \pi_c + \eta_i)\beta_S + G_i\beta_G + X_i\beta_X + \beta_c + \varepsilon_i \\ &= Z_i\pi_S\beta_S + G_i(\pi_G\beta_S + \beta_G) + X_i(\pi_X\beta_S + \beta_X) + (\pi_c\beta_S + \beta_c) + \underbrace{\eta_i\beta_S}_{\text{no interaction}} + \varepsilon_i \qquad (2.41) \end{aligned}$$

Equation 2.39 is the *first stage*, Equation 2.40 is the *second stage*, and Equation 2.41 is the *second stage reduced form.* The standard two-stage least squares estimation proceeds by predicting values for schooling, then regressing log earnings on these predicted schooling levels. Note that the first stage estimation consists of estimating a reduced form specification for each heterogenous treatment effect.

Another way to arrive at the same estimate is to regress *both* the outcomes *and* the treatments on the instruments, then "divide" the two coefficients. In matrix notation "dividing" the two coefficients involves inverting the matrix $(\pi'\pi)$. We have $\beta = E((\pi'\pi)^{-1}\pi'(\pi_{Y\text{on}Z}))$. Where $\pi_{Y\text{on}Z}$ is the regression of Y on the right hand side from the *first-stage.* This is why the relevance condition in Assumption 12 is about the invertibility of a matrix.

In the second stage reduced form (Equation 2.41), the first stage errors (η_i) do not interact with the geographic indicators. This is the key reason why we choose to estimate additional first stage parameters – we don't need to assume G_i is non-stochastic, or that G_i and η_i are independent. This also simplifies both the computational and analytical calculation of the conditional posterior distributions because we don't have to deal with heterogeneous reduced form variance-covariance matrices.

By estimating more parameters in the first stage, we drastically raise the threshold of "strength of the first stage" under which we would believe that our final two-stage least squares estimate is distributed normally. Staiger and Stock (1994); Stock and Yogo (2002) display how to check the F-statistic of the first stage for conditions when we can assume the two-stage least squares estimate is asymptotically normal. Sanderson and Windmeijer (2016) carries out the same large-sample-but-worsening-instruments analysis for the case with multiple endogenous variables. As noted by many authors (Nelson and Startz, 1990; Bekker, 1994; Phillips, 2009) the actual exact distribution of the two stage least squares estimate is not normal. In the just-identified case the exact distribution of the two-stage least squares estimate doesn't even have finite moments, which would negate the use of bootstrap procedures to estimate the mean.

Instead, we work with the posterior distribution of the two-stage least squares estimate and obtain estimates of confidence from the Bayesian posterior distribution. We follow homogeneous treatment effect examples (Lopes and Polson, 2014; Wiesenfarth et al., 2014) in choosing conjugate priors that are less informative about estimate. Conjugate priors allow us to easily sample from the posterior distribution, save computational time, and write out simple expressions for the conditional posterior distributions.

2.4.3 Error Structure

Assumption for the Errors/Omitted Variables:

We assume that the errors are normal, and specify priors for the covariances and the coefficients.

Assumption 13. *The errors are multivariate normal with mean zero and variance Σ:*

$$(\underset{1\times 1}{\varepsilon_i}, \underset{1\times(K_G)}{\eta_i}) \sim N(\vec{0}, \Sigma)$$

Assuming that the errors are normal is not as strong as assuming the final two-stage least squares estimate is distributed normally. As an example, recall that the most basic two-stage least squares estimate in the single treatment, single instrument case reduces to the ratio of two estimated coefficients. If the errors in the first and second stage are normal, then the two-stage least squares estimate is the ratio of two correlated normal random variables. Fieller (1932) calculates the exact probability density function of the ratio of two correlated normal random variables. In the just-identified case the two-stage least squares estimate doesn't have finite moments (Nelson and Startz, 1990).

2.4.4 Priors

Prior for the covariance:

We assume the variance Σ is distributed Inverse Wishart (the conjugate prior), with relatively uninformative choices for its parameters:

Prior 2. *Our prior for Σ is Inverse Wishart:*

$$\Sigma \sim IW(n_0 = 0.01, \Sigma_0 = (0.01) * \boldsymbol{I}_{(K_G+1)})$$

Where $\mathbf{I}$ represents th identity matrix, and recall that K_G is the number of geographic regions. Choosing small values for n_0 and Σ_0 are uninformative because if we were to observe N additional data: $\underset{N\times(K_G+1)}{e_N} = (\varepsilon_i, \eta_i)_{i=\{1,\ldots,N\}}$, then the posterior distribution for the variance would be $(\Sigma|e_N) \sim IW(n_0 + N, \Sigma_0 + N\hat{\Sigma}_N)$, where $\hat{\Sigma}_N$ is an estimate of the covariance from the data. Therefore, the small initial values for the shape parameters allow the data have a stronger impact on the posterior distribution of Σ.

We can also represent the variance-covariance matrix for the errors in reduced form:

$$
\begin{aligned}
\Omega &= \mathbf{B}\Sigma\mathbf{B}' \\
&= \begin{pmatrix} 1 & \beta_S' \\ \vec{0} & \mathbf{I}_{K_G} \end{pmatrix} \Sigma \begin{pmatrix} 1 & \vec{0}' \\ \beta_S & \mathbf{I}_{K_G} \end{pmatrix} \\
&= \begin{pmatrix} 1 & \beta_S' \\ \vec{0} & \mathbf{I}_{K_G} \end{pmatrix} \begin{pmatrix} \sigma_{\varepsilon\varepsilon} & \Sigma_{\varepsilon\eta} \\ \Sigma_{\eta\varepsilon} & \Sigma_{\eta\eta} \end{pmatrix} \begin{pmatrix} 1 & \vec{0}' \\ \beta_S & \mathbf{I}_{K_G} \end{pmatrix} \\
&= \begin{pmatrix} \sigma_{\varepsilon\varepsilon} + 2\Sigma_{\varepsilon\eta}\beta_S + \beta_S'\Sigma_{\eta\eta}\beta_S & \Sigma_{\varepsilon\eta} + \beta_S'\Sigma_{\eta\eta} \\ \Sigma_{\eta\varepsilon} + \Sigma_{\eta\eta}\beta_S & \Sigma_{\eta\eta} \end{pmatrix} && (2.42) \\
&= \begin{pmatrix} \omega_{\eta_0\eta_0} & \Omega_{\eta_0\eta} \\ \Omega_{\eta\eta_0} & \Omega_{\eta\eta} \end{pmatrix} && (2.43)
\end{aligned}
$$

Where I have expanded the variance-covariance matrices of the errors into convenient block form, and η_0 is the error from the reduced form regression of Y_i on the right hand side from the first stage. We have a choice of forming our prior for the covariance matrix in terms of Σ or Ω. Chao and Phillips (1998); Dreze (1976) discuss the 1-to-1 correspondence between priors on Σ and priors on Ω. Our current prior puts an identity matrix as the initial shape of the Inverse-Wishart distribution, which corresponds to an initial shape of $\mathbf{BB}'$ on the prior of the reduced form errors.

Prior for the coefficients:

I use a normal prior for the coefficients:

Prior 3. *Our prior for the coefficients is Multivariate Normal:*

$$(\beta, \pi) \sim N(\vec{0}, Diag([\lambda_B, \lambda_P])^{-1}) \qquad (2.44)$$

Where (β, π) are the set of coefficients in the structural model, and $Diag(v)$ corresponds to a matrix with the vector v on the diagonal and zero elsewhere. The main advantage of using a normal prior is because it is the conjugate prior to the specification that the errors are normal.

The normal prior with a diagonal covariance matrix corresponds to using ridge regression with the regularization penalties scaled by λ_B and λ_P. Recall that ridge regression minimizes the squared residuals plus a weighted sum of squares of the coefficients. If we were to simply use ridge regression of Y on X with constant penalties λ, then the analytical solution is: $\beta_{X,\text{ridgereg}} = (X'X + \mathbf{I}\lambda)^{-1}X'Y$, which is the standard analytical OLS solution with the λ added to the "denominator"[3]. Taking the regularization penalties towards zero is the same as increasing the variance of the normal priors to infinity, and this will lead us to the standard OLS analytical solution. Throughout this article, I refer to a "weak" prior as one that has variance set to 1 million, setting the ridge regression regularization penalties to $\frac{1}{(1 \text{ million})}$

Even without a Bayesian interpretation, some amount of regularization is still justified as the dimension of the treatment increases. If we were to attempt to estimate nonparametric returns to schooling, we would need to use regularization even if we did not want to have a Bayesian interpretation of the results because nonparametric instrumental variables are plagued by an "ill-posed inverse" problem. The estimation becomes the issue of regressing earnings on an infinite set of functions of our treatment (schooling). Unfortunately, our instrument doesn't impact all possible functions of the treatment. Newey and Powell (2003) shows that a nonparametric instrumental variables strategy can be consistent when one conducts ridge regression for the outcome on the predicted levels of schooling (the second stage), and reformulates ridge regression as a constrained maximization problem, for which the constraint on the magnitude of the coefficients is assumed to not bind.

2.5 Estimation

In order to obtain estimates of our confidence in the parameters, we sample from the conditional posterior distributions of each parameter (this is Gibbs sampling). Since all of our priors are conjugate distributions, the conditional posteriors take the same shape. First we have the posterior distributions of the variance-covariance matrices:

[3]This is why ridge regression is also called a shrinkage estimator.

2.5.1 Conditional Posterior for the Covariance Structure

$$(\Sigma|\beta, \pi, RH_{1N}, RH_{2N}, Y_N) \sim IW(n_0 + N, \Sigma_0 + N\hat{\Sigma}_N), \tag{2.45}$$

$$(\Omega|\Sigma, \beta, \pi, RH_{1N}, RH_{2N}, Y_N) = \mathbf{B}\Sigma\mathbf{B}' \tag{2.46}$$

$$\underset{N\times(2K_G+K_X)}{RH_{2N}} \equiv [\underset{N\times K_G}{\mathbf{S}_N}, \underset{N\times(K_G-1)}{G_N}, \underset{N\times K_X}{X_N}, 1] \tag{2.47}$$

$$\underset{N\times(K_Z+K_G+K_X)}{RH_{1N}} \equiv [\underset{N\times K_Z}{Z_N}, \underset{N\times(K_G-1)}{G_N}, \underset{N\times K_X}{X_N}, 1] \tag{2.48}$$

$$\hat{\Sigma}_N = [\hat{\varepsilon}_N, \hat{\eta}_N]'[\hat{\varepsilon}_N, \hat{\eta}_N] \tag{2.49}$$

$$\hat{\varepsilon}_N = Y_N - (RH)_{2N}\beta \tag{2.50}$$

$$\hat{\eta}_N = Y_N - (RH)_{1N}\pi \tag{2.51}$$

Where Y_N is the column of log weekly earnings and RH_{1N} and RH_{2N} are stacked rows of right hand side variables in the first and second stages respectively. We calculate Ω at each step of the Gibbs sampling process because this will make subsequent draws from the conditional posteriors of β and π much easier. As noted above, $\hat{\Sigma}_N$ is the estimated covariance matrix where $[\hat{\varepsilon}_N, \hat{\eta}_N]$ is an $N \times 2$ matrix of the two vectors of predicted errors horizontally stacked together.

2.5.2 Conditional Posterior for the Coefficients

Since we have chosen conjugate priors, the posterior distribution for β is also normal:

$$(\beta|\pi, \Sigma, \Omega, RH_{1N}, \widehat{RH}_{2N}, Y_N) \sim N(b_1, B_1) \tag{2.52}$$

$$\widehat{RH}_{2N} \equiv [RH_{1N}\pi, G_N, X_N, 1]$$

Now we need to specify what the parameters b_1, B_1 are. Note that the posterior for β only conditions on predicted schooling because we want use the variation induced in schooling by differences in the quarter-of-birth to identify β. This is the same reasoning as in any instrumental variables estimation process.

The posterior parameters for β are:

$$B_1^{-1} = Diag(\lambda_B) + \widetilde{\widehat{RH}}_{2N}' \widetilde{\widehat{RH}}_{2N} \tag{2.53}$$

$$B_1^{-1} b_1 = \widetilde{\widehat{RH}}_{2N}' \widetilde{Y}_N \tag{2.54}$$

$$\widetilde{\widehat{RH}}_{2N} = \widehat{RH}_{2N} \omega_{\eta_0\eta_0|\eta}^{-0.5} \tag{2.55}$$

$$\widetilde{Y}_N = (Y_N - \underbrace{(\mathbf{S}_N - RH_{1N}\pi)}_{\hat{\eta}_N} \Omega_{\eta\eta}^{-1} \Omega_{\eta\eta_0}) \omega_{\eta_0\eta_0|\eta}^{-0.5} \tag{2.56}$$

$$\omega_{\eta_0\eta_0|\eta} = (\omega_{\eta_0\eta_0} - \Omega_{\eta_0\eta} \Omega_{\eta\eta}^{-1} \Omega_{\eta\eta_0}) \tag{2.57}$$

The term $(\mathbf{S}_N - RH_{1N}\pi)$ our estimate of the reduced form errors.

Sampling for π looks quite similar to sampling from β, however we should should break π into the coefficients for their individual first stage equations. Recall from Equations (2.31,2.32) that we have one first stage equation for each interaction between the schooling and the geographic indicator because we want to avoid having our first stage reduced form errors interacting with the source of heterogeneity. Then we have K_G first stage equations, one for each geographic region, and π is a matrix of coefficients with dimension $(KZ + KG + KX) \times KG$. Let π_k represent a single column of π, then $\pi = [\pi_1, \pi_2, ..., \pi_{K_G}]$. Let $\pi_{(-k)}$ represent all the columns of π except for π_k. We can sample from the conditional posterior of π_k conditioning on $\pi_{(-k)}, \beta, \Omega, Y, RH1_N, R\hat{H}2_N$:

$$(\pi_k | \pi_{(-k)}, \beta, \Sigma, \Omega, RH_{1N}, \widehat{RH}_{2N}, Y_N) \sim N(b_1, B_1) \tag{2.58}$$

Since $\underset{(KZ+KG+KX)\times KG}{\pi}$ is a matrix of parameters, it will be easier to sample from the posterior distributions of each individual treatment. Let $\underset{(KZ+KG+KX)\times 1}{\pi_k}$ represent the kth column of π, and let π_{-k} represent the other columns. For example, π_1 represents the first stage coefficients that predict the first element of $\mathbf{S}_i$ from $RH1$.

Now it is easier to express the conditional posterior distribution of the one dimensional vector π_k.

$$(\pi_k|\pi_{-k}, \beta, \Sigma, \Omega, RH_{1N}, \widehat{RH}_{2N}, Y_N) \sim N(p_k, P_k) \tag{2.59}$$

We can express the posterior distribution of π are

$$P_k^{-1} = Diag(\lambda_{Pk}) + \widetilde{RH}'_{1kN}\widetilde{RH}_{1kN} \tag{2.60}$$

$$P_k^{-1}p_k = \widetilde{RH}'_{1kN}\widetilde{S}_{kN} \tag{2.61}$$

$$\widetilde{RH}_{1kN} = RH_{1N}\omega^{-0.5}_{\eta_k\eta_k|\eta_{(-k)}} \tag{2.62}$$

$$\widetilde{S}_{kN} = (S_k - \underbrace{([Y_N, \mathbf{S}_{(-k)}] - [\widehat{RH}_{2N}\beta, RH_{1N}\pi_{(-k)}])}_{[\hat{\eta}_0, \hat{\eta}_{-k}]}\Omega^{-1}_{\eta_{(-k)}\eta_{(-k)}}\Omega_{\eta_{(-k)}\eta_k})\omega^{-0.5}_{\eta_k\eta_k|\eta_{(-k)}} \tag{2.63}$$

$$\omega_{\eta_k\eta_k|\eta_{-k}} = \omega_{\eta_k\eta_k} - \Omega_{\eta_k\eta_{(-k)}}\Omega^{-1}_{\eta_{(-k)}\eta_{(-k)}}\Omega_{\eta_{(-k)}\eta_k} \tag{2.64}$$

Where $\mathbf{S}_{(-k)}$ are the treatments, leaving out the kth treatment, and $\omega_{\eta_k\eta_k}$ refers to the variance of the error in the kth first stage estimating equation. Since all the errors come from the same set of simultaneous equations, it makes sense that the posteriors take the exact same form. Once again, the term $([Y_N, \mathbf{S}_{(-k)}] - [\widehat{RH}_{2N}\beta, RH_{1N}\pi_{(-k)}])$ represents our estimates of the error from the reduced form second stage, and the errors from the first stages excluding the kth error.

The Gibbs sampling procedure using the Julia programming language. I drew $10,000$ samples from the Markov Chain Monte Carlo procedure, using the 2SLS estimates as the initial values and using $2,000$ initial burn-in simulations to allow the draws to stabilize. Although MCMC samples produces autocorrelated, we do not perform any thinning (only taking every nth observation), as measures of percentiles, medians, means, and variances are more precise without thinning (MacEachern and Berliner, 1994; Link and Eaton, 2012).

When we use relatively uninformative priors, we should obtain the same point estimates as the standard 2SLS approach. However, there is a literature on allowing the data to inform us about about the priors (Zellner et al., 2014; Zellner, 1978). Following recent methods Hartford et al. (2017), we use a k-fold cross validation determine the optimal variances in the normal prior (or the optimal regularization parameters.) K-fold cross validation consists of splitting the data into a training and validation set, then training the model on the training set and evaluates the mean squared error of the predicted outcomes on the validation set. I use Julia to discover the regularization penalties that minimize the out-of-sample mean-squared error. Recall that ridge regression regularization penalties correspond to variances for a normal prior.

2.6 Results

Table 2.4 summarizes the results of our exercise. We discuss each column from left to right. The first two columns are point estimates and 95% confidence intervals obtained by using standard techniques. All results are "significant". The OLS results indicate that there is a positive empirical relationship between education and earnings, although this relationship cannot be interpreted as causal. The positve point estimates for the IV estimates indicates that there might be a positive causal relationship between education and earnings. A standard interpretation of the 95% confidence interval as derived from the asymptotic approximation of the distribution of the IV-2SLS estimate would lead us to believe that the returns to education are "significantly" positive for all Census Regions.

The third column is the mean and 95% credible interval are gathered by Gibbs sampling 10,000 times from the posterior of the instrumental variables estimate. We use a "weak" prior that assumes the coefficients have variance 1 million (the ridge regression penalty is 1/(1 million)). The point estimates are the mean of the Gibbs samples, while the 95% credible intervals have endpoints determined by the 2.5 and 97.5 percentiles of the samples. Note that the point estimates for the case with weak priors are quite similar to the "standard" IV point estimates, but the 95% credible intervals are so much wider that only one out of nine intervals doesn't include zero. The wider intervals lead us to conclude that it is inappropriate to assume that the "standard" 2SLS estimate is normally distributed with the variance as derived from the delta method.

The final set of estimates is also has a mean and 95% credible interval obtained by Gibbs sampling 10,000 times from the posterior of the instrumental variables estimate. We use K-Fold cross validation to determine that a prior variance of 4.45 for the second stage and upwards of 1 million for the first stage performed the best at out-of-sample prediction (minimizing mean squared error using a two-stage approach).

The rightmost values display the percent of the posterior that lies above zero. This is a measure of our confidence that the returns to education are positive. We conclude that our identification strategy leads us to be 95% confident that the returns to education are postive for the following four census regions: East North Central, East South Central, South Atlantic, West South Central. Figure 2.4a shows a map of the regions and corresponding proportion of the posterior that lies above zero, representing the posterior chance that the returns to education are positive. Our identification strategy leads us to be 95% confident that the returns to education are positive for the following four census regions: East North Central, East South Central, South Atlantic, West South Central.

Table 2.4: Results

Census Region	Standard Methods: OLS Coefficient[1]	Standard Methods: IV Coefficient[1]	Bayesian Instrumental Variables: Posterior (Weak Prior)[2]	Bayesian Instrumental Variables: Posterior (Best Prior)[3][4]	
East North Central	0.059 (0.057,0.0611)	0.0259 (0.0239,0.028)	0.0262 (-0.1072,0.1583)	0.0715 (-0.0114,0.1562)	95.6%
East South Central	0.0606 (0.0581,0.0631)	0.0454 (0.0429,0.0479)	0.0452 (-0.0009,0.0898)	0.0453 (0.0013,0.0889)	97.8%
Middle Atlantic	0.0715 (0.0696,0.0734)	0.0364 (0.0344,0.0384)	0.0348 (-0.1656,0.2326)	0.0403 (-0.1188,0.2001)	69.5%
Mountain	0.0596 (0.0553,0.0639)	0.0245 (0.02,0.0289)	0.0235 (-0.1377,0.1862)	0.0288 (-0.1124,0.1704)	65.8%
New England	0.0687 (0.0654,0.072)	0.0532 (0.0498,0.0565)	0.0499 (-0.0754,0.173)	0.0504 (-0.0626,0.1645)	80.7%
Pacific	0.0542 (0.0504,0.0581)	0.0806 (0.0766,0.0847)	0.0736 (-0.1044,0.2516)	0.0701 (-0.0805,0.2209)	81.8%
South Atlantic	0.0653 (0.0633,0.0673)	0.0578 (0.0558,0.0598)	0.0549 (-0.0123,0.1253)	0.0558 (-0.0092,0.1227)	95.5%
West North Central	0.0627 (0.06,0.0653)	0.1483 (0.1449,0.1518)	0.1243 (-0.0519,0.3062)	0.1079 (-0.0419,0.2596)	92.1%
West South Central	0.0597 (0.0575,0.062)	0.0812 (0.0789,0.0835)	0.0798 (0.0366,0.1225)	0.08 (0.0389,0.1214)	99.99%

Notes:

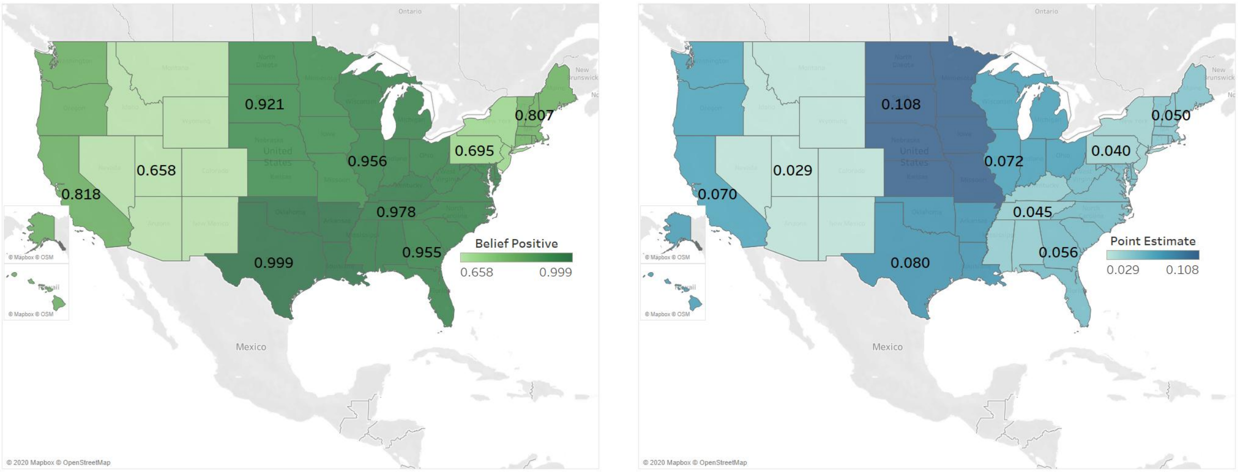
[1] These are point estimates and 95% confidence intervals obtained by using standard techniques. All results are "significant". The OLS results indicate that there is a positive empirical relationship between education and earnings, although this relationship cannot be interpreted as causal. The positve point estimates for the IV estimates indicates that there might be a positive causal relationship between education and earnings. A standard interpretation of the 95% confidence interval as derived from the asymptotic approximation of the dsitribution of the IV-2SLS estimate would lead us to believe that the returns to education are "significantly" positive for all Census Regions.

[2] This mean and 95% credible interval are gathered by Gibbs sampling 10,000 times from the posterior of the instrumental variables estimate. We use a "weak" prior that assumes the coefficients have variance 1 million (the ridge regression penalty is 1/(1 million)). The point estimates are the mean of the Gibbs samples, while the 95% credible intervals have endpoints determined by the 2.5 and 97.5 percentiles of the samples. Note that the point estimates for the case with weak priors are quite similar to the "standard" IV point estimates, but the 95% credible intervals are so much wider that only one out of nine intervals doesn't include zero. The wider intervals lead us to conclude that it is inappropriate to assume that the "standard" 2SLS estimate is normally distributed with the variance as derived from the delta method.

[3] As in [2], this mean and 95% credible interval are gathered by Gibbs sampling 10,000 times from the posterior of the instrumental variables estimate. We use K-Fold cross validation to determine that a prior variance of 4.45 for the second stage and upwards of 1 million for the first stage performed the best at out-of-sample prediction (minimizing mean squared error using a two-stage approach).

[4] The values in the last column display the percent of the posterior that lies above zero. This is a measure of our confidence that the returns to education are positive. We conclude that our identification strategy leads us to be 95% confident that the returns to education are postive for the following four census regions: East North Central, East South Central, South Atlantic, West South Central. Note that our largest point estimate (West North Central) is not 95% credible!

Figure 2.4: The Heterogenous Returns to Education



(a) Percent of Posterior Distribution Positive

(b) Mean of Posterior Distribution

These two maps show summary measures of the posterior distribution of the causal returns to education. We use K-Fold cross validation to determine that a prior variance of 4.5 for the second stage and a large prior variance upwards of 1000 for the first stage performed the best at out-of-sample prediction (using a two-stage approach) for log weekly earnings.

Figure 2.4a shows the proportion of the posterior that lies above zero, representing the posterior chance that the returns to education are positive. Our identification strategy leads us to be 95% confident that the returns to education are positive for the following four census regions: East North Central, East South Central, South Atlantic, West South Central.

Figure 2.4b shows the mean of the posterior distribution. Note that a larger mean does not correspond to more confidence in positive returns to education. Although West North Central had the largest point estimate, we aren't even 95% sure that the returns to education are positive for West North Central.

2.7 Conclusion

This article estimates the heterogeneity in the returns to education. We follow Angrist and Krueger (1991) in using the quarter-of-birth as an instrument for the level of schooling obtained by men born between 1930 and 1940. The quarter-of-birth is a "weak instrument" in the sense that we cannot assume the IV-2SLS estimate is distributed normal. This assumption of normality is an appeal to the asymptotic distribution of the two-stage least squares estimate, and calculations of the "standard errors" that are displayed in common analytical programs like Stata report values calculated from applying the Delta-method.

Instead of using the asymptotic distribution, we sample from the posterior distribution of the two-stage least squares estimate. When using a weak prior, our point estimates agree with the two-stage least squares instrumental variables estimates. However our "95%" confidence intervals are much wider, leading us to reject positive returns to schooling for certain regions, while the standard approach would lead to strong, "significant" results for all the regions. We also use cross-validation to determine the priors on the coefficients (same as ridge-regression regularization penalties) that perform the best for out-of-sample prediction using a two-stage least squares approach. The cross validation tells us to use an weak prior for the first stage, but use a stronger prior in the second stage. This empirical result agrees with the observation in Newey and Powell (2003) that some regularization is needed in the second stage to solve the ill-posed inverse problem.

Our specific application is complicated by the necessity of estimating heterogeneous treatment effects. In this article, we interact our proposed treatment (education level) with geographic characteristics. We show that the currently accepted method of only have one first stage equation - which instruments for education - can lead to biased results if the source of heterogeneity is correlated with the omitted factors in the first stage specification. In our case, instrumental variables estimate would still be biased if the region of birth is correlated with any omitted factor that determined the level of schooling. Since this assumption is too strong, instead we instrument for all the heterogeneous treatments, leading us to have multiple first stage estimating equations. Instead of deriving the actual posterior distribution, we iteratively sample from the conditional posterior distributions.

We have two main findings: for men born between 1930 and 1940 and who were induced to attend an additional year of schooling due to the the interaction between compulsory schooling laws and their quarter of birth, we are 95% sure that the returns to education were positive for these four regions, and the returns to education were probably highest for this region [4].

[4] Please refer to Figure 2.2 for a visual display of the regions, or Table 2.2 for the exact definitions.

2.8 Appendix A: A Diffuse Prior on the Denominator is a Strong Prior on Estimate

The instrumental variables two stage least squares estimate (IV-2SLS) divides two coefficients from the reduced form relationships of the outcome on the instrument and the treatment on the instrument (for a just-identified, one dimensional treatment). Since the two coefficients are calculated using maximum likelihood, the estimate implicitely assumes a "diffuse" prior on the coefficients in the two reduced form relationships. Unfortunately, the word "diffuse" does not mean "objective" or "unbiased." This section shows how only using maximum likelihood corresponds to making a strong prior assumption on an instrumental variables estimate. A "diffuse" prior on the denominator of a fraction corresponds to a "tight" prior on the entire fraction, which represents a strong stance on the value of the IV-2SLS estimate.

Here we compare the two methods for calculating the instrumental variables estimate. Table 2.5 shows that a "diffuse" prior a denominator actually corresponds to a strong prior for the entire fraction. We see from the rows labeled [2], the 95% confidence interval widens greatly when we decrease the variance of the prior on the first stage relationship between the treatment and the instrument. These two estimates keep the priors on β or π_Y large, but now tighten the variance on the denominator (π). The resulting posterior distribution explodes. This drives home the point that a "diffuse" or "weak" prior for the first stage coefficient actually corresponds to a strong prior for the overall fraction.

The rows labeled [1] shows that the IV2SLS maximum likelihood estimate and its asymptotic distribution have the same point estimates and confidence intervals as the posterior distribution of the IV estimate when we have wide variances for our priors. These wide variances correspond to the researchers being relatively unsure of their prior beliefs.

Rows [3] shows estimates that keep a wide variance for the prior on π. From the 95% posterior confidence intervals, we see that a wide prior for the denominator contributes to a tight prior for the entire fraction. The final two rows [4-5] show that emphasize the difference between placing a prior on β and a prior on π_Y. [5] shows that a small variance for the denominator (π) dominates a large variance for the numerator (π_Y). Since a prior for

β is actually a prior for the entire fraction - [4] shows that a small variance for β dominates the small variance for π.

Table 2.5: A Diffuse Prior for a Denominator is a Strong Prior for the Estimate

			Point Estimate	Point Estimate	95% Confidence Interval Lower	95% Confidence Interval Upper
	OLS MLE & Asymptotic Distribution		0.063	0.063	0.062	0.064
[1]	**IV MLE & Asymptotic Distribution**		**0.068**	**0.068**	**0.036**	**0.101**
	Posterior Distribution with Priors on β and π					
	SD for Prior on β	**SD for Prior on π**				
[1]	1000	1000	**0.068**	**0.068**	**0.036**	**0.102**
[2]	1000	0.001	0.648	0.485	-24.882	25.909
[3]	0.001	1000	0.000	0.000	-0.002	0.002
[4]	0.001	0.001	0.000	0.000	-0.002	0.002
	Posterior Distribution with Priors on πY and π					
	SD for Prior on πY	**SD for Prior on π**				
[1]	1000	1000	**0.068**	**0.068**	**0.034**	**0.102**
[2]	1000	0.001	0.260	0.467	-27.747	26.222
[3]	0.001	1000	0.010	0.010	0.004	0.024
[5]	0.001	0.001	-1.496	0.068	-9.969	9.721

[1] The IV2SLS maximum likelihood estimate and its asymptotic distribution have the same point estimates and confidence intervals as the posterior dsitribution of the IV estimate when we have wide variances for our priors. These wide variances correspond to the researchers being relatively unsure of their prior beliefs.

[2] These two estimates keep the priors on β or πY large, but now tighten the variance on the denominator (π). The resulting posterior distribution explodes. This drives home the point that a "diffuse" or "weak" prior for the first stage coefficient actually corresponds to a strong prior for the overall fraction.

[3] These estimates keep a wide variance for the prior on π. From the 95% posterior confidence intervals, we see that a wide prior for the denominator contributes to a tight prior for the entire fraction.

[4-5] The 95% posterior confidence intervals for these two rows emphasize the difference between placing a prior on β and a prior on πY.
[5] shows that a small variance for the denominator (π) dominates a large variance for the numerator (πY). Since a prior for β is actually a prior for the entire fraction - [4] shows that a small variance for β dominates the small variance for π.

CHAPTER 3

Estimates of the Impact of Pasadena's Minimum Wage Ordinance

Edward E. Leamer, Jonathan Gu, Mengshan Cui

3.1 Introduction

California Senate Bill No. 3, which was approved by Governor Brown on April 4, 2016, established a California minimum wage equal to $10.50 per hour for employers with 26 or more employees beginning on January 1, 2017, and stipulated annual increases in the California minimum wage up to $15 per hour on January 1, 2022. Prior to the passage of the California minimum wage, the City of Los Angeles had legislated its own minimum wage schedule with a level of $10.50 on July 1, 2016, six months earlier than the State of California, with increments that increase the LA City minimum wage to $15 on July 1, 2020, a year and a half before the California State minimum wage will reach $15.

The Pasadena Minimum Wage ordinance (Ordinance #7278) passed on March 14, 2016 adopts the City of LA minimum wage schedule through the end of June 2019. This paper studies the impact of minimum wage on important economic variables. Furthermore, we would like to provide policy implications of what are the possible benefits and risks of continuing on LA city's higher minimum wage schedule versus going back to California state minimum wage schedule. We have worked hard to distinguish the effect of the California minimum wage increases from the Pasadena increment since the City of Pasadena cannot call off the future increases in the California minimum wage and thus has discretion over only it's local increment. This is not easy to accomplish because the evidence so far is quite limited.

Figure 3.1: Pasadena Colorado Boulevard

An example of something that might be at stake in this local minimum wage decision is the location of restaurants along Colorado Boulevard illustrated in Figure 3.1. Colorado St/Blvd extends from Glendale through Eagle Rock and into Pasadena, with restaurants on all three segments. The Eagle Rock segment is governed by the higher minimum wage of the City of Los Angeles but Glendale has the lower minimum wage of the State of California. Eagle Rock may have the most at stake here, since if the City of Pasadena opts for the lower minimum wages of the State of California then Eagle Rock would face lower-wage competition both from the East (Pasadena) and from the West (Glendale), and jobs customers could move from Eagle Rock into both Pasadena and Glendale. On the other hand, if Pasadena continues to opt for the high-minimum-wage schedule of the City of Los Angeles, that puts enterprises within Pasadena in an adverse position compared with places like Glendale, La Cańada Flintridge and Alhambra and Monterey Park. The very limited experience with the Pasadena increment so far has not produced evidence of this kind of movement of jobs or enterprises, but the difference between the City and the State minimum wages is going to be larger in the years ahead, with possibly more impact.

The work described in this document is based primarily on the Quarterly Census of Employment and Wages collected by the State of California. We use these data to assess the impact of the California and Pasadena minimum wages on number of establishments, number of employees, and earnings per employee. We also use Pasadena and Los Angeles sales tax revenue to carry out a similar analysis to determine the impact of the minimum wage on sales tax revenue.

Solid conclusions regarding the impact of Pasadena minimum wages from 2011 to 2018 on earnings, employment, and number of establishments are difficult to make because of the limitations of the minimum wage "experiment" that has so far occurred, because the data we rely on only has labor earnings and number of workers but not hours worked, because the data are not individuals but enterprise based, because the geography of temporarily lower minimum wages surrounding Pasadena is complex, because the California minimum wage legislation dictates the precise dates when some workers must receive their wage increases but all other responses to this legislation may be made slowly over time possibly in anticipation of higher minimum wages to come, and because each industry has unobserved drivers that might mask the effects of minimum wage increments.

However, using several different econometric models for interpreting the data from 2011 to 2018, the evidence overall points to a positive impact of California/Pasadena minimum wages on the earnings of restaurant workers and of other low wage industries, confirming that the law is being obeyed. Our preferred model implies that a minimum wage increase of 10% would increase the average quarterly earnings per worker in limited-service restaurants by 8% and in full-service restaurants by 5%[1].

Our model also supports the conclusion that about half of the total increase in earnings resulting from a minimum wage increase occurs within the first quarter of the minimum wage increment. This response is consistent with the legislation which directly and immediately affects only part of each firm's employees but has lingering effects on the others.

While effects on average wages of employed workers are clear in the theory and clear in the data, employment effects are not a sure thing theoretically and are harder to detect in the data. The economists' favorite supply and demand model makes it a virtual certainty that job losses come with minimum wage increases. It is only a matter of when and how much. But

[1]This increase in average earnings does not mean necessarily that the low-wage workers are better off. An increase in earnings per worker might occur if the workers with the lowest earnings were laid off but we have not found evidence of job losses coincident with the earnings increases. It is also possible that the increase in average earnings per employee is a result of a reduction of hours worked by the low-wage employees and/or an increase in hours worked by the high-wage employees. Absent data on hours worked we are not in a position to comment on this possibility.

there are two other theoretical reasons why employment effects could be absent. One theory is that wages are determined not by competitive labor markets but by bilateral bargains between employers with many options and employees with few. For that reason a minimum wage might improve the bargaining power of workers and support higher wages with no loss of employment.The second theoretical reason why there may be small employment effects is that industry-wide increases in costs are normally passed on to customers in the form of higher prices. If these higher prices do not reduce sales, the level of employment required to provide those services also remains the same.

This discussion of the theory of employment effects of the minimum wage foreshadows the fact the evidence about employment effects is not so clear. Our preferred model only shows convincing negative employment effects of a minimum wage increase local to Pasadena for Limited Service Restaurants. Overall the traditional error bands around our estimates of the impact of either the State minimum wage and the Pasadena minimum wage on the 24 industries within our dataset are wide enough to include zero. To express this differently, the employment response to higher minimum wages is neither so sudden nor so great to make it transparent in the data we are studying, though a negative employment response appears present when viewed with the help of some models.

This work is closely related with literature that study the impact of minimum wage in Los Angeles. Both Beacon Economics (Thornberg et al. (2015)) and Reich et al. (2014) from UC Berkeley Institute for Research on Labor and Employment constructed reports for the City of Los Angeles evaluating the impact of future minimum wage increase on the workers, business, and economics in the city. However, these two reports make conflicting conclusions. The Berkeley group uses county level data to derive that increasing minimum wages to be beneficial to the city by increasing worker's earning. This study also finds no significant employment effect due to the minimum wage increase. The report by Beacon Economics argues that minimum wage increases would fail on a cost-benefit basis. This report uses ACS data and estimation result from Meer and West (2016) to conclude that only one in every four dollars of increasing cost goes to low income workers.

Literature that studies minimum wage have conflicting results in general. Meer and West

(2016) finds that minimum wages reduces job growth. Jardim et al. (2018) evaluates the effects of the Seattle minimum wage ordinance. It concludes that a higher minimum wage reduced hours worked in low-wage jobs by 9%, while hourly wages in such jobs increased by 3% . Baskaya and Rubinstein (2012) and Congressional budget Office (CBO (2014)) both studied the impact of federal minimum wage. The former one finds substantial disemployment effect of minimum wages on teenagers, while the latter one finds increase in earnings would not go to low income families. Sabia et al. (2012) estimate the effect of the New York State minimum wage increase. They found the minimum wage increases is associated with 20.2% to 21.8% decrease in the employment of low-skilled workers.

Other literatures such as Dude et al. (2010), Allegretto et al. (2011), Allegretto et al. (2017), Card and Krueger (1994), Dude et al. (2007), Addison et al. (2009), Giuliano (2013), and Hirsch et al. (2015) find no effect of minimum wage on employment. Neumark and Wascher (2011) study the effects of the interactions between the Earned Income Tax Credit and minimum wages on labor market outcomes. Wicks-Lim (2006) documents ripple effects of minimum wage. Brochu and Green (2013) uses Canadian data to study the labor market transition rate.

There are relatively few research papers that examine a minimum increase at a scale so specific to one city. Our research uses data from the individual zipcodes within and around Pasadena to examine the impact of minimum wage. The study that closest matches our detailed geographic examination is the study by Jardim et. al (2017), but this study still uses data that compares the employment across two counties. However, there could be large variance across cities within a county and a wealthier city with higher average wages and home prices may be less susceptible to disemployment effects. There are areas in Pasadena that are filled with wealthy households, and there are also zipcodes within Pasadena that are filled with college students attending Caltech or other institutions. We break the city of Pasadena down into five different zones based on the average level of income. For each zone we create a comparison group from the zipcodes in the surrounding region that most closely mimics the conditions of the region around Pasadena. The impact of minimum wage could be different between these areas. However our data does not offer any significant differences in impact by the different wealth levels within Pasadena.

Furthermore, the impact of minimum wages can be broken down by industry as well. Many recent studies have focused on the fast-food industry, and indeed we do as well, but we also highlight the potential impact of minimum wages on other vulnerable industries as well. For example, our data shows that the opening of new hair, nail, and skin care services have dropped off in Pasadena since the minimum wage increase.

The rest of the paper is organized as follows: Section 3.2 describes the minimum wage schedule in California, Los Angeles County, and Pasadena. We further discuss the limitation of this study cause by the policy design. Section 3.3 contains data source and summary statistics. Section 3.4 discusses main model and specification. Section 3.5 presents findings from the main model. Section 3.6 offers robustness check. Section 3.7 concludes.

3.2 Minimum Wage Schedule

An ideal minimum wage experiment would be a randomized controlled trial in which a group of identical regions is randomly divided into two groups: one group with an increase in the minimum wage and the other with no increase. Then the data on employment, for example, can be summarized in four numbers: the levels of employment in the two groups, both before and after the minimum wage increase. If the communities that experienced the minimum wage increase had a smaller increment in employment than the communities that did not have the minimum wage increase, we would conclude that the minimum wage was suppressing employment. That is what economist call a "difference in differences" estimate.

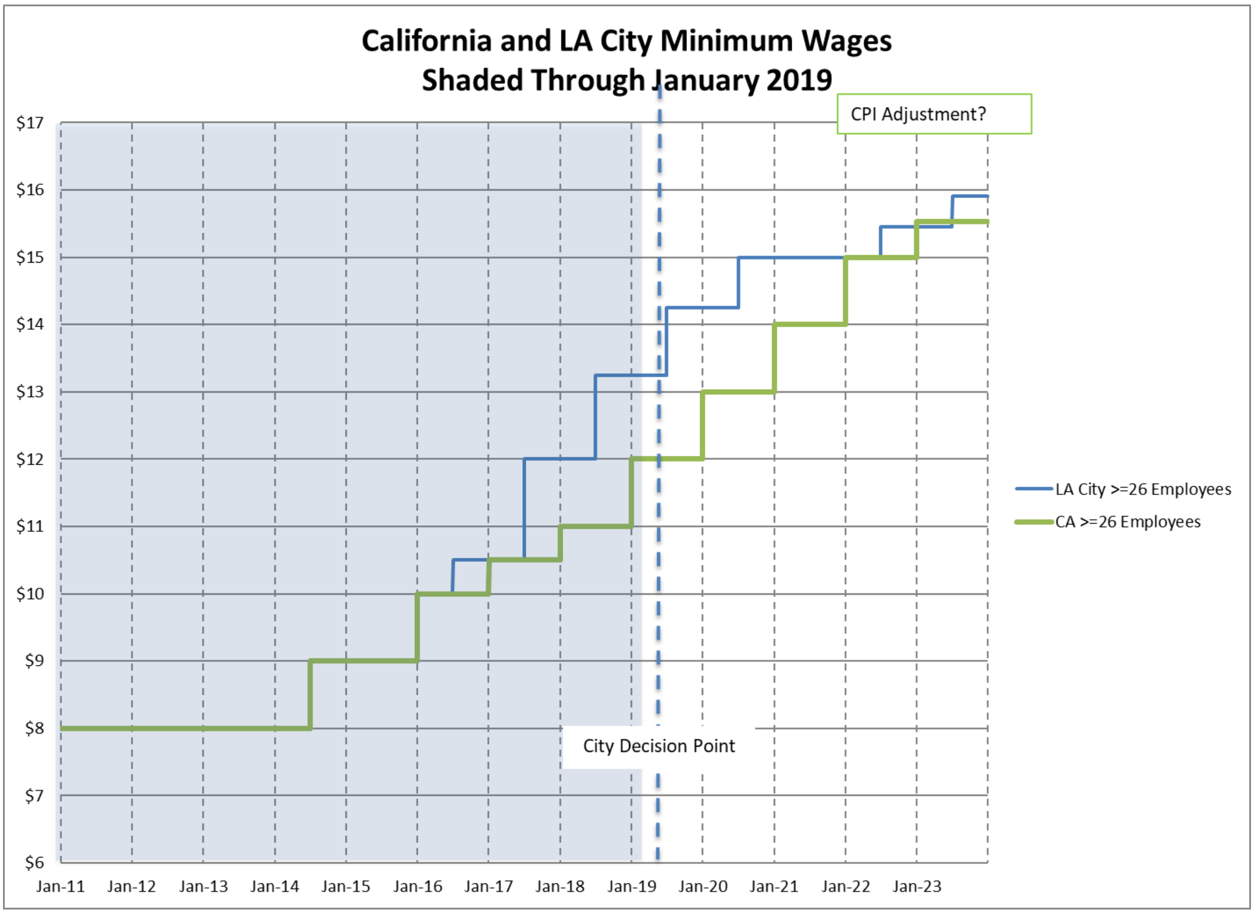
Unfortunately, there are no such experiments. There are no identical regions that have adopted different minimum wages. The level of local minimum wage was never chosen randomly but was determined by a political process that is presumably sensitive to the possibility that a minimum wage set too high can have adverse employment outcomes. If we discover that the sickest people take the most medicine, that is not proving that the medicine has adverse effects. Likewise, if we discover that the communities with the highest minimum wages have the greatest increases in employment, that is not proving that higher minimum wages increase employment. What we are saying is that it's complicated to pull from the data convincing evidence about the effects of the minimum wage. But we have to do the best with what we have, providing appropriate caveats when needed. The first step in that journey is to think clearly about the nature of the experiment we are observing.

We think that the two major problems with the data that we have available are: (1) the whole schedule of minimum wage increases was announced in advance, allowing firms to react in anticipation of minimum wage increments yet to come. (2) the Pasadena minimum wage increment creates a complex local geography of business competition, allowing enterprises to escape the Pasadena increment with a fairly short move to a different jurisdiction. These two issues are now discussed.

3.2.1 The Minimum Wage Increases are determined years in advance

The California and City of LA minimum wage schedules beginning in 2011 (the first year of the Pasadena data that we are studying) are illustrated in the Figure 3.2 which has a shaded region representing the history ending in 2019Q1, the last quarter of our data, and a dashed vertical line indicating the limit of Ordinance #7278, at which point Pasadena will either revert to the California minimum or stick with the LA minimum or something else.

Figure 3.2: California and City of Los Angeles Minimum Wages



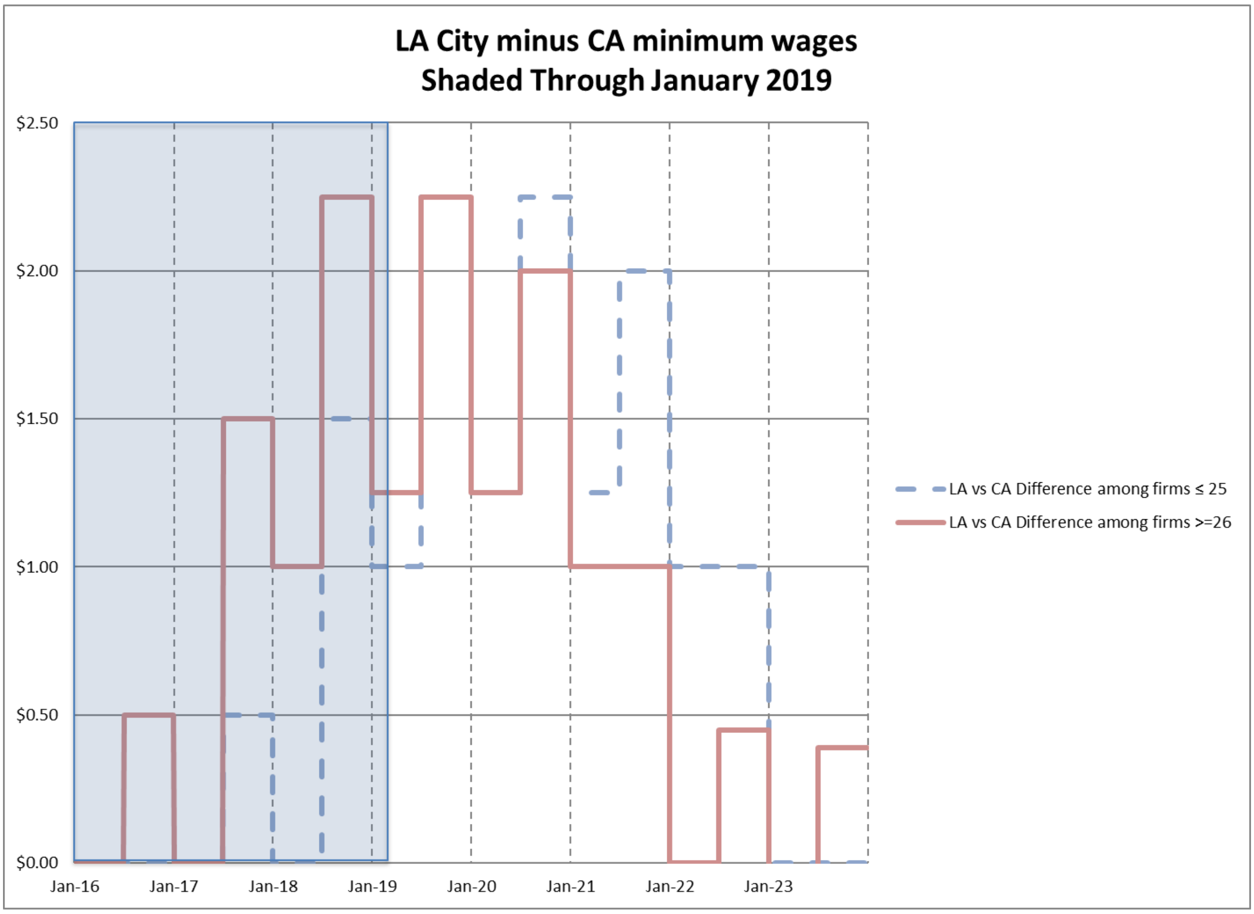
The legislation adopted by the State of California and by the City of Los Angeles firmly established increases in the minimum wage for six or seven years into the future and even indefinitely because of the inflation adjustment that commences in 2022/2023. The best way to summarize this graph in words is that California and Los Angeles/Pasadena have adopted two different but parallel paths toward $15, which means that the impact of the Pasadena ordinance might be only to accelerate by a year or two the impact of the California minimum wage.

But it's more complicated than that. This legislation gives businesses plenty of advance warning and plenty of time to plan how to respond, such as by moving to another location or not opening a new enterprise, by changing the nature of the service provided, by adopting human resources systems that weed out the less productive workers, by automating, by passing the incremental costs on to customers via higher prices or onto building owners via lower rents, or by owners absorbing part of the cost increase. The possible reactions are quite diverse and many are hard or impossible to identify in the data that we have. In particular, it may be difficult to identify an employment effect because any employment reductions that occur can be more a consequence of the whole schedule of minimum wage increments rather than the year-by-year increments. The data analysis that we carry out focuses on the year-by-year increments and only incidentally picks up the effect of the whole schedule. This is quite different from the likely evidence about wage effects since the legislation stipulates exactly when wages have to increment, which is something we should be able to see in the data, and do.

3.2.2 The Local Increment to the California Minimum Wage is Small and Variable

We will be studying the possibility that the Pasadena increment has a different effect than the California minimum wage. Our models will include two variables: (1) the prevailing minimum wage equal to the California minimum wage plus the local increment and (2) the local increment which is the amount by which the Pasadena minimum wage exceeds the California minimum wage. The second variable has a zero coefficient if all that matters is the prevailing minimum wage but a nonzero coefficient if the effect of the local increment is different. For wages we expect the first coefficient to be positive and the second zero, meaning that what matters for setting wages is the prevailing minimum wage not how much of it is dictated by local legislation. For employment, we expect negative coefficients on both, meaning the adverse employment effect is greater for the local components of the minimum wage because it encourages firms to move to close locations with lower minimum wages. In contrast, escaping the California minimum wage requires a move out-of-state. On the other hand, moving from Pasadena to one of the surrounding communities would only delay the minimum wage increment by about a year and a half, and that short delay might not justify the cost of moving. In that case, responses like automation at the Pasadena location might be preferred to moving in pursuit of a temporarily lower minimum wage.

Figure 3.3: City of Los Angeles Increments to the California Minimum Wage



The local increment for the City of Los Angeles is illustrated in Figure 3.3, which distinguishes enterprises with more than 25 employees from smaller enterprises. Here we see a problem for our study: through January 2019 the Pasadena increment was only $0.50 in the second half of 2016 and $1.50 for the second half of 2017 and then $2.25 in the second half of 2018 for firms with 26 or more employees, but much less for firms with 25 or fewer employees. That difference should show up in wages but maybe not so clearly in employment.

3.2.3 Geographical Variability of Minimum Wages

The Pasadena/City of LA increment to minimum wages creates a geographical aspect to the minimum wage experiment by establishing adjacent or close communities with different minimum wages. The local geography is illustrated in the four images in Figure 3.4. Figure

3.4(a) has Pasadena shaded in blue and adjacent or close regions that are subject only to the California minimum wage shaded in light red. (La Canada, Glendale, South Pasadena, Alhambra, San Gabriel, Temple City, San Marino, Arcadia and Sierra Madre.) The lighter regions to the northeast and southwest of Pasadena are Altadena and the City of LA, both with the same minimum wage schedule as Pasadena.

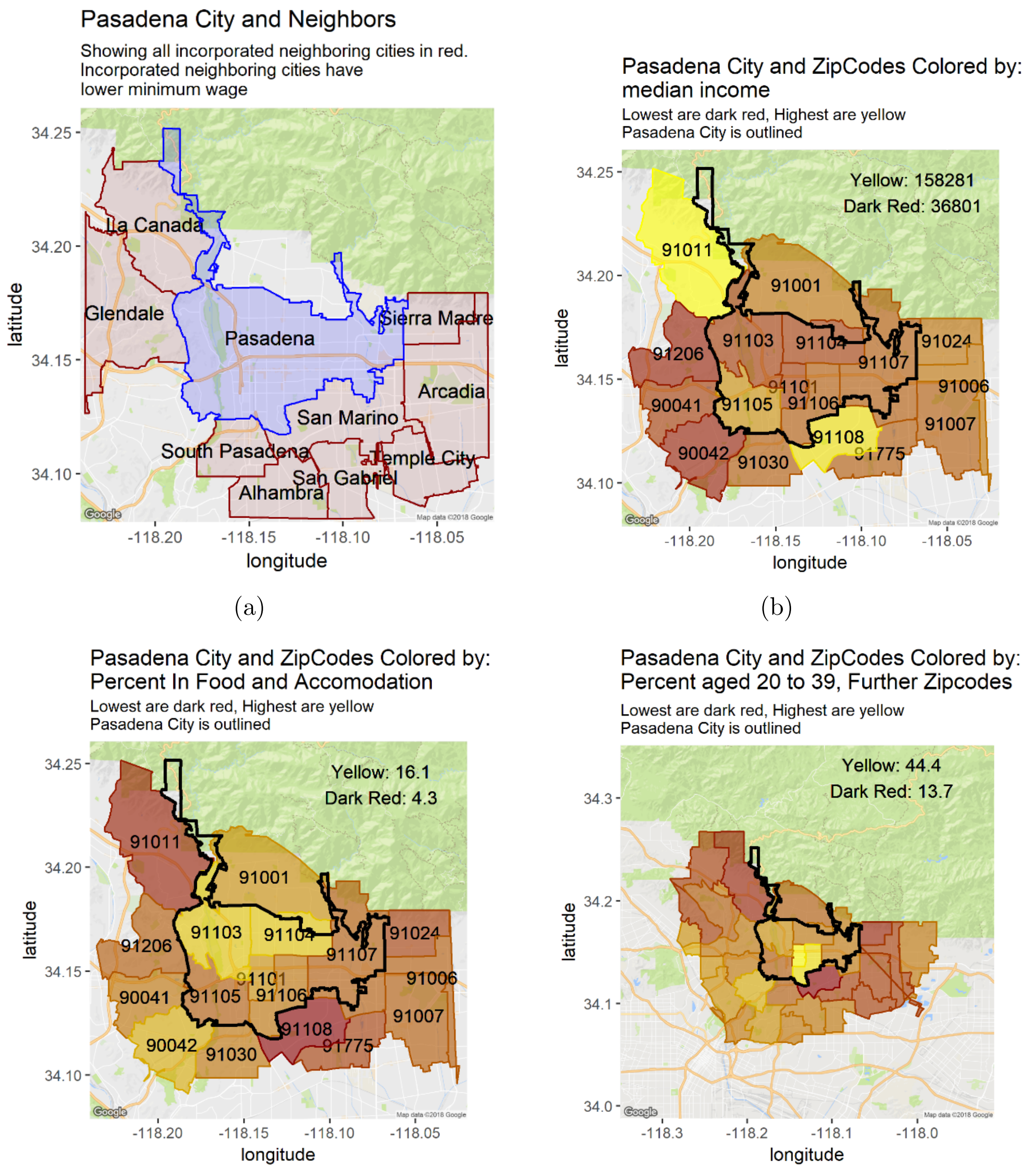
A special risk created by the Pasadena minimum wage is that jobs might leave Pasadena in favor of one of the close cities with a lower minimum wage. That could make the effect of the Pasadena increment on employment greater than the effects of the California increments. It also raises the possibility that we will double-count the employment effects if we use regions close to Pasadena as a control group for Pasadena since we would count the job loss in Pasadena and also the job gain in the neighboring community.

This image captures the difficult question that confronts the Pasadena City Council: should Pasadena align itself with the City of LA and Altadena, which would encourage the movement of jobs to the region shaded red (Glendale, La Canada, South Pasadena, San Marino and so on), or should Pasadena align itself with the red region, thus encouraging a job flow into Pasadena or other red cities out of the City of LA and Altadena.

The three other images in Figure 3.4 help understand what is at risk in this decision. Figure 3.4(b) has the zip codes color-coded by median income of the residents. The highest median incomes are in La Cańada Flintridge and San Marino. Within Pasadena the south-western zipcode 91105 has a high median income but the rest of the zipcodes have lower and comparable income levels. Figure 3.4(c) illustrates the percent of the residents who work in food service and accommodations. It is the northern zipcodes of Pasadena, 91103 and 91104, that have high fractions of residents in this sector. Outside of Pasadena the region with a high fraction of the residents in food services and accommodations is Highland Park (90042).

Another geographic complexity is that Pasadena has neighborhoods that are quite different in terms of income, age, and sectoral job mix. Per the data reported in the Table 3.1, median incomes within Pasadena vary from a low of $61,473 in 91101 to a high of $107,284 in 91105. Among the other differences are: 48.7% of workers in 91101 were young (20-39)

Figure 3.4: Map of Pasadena and surrounding cities



(a)

(b)

(c)

(d)

while 27% were young in 91105; 27% earned less than $25,000 in 91101 but only 11.1% in 91105. It is likely that the younger lower-paid workers from 91101 would be more impacted by the minimum wage than older better paid workers who live in 91105, but our data sets are based on location of work not location of residence.

Table 3.1: Check for Balanced Characteristics of Comparison Groups

Group		1			2	
city	Glendale	Alhambra	Pasadena*	Temple City	Monrovia	Pasadena*
Zipcode	91202	91803	91101	91780	91016	91103
Total Population	23219	29502	20761	35674	41901	28124
Number of Households	8768	9566	10745	11305	14699	8381
Median Income	62104	57380	61473	62461	67868	62697
Age 20-39	33.8%	32.3%	48.7%	29.2%	32.7%	36.4%
High School or less	23.2%	32.7%	13.8%	26.9%	26.5%	28%
Earning less than $25,000	23.6%	21.5%	27.4%	20.2%	17.1%	23.4%
Labor Force Participation	61%	60.7%	68.3%	59.1%	71.1%	63.6%
Unemployment rate	8.7%	5%	7.2%	7%	9.3%	7.2%
Occ—Ind						
Service	14.2%	21.9%	11.9%	16.9%	18.1%	25.1%
Sales	29.1%	26.9%	17.9%	32.3%	24.7%	21.4%
Construction	3.2%	4.8%	4.3%	4.4%	6.1%	8.6%
Retail	12.7%	11.2%	5.9%	11%	10.5%	10.3%
Accommodation and Food	6.9%	12.3%	10.4%	10.9%	10.6%	13.2%
Group			3			
city	Arcadia	Montrose	Pasadena*	Pasadena*		
Zipcode	91007	91020	91106	91104		
Total Population	34619	8448	24875	38725		
Number of Households	11647	3345	10540	13081		
Median Income	75353	70014	75160	70208		
Age 20-39	25.3%	33.6%	44.9%	33%		
High School or less	21.7%	20.8%	12.5%	22.3%		
Earning less than $25,000	17.7%	18.7%	16.9%	21.6%		
Labor Force Participation	58.5%	68%	70.4%	66.2%		
Unemployment rate	7%	7.4%	5.5%	8.5%		
Occ—Ind						
Service	11.9%	11.8%	11.7%	19.7%		
Sales	28%	28.5%	18.4%	20%		

Continued on next page

Table 3.1 – continued from previous page

Construction	3%	3.8%	3.8%	3.9%		
Retail	8.6%	8.5%	8.1%	8.6%		
Accommodation and Food	8.6%	6.3%	10%	13.3%		
Group		4			5	
city	South Pasadena	San Gabriel	Pasadena*	Glendale	Sierra Madre	Pasadena*
Zipcode	91030	91775	91107	91208	91024	91105
Total Population	25905	25389	32027	17180	11067	11728
Number of Households	10150	8164	12502	5876	4403	5485
Median Income	84683	79637	84663	111563	95256	107284
Age 20-39	30.7%	27.2%	31.3%	28.7%	23.1%	27%
High School or less	11.8%	24.6%	15%	14.4%	11.6%	10.8%
Earning less than $25,000	13.2%	15.6%	15.1%	9.6%	10.7%	11.1%
Labor Force Participation	70.8%	61.2%	64.9%	65.3%	66.2%	64.5%
Unemployment rate	6.1%	3.8%	6.6%	4.5%	5.3%	5.9%
Occ—Ind						
Service	10%	13.8%	12.3%	11.3%	5.5%	6.9%
Sales	19.9%	23.1%	23%	25.9%	25.1%	17.8%
Construction	4%	4.6%	4.2%	3.9%	3.4%	5.3%
Retail	6.7%	8.9%	9.5%	8.5%	7.5%	5.6%
Accommodation and Food	9.3%	6.9%	8.3%	9.1%	6.7%	8.7%

*: Above State Minimum Wage

3.3 Data

3.3.1 Data Source

Two main data sets we use are Quarterly Census of Employment and Wages and Sales Tax Revenue.

We rely primarily on data collected by the Quarterly Census of Employment and Wages. Every enterprise in the United States is required to report quarterly the total wages paid in the quarter and the number of employees in each month of the quarter.

The sales tax data has been assembled by HdL Companies and contains quarterly city level data for sales tax revenue for apparel, fast casual dining, casual dining, quick-service dining, and specialty stores. This data set includes the city of Pasadena, Glendale, Monrovia, Burbank, Arcadia, Temple City, Sierra Madre, West Hollywood, Santa Monica, the city of Los Angeles, and Los Angeles County. It covers the period from 2011 quarter 1 to 2018 quarter 1.

Other data sets includes American Community Survey and Current Population Survey.

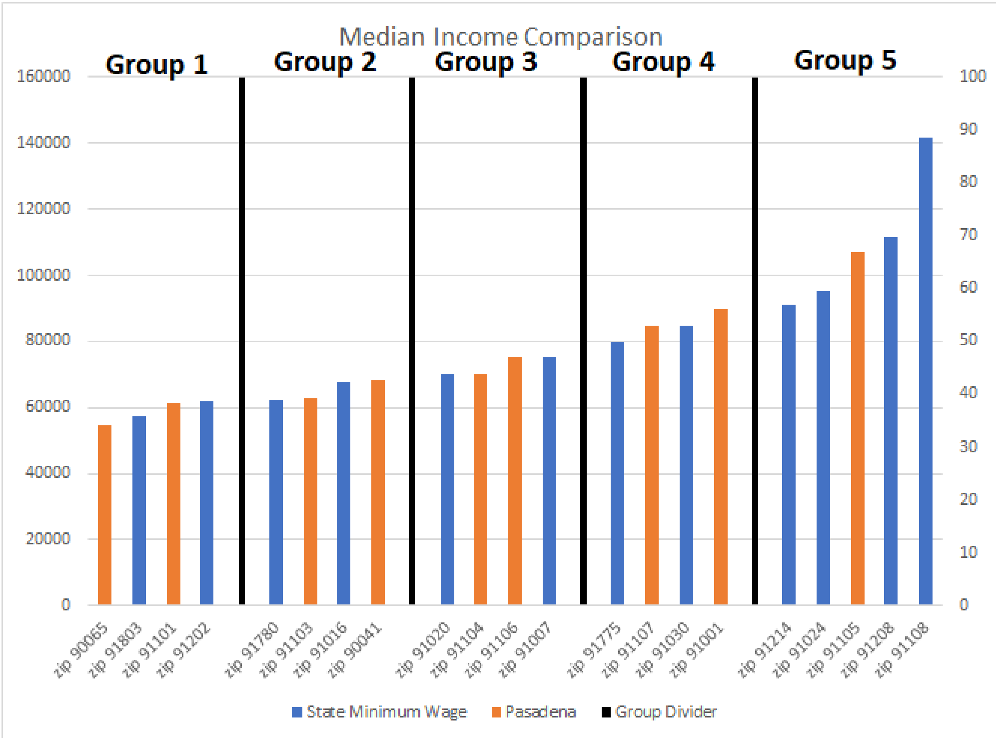
3.3.2 Group

Our strategy for estimating the impact of the Pasadena minimum wage is to compare pairs of regions that are similar to each other but have different minimum wage schedules. We have split Pasadena and its surrounding regions up into ten areas, with five areas consisting of a distinct section of Pasadena and five areas capturing economically similar areas around Pasadena. These areas capture much variation in income within the Pasadena: for example the neighborhood of Pasadena to the southeast near San Marino is quite wealthy, and we would like to compare this wealthy Pasadena neighborhood with another relatively wealthy district nearby that is not impacted by the Pasadena minimum wage ordinance. As another example, the area around Cal tech is populated by many residents between the ages of twenty and thirty, and we would have found two other zipcodes near Pasadena that has the most similar economic and demographic characteristics. Several of the groups also include close zipcodes outside Pasadena with the same minimum wage as the Pasadena zipcodes.

Table 3.2: Zipcodes with similar median incomes

Group	Far Option
G1:	Low MW: Alhambra 91803, Glendale 91202 High MW: Pasadena 91101, LA 90065
G2:	Low MW: Temple City 91780, Monrovia 91016 High MW: Pasadena 91103, LA 90041
G3:	Low MW: Montrose 91020 Arcadia 91007 High MW: Pasadena 91104, 91106
G4:	Low MW: San Gabriel 91775, South Pasadena 91030 High MW: Pasadena 91107, Altadena 91001
G5:	Low MW: Sierra Madre 91024, Glendale 91208, San Marino 91108, La Crescenta 91214 High MW: Pasadena 91105

Figure 3.5: Median Income Comparisons



Our groups are reported in Table 3.2 which begins with Group 1 which has a high minimum wage region composed of Pasadena 91101 and the City of LA 90065, contrasted with the low MW zipcodes in Alhambra and Glendale. Figure 3.5 illustrates the median incomes in each of these zipcodes by groups, which was the basis for our groups. We try to group zipcodes with similar income together. Table 3.1 provides balanced checks of control zipcodes (CA minimum wage) and treatment zipcodes(Pasadena and LA city). For the balance checks, we examine variables that are relevant to the impact of minimum wage. In Group 1, Pasadena 91101 (which surrounds Caltech) has more young people, more educated people, more people earning less than 25,000. In terms of occupation, Pasadena has less Service, Sales, and Retail than their proposed controls in Alhambra and Glendale. In Group 2, Pasadena 91003 has more young people, but also it has more less educated people and people with low earnings. Here we can see the benefits of including more zipcodes. Temple City and Monrovia are large zipcodes with population above the median of our sample. In terms of occupation, Pasadena has more service and less sales. In Group 3, we have a very small proposed control zipcode in Montrose. Montrose is tiny city, with only one zipcode and a population of 8500. Pasadena 91106 has more young people and is more educated than the proposed controls. Pasadena 91104 is actually quite similar to the proposed controls, although it twice the percentage of people working in service occupations. Among the rich counties, all the zipcodes are quite similar in terms of characteristics are a likely to affect the impact of minimum wage.

Figure 3.6: Map of Five Comparison Groups

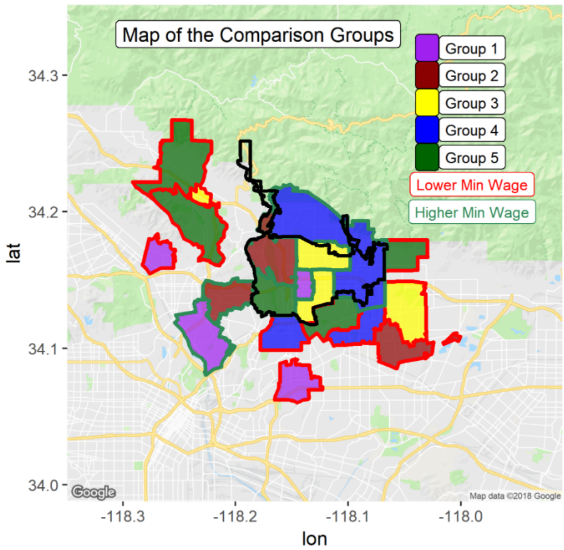


Figure 3.6 is a color coded map of these regions. If we were to do a complete local city comparison, we would simply compare the blue regions with the red regions. Further analysis show that there is strong zipcode level heterogeneity within the cities. We would be better off comparing zipcodes that are similar with each other. Figure 3.4(b) shows the variation in income. We can see the pitfalls of comparing the zipcode 91105 in Pasadena with zipcode 91206 in Glendale. The Pasadena zipcode has a much higher median income. Figure 3.4(c) shows that Pasadena zipcodes 91103 and 91104 have the highest percentages of people working in the food and accommodation occupation. Finally we can see that Pasadena zipcodes 91101 and 91106 have 44% of their population aged between 20 and 40. For reference, classic Old Town Pasadena and Caltech are in zipcode 91101. The administrative buildings and dormitories of Caltech actually have their own zipcode (91126).

Figure 3.7: Group Industry Composition

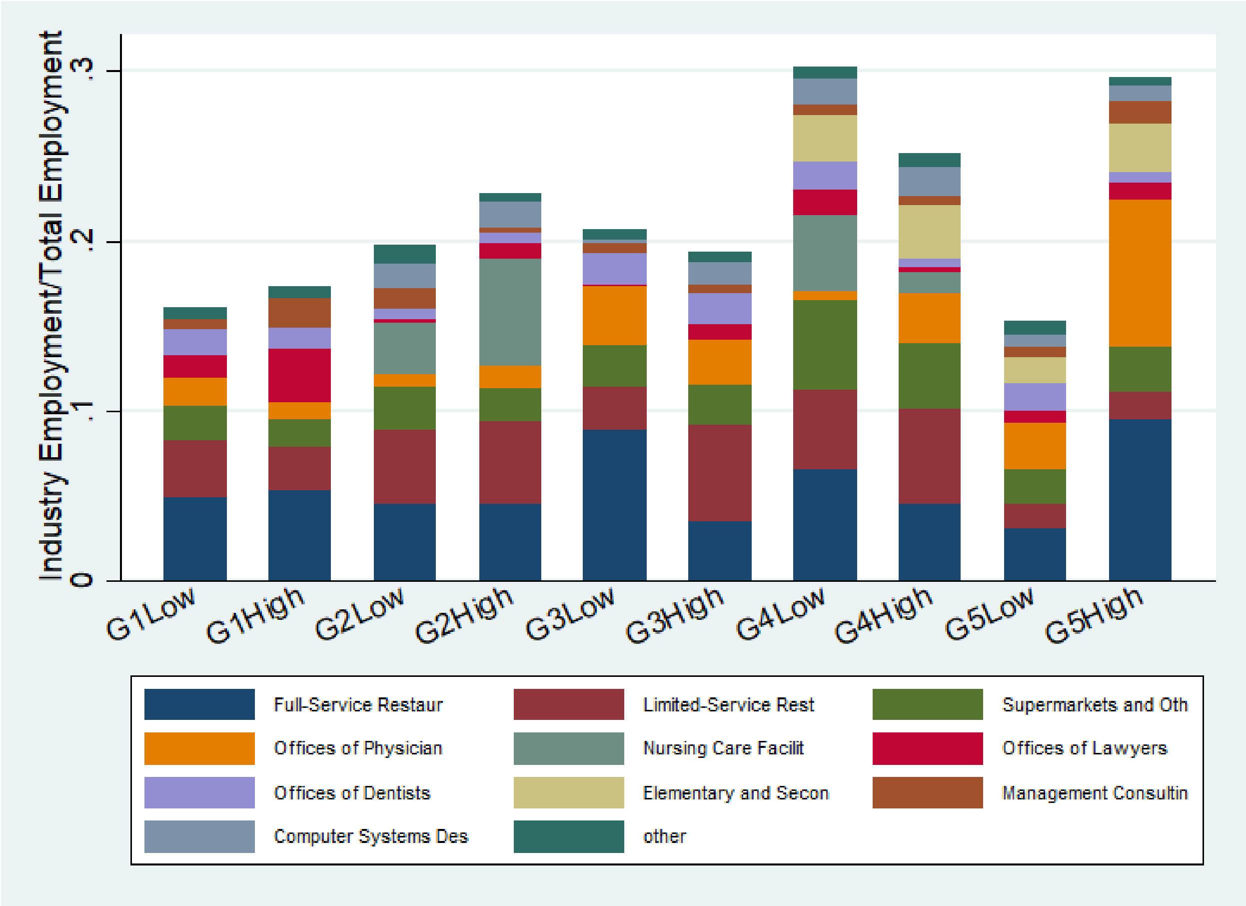


Figure 3.7 presents city (zipcodes within a group with same minimum wage schedule) industry composition difference. We present the top-ten employed industries from QCEW non-confidential data. The "other" means all other industries that are non-confidential. This figure shows that cities within the same group have relatively similar industry compositions. However, each city has very different industry compositions. For example, Group 1 cities have relatively higher ratio of employments in offices of lawyers without any employment in nursing care or computer system. Group 2 cities have high level of employment in nursing care facilities. Group 4 and 5 have higher employment in Elementary and Secondary Schools. This figure emphasizes the importance of including a city-industry fixed effect in regression to better control for industry composition difference.

3.3.3 Summary Statistics

Table 3.3 includes all the industries for which the Quarterly Census of Employment and Wages has data for going back to 2011. The values are average employment numbers during the period 2011 quarter 1 to 2017 quarter 4 for the region with the Pasadena minimum wage composed of Pasadena, Altadena zip code 91001, and LA zip codes 90041, 90065. The sectors are sorted by employment levels and each column shaded with the largest numbers dark and the smallest light.

Table 3.3: Pasadena Industry Detail

Industry	Employment	Firms	Earnings Per Person Per Quarter
Full-Service Restaurants	6361	257	$6,379
Limited-Service Restaurants	4662	235	$5,298
Physician Offices	3139	502	$17,531
Supermarkets and Groceries	2488	42	$7,385
Lawyer Offices	1550	401	$20,082
Elementary and Secondary Schools	1441	24	$13,627
Nursing Facilities	1363	19	$8,364
Computer Systems Design	1215	130	$23,756
Management Consulting Services	1126	187	$22,024
Dentist Offices	1060	173	$11,545
Insurance Agencies and Brokerages	1017	138	$20,631
Accounting, Tax Preparation	759	109	$14,411
Pharmacies and Drug Stores	583	61	$11,442
Child Day Care Services	563	49	$6,858
Residential Building Construction	560	114	$14,925
Hair, Nail, and Skin Care Services	539	105	$5,240
Home Health Care Services	475	16	$9,179
Other Technical Consulting Services	264	138	$16,373
Veterinary Services	256	22	$8,979
Commercial Banking	237	21	$17,635
Fitness and Recreational Sports Centers	212	10	$4,769
Hotels and Motels	188	16	$5,416
Continuing Care Retirement Communities	185	5	$6,073
Janitorial Services	61	11	$5,894

Figure 3.8 illustrates the fractions of minimum wage workers in various industries. At the top are hair and nail salons with 60% of the workers paid less than $12 per hour, and restaurants with 50% of their workers in that category. These are sectors which require

special scrutiny.

Figure 3.8: Prevalence of Minimum Wage Workers

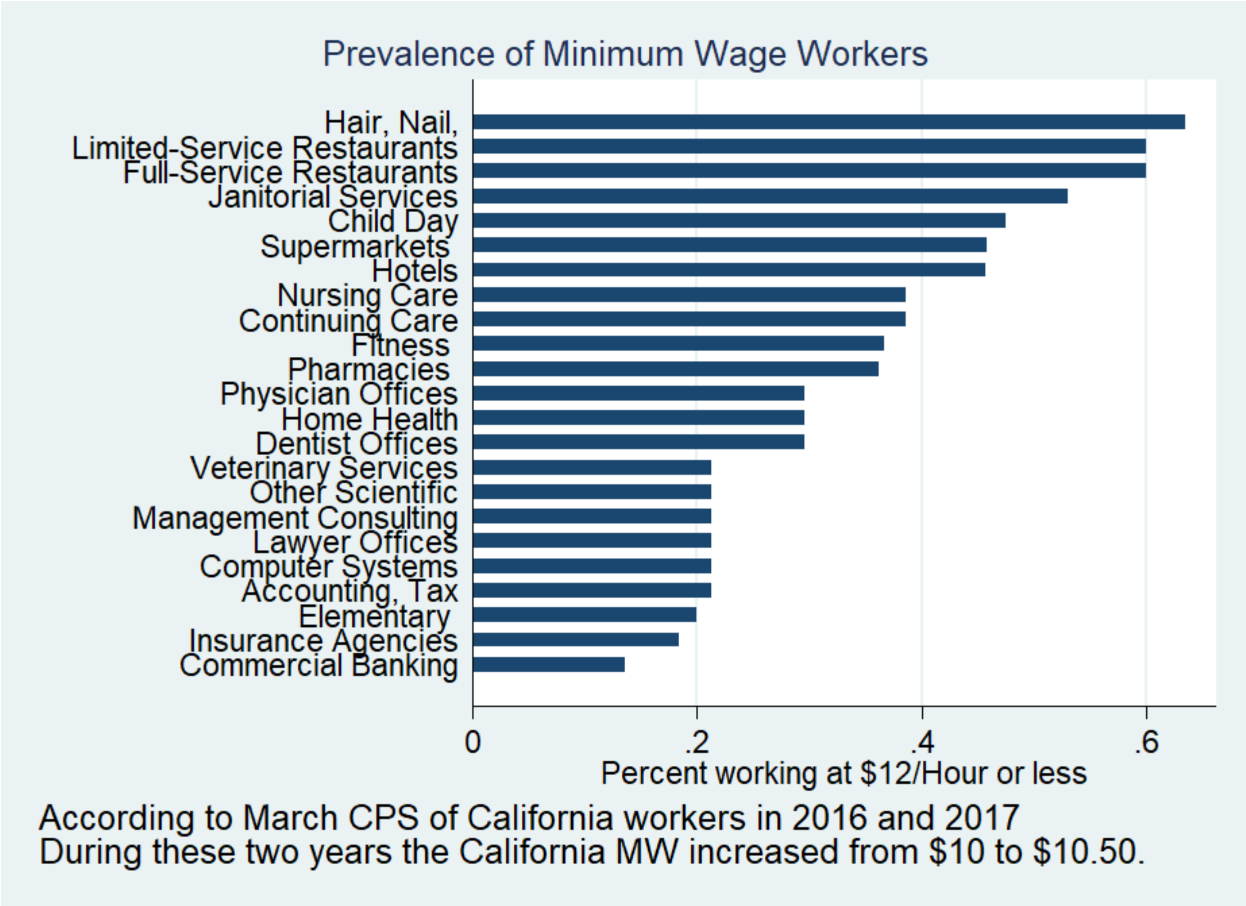


Table 3.4 compares the sales tax revenue in the whole of Los Angeles county with the City of Pasadena in year 2011 and 2017. From this table we can see that the biggest source of sales tax revenue in the county is quick-service dining with around $66 million in sales taxes in 2011 and almost $93 million in 2017. However, in Pasadena, both casual dining and apparel have larger sales than quick-service dining in both 2011 and 2017. However, quick-service dining grew 41.49% in Pasadena from 2011 to 2017, while apparel has almost no growth during this period. The standout industry in terms of growth of revenue in both LA County and Pasadena is fast casual dining. From this table, it does not appear that the increases in minimum wages are reducing tax revenue, but more on this below when this data is filtered through an econometric model.

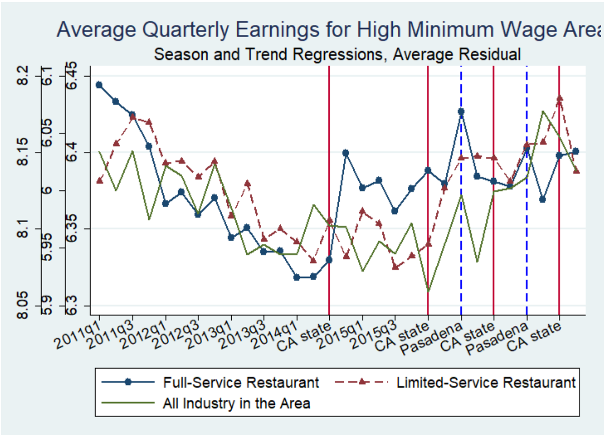
Table 3.4: Industry Annual Sales Tax Revenue Comparasion between LA county and Pasadena

	2011		2017		Growth Rate 2011-2017	
Industry	LA County	Pasadena	LA County	Pasadena	LA County	Pasadena
Quick-Service Dining	$66,455,360	$1,084,072	$92,918,480	$1,533,899	39.82%	41.49%
Apparel	$66,382,680	$1,755,670	$84,285,120	$1,767,651	26.97%	0.68%
Casual Dining	$54,843,800	$2,027,759	$86,950,080	$2,942,134	58.54%	45.09%
Specialty Stores	$32,802,512	$753,472	$40,547,560	$863,900	23.61%	14.66%
Fast Casual Dining	$7,598,740	$285,851	$17,678,712	$712,750	132.65%	149.34%

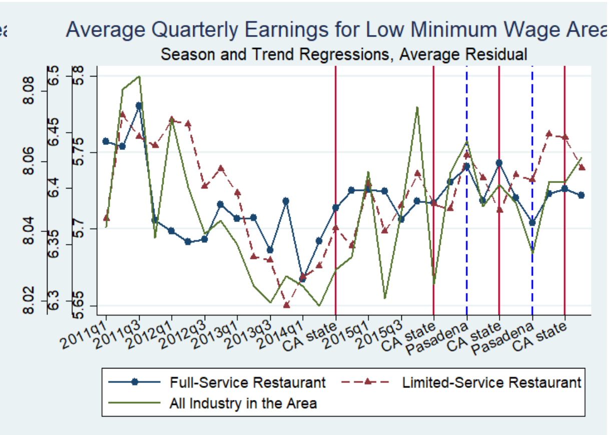
3.3.4 Industry Summary Statistics: Restaurants

In order to study the impact of minimum wage, we will focus on industries that are low skill labor intensive. Restaurant is one of the most important industry with low income workers. We further divide restaurants into limited and full service restaurants. Figure 3.9 present inter-temporal patterns (controlling for seasonal fixed effects and a time trend) of (1) average earnings, (2) employment, and (3) number of establishments. Each figure includes the data for all-industries, and for full-service and limited-service restaurants. Figures are presented for high minimum wage areas (Pasadena, and zipcodes 91001, 90041, and 90065 in Los Angeles, and Altadena) and for low minimum areas. All figures include vertical lines that indicate when either the California or the Pasadena minimum wage was increased. The removal of trends from all these figures supports visual displays that mimic the model-based analysis that also includes trends. These images are different if the trends are not removed, just as our estimates are different if the trend variables are not included. The main take-aways fro these figures are: (1) high minimum areas and low minimum areas have similar patterns for average of all industries, but very different patterns for restaurants. (2) Restaurants react to minimum wage changes very differently than the average of all industries. This indicates that minimum wages have heterogenous impact depending on the industry. The source of the heterogenous response in minimum wages could be due to the prevalence of minimum wage workers in each industry, whether low-wage workers are easy to substitute by technological capital, the average turnover of employees, etc. (3) Full and limited service restaurants react differently to minimum wage changes. This emphasizes the importance of looking at finer detail industry level. The finest detail that we have obtained from the QCEW is at the 5 digit NAICS level. The higher number of digits indicates a finer

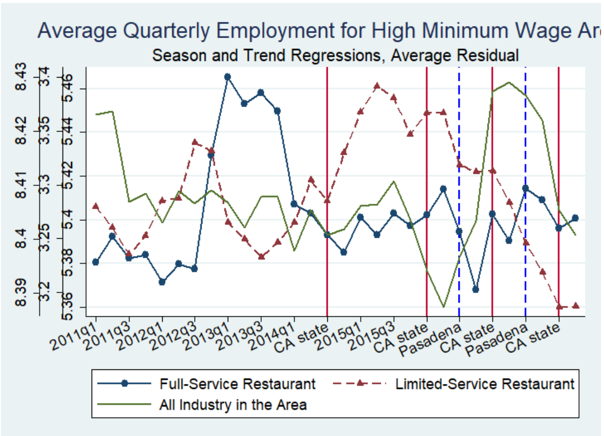
Figure 3.9: Average Quarterly Earnings, Employment, and Establishment for Restaurants



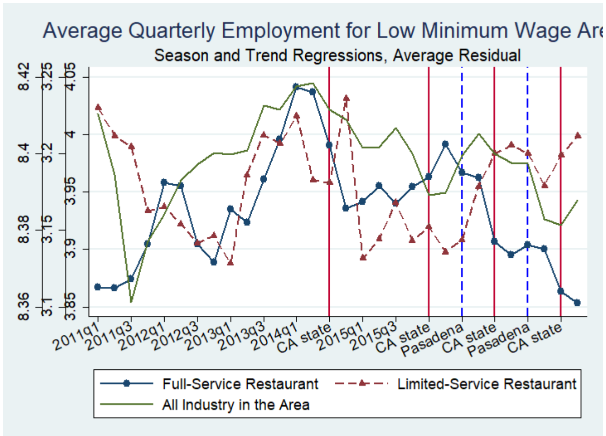
(a)



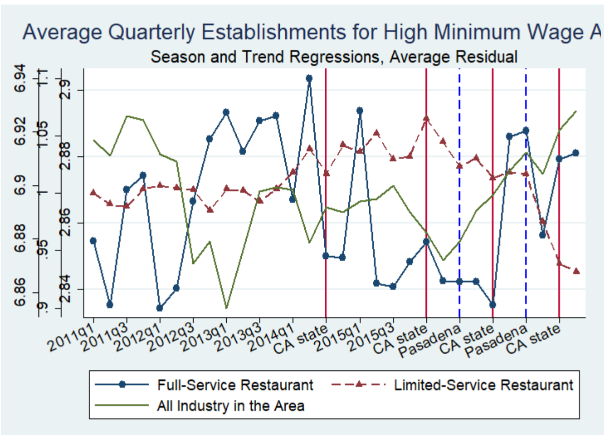
(b)



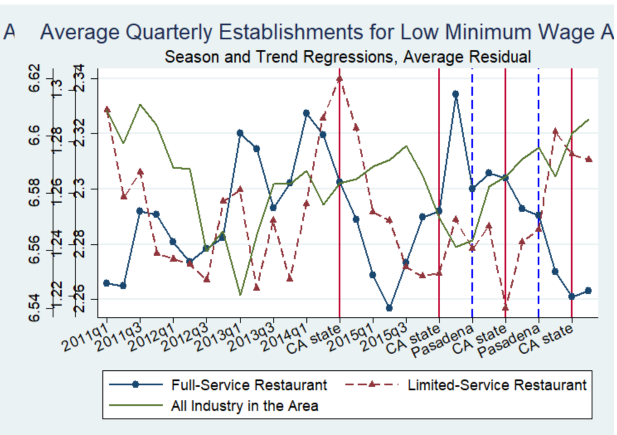
(c)



(d)



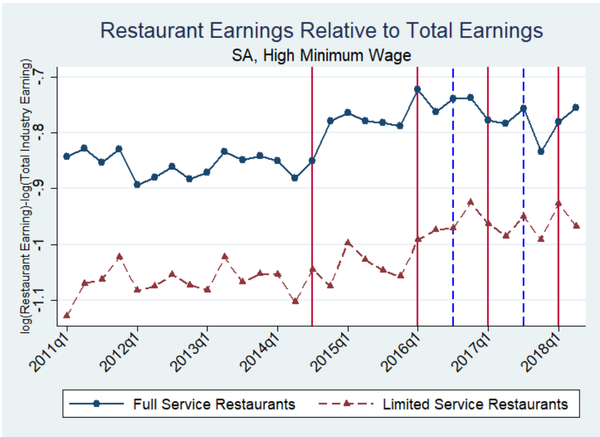
(e)



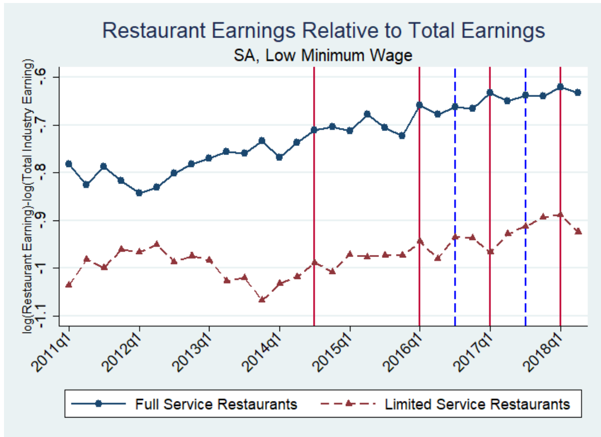
(f)

level of detail of the classification of businesses.

Figure 3.10: Restaurant Earnings, Employment, and Establishments Relative to Total

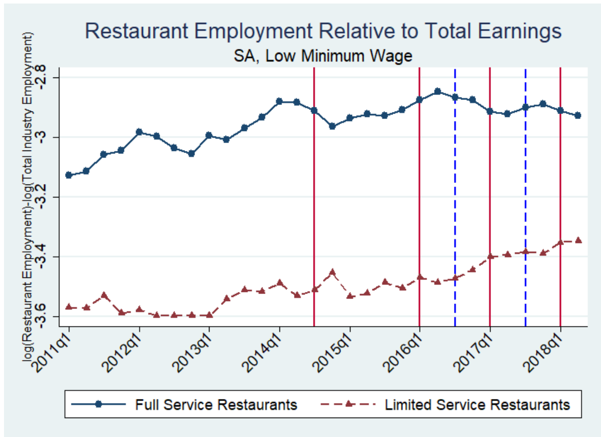


(a)



(b)

(c)



(d)

Figures 3.10 show seasonal adjusted quarterly fixed effect of

$$log(\frac{\text{restaurant earnings}}{\text{average earning of all industries in the area}})$$

and of

$$log(\frac{\text{restaurant employment}}{\text{employment of all industries in the area}}).$$

They are also presented separately for high and low minimum wage areas. A decreasing trend means the restaurant earnings or employment grows slower than the whole economy. An increasing trend means restaurant earnings or employment grows faster than the whole economy. The main messages from these figures are: (1) Restaurants earnings are slightly increased compared to the economy. The increase is more consistent after the first California minimum wage increase in July 2014. (2) Restaurants employment are increasing compared to the economy. This increase pattern seems to be unaffected by minimum wage change. (3) Employment change pattern varies across high and low minimum wage areas and across industry. In high minimum areas, full-service restaurant employment moves very closely with the whole economy, while limited-service restaurant employment increases.

3.4 Models and Specification

The main regression being used in this research is:

$$
\begin{aligned}
log(y_{ict}) = & \beta_{0i} + \delta_i log(\text{City Minimum Wage}_{ct}) + \gamma_i log(\frac{\text{City Minimum Wage}}{\text{CA State Minimum Wage}}_{ct}) \\
& + \beta_{1i} log(y_{ic(t-1)}) + \beta_{2i} log(Y_{ct}) + \beta_{3i} log(Y_{c(t-1)}) + \beta_{4i} Quarter_i + \theta_{si} + \epsilon_{ict} \quad (3.1)
\end{aligned}
$$

where:

- i: industry; c: city (zipcodes within same group with same minimum wage schedule); t: quarter.
- y_{ict} is the dependent variable for industry i in city c in quarter t.
- City Minimum Wage$_{ct}$ is the minimum wage (per hour) that the city c is on in quarter t.
- CA State Minimum Wage$_{ct}$ is the California minimum wage (per hour) in quarter t.
- $y_{ic(t-1)}$ is the dependent variable for industry i in city c in one quarter before.
- Y_{ct} is the total number of a variable in the city c in quarter t : $Y_{ct} = \sum_i Y_{ict}$.
- $Y_{c(t-1)} = \sum_i Y_{ic(t-1)}$.
- $Quarter_i$ represents industry level time trend.
- θ_{si} is being used to control for the industry-seasonal fixed effect.

The above regression is estimated for each industry separately. Our models use three "dependent" variables observed quarterly at the level of an industry in a particular region: Earnings per employee, Employment, and Number of Establishments. To explain the movements in these three dependent variables we have used "dynamic" models that allow the impact of an increment in the minimum wage to be spread over time. We include as explanatory variables two minimum wage variables, the prevailing minimum wage and the part of the prevailing wage that is due to the local legislation. We also include explanatory variables that reflect overall area-wide changes like the total employment and overall average earnings per employee which we take to be unaffected by minimum wages.

Each variable we have included in our models captures the effect one of the key factors mentioned above. Previous literature on minimum wage has mainly used the "two-way fixed effects" approach. Our model deviates from the previous literature in a number of ways, most notably, by taking into account the dynamic nature of our data: we are able to say how much of the impact of minimum wage we expect to occur in the first quarter. This difference is essential when analyzing dynamic data with measurements of the same quantity (such as employment in Supermarkets in Pasadena) over multiple periods. Without taking into account the dynamic nature of the data, some other researchers may assume that the number of employees on the payroll at Ralphs on Monday is completely independent of the number of employees on the payroll on the following Tuesday. In order to account for the correlation between outcomes we have included lagged dependent variables. These lagged dependent variables will also tell us how much of the minimum wage impact is expected to occur in the first quarter of a minimum wage increase. To study the long run impact of minimum wage, we use the fact the in the long run stable state, $y_{ict} = y_{ic(t-1)}$ and rewrite the regression equation into:

$$
\begin{aligned}
log(y_{ict}) =& \frac{\beta_{0i}}{1-\beta_{1i}} + \frac{\delta_i}{1-\beta_{1i}} log(\text{City Minimum Wage}_{ct}) \\
&+ \frac{\gamma_i}{1-\beta_{1i}} log(\left\{\frac{\text{City Minimum Wage}}{\text{CA State Minimum Wage}}\right\}_{ct}) + \frac{\beta_{2i}}{1-\beta_{1i}} log(Y_{ct}) \\
&+ \frac{\beta_{3i}}{1-\beta_{1i}} log(Y_{c(t-1)}) + \frac{\beta_{4i}}{1-\beta_{1i}} Quarter_i + \frac{\theta_{si}}{1-\beta_{1i}} + \epsilon_{ict}
\end{aligned} \tag{3.2}
$$

where $\frac{\delta_i}{1-\beta_{1i}}$ measures the long-run impact of minimum wage increase and $\frac{\gamma_i}{1-\beta_{1i}}$ estimates long-run impact of Pasadena increase its minimum wage above the state level.

In order to account for underlying forces that affect out outcomes separate from the minimum wage we have included a time trend and also the sum total of the outcome variable across all industries. The sum total outcome variable (such as the total number of employees in all industries in Pasadena) is included to reflect the changes in the economy that are local to the city.

The time trend is included to capture factors that may affect the real price of labor in the economy such as the constantly increasing technological progress, increasing availability of capital, or increasing rates of educated eligible workers. Without adding the time trend our results would actually be quite similar, indeed, without adding time trends we do find more results that are individually significant. However without a time trend the minimum wage is the only variable that documents the passage of time in our model, so any underlying force that is changing over the time of our study could be attributed to minimum wage, therefore we add time trends so that our results will indicate the impact of minimum wage above and beyond the time trend. As we can see in the data display of the number of establishments of hair, nail, and skin care services, including a time trend would lead us to expect that without minimum wages, the growth in the number of salons would have continued. This can be seen as both a positive and a negative attribute of the time trend: Positive if it were actually the case the hair, nail, and skin care services is a booming industry that would have continued it growth without minimum wages, and negative if we believe that the timing of number of salon establishments reaching an equilibrium level coincided with the implementation of California state minimum wage.

It is important to note that our analysis does give what we deem to be false positives because the industries that our model and our data report to be impacted by the minimum wage are not low-wage industries. Specifically, we see positive earnings impact of the minimum wage on veterinary services and dentist's offices even though the average employee at a veterinary clinic or a dentist's office makes twice as much as an average restaurant worker. These false positives highlight a caveat of our model: adding a linear time trend and total industry outcome variables into our model does not capture all of the underlying forces that can drive changes in earnings. If a sudden boom in dog ownership and dental hygiene occurred in 2014, then we cannot disentangle the sudden boom with the increasing California state minimum wage in 2014.

A third problem industry we have is the industry known as "Other Technical and Consulting Service" which is an amalgamation of consulting services that have not been classified into a specific industry. This sector is highly paid and ranks among the lowest in the proportion of employees that are working at minimum wage. This sector also happens to experience a nationwide decline in employment near the end of 2013, which precedes the California state minimum wage increase. This decline is likely simply a transfer of jobs from one industry code to another: on the aggregate level, there has actually been no change in the number of consulting jobs over this time, and management and business consulting (which have their own industry code) is on the rise during our data.

3.5 Main Results

3.5.1 Main Specification

For each industry and each dependent variable we have estimated a total of 24 different models. We report in this section the results generated by the one specification that we think yields the most reliable results. This model includes time trends, utilizes the data from all the five groups of regions together, and includes the Pasadena increment to the

Table 3.5: Regression Result for Predicting Impact of Minimum Wage on Earnings per person

Industry	log(MW)	log(Incre*)	$y_{(t-1)}$	$log(Y_t)$	$log(Y_{t-1})$	Quarter	R^2
Accounting, Tax Preparation, Bookkeeping	0.306	0.698	0.242	0.265	0.583	-0.004	0.758
Child Day Care Services	0.156	0.158	0.277	0.072	-0.155	0.002	0.681
Commercial Banking	0.09	0.943	0.527	0.674	-0.389	0	0.779
Computer Systems Design and Related Services	-0.244	0.302	0.46	0.109	0.306	0.005	0.704
Continuing Care Retirement Communities	0.069	0.854	-0.187	-0.103	-0.003	0.005	0.359
Dentist Offices	0.381	-0.891	0.045	-0.022	-0.29	0	0.825
Elementary and Secondary Schools	0.317	-0.558	-0.222	0.023	0.016	0.007	0.8
Fitness and Recreational Sports Centers	-0.949	0.524	-0.181	0.355	1.937	-0.006	0.273
Full-Service Restaurants	0.246	-0.318	0.493	0.053	-0.015	0.003	0.896
Hair, Nail, and Skin Care Services	0.153	0.162	0.488	0.055	0.182	-0.001	0.816
Home Health Care Services	-0.33	0.491	0.435	-0.011	0.365	0.001	0.575
Hotels (except Casino Hotels) and Motel	0.159	0.562	0.676	-0.471	0.798	-0.002	0.636
Insurance Agencies and Brokerages	0.798	-0.061	-0.033	0.183	-0.236	-0.003	0.639
Janitorial Services	0.571	-0.117	-0.195	0.226	0.252	0.003	0.839
Lawyer Offices	0.028	0.19	0.13	0.139	0.109	0.002	0.685
Limited-Service Restaurants	0.461	-0.088	0.433	0.214	-0.096	-0.002	0.846
Management Consulting Services	0.104	0.12	0.565	0.058	-0.098	0.002	0.594
Nursing Care Facilities	-0.191	0.157	-0.02	0.008	-0.294	0.013	0.494
Other Scientific and Technical Consulting	-0.321	0.714	0.422	-0.066	-0.892	0.015	0.594
Pharmacies and Drug Stores	0.236	0.148	0.227	0.035	-0.076	0	0.635
Physician Offices	0.035	0.609	0.332	0.406	0.422	-0.005	0.776
Residential Building Construction	0.086	0.786	0.533	-0.056	0.296	0.003	0.636
Supermarkets and Other Grocery	0.108	0.299	0.422	-0.002	-0.029	0.001	0.74
Veterinary Services	0.46	-0.607	0.489	0.37	-0.49	-0.003	0.856

* The increment is the ratio of Pasadena MW to the California state MW
Green or Red: This result is individually significant

minimum wage.

We will first examine the findings of an increase in minimum wages inclusive of the Pasadena increment. It is important to note that these results may be driven primarily by increases in the California state minimum wage because the California minimum wage rose by $4 from $8 per hour in 2011 to $12 per hour in 2019, while the Pasadena minimum wage has risen above the California minimum wage by 50 cents in the second half of 2016, and by

$1.50 during the second half of 2017 and by $2.25 in the second half of 2018 and the second half of 2019.

We find significant impact of the rising California state minimum wage on earnings per quarter for many industries. We have highlighted four industries because they form a relatively large part of the Pasadena labor force, they have a high proportion of workers working within $2 of the minimum wage, and our model specification suggests that the rise in minimum wages has a positive impact on earnings: full and limited service restaurants,

Table 3.6: Regression Result for Predicting Impact of Minimum Wage on Employment

Industry	log(MW)	log(Incre*)	$y_{(t-1)}$	$log(Y_t)$	$log(Y_{t-1})$	Quarter	R^2
Accounting, Tax Preparation, Bookkeeping	-0.263	0.413	0.803	0.338	-0.109	0.003	0.977
Child Day Care Services	-0.017	-0.055	0.818	1.016	-1.092	0.002	0.948
Commercial Banking	0.079	0.082	0.813	-0.621	0.131	0.003	0.975
Computer Systems Design and Related Services	-0.134	-0.021	0.927	0.464	0.009	0	0.967
Continuing Care Retirement Communities	-0.391	1.071	0.865	-0.85	-0.563	0.014	0.977
Dentist Offices	-0.032	0.124	0.879	0.299	-0.068	0	0.988
Elementary and Secondary Schools	-0.118	0.6	0.893	0.67	-0.808	0.001	0.975
Fitness and Recreational Sports Centers	0.044	-0.095	0.285	-1.004	-2.509	0.051	0.977
Full-Service Restaurants	-0.129	0.157	0.819	0.313	-0.202	0.002	0.989
Hair, Nail, and Skin Care Services	-0.039	-0.126	0.847	0.148	0.224	-0.001	0.955
Home Health Care Services	-0.625	1.089	0.809	0.573	-1.081	0.008	0.978
Hotels (except Casino Hotels) and Motel	-0.01	0.069	0.733	0.763	-0.486	-0.001	0.981
Insurance Agencies and Brokerages	-0.045	0.295	0.827	0.293	-0.315	0.002	0.962
Janitorial Services	-0.686	-0.247	0.681	4.938	-4.16	0.007	0.994
Lawyer Offices	-0.102	-0.051	0.895	-0.024	-0.02	0.002	0.991
Limited-Service Restaurants	-0.005	-0.397	0.753	0.508	-0.483	0.004	0.986
Management Consulting Services	0.445	-0.359	0.764	0.801	-0.71	-0.004	0.921
Nursing Care Facilities	0.062	0.271	0.825	0.247	-0.298	-0.003	0.986
Other Scientific and Technical Consulting	-0.807	1.099	0.85	-0.648	0.377	0.005	0.89
Pharmacies and Drug Stores	-0.136	0.232	0.902	0.321	-0.391	0.002	0.946
Physician Offices	-0.281	0.551	0.724	0.752	-0.672	0.004	0.981
Residential Building Construction	-0.073	0.324	0.842	-0.324	-0.373	0.006	0.94
Supermarkets and Other Grocery	-0.06	0.02	0.76	0.319	-0.228	0.001	0.955
Veterinary Services	-0.306	0.384	0.857	2.31	-1.939	0.005	0.972

* The increment is the ratio of Pasadena MW to the California state MW
Green or Red: This result is individually significant

Table 3.7: Regression Result for Predicting Impact of Minimum Wage on Establishments

Industry	log(MW)	log(Incre*)	$y_{(t-1)}$	$log(Y_t)$	$log(Y_{t-1})$	Quarter	R^2
Accounting, Tax Preparation, Bookkeeping	-0.342	0.449	0.844	-0.112	0.512	0.004	0.991
Child Day Care Services	-0.018	-0.023	0.906	-0.335	0.408	0.001	0.963
Commercial Banking	0.165	-0.461	0.746	1.392	-0.191	-0.006	0.962
Computer Systems Design and Related Services	-0.15	0.503	0.851	0.322	-0.001	0.001	0.968
Continuing Care Retirement Communities	-0.497	1.337	0.934	1.006	0.185	-0.001	0.857
Dentist Offices	-0.11	0.135	0.888	-0.086	0.074	0.002	0.994
Elementary and Secondary Schools	-0.193	0.523	0.881	0.178	-0.205	0.001	0.979
Fitness and Recreational Sports Centers	0.108	0.105	0.812	0.009	0.63	0	0.935
Full-Service Restaurants	-0.061	0.04	0.838	0.089	-0.046	0.001	0.976
Hair, Nail, and Skin Care Services	-0.373	0.23	0.845	-0.586	0.577	0.007	0.974
Home Health Care Services	-0.683	0.322	0.597	-0.289	0.4	0.009	0.967
Hotels (except Casino Hotels) and Motel	-0.23	0.335	0.745	-0.924	1.188	0.004	0.978
Insurance Agencies and Brokerages	0.044	-0.026	0.889	0.17	0.036	-0.001	0.989
Janitorial Services	-0.529	0.142	0.767	-0.049	0.15	0.01	0.957
Lawyer Offices	-0.232	0.232	0.846	0.357	-0.115	0.003	0.997
Limited-Service Restaurants	-0.106	-0.091	0.823	-0.14	0.193	0.003	0.988
Management Consulting Services	-0.649	0.98	0.777	0.831	-0.047	0.01	0.975
Nursing Care Facilities	0.133	-0.109	0.891	0.487	-0.436	-0.004	0.94
Other Scientific and Technical Consulting	-0.609	0.645	0.843	0.152	-0.7	0.007	0.976
Pharmacies and Drug Stores	0.068	0.112	0.786	0.063	-0.015	0.001	0.951
Physician Offices	-0.258	0.302	0.73	0.064	0.106	0.004	0.998
Residential Building Construction	-0.212	0.418	0.809	0.377	0.351	0	0.986
Supermarkets and Other Grocery	0.024	-0.063	0.933	-0.582	0.719	-0.001	0.975
Veterinary Services	-0.159	0.226	0.847	0.096	0.221	0.003	0.985

* The increment is the ratio of Pasadena MW to the California state MW
Green or Red: This result is individually significant

supermarkets, and hair, nail, and skin care services.

Table 3.5,3.6, and 3.7 present regression result of the main specification using earnings per person, employment, and number of firms as dependent variable. Table 3.8 shows long run impact of state minimum wage change and increment change on earnings per person, employment, and number of establishment. We found significant positive impact of minimum wage on earnings for both full and limited service restaurants, janitorial services, dentist office, elementary and secondary schools, and insurance agencies and brokerages. Negative coefficient on increment ratio means that if Pasadena increases its minimum wage above state minimum wage, the impact of minimum wage on earning is going to be less positive than increase the state minimum wage. We found negative coefficient for increment for continuing care retirement, dentist office, and full-service restaurants. There is less significant impact of minimum wage on employment and number of establishments. The coefficient on lag term is also different across three variables. We can see that earnings per person has least significant coefficient on lag terms compared with the coefficient of lag term on employment and number of firms. This is consistent with our expectation since number of employment and firms need longer time to adjust. Therefore, we expect high correlation across time for

Table 3.8: Regression Result for Long Run Impact of Minimum Wage on Earnings per person, Employment, and Number of Establishments

	Earnings per person		Employment		Establishments	
Industry	LR MW	LR Incre	LR MW	LR Incre	LR MW	LR Incre
Accounting, Tax Preparation, Bookkeeping	0.404	-0.921	-1.336	2.099	-2.195	2.882
Child Day Care Services	0.215	-0.219	-0.0957	-0.3	-0.187	-0.251
Commercial Banking	0.191	1.995	0.423	0.441	0.651	-1.817
Computer Systems Design and Related Services	-0.451	-0.56	-1.82	-0.283	-1.012	3.385
Continuing Care Retirement Communities	0.058	-0.719	-2.884	7.907	-7.541	20.28
Dentist Offices	0.398	-0.932	-0.261	1.027	-0.988	1.207
Elementary and Secondary Schools	0.26	-0.457	-1.098	5.605	-1.624	4.414
Fitness and Recreational Sports Centers	-0.803	0.443	0.0609	-0.133	0.576	0.559
Full-Service Restaurants	0.486	-0.627	-0.71	0.867	-0.373	0.248
Hair, Nail, and Skin Care Services	0.299	-0.317	-0.253	-0.822	-2.4	1.482
Home Health Care Services	-0.584	0.87	-3.273	5.709	-1.694	0.799
Hotels (except Casino Hotels) and Motel	0.491	-1.733	-0.0372	0.258	-0.899	1.311
Insurance Agencies and Brokerages	0.773	-0.0587	-0.261	1.701	0.392	-0.234
Janitorial Services	0.477	-0.0982	-2.152	-0.774	-2.269	0.607
Lawyer Offices	0.032	0.218	-0.97	-0.487	-1.509	1.506
Limited-Service Restaurants	0.812	-0.156	-0.0192	-1.606	-0.6	-0.517
Management Consulting Services	0.24	0.275	1.882	-1.52	-2.906	4.386
Nursing Care Facilities	-0.188	-0.154	0.354	1.548	1.221	-1.002
Other Scientific and Technical Consulting	-0.556	1.235	-5.391	7.34	-3.872	4.1
Pharmacies and Drug Stores	0.306	0.192	-1.387	2.361	0.319	0.523
Physician Offices	0.0531	-0.912	-1.016	1.994	-0.955	1.12
Residential Building Construction	0.184	-1.683	-0.462	2.049	-1.109	2.19
Supermarkets and Other Grocery	0.187	-0.518	-0.248	0.0835	0.36	-0.932
Veterinary Services	0.901	-1.188	-2.144	2.691	-1.042	1.479

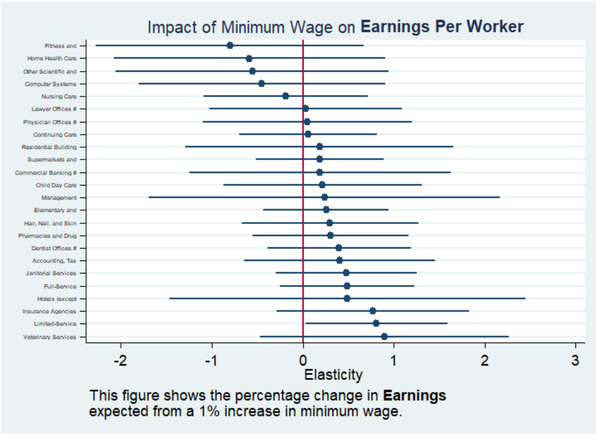
* The increment is the ratio of Pasadena MW to the California state MW
Green or Red: This result is individually significant

these two variables, but not for earnings, since it is easier to adjust.

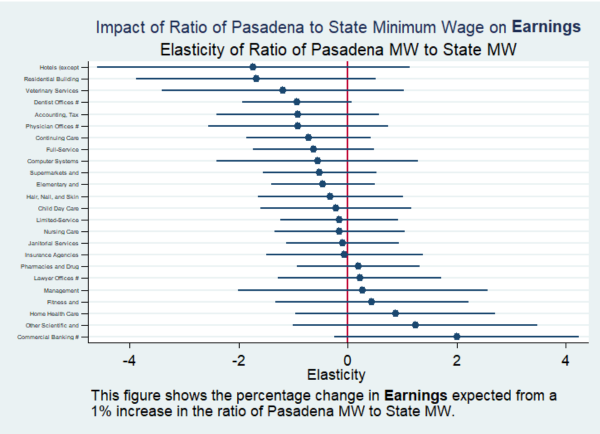
We also display the estimated effects of both the prevailing minimum wage inclusive of any local increment and also effect of the local increment on earnings per worker, number of workers and number of establishments for each industry. In Figure 3.11, each estimate is surrounded with corresponding 95% confidence interval. These estimates are based on the preferred model described in detail above. The estimated impacts below include the estimated impact of minimum wages on all the industries for which we have a complete set of data points over our time period. Many of these industries are not comprised of many minimum wage workers, therefore we would not expect to find a strong impact of minimum wages on these industries at all.

Figure 3.11 (a) presents the impact of minimum wage on earnings per worker by industry. Veterinary Service, Hotels, and Limited-Service restaurants have the largest point estimates. Figure 3.11 (b) shows that differential impact of a local Pasadena increment. A negative estimate indicates a smaller impact of a Pasadena minimum wage increase on earnings. Notice that nearly all of our results are not individually statistically significant. Figure 3.11 (c) presents the impact of minimum wage on employment by industry. Other scientific and technical consulting, Home health care, and Continuing care have the largest negative point estimates. The negative impact of minimum wage on the two consulting industries are quite surprising because they are do not have a large proportion of their workforce working at minimum wage. Our analysis shows that these two industries have been shrinking nationwide as well. Furthermore, the broader category of consulting firms in general (NAICS code 541) has remained stable over this time period. Therefore there is evidence that the decrease is due to the reclassification of many firms in the "Other Technical Consulting" sector to a different consulting designation. Figure 3.11 (d) shows that differential impact of a local Pasadena increment. A negative estimate indicates a stronger negative impact of a Pasadena minimum wage increase on employment. Figure 3.11 (e) presents the impact of minimum wage on establishments by industry. Continuing care, Other scientific and technical consulting, and Management Consulting have the largest negative point estimates. Nursing and continuing care also exhibit negative establishment effects. Figure 3.11 (f) shows that differential impact of a local Pasadena increment. A negative estimate indicates a stronger negative impact of

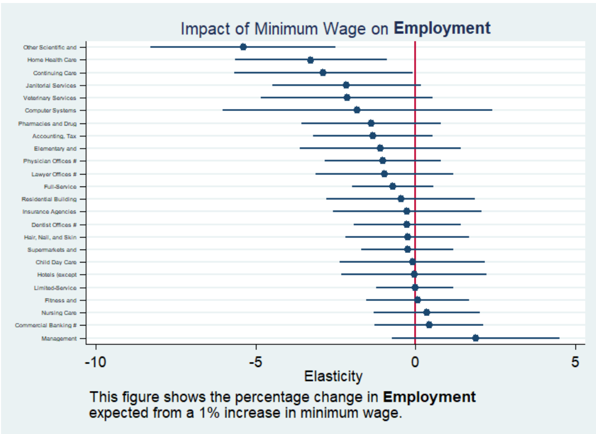
Figure 3.11: Impact of Minimum Wage on Earnings per person, Employment, and Number of Establishments



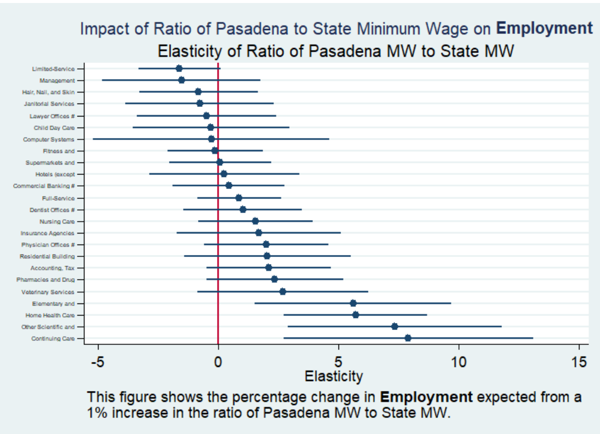
(a)



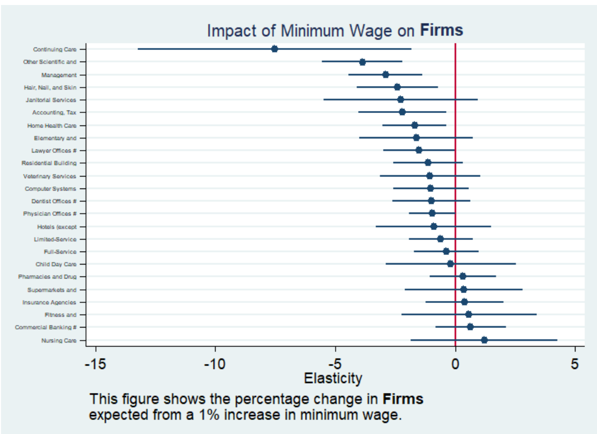
(b)



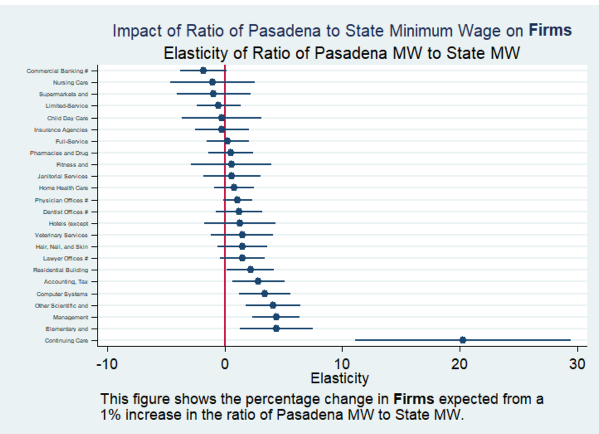
(c)



(d)



(e)



(f)

a Pasadena minimum wage increase on establishments.

3.5.2 Impact of Minimum Wage on Sales Tax Revenue

Sales tax revenue data that have been provided to us by the city of Pasadena can also be explored for minimum wage effects. Although this dataset does not break Pasadena and the surrounding regions apart into smaller pieces like the QCEW data, it does include data from nearby other cities that are similar to Pasadena in terms of income. These other cities are: Glendale, Monrovia, Santa Monica, and West Hollywood.

Figure 3.12: Pasadena Sales Revenue

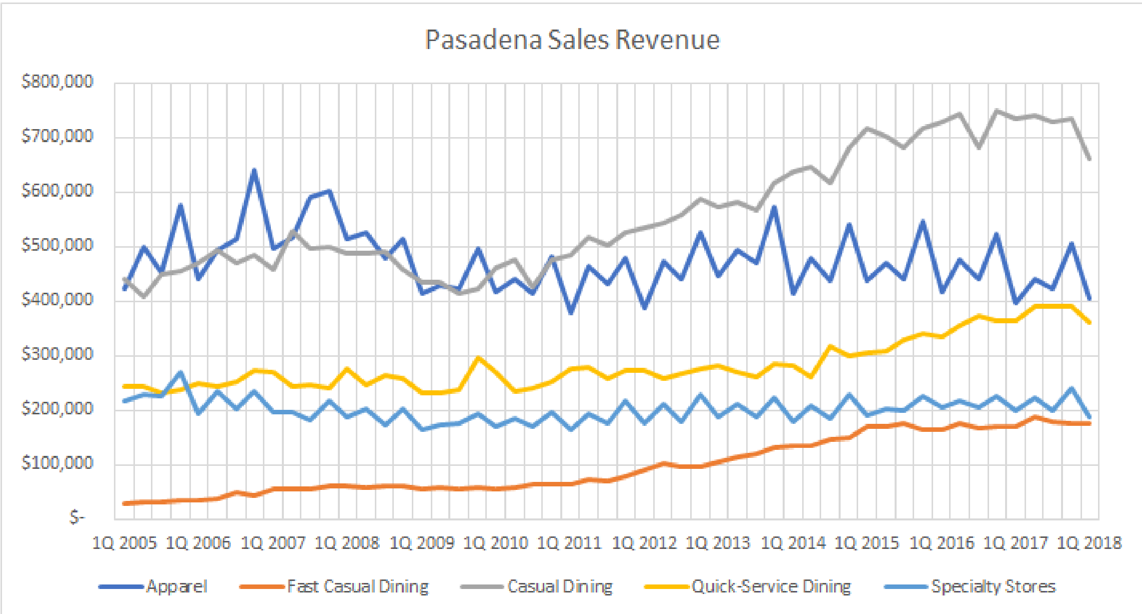
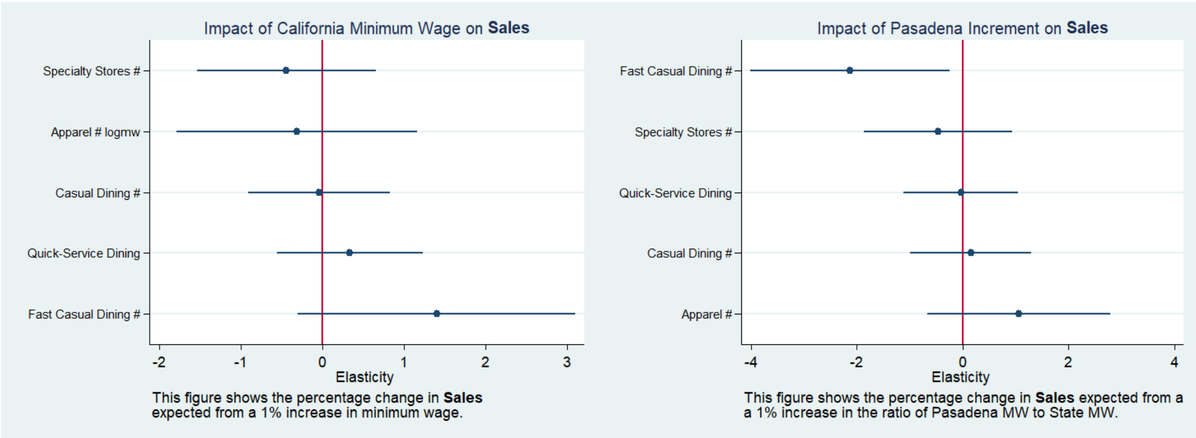


Figure 3.12 illustrates the increasing importance of food services as a source of tax revenue for Pasadena since 2011. Casual dining, quick-service dining, and fast casual dining have all experienced substantial increases in tax revenue since 2011 while apparel and specialty stores have been quite stable.

The timing of the rise in tax revenue from the restaurant sectors after 2011 suggests that the tax revenue is favorably affected by the rise in the minimum wage. A positive impact of minimum wages on sales revenue can occur either because more quantity is sold or because prices rise. A reason why more quantity might be sold is that the added income of restaurant workers allows them to buy more of their own product. A more likely story

Figure 3.13: Impact of Minimum Wage on Sales Tax Revenue



(a) (b)

is that the increase in minimum wages is passed on to customers via higher prices. And of course there may be reasons for increases in price or increases in sales volumes that have nothing to do with the minimum wage.

We can use the same specifications as we have in our previous analysis of earnings, employment, and establishments to examine the impact of minimum wage on sales revenue in these five industries. Figure 3.13 (a) shows the impact of an increase in minimum wages inclusive of the local increment. The solid dots are our point estimates, which show that for Fast Casual Dining, (for example: McDonalds), a 1% increase in minimum wage would result in a 1% increase in sales revenue. The line intervals indicate a 95% confidence interval of our point estimates, and if the lines intersect the solid red line at 0, then our point estimates are not statistically significant at the 5% confidence level. We can see that none of our point estimates of the impact of minimum wage on sales revenue is statistically significant, although we could say that cheaper restaurants seem to have a stronger response than more expensive restaurants and clothing stores. For the restaurants classified as fast casual dining, the evidence says that we would see only 30% of the increase in sales revenue in response to minimum wages would occur within three months.

Figure 3.13 (b) illustrates the separate impact of the Pasadena increment to the minimum wage. Our model includes two minimum wage variables, one is the prevailing minimum wage inclusive of the Pasadena increment and the other is the Pasadena increment separately. If the Pasadena increment behaves just like the California increment, then this second variable would have a zero effect. Once again we would like to stress that we do not have much evidence of this second effect because the Pasadena minimum wage has only risen above the California minimum wage briefly three times in our dataset (which spans to the 2018q1). From the line intervals displayed we can see that only the fast casual dining effect is bounded away from zero, suggesting that the Pasadena increment has much less of an impact than a statewide increase in minimum wages.

Overall our evidence says that sales revenue has a stronger response to minimum wages for restaurants that are cheaper and faster, while restaurants that are more expensive, clothing, and specialty stores do not show evidence of a response.

3.5.3 Impact by January 2021

This part provides a forecast of what happens to earnings, employment, and firms in four specific industries through January 2021 under two different choices for Pasadena's minimum wage, \$15 if Pasadena opts for the \$1 local increment dictated by the higher minimum wage schedule of the City of Los Angeles (Scenario 1), or \$14 if the California minimum wage is used (Scenario 2). (In early February 2019. Pasadena voted in favor of continuing on the same minimum wage schedule as the City of Los Angeles. Therefore we will see if our predictions bear out.)

For Scenario 1, in January 2021, the Pasadena will have a prevailing minimum wage of \$15, equal to the California level of \$14 plus the \$1.00 Pasadena increment. This involves a rise in the prevailing minimum wage from \$13.25 to \$15 and a fall in the Pasadena increment from \$1.25 to \$1.00. Table 3.9 shows the forecasted long run impact of this rise from 13.25 to \$15 with the local increment equal to \$1. For earnings, we expect to see 4% to 6% increase. We also expect some negative impact on employment and number of establishments, especially in full-service restaurants and hair, nail, and skin care services.

In Scenario 2, the California state minimum wage will increase to \$14 by January 2021. Table 3.9 report the estimated impact of \$13.25 to \$14 (+5.7%) while dropping the Pasadena increment from \$1.25 to \$0.00 (-7.5%). This is predicted to increase average quarterly

Table 3.9: Estimated Impact of Minimum Wage by January 2021

Industry	Average Earnings per Quarter	Scenario 1 Potential Increase	Scenario 1 Percent Increase	Scenario 2 Potential Increase	Scenario 2 Percent Increase
Full-Service Restaurants	6379	485	7.60%	553	8.67%
Limited-Service Restaurants	5298	583	11.00%	322	6.07%
Supermarkets and Groceries	7385	255	3.45%	439	5.95%
Hair, Nail, and Skin Care Services	5240	238	4.55%	246	4.69%

Industry	Average Total Employment	Scenario 1 Jobs at risk	Scenario 1 Percent at risk	Scenario 2 Jobs at risk	Scenario 2 Percent at risk
Full-Service Restaurants	6361	-700	-11.00%	-776	-12.20%
Limited-Service Restaurants	4662	130	2.78%	699	15.00%
Supermarkets and Groceries	2488	-85	-3.43%	-54	-2.19%
Hair, Nail, and Skin Care Services	539	-10	-1.79%	34	6.32%

Industry	Average Total Firms	Scenario 1 Firms at risk	Scenario 1 Percent at risk	Scenario 2 Firms at risk	Scenario 2 Percent at risk
Full-Service Restaurants	257	-14	-5.40%	-11	-4.45%
Limited-Service Restaurants	235	-16	-6.95%	4	1.49%
Supermarkets and Groceries	42	3	6.52%	5	10.80%
Hair, Nail, and Skin Care Services	105	-36	-34.50%	-29	-27.60%

earnings in limited-service restaurants by 6.07% and full service restaurants by 8.67%, and to reduce the number of hair, nail and skin salons by 27.60%.

3.6 Robustness Check

3.6.1 Without Pasadena Increment

Adding an additional term representing Pasadena increment does not change the impact of minimum wage on earnings, employment, or number of establishments, except for limited-service restaurants. Adding the local increment of minimum wage would help us understand a local Pasadena minimum wage differs from a statewide minimum wage. The negative impact of Pasadena increment on employment level for limited-service restaurants shows evidence that when local minimum wage increase, minimum wage jobs may migrate to nearby areas.

3.6.2 Without Time Trend

When time trend is not included in the specification, we observe more industries with statistically significant impact from minimum wage on earnings. This is because both earnings and minimum wages are generally increasing over time. Even without increasing minimum wages, historically we observe earnings increase over time due to inflation. Without controlling for time trend, we would mix the increase of earnings due to inflation with the impact of minimum wage.

There is little evidence of impact of minimum wage on employment with or without the time trend. More industries have significant negative impact of minimum wage on establishments when time trends are added. As the figure to the right shows, some industries exhibit increasing establishments until minimum wages are increased. Therefore adding a time trend allows us to project the number of establishments that would have been there had there not been a minimum wage increase.

3.6.3 Analysis for Groups Separately

The results we have discussed so far are using all the groups (The groups are separated by income level. Group 1 has the least income, and Group 5 has the most). We also conduct analysis for each group separately. The purpose of doing group-wise analysis is to examine whether a change in minimum wage has different impact depending on the income level of the affected area. We find that the impact of minimum wage differs little across groups but there is no obvious pattern. Groups 2 and 3 provide the most significant evidence that increase in minimum wage would increase earnings. Group 4 presents the weakest evidence. Most industries show different results across different groups. However, for Full and Limited service restaurants, there are consistent results across all groups showing that an increase in minimum wages would increase earnings. There is little evidence of the impact of minimum wages on employment or number of establishments.

3.7 Conclusion

We have used the data available to us to analyze the impact of minimum wages on earnings, employment, establishments, and sales tax revenue. We find that minimum wages have a measurable impact on earnings for low wage industries (such as full and limited service restaurants), and our preferred model shows a significant negative impact of minimum wages on the number of Hair/Nail Salons and also a negative impact of the Pasadena increment on the number of jobs in Limited Service Restaurants. However, we do obtain negative estimates of the impact of minimum wages on employment and establishments in most industries.

This study has difficulty detecting the impact of minimum wages on employment and establishments because firms may anticipate upcoming changes in minimum wages, and also may adjust everything but wages slowly over time. Indeed our own estimates show that only one-fifth of the impact of an increase in minimum wages would show in the data within three months. Data from the rest of 2018 would be quite helpful because 2018 and 2019 are the years during which the Pasadena minimum wage is highest above the California state minimum wage.

We find evidence that 50% of the impact of minimum wages on earnings is realized in the first quarter, while only 20% of the impact of minimum wages on employment of on establishments is realized in the first quarter.

We find smaller estimates in general of the impact of a Pasadena increment than a Statewide increment, however jobs in limited service restaurants show evidence of leaving Pasadena in higher numbers when the difference between the Pasadena minimum wage and the California minimum wage is greater. An additional year of data and the corresponding greater time and greater difference between the Pasadena and Statewide minimum wage levels would allow us to more accurately estimate the separate effect of the Pasadena increment.

(2014, February). The effects of a minimum-wage increase on employment and family income. Technical report, Congressional Budget Office.

Addison, J. T., M. L. Blackburn, and C. D. Cotti (2009). Do minimum wages raise employment? evidence from the u.s. retail-trade sector. *Labour Economics 16*, 397–408.

Allegretto, S., A. Dube, and M. Reich (2011). Do minimum wages really reduce teen employment? Accounting for heterogeneity and selectivity in state panel data. *Industrial Relations 50*(2), 205–240.

Allegretto, S., A. Dube, M. Reich, and B. Zippere (2017). Credible research designs for minimum wage studies: a response to neumark, salas, and wascher. *ILR Review 70*(3), 559–592.

Andrews, I., J. H. Stock, and L. Sun (2019). Weak instruments in instrumental variables regression: Theory and practice. *Annual Review of Economics 11*, 727–753.

Angrist, J. D. and A. B. Krueger (1991). Does compulsory school attendance affect schooling and earnings? *The Quarterly Journal of Economics 106*(4), 979–1014.

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

Baskaya, Y. S. and Y. Rubinstein (2012). Using federal minimum wages to identify the impact of minimum wages on employment and earnings across the u.s. states. *Working Paper*.

Bekker, P. A. (1994). Alternative approximations to the distributions of instrumental variable estimators. *Econometrica: Journal of the Econometric Society*, 657–681.

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.

Bound, J., D. A. Jaeger, and R. M. Baker (1995). Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak. *Journal of the American statistical association 90*(430), 443–450.

Brochu, P. and D. A. Green (2013). The impact of minimum wages on labour market transitions. *The Economic Journal 123*, 1203–1235.

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.

Card, D. and A. B. Krueger (1994). Minimum wages and employment: a case study of the fast-food industry in new jersey and pennsylvania. *The American Economic Review 84*(4), 772–793.

Chamberlain, G. and G. Imbens (1996). Hierarchical bayes models with many instrumental variables. *NBER Technical Working Paper*.

Chao, J. C. and P. C. Phillips (1998). Posterior distributions in limited information analysis of the simultaneous equations model using the jeffreys prior. *Journal of Econometrics 87*(1), 49–86.

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.

Dreze, J. H. (1976). Bayesian limited information analysis of the simultaneous equations model. *Econometrica: Journal of the Econometric Society*, 1045–1075.

Dude, A., T. W. Lester, and M. Reich (2010). Minimum wage effects across state borders: estimates using contiguous counties. *Working Paper*.

Dude, A., S. Naidu, and M. Reich (2007). The economic effects of a citywide minimum wage. *ILR Review 60*(4), 522–543.

EconML, M.-R. (2019). EconML: A Python Package for ML-Based Heterogeneous Treatment Effects Estimation. https://github.com/microsoft/EconML. Version 0.x.

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.

Fieller, E. C. (1932). The distribution of the index in a normal bivariate population. *Biometrika*, 428–440.

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

Giuliano, L. (2013). Minimum wage effects on employment, substitution, and the teenage labor supply: evidence from personnel data. *Journal of Labor Economics 31*(1), 155–194.

Hahn, J. and J. Hausman (2002). A new specification test for the validity of instrumental variables. *Econometrica 70*(1), 163–189.

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

Hartford, J., G. Lewis, K. Leyton-Brown, and M. Taddy (2017). Deep iv: A flexible approach for counterfactual prediction. In *Proceedings of the 34th International Conference on Machine Learning-Volume 70*, pp. 1414–1423. JMLR. org.

Hirsch, B. T., B. E. Kaufman, and T. Zelenska (2015). Minimum wage channels of adjustment. *Industrial Relations 54*(2), 199–239.

Hoogerheide, L., F. Kleibergen, and H. K. van Dijk (2007). Natural conjugate priors for the instrumental variables regression model applied to the angrist–krueger data. *Journal of Econometrics 138*(1), 63–103.

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

Jardim, E., M. C. Long, R. Plotnick, E. van Inwegen, J. Vigdor, and H. Wething (2018). Minimum wage increases, wages, and low-wage employment: evidence from seattle. *NBER Working Paper*.

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

Leamer, E. E. (2010). Tantalus on the road to asymptopia. *Journal of Economic Perspectives 24*(2), 31–46.

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

Link, W. A. and M. J. Eaton (2012). On thinning of chains in mcmc. *Methods in ecology and evolution 3*(1), 112–115.

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.

Lopes, H. F. and N. G. Polson (2014). Bayesian instrumental variables: priors and likelihoods. *Econometric Reviews 33*(1-4), 100–121.

MacEachern, S. N. and L. M. Berliner (1994). Subsampling the gibbs sampler. *The American Statistician 48*(3), 188–190.

Meer, J. and J. West (2016). Effects if the minimum wage on employment dynamics. *Journal of Human Resources 51*(2), 500–522.

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

Nelson, C. R. and R. Startz (1990). The distribution of the instrumental variables estimator and its t-ratio when the instrument is a poor one. *Journal of business*, S125–S140.

Neumark, D. and W. Wascher (2011). Does a higher minimum wage enhance the effectiveness of the earned income tax credit? *Industrial and Labor Relations Review 64*(4), 712–746.

Newey, W. K. and J. L. Powell (2003). Instrumental variable estimation of nonparametric models. *Econometrica 71*(5), 1565–1578.

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

Phillips, P. C. (2009). Exact distribution theory in structural estimation with an identity. *Econometric Theory 25*(4), 958–984.

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

Reich, M., K. Jacobs, A. Bernhardt, and I. Perry (2014, September). The mayor of los angeles' proposed city minimum wage policy: A prospective impact study. Technical report, Institute for Research on Labor and Employment University of California Berkeley.

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.

Sabia, J. J., R. V. Burkhauser, and B. Hansen (2012). Are the effects of minimum wage increases always small? New evidence from a case study of new york state. *ILR Review 65*(2), 350–376.

Sanderson, E. and F. Windmeijer (2016). A weak instrument f-test in linear iv models with multiple endogenous variables. *Journal of Econometrics 190*(2), 212–221.

Staiger, D. and J. H. Stock (1994). Instrumental variables regression with weak instruments. Technical report, National Bureau of Economic Research.

Stock, J. H. and M. Yogo (2002). Testing for weak instruments in linear iv regression. Technical report, National Bureau of Economic Research.

Thornberg, C., J. Levine, D. Schrader, B. Vanderplas, and M. Saia (2015, March). Cost-benefit analysis: Los angeles minimum wage proposal. Technical report, Beacon Economics.

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

Wicks-Lim, J. (2006). Mandated wage floors and the wage structure: new estimates of the ripple effects of minimum wage laws. *PERI working paper*.

Wiesenfarth, M., C. M. Hisgen, T. Kneib, and C. Cadarso-Suarez (2014). Bayesian nonparametric instrumental variables regression based on penalized splines and dirichlet process mixtures. *Journal of Business & Economic Statistics 32*(3), 468–482.

Young, A. (2019). Consistency without inference: Instrumental variables in practical application.

Zellner, A. (1978). Estimation of functions of population means and regression coefficients including structural coefficients: A minimum expected loss (melo) approach. *Journal of Econometrics 8*(2), 127–158.

Zellner, A., T. Ando, N. Baştürk, L. Hoogerheide, and H. K. Van Dijk (2014). Bayesian analysis of instrumental variable models: Acceptance-rejection within direct monte carlo. *Econometric Reviews 33*(1-4), 3–35.