UNIVERSITY OF CALIFORNIA, SAN DIEGO

Three Essays in Applied Microeconomics

A Dissertation submitted in partial satisfaction of the requirements
for the degree Doctor of Philosophy

in

Economics

by

Christopher Paul Steiner

Committee in charge:

Professor Richard Carson, Chair
Professor Jennifer Burney
Professor Julie Cullen
Professor Steven Erie
Professor James Hilger
Professor Mark Jacobsen

2015

The Dissertation of Christopher Paul Steiner is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

Chair

University of California, San Diego

2015

DEDICATION

I would like to dedicate the education chapter to Graham Hereford, a friend of mine who passed away from lymphoma. He was pursuing a college degree when he was diagnosed and faced immense bureaucratic challenges both fighting cancer and attempting to finish his education. I was unable to attend a portion of his funeral because it coincided with my undergraduate graduation – it was a really tough weekend. We have to do a better job doing medical research, providing quality college education, and allocating health care effectively to our neighbors.

EPIGRAPH

☺

TABLE OF CONTENTS

Signature Page …………………………………..……………………..... iii

Dedication ……………….…………………………………………….... iv

Epigraph …………………………………………………………………. v

Table of Contents …………….……………………………………….… vi

List of Figures ………………………………………………………….. vii

List of Tables …………………………………………….……………... ix

Acknowledgements ……………………………………………..…… xi

Vita …………...……………………………………………………….. xiv

Abstract of the Dissertation ………………………………………….. xv

Chapter 1: "Hitting Capacity: Implications for the Valuation of Outdoor Recreation" ……………………………………………………………… 1

Chapter 2: "An Analysis of the Cost of an Undergraduate Degree and the Incentives of the State, the University, and the Student" ……………………. 48

Chapter 3: "Pollution Whack-a-Mole: Ambient Acetaldehyde and the Introduction of E-10 Gasoline in the Northeast" …………………………… 119

LIST OF FIGURES

Figure 1.1: Trip Types – Fish Caught ……………………………………………... 28

Figure 1.2: Figures of WTP with Sellouts …………………………………… 29

Figure 1.3: Formal Test of Model …………………………………………… 32

Figure 2.1: The University Preference Parameter and Modified Funds-Per-Student ……………………………………………………………......... 59

Figure 2.2: The University Preference Parameter and Student Major Choices ………………………………………………………………………….. 60

Figure 2.3: The University Preference Parameter and Funds Given to Each Department ……………………………………………………………….. 61

Figure 2.4: Modified Funds-per-Student Based on Economics Cost Parameter (Sigma) …………………………………………………………………… 63

Figure 2.5: Number of Students in Each Major Based on Economics Cost Parameter (Sigma) ………………………………………………………... 64

Figure 2.6: Number of Students in Each Major Based on Economics Cost Parameter (Sigma) ………………………………………………………... 65

Figure 2.7: Modified Funds-per-Student Based on Student Preference Parameter (Gamma) ……………………………………………………………….. 66

Figure 2.8: Number of Students in Each Major Based on Student Preference Parameter (Gamma) ………………………………………………………… 67

Figure 2.9: Proportion of Funds Given to Major Based on Student Preference Parameter (Gamma) ………………………………………………………… 68

Figure 2.10: Modified Funds-per-Student Based on Salary of Economics (Salary) …………………………………………………………………… 69

Figure 2.11: Number of Students in Each Major Based on Salary of Economics (Salary) ………………………………………………………………… 70

Figure 2.12: Proportion of Funds Given to Major Based on Salary of Economics (Salary) ………………………………………………………………… 71

Figure 2.13: Simulation: Level of Funding and Total Funding Based on Per-Student State Bonus Payment for Engineering ………………………………. 75

Figure 2.14: Simulation: Level of Funding and Total Funding Based on Per-Student State Bonus Payment for Engineering ………………………………. 76

Figure 2.15: Simulation: Number of Students in Each Major Based on Bonus Payment for Engineering ……………………………………………………. 77

Figure 3.1: Percentage of Ethanol Present in Fuel by Metro Area ………….. 136

Figure 3.2: U.S. Ethanol Plants in 2006 …………………………………….. 137

Figure 3.3: Acetaldehyde Monitors ………………………………………… 138

Figure 3.4: Median Acetaldehyde Measures in Each State …………………. 139

Figure 3.5: Median Acetaldehyde Measures Outside of New Jersey ……….. 140

LIST OF TABLES

Table 1.1: Medians of Sellout Model vs. Standard WTP Model …………….. 18

Table 1.2: Willingness to Pay Estimates in the Naïve Model (Parameter Estimates) ………………………………………………………………… 23

Table 1.3: Willingness to Pay Estimates Derived from the RUM on the Previous Page ……………………………………………………………………… 25

Table 1.4: First Stage Sellout Model ………………………………………… 26

Table 1.5: One Draw of the Soldout Model ………………………………….. 27

Table 2.1: Cost Per Credit Hour (Department Code) ………………………... 92

Table 2.2: Cost Per Credit Hour (Subject Code) …………………………..... 94

Table 2.3: Correlation Between Measures …………………………………... 98

Table 2.4: Regression of Log Cost, Including Adjusted Hours as a Variable .. 99

Table 2.5: Regression of Log Cost without Hours ………………………….. 100

Table 2.6: Regression of Mean-Hour Adjusted Cost on Variables …………. 101

Table 2.7: Regression of Log Time in Years ……………………………….. 102

Table 2.8: Regression of Log Penner Ratio on Other Variables ……………. 103

Table 2.9: Cost per Degree …………………………………………………. 104

Table 3.1: Regression on Levels ……………………………………………. 141

Table 3.2: Log Levels ………………………………………………………. 144

Table 3.3: Log Levels with Other Variables ………………………………... 146

Table 3.4: Log Levels with Other Variables ………………………………... 148

Table 3.5: Weighted Regressions ………………………………………….. 150

Table 3.6: Poisson Regression ……………………………………………… 153

Table 3.7: Controlling the Controls ………………………………………… 155

ACKNOWLEDGEMENTS

I would like to thank my family and my partner for their unwavering support through this dissertation.

I would like to acknowledge my entire committee for their support through this dissertation. My chair, Richard Carson, has been extremely generous with his time and positivity – it takes a skilled expert to know how to guide graduate students through the relatively new journey of economics research.

I would like to thank Melissa Famulari both for her assistance in getting the education chapter completed and for her expertise as a teaching mentor.

I would like to thank James Hilger for giving me the opportunity to work at the National Oceanic and Atmospheric Administration, where I had access to the data necessary for Chapter 1 on the Commercial Passenger Fishing Vessel Industry in San Diego.

I would like to thank Julie Cullen and Mark Jacobsen for their support throughout my dissertation writing process and invaluable feedback. Prof. Cullen was also generous with her time as a job market coordinator this year, and I also thank her for allowing me to do research with her my first year here.

Chapter 1 ("Hitting Capacity: Implications for the Valuation of Outdoor Recreation") is joint with James Hilger and is being presented at the Association of Environmental and Resource Economists conference in San Diego, 2015. An abstract of this paper is already public, and it will be prepared for publication.

Chapter 2 ("An Analysis of the Cost of an Undergraduate Degree and the Incentives of the State, the University, and the Student") is joint with Richard Carson and Melissa Famulari, and it will be prepared for publication.

Chapter 3 ("Pollution Whack-a-Mole: Ambient Acetaldehyde and the Introduction of E-10 Gasoline in the Northeast") is available in different form on AgEcon Search and was presented at the Agricultural and Applied Economics Association Meetings in Minneapolis, 2014. It will be prepared for publication.

For the fisheries chapter, I would like to thank Richard Carson, Mark Jacobsen, Yixiao Sun, Jeffrey Shrader, Johnathan Sweeney, two preliminary reviewers at NOAA, and participants at the Center for Environmental Economics at UCSD seminar series. This paper is the work of the authors and not the opinion of the National Oceanic and Atmospheric Administration.

For the education chapter, I would like to thank Devaney Kerr, Financial Manager in the Mechanical & Aerospace Engineering Department, and Sue King, Chief Administrative Officer in the Economics Department, with help procuring and understanding the administrative data. Further, I would like to thank Christine Hurley and Kirk Belles. Christopher Beliare from the real estate firm Newmark Grubb Knight Frank assisted me in locating lease data for the UCSD area. Further, input from department chair Valerie Ramey was invaluable.

For the air chapter, I would like to thank several NOAA NCDC and EPA AQS Datamart personnel for extensive help with downloading and interpreting data. I am also grateful to the Center for Environmental Economics for feedback. Feedback from Julie Cullen was invaluable.

The author requests that readers understand that before publication, all work in the dissertation here is work nearly finished but is subject to change before publication. The author has made every attempt to reduce errors before dissertation submission but will review code and figures again before publication. Mistakes are his own.

VITA

2008	Bachelors of Science, Mathematics and Economics University of Illinois at Urbana-Champaign
2009	Masters of Arts, Economics University of California, San Diego
2013	Candidate of Philosophy, Economics University of California, San Diego
2015	Doctor of Philosophy, Economics University of California, San Diego

ABSTRACT OF THE DISSERTATION

Three Essays in Applied Microeconomics

by

Christopher Paul Steiner

Doctor of Philosophy in Economics

University of California, San Diego, 2015

Professor Richard Carson, Chair

These three essays investigate three different cases where naïve good intentions – policy or econometric – actually lead to suboptimal policy or measurement outcomes.

In the first chapter, James Hilger and I investigate bias in the commercial passenger fishing vessel (CPFV) industry when a naïve researcher estimates willingness to pay estimates (WTP), derived from random utility models (RUM), in the context of vessel sellouts. Using incorrectly estimated WTP measures may lead to undervaluation of natural resources.

In the second essay, Richard Carson, Melissa Famulari, and I simulate a university with a benevolent higher level administrator who wants to keep per-

student funding roughly the same, or same with adjustment for preferences, across the university in a CES-style fashion. If students also prefer to major in departments with high per-student funding, these two goals are in conflict and necessitate the higher-level administrator to lower per-student funding for popular departments. Using data from UCSD, we find that departments with large numbers of students are less expensive per degree, have higher modified student-to-faculty ratios, and graduate students *sooner* than other departments.

In the third essay, I investigate the transition from methyl tertiary-buthyl ether- (MTBE-) enhanced to ethanol-enhanced (E-10) fuel in the Northeastern United States. Using a complicated set of phase-ins and phase-outs, I use difference-in-difference estimation to show that ambient acetaldehyde pollution substantially increased in percentage terms because of E-10 – although this is a small level increase, since the level of acetaldehyde is low in the area. Using a stylized calculation based on cancer risk still shows damages of this pollution are levels of magnitude lower than the billion dollar water pollution cleanup costs from MTBE additive.

Chapter 1

Hitting Capacity: Implications for the Valuation of Outdoor Recreation

1.1 Chapter Abstract

Choices are often limited as the most popular alternatives reach capacity and sell out; thereafter, selection is over less preferred choices. In the context of nonmarket goods, willingness to pay (WTP) welfare measures provide an estimate of the value of nonmarket characteristics as calculated through the modeling of preferences using a random utility model framework. Random utility calculations are based on the choice attributes and the observed choices consumers make from a set of options. Such models are estimated under the implicit assumption that all options are available to all consumers. If choices can "sell

out," the properly specified choice model would drop unavailable alternatives from the set of options; however, actual availability is almost never observed at the individual consumer level. Ignoring capacity constraints can result in biased parameter and WTP estimates. A solution to this problem that can be implemented using only aggregate level data is provided. We provide an empirical application of modeling vessel choice in the recreational overnight fishing trip market in San Diego – where particular boats are often sold out.

1.2 Introduction

Outdoor recreators are often limited in their recreational site choice because more popular alternatives sell out. Consumer passenger fishing vessel (CPFV) trips, the focus of this particular paper, sell out. Failure to account for these sellouts leads to bias in the parameters of random utility models (RUMs) used to analyze consumer choice from multiple alternatives. For while the aggregate choice-set is known, choice-set data at the individual choice occasion is rarely available. In the absence of information on the actual choice-set at the time of decision, the analyst may incorrectly conclude that, for some, the characteristics of the next best alternatives are preferred over the characteristics of the choices that sell out.

In the basic RUM framework, the consumer's choice set is constructed of a static set of alternatives. While early modeling efforts allowed consumers to face individual choice sets constructed of different alternatives (see Haab and Hicks (1999) for survey of the literature), they did not account for the possible impact that the choices of a subset of consumers may have on the feasible choice set of the remaining consumers.

For example, let's assume that there are two recreational fishing vessels that sportfishers ride for fishing trips; further, they are identical except for one has a historically higher catch rate for prized fish. If the neither boat sells out, then the recovered WTP estimates are correct. Here, the analyst assumed that that both vessels were available to all consumers. However, if the vessel with the higher catch rate reaches its capacity early, then anglers arriving later will select the less preferable boat. This will result in a WTP for the catch rate which has a downward bias since many anglers appear to prefer less catch for more money.

The lack of a price mechanism used by government agencies often allows for excess demand in many recreational activities. For instance, at Yosemite National Park, campers are warned that reservation campgrounds can often fully book minutes after booking is opened online – and that non-reservation campgrounds often fill up before noon[1]. At the Grand Canyon, campers at the South Rim are warned that, during the summer, "campgrounds hustle and bustle and are often filled to capacity.[2]" While queues and full bookings are not an economically efficient way of allocating space, the National Park Service serves the public and is not a profit-maximizing entity. While the application presented in this paper happens to be a market-driven sellout largely associated with advertising practices and landings controlling multiple boats – sellouts also occur frequently in many outdoor recreation literatures.

[1] http://goo.gl/dpCVSZ
[2] http://goo.gl/jlt8O8

Similarly, preferred beaches often do not run out of space, but they frequently run out of parking – effectively limiting the number of visitors to the beach. Popular sporting events often sell out – with laws or transactions costs effectively preventing prices from increasing through scalping. All of these common, capacity hitting events have implications for choice. In our application, vessels typically do not do last-minute price adjustments to avoid sellouts.

The operations research literature has long been concerned with stockouts from an inventory and marketing perspective, often with papers using modified RUM's (i.e., Conlon and Mortimer 2013, Musalem et.al. 2010). This is the first paper that models sellouts in a recreational setting. It is useful to distinguish between crowding and sellouts: Many recreational papers look at crowding – crowded beaches, for instance (i.e., McConnell 1977). Crowding differs from sellouts in that that crowding is a recreational characteristic – and sellouts represent a hard constraint forcing the activity out of the choice set entirely.

This paper reports on the modeling of consumer recreational behavior in the context of the San Diego based sport fishery, which is characterized by a large number of sold-out boats. We first demonstrate that the standard RUM parameter estimates and the corresponding WTP measures are generally biased under conditions that fail to account for sellouts. We then propose a RUM estimator which accounts for sellouts and examine its econometric properties. Lastly, we estimate the proposed RUM that accounts for sellouts and report statistically significant and large differences between RUM parameter and WTP estimates when sellouts are taken into account versus the traditional model.

Our focus is on five highly migratory species (HMS)[3] – and whether vessels are successful in targeting these species. In order to look at these species, we limit our analysis to overnight to two-day trips (at least one night out at sea), as catching HMS on a day trip is rare[4].

We look at the proportion of HMS species caught the previous season as a proxy for the type of fish species anglers are likely catch on the trip. Empirical results indicate the per trip WTP for successful HMS targeting in the hundreds of dollars. Parameter and WTP estimates have large biases when sellouts are not accounted for in the RUM.

The rest of the paper proceeds as follows. Section 1.3 reviews relevant literature. Section 1.4 presents a sellout simulation. Section 1.5 develops the framework for empirical estimation. Section 1.6 presents the data. Sections 1.7 and 1.8 present the empirical results for the standard and proposed models. Section 1.9 concludes.

1.3 Literature

A pair of papers in the recent marking literature has pointed out that estimates using data from store shelves or vending machines can be biased if a sellout occurs and this is not taken into account. Musalem, et.al. (2010) uses a Bayesian framework to simulate stock-outs of shampoo. Conlon and Mor-

[3]Bluefin tuna, yellowfin tuna, albacore tuna, dolphinfish, and yellowtail (see Figure 1.1).
[4]See Figure 1.1. The industry bundles trips into different trip types based on length of trip, ranging from "half day" trips all the way up to extremely long, multiday trips. Longer multiday trips are not considered here as they often involve specialized itineraries.

timer (2013) looks at vending machines with incomplete product availability and average all possible combinations of potential choice sets. Both Conlon and Mortimer and Musalem, et.al. are focused on estimating lost revenues from stock-outs. The approaches in these papers are both computationally intensive and difficult to implement[5]. Further, Conlon and Mortimer require each potential choice set to be a factor in the likelihood. This leads to 2^N different combinations of sellouts.

The recreational nonmarket valuation literature has undergone considerable methodological advances as analysts have addressed a myriad of distinct markets and local conditions. While these models are applied to many areas of recreation, sportfishing is a particularly well-studied phenomenon (Boyle, et.al. 1998). Johnston, et.al. (2006) is a meta-analysis of 48 fishery studies in the United States and Canada. It determined that the methodology of studies influenced WTP values.

Sportfishing continues to be researched because of the relevant policy particulars. Sportfishing is an extremely popular American pastime that encompasses a myriad of issues such as recreational-commercial allocation, impacts to pollution, climate change, and job creation. For instance, Carson, Hanemann, and Wegge (2009) discuss their assistance to Alaska to help policy makers determine the consequence of closing fishing sites to avoid overfishing. In the Alaskan fresh-water salmon fishery, closing one fishery may lead to spillovers into other areas – the authors use a nested logit framework to predict where the

[5] An alternative to these methods include maximum score estimation, as suggested by Fox (2007).

fishermen would go. Despite all of these advances, this is the first paper we are aware of in the general areas of recreational economics and non market valuation that account for sellout bias in parameter and welfare estimates.

Few fisheries studies look at overnight trips because of the challenge posed to researchers. Many explicitly exclude overnight trips (for instance, Haab et.al. 2012). McConnell and Strand (1999) is one of the few studies to look at overnight trips. In Southern California, excluding overnight trips ignores the recreationally, economically, and environmentally important HMS fishery.

An additional distinction between this study and the fisheries literature is that we use a census of CPFV trips in Southern California from the California Department of Fish and Wildlife. This census data comes from a mandatory logbook program. We pair the vessel census with pricing from the Internet Archive. Most studies use angler surveys (Shaw and Ozog (1999), Larson and Lew (2013), and Hauber and Parsons (2000)). Carter and Liese (2010) take a different approach; they use a hedonic pricing analysis and a survey of boat pricing.

1.4 A Sellout Simulation

Not accounting for sellouts can lead to biased RUM parameter estimates and their corresponding WTP estimates. This section briefly simulates a hypothetical example to demonstrate the severity of this bias. Assume that a group of anglers have the following utility for a fishing trip:

$$U_{bit} = \beta_{price} p_{bt} + \beta_{catch} f_{bt} + \beta_X c_{bt} + \epsilon_{bit}, \tag{1.1}$$

where $i = 1, \ldots$,approximately 100,000 denotes fishermen, $b = 1, \ldots, 6$ denotes the vessels in the choice set, p_{bt} and f_{bt} denote the price and predicted fish catch, respectively, for vessel b and for trip t, and c_{bt} denotes an additional vessel characteristic (such as the age of the vessel). We set $\beta_{price} = -2.5$, $\beta_{catch} = 7.5$, and $\beta_X = 1.25$, and we denote the random error as $\epsilon_{bit} \sim$ Type I Extreme Value. We draw 100,000 customers and divide them into T time periods/queues; we order them and change the time period every Y people, where Y is a random variable $Y \sim N(70, 15^2)$. The last trip, which truncates, is discarded. The vessels have capacities $[20, 15, 10, 100, 100, 100]'$.

We assign different prices and different predicted fish catch for each time period x vessel combination. Each time period has three random variables, $F_t \sim U(1.5, 2.5)$, $P_t \sim U(0.9, 1.1)$, and $C_t \sim U(0.95, 1.05)$. Parameters $u_{\bullet bit} \sim U(0, 1)$ are assigned in the following way: $C_{bit} = C_t[(2.5 - 0.25i) + u_{Cbit}]$, $F_{bit} = F_t[(18 - 1.5i) + 4u_{Fbit}]$, and $P_{bit} = P_t[(60 - 5i) + 2u_{Pbit}]$. Vessel 1 has the highest expected price and the highest catch, and the vessels continue to decline in price and quality as the index gets higher. However, sometimes, vessel 2 may be more preferable for most people – and for any one angler, the random component ϵ_{bit} potentially makes any vessel the preferred choice.

Estimation of WTP follows maximum likelihood estimation; following

McFadden (1974), the probability of selecting boat i is:

$$\frac{\exp\left(v_{bit}\right)}{\sum_{j=1}^{6} \exp\left(v_{bjt}\right)} \tag{1.2}$$

Using the above utility function (1.1), its parameters, and the distribution of vessel attributes, the expected number of passengers that select each vessel as the utility maximizing choice can be generated by means of maximum likelihood estimation of (1.2). For a moment, assume that there are no sellouts; i.e., the capacities are $[\infty, \infty, \infty, \infty, \infty, \infty]'$. The following table reports the simulated number of passengers aboard each vessel:

Vessel	Passengers
1	54%
2	28%
3	12%
4	4%
5	1%
6	<1%

However, given the capacities are $[20, 15, 10, 100, 100, 100]'$, the actual boarding is as follows:

Vessel	Passengers
1	25%
2	16%
3	8%
4	30%
5	14%
6	6%

The point estimate of the WTP measure for f_{bt} is denoted as WTP $= -\beta_{catch}/\beta_{price}$. If sellouts do not occur (the capacities are $[\infty, \infty, \infty, \infty, \infty, \infty]'$), the estimates of the utility function are accurate:

Psuedo $R^2 = 0.95$	β_{price}	β_{catch}	β_X	WTP
Estimate	-2.50	7.50	1.19	3.01
95% CI	$(-2.55, -2.44)$	$(7.34, 7.66)$	$(1.09, 1.29)$	$(3.00, 3.01)$

Now assume that capacities are $[20, 15, 10, 100, 100, 100]'$. The analyst naïvely estimates the model without adjusting the choice set, as he does not observe which people had a reduced choice. The parameters are biased.

Psuedo $R^2 = 0.13$	β_{price}	β_{catch}	β_X	WTP
Estimate	-0.17	0.32	0.12	1.91
95% CI	$(-0.17, -0.17)$	$(0.32, 0.32)$	$(0.09, 0.14)$	$(1.89, 1.93)$

One might be tempted to make some type of post hoc correction using fixed effects. The appendix, however, shows that even making such a correction will not generally eliminate the bias in willingness to pay estimates.

1.5 Framework for Estimation

We start by following a standard random utility model (McFadden 1974). Denote $U_{\mathbf{c}it}$ to be the utility of a customer i facing a matrix of vessels, trip characteristics, and trips $\mathbf{C}_t = \begin{bmatrix} \mathbf{c}_1 & \cdots & \mathbf{c}_C \end{bmatrix}$ during a time period, t (each trip is denoted as a vector including a trip identifier and characteristics). Each person decides between trips on a set of connecting weekdays or the weekend. Each set of weekdays or days on the weekend are considered one choice occasion. In this setup, $U_{\mathbf{c}it} = v_{\mathbf{c}it} + \epsilon_{\mathbf{c}it}$, where $v_{\mathbf{c}it}$ is the deterministic component and $\epsilon_{\mathbf{c}it}$ is random (Type I Extreme Value) error. The probability of selecting trip $\mathbf{c} \subset \mathbf{C}_t$ is:

$$\frac{\exp\left(U_{\mathbf{c}it}\right)}{\sum\limits_{\hat{\mathbf{c}} \in \mathbf{C}_t} \exp\left(U_{\hat{\mathbf{c}}it}\right)} \tag{1.3}$$

The deterministic component, $v_{\mathbf{c}it} = \mathbf{x}'_{\mathbf{c}it}\boldsymbol{\beta}$ includes price and fish catch variables. The willingness to pay for a particular trip attribute, j, is given by $-\beta_j/\beta_{\text{price}}$.

A standard framework would estimate (1.3) using maximum likelihood estimation on the available choice set. However, we do not have information on the full choice set, so we must predict whether a vessel sells out. Let N_t equal the number of anglers in the market during the time-frame t with vector of trips and trip characteristics $\mathbf{c}$. A first stage linear probability model for predicting sellouts is:

$$\text{soldout}_{\mathbf{c}t} = \boldsymbol{\beta}'\mathbf{X}_{\mathbf{c}t} + \delta N_t + \epsilon_{\mathbf{c}t} \tag{1.4}$$

Denote the fitted value as $\widehat{\text{soldout}}_{\mathbf{c}t}$, and the standard error of the forecast as $\hat{\sigma}$.

Draw a parameter $\hat{g}_i$ from $U(0,1)$. Define p as:

$$p_{\mathbf{c}it} = \widehat{\text{soldout}}_{\mathbf{c}t} - \delta \hat{g}_i N_t \tag{1.5}$$

Additionally, draw $q_{1\mathbf{c}it} \sim U(0,1)$ and $q_{2\mathbf{c}it} \sim N(0,1)$. Trip **c** is eliminated from the choice set for person i if $q_{1\mathbf{c}it} \leq p_{\mathbf{c}it} + \hat{\sigma} q_{2\mathbf{c}it}$. We then run equation (1.3) with the new choice set, and repeat this process 1000 times. We then look at the distribution of point estimates and their confidence intervals.

1.6 Sportfishing Data

1.6.1 Data

Sportfishers from around the world access near- and off-shore fishing grounds off Southern California and Baja California through the Commercial Passenger Fishing Vessel (CPFV) fleet operating out of San Diego. The CPFV industry in Southern California is an important recreational and economic resource; over 500,000 angler days in 2013 alone contributed to over 1500 jobs (NOAA District 1 only) (Hilger 2014). The recreational fishery (as a whole) in California is estimated to have added \$1.0 billion of value to the economy in 2012 (National Marine Fisheries Service 2014). The industry bundles trips into different trip types based on length of trip. We will focus on the high season trips stratified at 24, 36, and 48 hours, which are marketed as overnight, 1½ day, and 2-day trips, respectively. These longer trips specialize in relatively more prized species, including highly migratory species (HMS), the focus of this analysis (see Figure 1.1).

We apply the preceding model to a unique data set of overnight to two day CPFV trips in San Diego, California. The primary data consists of mandatory trip records compiled from the *Skipper's Log Book*, a data collection program from the California Department of Fish and Wildlife. For each trip, skippers record the number of passengers, the number of fish kept by species, the number of hours spent fishing, the days spent fishing, and the departure and arrival times. The U.S. Coast Guard and the California Department of Fish and Wildlife have provided vessel characteristics. Our focus will be overnight to two-day trips porting fish in San Diego, fishing from July 2-September 30, 2012[6]. Each weekend or set of connected weekdays is considered one time period[7].

To determine the price of the trip, we build a dataset from the Internet Archive and other sources[8]. In many cases, the prices for each vessel-trip combination exhibited low price variability, in which case we averaged the prices. In other cases, there were more nuanced price schemes, in which case we brought the temporal schemes into the data.

Vessels must specialize to some extent. Not all vessels can accommodate anglers seeking to catch HMS species because long-range trips require sleeping

[6]July 1 is excluded since it would be the only day in July in the corresponding weekend.

[7]Friday is considered part of the weekend. For the purposes of choice-set bundling, for 2 day trips, the weekday or weekend status is based off of the first day fished. For purposes of prices, in general, the price is a combination of the fishing days.

[8]The Internet Archive's Wayback Machine is a large database that crawls and stores websites as they appear during the crawl (archive.org). Most vessels in 2012 booked through websites. We attempted to use 2012 data to the extent possible; in some cases, we imputed 2012 prices from inflation-adjusted 2013 or 2014 prices.

quarters and larger fuel tanks. Furthermore, vessels may reduce passenger capacity for longer trips for passenger comfort or to take on additional supplies.

The length of the trip is one of the most important variables in terms of predicting fish catch: Shorter trips are different products than longer trips. Under normal circumstances, anglers interested in catching an HMS species look beyond day trips. For instance, HMS species are rarely caught aboard single day trips (Figure 1.1). However, these species make up a large percentage of fish catch for trips that stay out for longer periods of time.

This analysis covers 84% of reported overnight to 2 day trips that we determined were actually overnight to 2 day trips. Vessels were dropped from the data if they had no sleeping quarters or did not offer trips as described by this data subset. Vessels occasionally appeared as offering a 2 day trip due to an error filling out the form, a data entry error, or another error. For a small number of trips, no price information was available, and these vessels were dropped from the analysis. The actual price paid is not directly observed, so we assume that the final price paid was the posted price listed on the ticket for a typical trip or that specific trip on the vessel-trip-type combination. If the vessel mostly books charters, and we did not have individual price-level data, the typical charter rate divided by the number of people on the current trip is used as the price the individual faced in choosing between trip options.

1.6.2 Trip Choice Options and Sellouts

A sizable portion of overnight to two-day trip openings sell out. The marginal cost of adding an angler on a nearly full boat is quite small, so vessels can increase profits by making sure that trips are fully booked. The CPFV industry is heavily capitalized and very competitive.

The number of available spots on a vessel-trip combination are not recorded in the log-book data; as such, we need to make an inference about capacity using observable data. Fortunately, most of the boats involved are rated for their maximum capacity in terms of passengers for each trip type. We consider a vessel to be sold out if it is within two of its maximum capacity to allow for occasional no shows – and, because when there is only one or two spots available on a boat, many interested groups such as a husband and wife or a husband and wife and one child will be unable to book a trip on the vessel. We look at all of the boardings on the vessel-trip type combination. To eliminate a small number of outliers directly caused by data-entry (i.e., 100 instead of 10), we look at the highest boarding level for which more than 3% of trips on the trip type-vessel combination have greater than or equal to that boarding level.

The second stage of our analysis uses the predicted probability that the vessel was not available in choice sets faced by some anglers. This has the advantage that many vessels usually have a positive probability of not being available on a particular choice occasion. The possibility of last minute cancellations effectively makes and erroneous counts effectively make sellouts stochastic from the perspective of our analysis.

1.7 Preliminary Analysis

We first estimate model (1.3) *without the sellout adjustment* and calculate the WTP between types of trips in a naïve model. Let v_{ibrt} be the deterministic portion of the utility of customer i on a trip r on a vessel b during time period t.

$$v_{ibrt} = \beta_{price} p_r + \gamma_{1½} 1\{r \text{ is } 1½ \text{ day}\} + \gamma_2 1\{r \text{ is 2 day}\} \tag{1.6}$$

As shown in Table 1.3, the WTP for a 1½ day trip, given by $-\beta_{\text{price}}/\gamma_{1½}$ is \$138 over the overnight trip base, and the value of a two-day trip is \$257 over the overnight trip base. However, these prices include all components typical of a 1½ day or 2 day trip. In other words, we have not yet accounted for the possible role of HMS species catch.

Let's look at two specifications of the RUM – first without adjusting for sellouts. These two specifications are (with additional variables for the longer specification in parenthesis) below in equation (1.7). Let the proportion of HMS fish caught on vessel b for trip-type e during the entire season *before* time period t be pr_{bet}. If the average for the trip types in the previous year-vessel-trip type combination is not available, we use the average. In this case, we allow $h_{bet} = 1$ (otherwise 0). Let i represent an individual, r represent a trip, and bl_b equal the beam times the length of the vessel. The variable age is the age of the vessel, in years. Specification (1.7) is:

$$v_{ibert} = \sum_{\iota \in 1,1.5,2} \tau_\iota \mathrm{pr}_{bet} 1_{\iota\mathrm{day}} + \left(\sum_{\iota \in 1.5,2} \psi_\iota 1_{\iota\mathrm{day}} \right) + \beta_{price} p_r + \beta_h h_{bet} + \mathrm{bl}_b + \sum_{\iota \in 1,2} \mathrm{age}_b^\iota \tag{1.7}$$

As shown in Table 1.3, WTP is \$37, \$82, and \$232, respectively, for 100% catch on overnight, 1½ day, and 2 day trips. This means a change from 40% HMS to 50% HMS would have an additional WTP of \$3.71 on overnight trips. Once we take into account the type of trip taken, however, by adding the terms in parenthesis above, overnight trips get an additional WTP of \$81 for 100% HMS trips, and there is no additional WTP for 1½ day and 2 day trips. However, these trips already have an additional WTP of \$122 and \$252, respectively (over overnight trips) – these are the coefficients for $\psi_{1½}$ and ψ_2.

1.8 Analysis Accounting For Sellouts

In this section, we report on the specification *with a sellout correction*. Because of the sellouts, we run specification (1.7) through RUM (1.3) but with the sellout adjustment described in section 1.4, equation (1.4). Let M_m be a fixed effect for month m, E_e be a fixed effect for trip-type e, and V_b be a fixed effect for vessel b. The first-stage specification of this sellout model[9] is:

$$\mathrm{soldout}_{bemrt} = \delta_1 N_t + \alpha_1 V_b + \alpha_2 M_m + \alpha_3 E_e + \alpha_4 p_r E_e + \epsilon_{bemrt} + \left(\delta_2 N_t^2\right) \tag{1.8}$$

[9]Charters are not included in the first-stage regression and are considered an available choice in the second stage.

The term $\left(\delta_2 N_t^2\right)$ is only used in a robustness check, with the selected probability equal to $\widehat{\text{soldout}}_{ibemrt} - \delta_1 \hat{g}_i N_t - \delta_2 \left(\hat{g}_i N_t\right)^2$. The R^2 of the regression with the restriction $\delta_2 N_t^2 = 0$ is 0.28 and is reported in Table 1.4; very little additional explanatory power is gained by the square term. Additionally, the AIC is slightly lower for the restricted model.

The first draw is shown in Table 1.5, and the results of the the two specifications in (1.7) with adjustment (1.8) are shown in Figure 1.2. The medians are summarized in the table below, and they are compared to the WTP estimates in the standard model.

Table 1.1: Medians of Sellout Model vs. Standard WTP Model.

	Short Naïve Model	Short Sellout Model	Long Naïve Model	Long Sellout Model
Pr X Overnight	\$37	\$161	\$81	\$198
Pr X 1.5 Day	\$82	\$209	-\$19	\$52
Pr X 2 Day	\$232	\$434	-\$10	\$274
WTP 1.5 Day			\$122	\$166
WTP 2 Day			\$252	\$166

The medians for the runs on the short specification (1.7) are \$161, \$209, and \$434 for $-\tau_1/\beta_{price}$, $-\tau_{1½}/\beta_{price}$, and $-\tau_2/\beta_{price}$, respectively. In the naïve model, these estimations were \$37, \$82, and \$232. When we run the long specification (1.7), thus accounting for trip type, the medians change to \$198, \$52, and \$274, respectively. The \$52 is statistically non-different than zero; $-\tau_1/\beta_{price}$

and $-\tau_2/\beta_{price}$ are statistically different than zero. When we account for sell-outs as opposed to running the naïve model, the premium for the trip-type (over overnight trips) jumped from \$122 to \$166 for 1½ day trips and dropped from \$252 to \$166 for 2 day trips. We run a formal test of the sellout model, as contrasted to the naïve model, in section 1.8.2.

1.8.1 Interpretation of Coefficients

The coefficients τ_1, $\tau_{1\frac{1}{2}}$, and τ_2 in specification (1.7) tell us the additional utility of a trip that spends overnight, 1½ days, and 2 days on the water and spends 100% of the time targeting – and succeeding to catch – HMS species. First, 100% catch means something different for each type of trip. A 100% HMS target means more hours targeting HMS on a two day trip, which means more HMS fish. Thus, even if every vessel only targeted HMS fish, τ_2 would inherently have a different utility value – as it would represent a catch of more fish than τ_1.

When we take into account trip type, the value of 100% HMS catch goes down in 1½ and 2 day trips – but goes up for overnight trips. As shown in Figure 1.1, the targeting of HMS species is part of the journey and is likely part of the included value of \$170 and \$161. For overnight trips, catching a high proportion of HMS species is likely a strong signal that the vessel does well and is capable of catching prized fish.

The variable pr_{bet} has a few advantages over using just fish catch. First, crowding on vessels is correlated with number of customers (and thus WTP) –

but is negatively correlated with per-capita fish catch. Using a targeting metric such as pr_{bet} reduces these correlations – targeted fish will show up in the proportion regardless of how crowding impacts catch. Additionally, different vessels may target different angler skills – if we assume that catch is proportional only to angler skill and targeted location, then the appropriate measure is pr_{bet}.

1.8.2 Formal Test of Model

We combine the results of the naïve estimation with those of the soldout adjustment. First, we combine the RUM model results in a seemingly unrelated estimate framework. Denote the estimates on the coefficient for characteristic j in the naïve regression as α_j and in each of the draws i of the soldout adjustment as β_{ij}. We estimate the following statistic for characteristic j 1000 times – with standard errors from the delta method.

$$\frac{\beta_{ij}}{\beta_{price}} - \frac{\alpha_j}{\alpha_{price}} \tag{1.9}$$

The results of the draws are in Figure 1.3. All values are statistically different when we account for sellouts.

1.9 Conclusions

We have developed a predictive two-stage RUM model that controls for sellouts in the second stage RUM using first stage sellout predictions. We have demonstrated through a simulation exercise the left unaccounted for, sellouts can lead to biased RUM parameter estimates and the corresponding welfare measures. We have applied the two-stage model to analyze consumer choice for overnight to multi-day recreational fishing trips and report WTP estimates for increases is the catch proportion of preferred HMS fish that are significantly higher that those from the standard naive model.

While many activities experience frequent sellouts, these sellouts have often not been accounted for in empirical models used to estimate WTP. While recent advances have been made in the marketing literature, these advances are currently computationally burdensome for large choice set with a large number of sellouts. With no information on sellouts, the analyst may incorrectly assume that the consumer has chosen their utility maximizing choice, when in fact they may be choosing from a less-preferred subset of the choice set. Left unaccounted for, sellouts may lead to biased RUM parameter estimates and the corresponding welfare measures. In extreme cases, the sign of the estimated welfare measure may flip. This paper is the first we are aware of that tackles this issue in the resource economics context of nonmarket valuation. As welfare measure estimates are used frequently in resource management as a component of cost-benefit analysis or damage assessment, it is important to get them correct.

The method presented identifies a probabilistic model on sellouts that has several advantages over other alternatives. The proposed method works with both a large number of choice alternatives and a large number of sellouts, while the methods employed in the marketing literature have been limited to applications with a small number of infrequently changing sellouts (or “stock-outs”).

Additionally, the proposed method is computationally straightforward relative to competing models.

As an empirical application we model recreational sportfishing CPFV trip choice as a function of the proportion of catch that is comprised of high value HMS target fish while controlling for vessel and other trip characteristics. Estimates of the amount that consumers would be willing to pay for increases in the proportion of high value HMS species landed from the naïve model were generally lower than those from the proposed model which accounts for sellouts. WTP estimates in a naïve model were different from the sellout model by \$284 for a two-day trip only catching HMS species versus a two-day trip only catching non-HMS species (i.e., \$28.40 for a 10 percentage point increase). This large discrepancy can be explained by the fact that these vessels are very popular – so popular that they often sellout. Accounting for sellouts with the proposed two-stage model resulted in statistically significant differences in the parameter estimates from the naïve model and statistically different WTP estimates for changes in the catch proportion.

Areas for future research include measuring the comparisons of different methods of accounting for sellouts, and analyzing the relationship between the size of the sellout bias using the naive model and the size of the choice-set and the proportion of sellouts in the market. Additional research can analyze business structures to identify the underlying factors which characterize markets with a high proportion of sellouts.

1.10 Tables

Table 1.2: Willingness to Pay Estimates in the Naïve Model (Parameter Estimates). The next two tables show WTP for different types of trips, *without a sellout adjustment*. The first table shows parameter estimates. Columns (4) and (5) are numerically identical. **Pr not available described in text; if a Pr was not found, we used the average for Pr and a binary variable. The coefficient for the binary variable is shown. (WTP next page.)

	(1) RUM	(2) RUM	(3) RUM	(4) RUM	(5) RUM
One and One Half Day Binary	0.337***			0.229**	0.229**
	(17.90)			(3.25)	(3.25)
Two Day Binary	0.627***			0.474***	0.474***
	(23.55)			(4.87)	(4.87)
Price	-0.00244***	-0.00223***	-0.00174***	-0.00188***	-0.00188***
	(-33.44)	(-31.71)	(-23.02)	(-23.21)	(-23.21)
Pr X Overnight		-0.162***	0.0647	0.153**	
		(-3.91)	(1.48)	(3.25)	
Pr X 1.5 Day		0.254***	0.143***	-0.0354	-0.188*
		(8.56)	(4.66)	(-0.48)	(-2.17)
Pr X Two Day		0.558***	0.404***	-0.0190	-0.172
		(16.21)	(11.16)	(-0.19)	(-1.55)
Last Year Pr Not Available**		0.0697***	0.0981***	0.0915***	0.0915***
		(4.30)	(5.83)	(5.40)	(5.40)
Beam X Length			0.000648***	0.000648***	0.000648***
			(32.34)	(32.30)	(32.30)
Vessel Age			0.0116**	0.0138***	0.0138***
			(3.24)	(3.60)	(3.60)
Square Vessel Age			-0.000111**	-0.000136**	-0.000136**
			(-2.68)	(-3.08)	(-3.08)
Pr					0.153**
					(3.25)
N	707277	707277	707277	707277	707277
Pseudo R^2	0.0124	0.0122	0.0203	0.0206	0.0206
AIC	130179.1	130203.5	129141.3	129114.9	129114.9
Log Lik.	-65086.6	-65096.8	-64562.7	-64547.4	-64547.4

t statistics in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 1.3: Willingness to Pay Estimates Derived from the RUM on the Previous Page. Based on the random utility model in the basic specification, willingness to pay for a 1½ day trip is \$138, and willingness to pay for a 2 day trip is \$257 – over an overnight base. Columns (4) and (5) are numerically identical.

	(1)	**(2)**	**(3)**	**(4)**	**(5)**
WTP 1.5 Day	138.1***			121.9**	121.9**
	(18.90)			(3.28)	(3.28)
WTP 2 Day	256.8***			252.1***	252.1***
	(32.05)			(5.11)	(5.11)
Pr X Overnight		-72.57***	37.13	81.27**	
		(-3.92)	(1.47)	(3.23)	
Pr X 1.5 Day		114.0***	82.23***	-18.83	-18.83
		(8.53)	(4.73)	(-0.48)	(-0.48)
Pr X 2 Day		250.2***	231.8***	-10.09	-10.09
		(17.81)	(12.73)	(-0.19)	(-0.19)

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

Table 1.4: First Stage Sellout Model. This is the result of the first stage of the sellout model – in-text specification (1.8). Regressions include robust standard errors. The standard deviation of the forecast does use regular standard errors; however, both standard errors are very close, and the standard deviation of the forecast is larger than the robust alternative standard deviation of prediction. The table on the right includes the square of the numbers of anglers in the market.

	(1) **Soldout Binary**	**(2)** **Soldout Binary**
Number of Anglers in Market	0.000248	0.000565
	(1.88)	(1.06)
Sq. Number of Anglers in Market		-0.000000195
		(-0.61)
July Binary	-0.238**	-0.209*
	(-3.06)	(-2.26)
August Binary	-0.0553	-0.0498
	(-1.12)	(-0.99)
One and One Half Day Binary	-0.418	-0.421
	(-0.94)	(-0.95)
Two Day Binary	0.809	0.809
	(1.39)	(1.39)
Price X Overnight	-0.00163	-0.00159
	(-1.62)	(-1.56)
Price X 1.5 Day	0.000101	0.00015
	(0.06)	(0.09)
Price X 2 Day	-0.0022	-0.00219
	(-1.66)	(-1.64)
Constant	0.212	0.0649
	(0.84)	(0.18)
N	557	557
R2	0.277	0.277
AIC	681.9	683.5
Log Lik.	-303.9	-303.8

t statistics in parentheses
Vessel Fixed Effects Included
* p<0.05, ** p<0.01, *** p<0.001

Table 1.5: One draw of the soldout model. All draws (with confidence intervals) are available in Figure 1.2.

	RUM Coeff.	WTP	RUM Coeff.	WTP	RUM Coeff.	WTP
Price	-0.00274***		-0.00189***		-0.00200***	
	(-40.26)		(-25.31)		(-25.22)	
Pr X Overnight	-0.145***	-52.99***	0.306***	161.7***	0.397***	198.6***
	(-3.51)	(-3.53)	(6.85)	(6.44)	(8.29)	(7.87)
Pr X 1.5 Day	0.568***	207.1***	0.396***	209.5***	0.0938	46.92
	(19.17)	(17.93)	(12.84)	(12.25)	(1.26)	(1.25)
Pr X 2 Day	0.927***	338.2***	0.824***	435.5***	0.559***	279.6***
	(27.38)	(28.07)	(22.76)	(22.79)	(5.42)	(5.18)
Pr N/A, Used Avg	0.124***		0.235***		0.226***	
	(7.56)		(13.40)		(12.79)	
Beam X Length			0.00113***		0.00113***	
			(55.55)		(55.58)	
Vessel Age			0.0274***		0.0317***	
			(7.58)		(8.29)	
Sq. Vessel Age			-0.000303***		-0.000350***	
			(-7.29)		(-8.03)	
1.5 Day Binary					0.341***	170.4***
					(4.82)	(4.86)
2 Day Binary					0.325**	162.4***
					(3.26)	(3.37)
N	485919	485919	485919	485919	485919	485919
Psuedo R^2	0.0284		0.0548		0.0551	
AIC	115015.5	.	111897.9	.	111871.7	.
LL	-57502.7		-55941.0		-55925.9	

t statistics in parentheses
* p<0.05 ** p<0.01 *** p<0.001

1.11 Figures

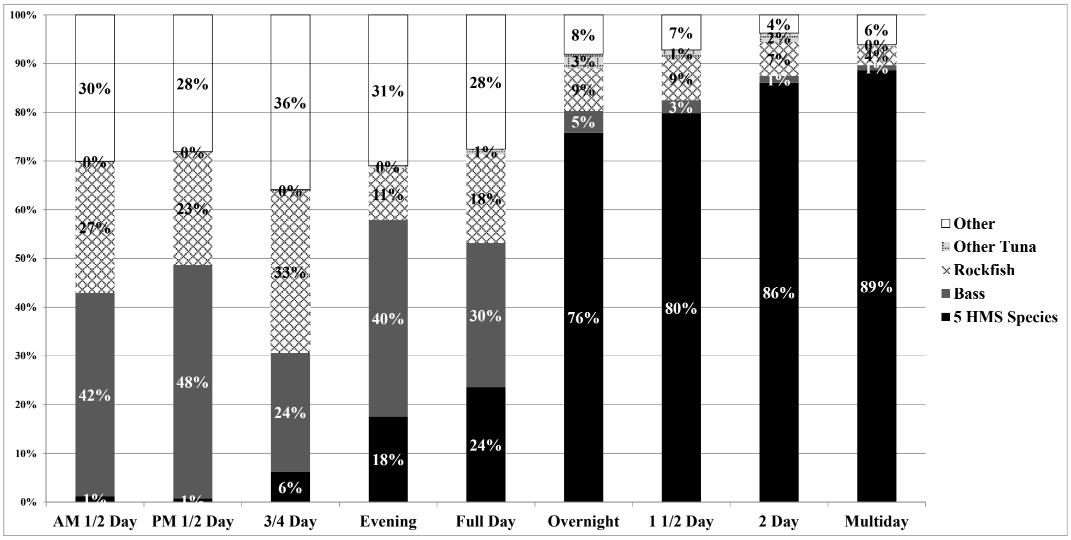


Figure 1.1: Trip Types – Fish Caught. Here, we illustrate the proportion within species group of fish caught by trip type. Shorter trips catch more rockfish and bass, while longer trips carry highly prized HMS species. (Other tuna includes bigeye, longtail, skipjack, blackskip jack, and unspecified.)

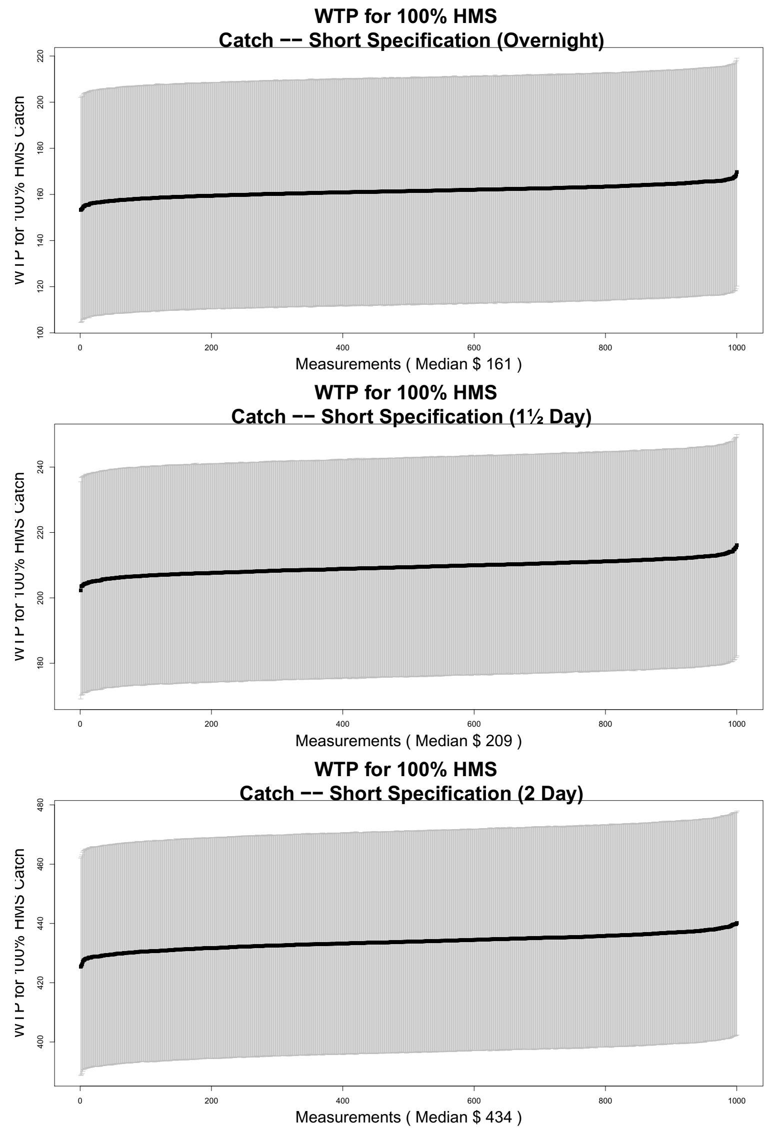


Figure 1.2: Figures of WTP with Sellouts. Above and on the following pages, we show the 1000 draws of specifications of the sellout model with long or short versions of in-text specification (1.7) and short specification (1.8).

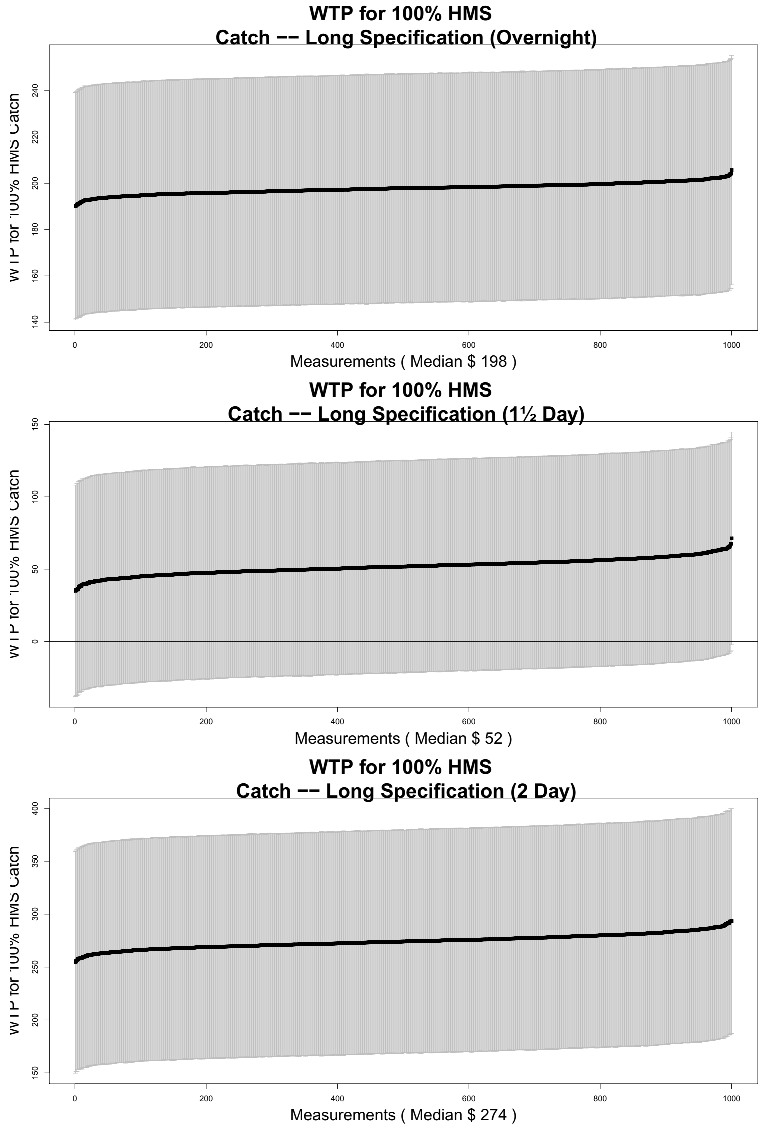
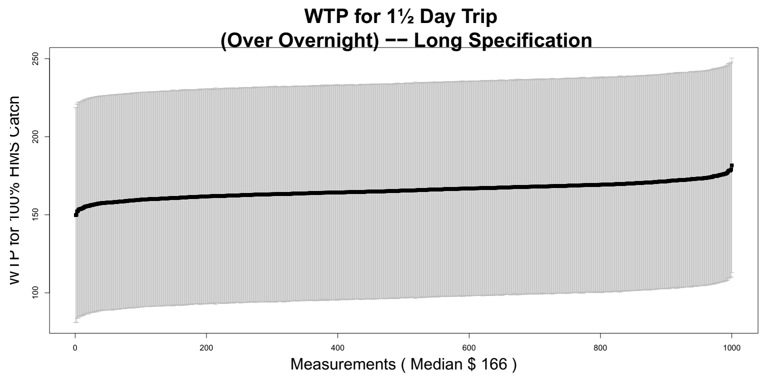


Figure 1.2: Figures of WTP with Sellouts, *continued.*



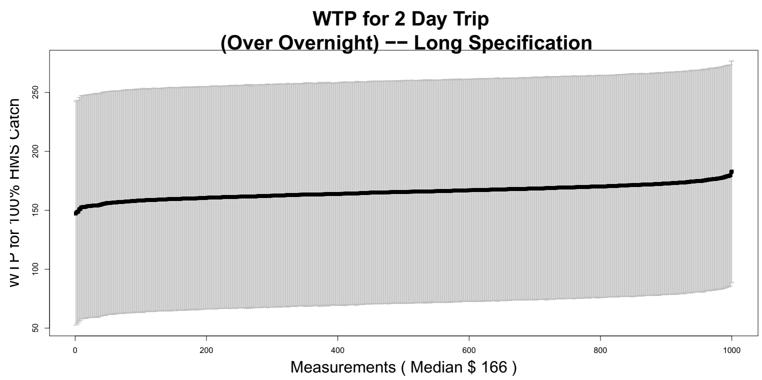


Figure 1.2: Figures of WTP with Sellouts, *continued.*

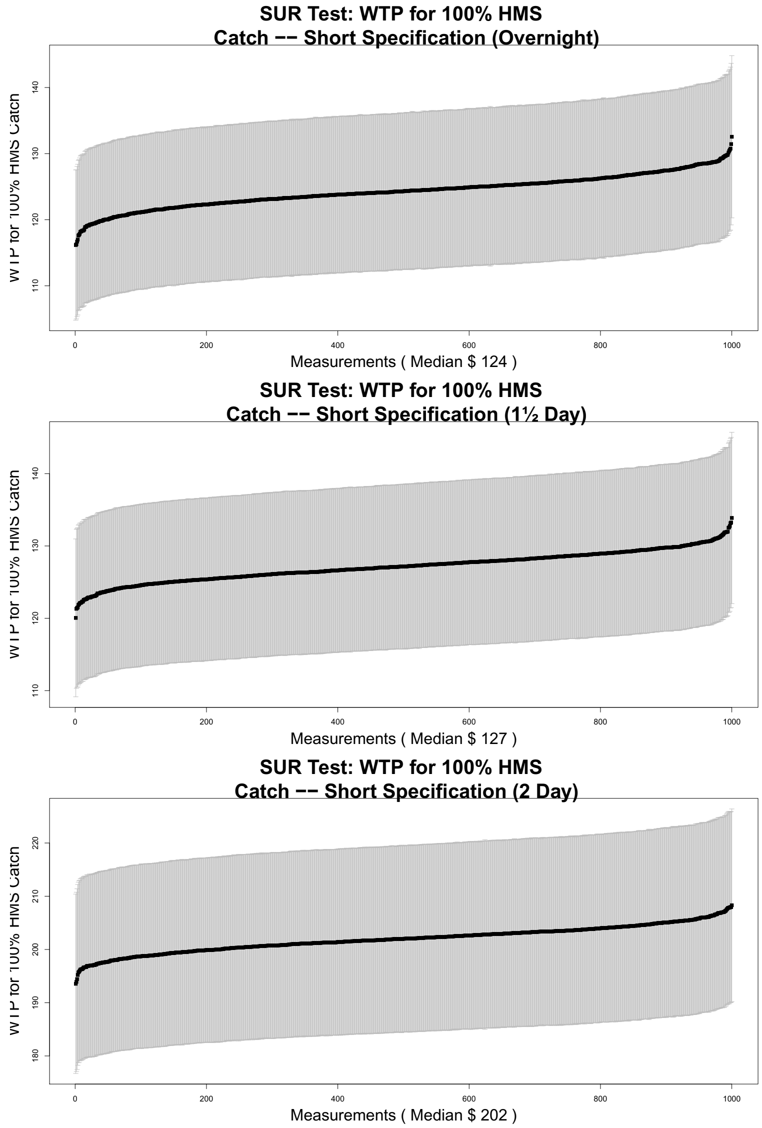


Figure 1.3: Formal Test of Model. Above and on the following pages, we show statistic (1.9) using long and short specification (1.7). All estimates are significant different from zero. We note that this process is different than testing whether the sellout model estimate is different than a particular number without standard deviation – for these estimates, see Figure 1.2.

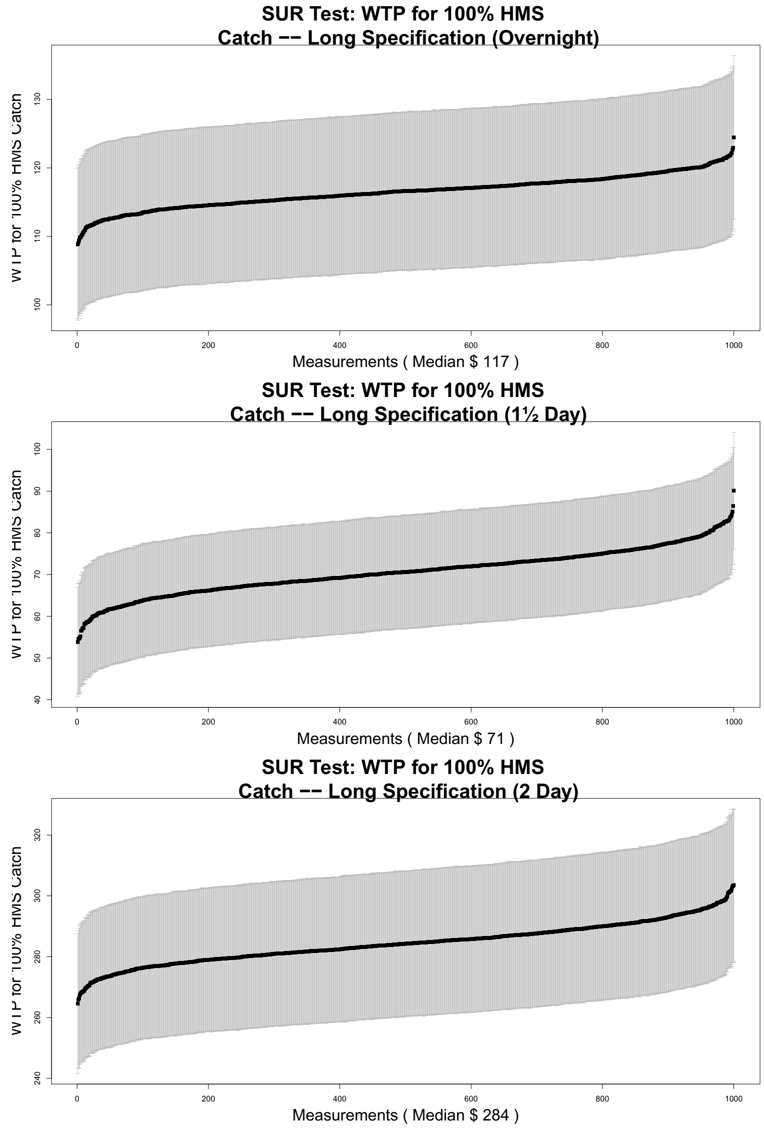


Figure 1.3: Formal Test of Model, *continued*.

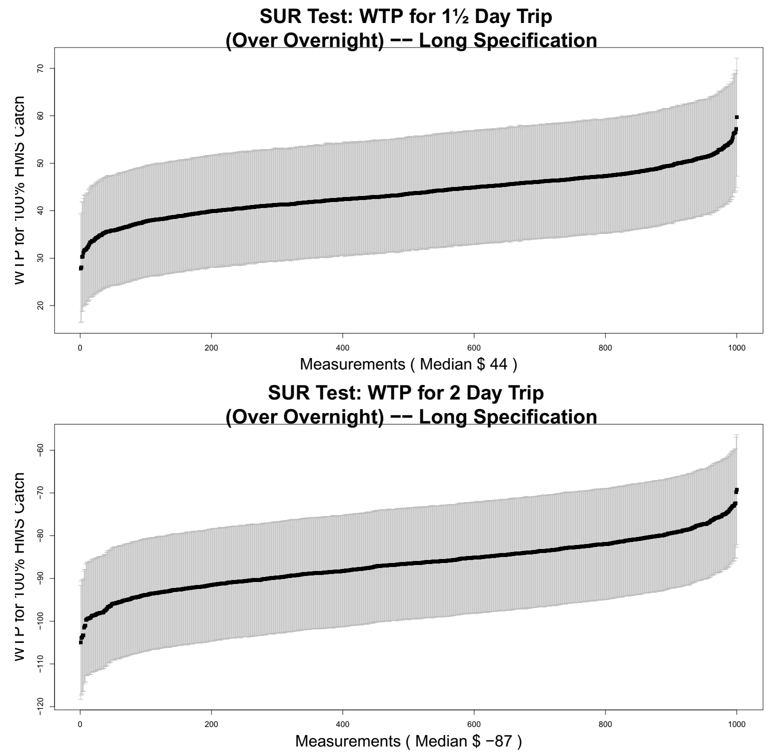


Figure 1.3: Formal Test of Model, *continued*.

1.12 Chapter 1 Appendix

1.12.1 Appendix

An alternative approach to approximating a sellout event would be to adjust for the sellout using fixed effects. Empirical justification follows.

1.12.1.1 Approximation of the Log Likelihood Function

In one respect, adding fixed effects for sellouts would be an approximation of the log-likelihood function. Following Conlon & Mortimer (2013), given $i = 1, \ldots, N$ possible choice set probabilities with probabilities $\alpha_1, \ldots, \alpha_N$ for a person who selects vessel q from vessels $v = 1, \ldots, V$, the log likelihood contribution is:

$$
\begin{aligned}
\ell_i &= \sum_{i=1}^{N} \alpha_i \log\left(\frac{\exp\left[\mathbf{X}_q'\boldsymbol{\beta}\right]}{\sum_{j\in i} \exp\left[\mathbf{X}_j'\boldsymbol{\beta}\right]} \right) \\
&= \sum_{i=1}^{N} \log\left(\frac{\exp\left[\mathbf{X}_q'\boldsymbol{\beta}\right]}{\sum_{j\in i} \exp\left[\mathbf{X}_j'\boldsymbol{\beta}\right]} \right)^{\alpha_i} \\
&= \log \prod_{i=1}^{N} \left(\frac{\exp\left[\mathbf{X}_q'\boldsymbol{\beta}\right]}{\sum_{j\in i} \exp\left[\mathbf{X}_j'\boldsymbol{\beta}\right]} \right)^{\alpha_i} \\
&= \log \prod_{i=1}^{N} \left[\frac{\left(\exp\left[\mathbf{X}_q'\boldsymbol{\beta}\right]\right)^{\alpha_i}}{\left(\sum_{j\in i} \exp\left[\mathbf{X}_j'\boldsymbol{\beta}\right]\right)^{\alpha_i}} \right]
\end{aligned}
$$

$$= \log \frac{\exp\left[\mathbf{X}_q'\boldsymbol{\beta}\right]}{\prod_{i=1}^{N}\left(\sum_{j\in i}\exp\left[\mathbf{X}_j'\boldsymbol{\beta}\right]\right)^{\alpha_i}}$$

The contribution to the denominator from each utility is linearly approximated. If the vessel does not sell out, it appears in each factor of the denominator, and will be approximated as equal to 1. Otherwise, it receives a formal correction term. Denote the set of possibly sold out vessels as S. Then, an approximation of the above equation comes to:

$$\begin{aligned}
\ell_i &= \log \frac{\exp\left[\mathbf{X}_q'\boldsymbol{\beta}\right]}{\prod_{i=1}^{N}\left(\sum_{j\in i}\exp\left[\mathbf{X}_j'\boldsymbol{\beta}\right]\right)^{\alpha_i}} \\
&\approx \log \frac{\exp\left[\mathbf{X}_q'\boldsymbol{\beta}\right]}{\sum_{w\in\{1,\ldots,V\}\setminus S}\exp\left[\mathbf{X}_w'\boldsymbol{\beta}\right] + \sum_{u\in S}\exp\left[\mathbf{X}_u'\boldsymbol{\beta}\right]e^{-\gamma_u}} \\
&= \frac{\exp\left[\mathbf{X}_q'\boldsymbol{\beta}\right]\prod_{u\in S}\exp\left(\gamma_u\right)}{\sum_{w\in\{1,\ldots,V\}\setminus S}\exp\left[\mathbf{X}_w'\boldsymbol{\beta}\right]\prod_{u\in S}\exp\left(\gamma_u\right) + \sum_{u\in S}\exp\left[\mathbf{X}_u'\boldsymbol{\beta}\right]\prod_{t\in S, t\neq u}\exp\left(\gamma_t\right)}
\end{aligned}$$

1.12.1.2 Allude to McFadden (1974)

There are many choices of vessels, $V_1, \ldots, V_N \in \mathbf{V}$. Each vessel has characteristics $\sigma_1, \ldots, \sigma_N \in \boldsymbol{\sigma}$. The *modified independence of irrelevant alternatives* states that:

$$\frac{P\left(V_2|\boldsymbol{\sigma}, \{V_1, V_2\}\right)}{P\left(V_1|\boldsymbol{\sigma}, \{V_1, V_2\}\right)} = \frac{\gamma_2}{\gamma_1}\frac{P\left(V_2|\boldsymbol{\sigma}, \mathbf{V}\right)}{P\left(V_1|\boldsymbol{\sigma}, \mathbf{V}\right)} \tag{1.10}$$

Here, $\gamma_2 < 1$ is a modification to correct for whether V_2 is in the choice set. Denote $P(V_m \mid \boldsymbol{\sigma}, \{V_m, V_n\})$ as p_{mn}. Then, rearranging the above equation,

$$P(V_2|\boldsymbol{\sigma},\mathbf{V}) = \frac{p_{21}}{p_{12}}\frac{\gamma_1}{\gamma_2}P(V_1|\boldsymbol{\sigma},\mathbf{V}) \tag{1.11}$$

Continuing with McFadden (1974), this can be arranged thusly, if we assume that there is a third vessel:

$$\begin{aligned}
P(V_1|\boldsymbol{\sigma},\mathbf{V})\sum_{V_i\in\mathbf{V}}\frac{p_{i1}}{p_{1i}}\frac{\gamma_1}{\gamma_i} &= 1\\
P(V_1|\boldsymbol{\sigma},\mathbf{V}) &= \frac{1}{\sum_{V_i\in\mathbf{V}}\frac{p_{i1}}{p_{1i}}\frac{\gamma_1}{\gamma_i}}\\
P(V_1|\boldsymbol{\sigma},\mathbf{V}) &= \frac{1}{\sum_{V_i\in\mathbf{V}}\frac{p_{i3}/p_{3i}}{p_{13}/p_{31}}\frac{\gamma_1}{\gamma_i}}\\
P(V_1|\boldsymbol{\sigma},\mathbf{V}) &= \frac{p_{13}/p_{31}}{\sum_{V_i\in\mathbf{V}}\frac{p_{i3}}{p_{3i}}\frac{\gamma_1}{\gamma_i}}
\end{aligned}$$

From the McFadden assumption of additive separability of the utility function and denoting utility appropriately, we have:

$$P(V_1|\boldsymbol{\sigma},\mathbf{V}) = \frac{\exp(U(V_1,\sigma_1))}{\sum_{V_i\in\mathbf{V}}\frac{\gamma_1}{\gamma_i}\exp(U(V_i,\sigma_i))} \tag{1.12}$$

The denominator here reflects the choice set availability. Denote the soldout vessels as the set $\mathbf{S}$. In the log likelihood function, the contribution of a person p who selects vessel p is:

$$\begin{aligned}
\ell_p &= \frac{\exp\left(U\left(V_p,\sigma_p\right)\right)}{\sum\limits_{V_i\in\mathbf{V}}\frac{\gamma_p}{\gamma_i}\exp\left(U\left(V_i,\sigma_i\right)\right)} \\
&= \frac{\exp\left(U\left(V_p,\sigma_p\right)\right)}{\sum\limits_{V_j\in\mathbf{V}\setminus\mathbf{S}}\exp\left(U\left(V_j,\sigma_j\right)\right)+\sum\limits_{V_i\in\mathbf{S}}\gamma_i^{-1}\exp\left(U\left(V_i,\sigma_i\right)\right)} \\
&= \frac{\exp\left(U\left(V_p,\sigma_p\right)\right)\prod\limits_{V_i\in\mathbf{S}}\gamma_i}{\sum\limits_{V_j\in\mathbf{V}\setminus\mathbf{S}}\exp\left(U\left(V_j,\sigma_j\right)\right)\prod\limits_{V_i\in\mathbf{S}}\gamma_i+\sum\limits_{V_i\in\mathbf{S}}\exp\left(U\left(V_i,\sigma_i\right)\right)\prod\limits_{V_k\in\mathbf{S}\setminus V_i}\gamma_k}
\end{aligned}$$

Let $U\left(V_i,\sigma_i\right)=\mathbf{X}_i'\boldsymbol{\beta}$, where $\mathbf{X}_i=\sigma_i$ are the characteristics of the vessel, and let $\gamma_i=e^{-\theta_i}$. Then, this equation becomes:

$$\frac{\exp\left(\mathbf{X}_p{}'\boldsymbol{\beta}-\sum\limits_{V_i\in\mathbf{S}}\theta_i\right)}{\sum\limits_{V_j\in\mathbf{V}\setminus\mathbf{S}}\exp\left(\mathbf{X}_j{}'\boldsymbol{\beta}-\sum\limits_{V_i\in\mathbf{S}}\theta_i\right)+\sum\limits_{V_i\in\mathbf{S}}\exp\left(\mathbf{X}_i{}'\boldsymbol{\beta}-\sum\limits_{V_k\in\mathbf{S}\setminus V_i}\theta_k\right)} \tag{1.13}$$

The issue regarding the approximation is that of the mathematical structure of the approximation. For those who do not have a vessel in their choice set, the approximation requires that the modification fixed effects term drives the sold out choice set to zero. This is a heavy burden. Let's denote the denominator as in Section 1.12.1.1:

$$\sum_{w\in\{1,\ldots,V\}\setminus S}\exp\left[\mathbf{X}_w'\boldsymbol{\beta}\right]\prod_{u\in S}\exp\left(\gamma_u\right)+\sum_{u\in S}\exp\left[\mathbf{X}_u'\boldsymbol{\beta}\right]\prod_{t\in S,t\neq u}\exp\left(\gamma_t\right) \tag{1.14}$$

Here, if we look at the likelihood in a probabilistic sense, as in Bronnenberg and Vanhonacker (1996), we could interpret γ as the log of the probability that someone "notices" a vessel on the market. However, with large number of sellouts, this has a distinct disadvantage... for each individual person, the ratio between the terms is infinite. For vessels that sell out regularly and right away, this will lead to bad behavior. Further, in this application, this approximation requires a large amount of power, and given the instability of the estimates, an estimation using this approximation was not successful.

1.12.2 Simulation

To look at the bias from taking this approach, we run a simulation.

Initially, $N_{init} = 30,000$ observations are made. A random draw from $N(0,1)$ is given for each observation, and, iteratively, from $1,\ldots,N_{init}$, a new trip is created if the PDF value is over 1.975. Then, trips are capped at 200 passengers. Trips below 15 passengers are dropped. Trips between 15-30 passengers are duplicated four times.

Four vessels, A, B, C, and D are available for passengers to choose. The spots on vessels B, C, and D are 200, but the spots available on A varied. The catch per unit effort (CPUE) and price (p) for trip t and vessel v is given by:

$$\begin{aligned} CPUE_{tv} &= A_v \alpha_t + e_{tv} \\ p_{vt} &= \rho_{vt} \end{aligned}$$

With the following parameters:

Simulation 1	**Simulation 2**
$A_A = 2.7$	$A_A = 3.1$
$A_B = 2.1$	$A_B = 2.7$
$A_C = 1.7$	$A_C = 2.5$
$A_D = 1.4$	$A_D = 2.3$
$\alpha_t = \max(0, k); k \sim N\left(1, .017^2\right)$	$\alpha_t = \max(0, k); k \sim N\left(1, .017^2\right)$
$e_{tv} = \max(0, j); j \sim N\left(0, .01^2\right)$	$e_{tv} = \max(0, j); j \sim N\left(0, .01^2\right)$
$\rho_{At} \sim N\left(1.2, .05^2\right)$	$\rho_{At} \sim N\left(1.2, .1^2\right)$
$\rho_{Bt} \sim N\left(1.1, .05^2\right)$	$\rho_{Bt} \sim N\left(1.1, .1^2\right)$
$\rho_{Ct} \sim N\left(1.0, .05^2\right)$	$\rho_{Ct} \sim N\left(1.0, .1^2\right)$
$\rho_{Dt} \sim N\left(0.9, .05^2\right)$	$\rho_{Dt} \sim N\left(0.9, .1^2\right)$

Customer i chooses vessel V if it is available and gives her the most utility out of all other available choices. Let $\epsilon_{tv} \sim$ Type I Extreme Value. The utility is given by:

Simulation 1	**Simulation 2**
$-2p_{vt} + 2CPUE_{tv} + 1.3\epsilon_{tv}$	$-0.5p_{vt} + 1.5CPUE_{tv} + \epsilon_{tv}$

The simulation goes through each individual, in queue order, and selects their most favored vessel. However, the spots on vessel A vary from 25 to 60. The probability of the vessel being soldout across individuals is recorded. We run the following estimations:

Naïve Estimation: The estimation run is a standard conditional logit model with no fixed effects.

Naïve Estimation with Fixed Effect: This estimation is run with a fixed effect for boat 1 to try to alleviate the sellout error.

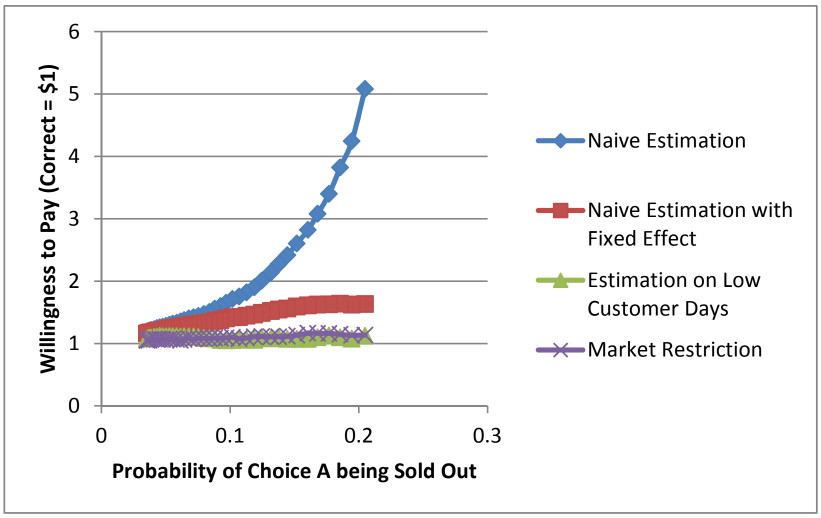
Estimation on Low Customer Days: This estimation is the correct estimation on days that are not sold out.

Market Restriction: We run the regression with modified independence of irrelevant alternatives (see attached). Since the distribution of the error on the utility may not be Type I Extreme Value in this case, there may be some theoretical difference in the estimate. However, the estimate does well. See below for some comparisons to traditional logit models. In this setup, we define the sellout modification γ to be determined by the size of the market (divided by a mean market measure, which WLOG is not necessary; for ease of programming, this mean was done across individual selection decisions and not across trips), whether there were no empty spots on vessel 1, and whether the market was larger than the number of spots on vessel 1. Thus, a conditional error term was added equal to:

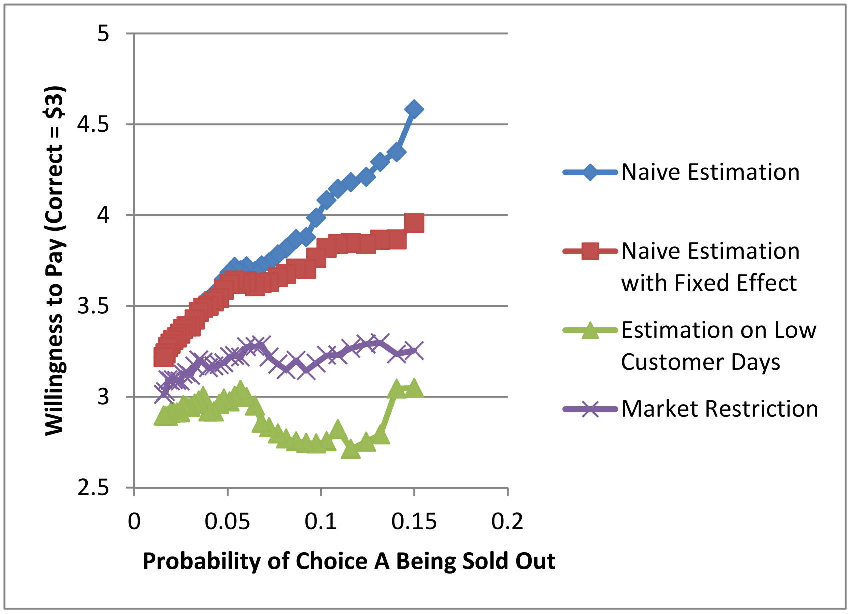
$$\frac{\text{Total Anglers}_t}{\overline{TA}} \times 1\{\text{Total Anglers that Day} > \text{Total Spots Boat 1}\}$$
$$\times 1\{\text{Vessel A Sold Out}\} \times 0\{\text{Vessel 1}\}$$

1.12.2.1 Simulation Results

The following chart plots the probability of being sold out against the estimate of WTP for the four methods under simulation 1. The correct WTP estimate is $1.



The following chart plots the probability of being sold out against the estimate of WTP for the four methods under simulation 2. The correct WTP estimate is $3.



1.13 Acknowledgments for "Hitting Capacity: Implications for the Valuation of Outdoor Recreation"

This paper is co-authored with James Hilger (National Oceanic and Atmospheric Administration). I would like to thank James Hilger for giving me the opportunity to work at the National Oceanic and Atmospheric Administration, where I had access to the data necessary for this chapter. It is being presented at the Association of Environmental and Resource Economists conference in San Diego, 2015.

I would like to thank Richard Carson, Mark Jacobsen, Yixiao Sun, Jeffrey Shrader, Johnathan Sweeney, two preliminary reviewers at NOAA, and participants at the Center for Environmental Economics at UCSD seminar series. This paper is the work of the authors and not the opinion of the National Oceanic and Atmospheric Administration.

1.14 Bibliography

Boyle, K., R. Bishop, J. Caudill, J. Charbonneau, D. Larson, M.A. Markowski, R.E. Unsworth, and R.W. Paterson. 1998. A Database of Sportfishing Values. Economics Division – Fish and Wildlife Service, Department of the Interior.

Bronnenberg, B.J., and W.R. Vanhonacker. 1996. "Limited Choice Sets, Local Price Response and Implied Measures of Price Competition." *Journal of Marketing Research* 11:163-173.

Carson, R.T., W.M. Hanemann, and T.C. Wegge. 2009. "A Nested Logit Model of Recreational Fishing Demand in Alaska." *Marine Resource Economics* 24:101-129.

Carter, D.W., and C. Liese. 2010. "Hedonic Valuation of Sportfishing Harvest." *Marine Resource Economics* 25:391-407.

Conlon, C.T., and J.H. Mortimer. 2013. "Demand Estimation Under Incomplete Product Availability." *American Economic Journal: Microeconomics* 5(4):1-30.

Fox, J.T. 2007. "Semiparametric Estimation of Multinomial Discrete-Choice Models Using a Subset of Choices." *RAND Journal of Economics* 38:1002-1019.

Haab, T.C., and R.L. Hicks. 1999. "Choice Set Considerations in Models of Recreation Demand: History and Current State of the Art." *Marine Resource Economics* 14:271-81.

Haab, T., R. Hicks, K. Schnier, and J.C. Whitehead. 2012. "Angler Heterogeneity and the Species-Specific Demand for Marine Recreational Fishing." *Marine Resource Economics* 27:229-251.

Hauber, A.B., and G.R. Parsons. 2000. "The Effect of Nesting Structure Specification on Welfare Estimation in a Random Utility Model of Recreation Demand: An Application to the Demand for Recreational Fishing." *American Journal of Agricultural Economics* 82:501-514.

Hilger, J. 2014. "2013 California Marine Recreational Fishing Trip Effort and Economic Impact Estimates". La Jolla, CA: National Marine Fisheries Service, Southwest Fisheries Science Center.

Hilger, J.R., and J.R. Sweeney. 2013. "CA CPFV Skipper Logbook Trip Type Classification: 1995-2012." La Jolla, CA.

Johnston, R.J., M.H. Ranson, E.Y. Besedin, and E.C. Helm. 2006. "What Determines Willingness to Pay per Fish? A Meta-Analysis of Recreational Fishing Values." *Marine Resource Economics* 21:1-32.

Larson, D.M., and D.K. Lew. 2013. "How Do Harvest Rates Affect Angler Trip Patterns?" *Marine Resource Economics* 28:155-173.

McConnell, K.E. 1977. "Congestion and Willingness to Pay: A Study of Beach Use." *Land Economics* 53:185-95.

McConnell, K.E., and I.E. Strand. 1999. "Overnight Trip Choice for Marine Anglers Report 40ANF804203." Available Online through Internet Archive. https://web.archive.org/web/20060929190413/http://www.st.nmfs.gov/st1/econ/mcconnellandstrand.pdf.

McFadden, D. 1974. "Conditional Logit Analysis of Qualitative Choice Behavior." *Frontiers in Econometrics* 105-142.

Musalem, A., M. Olivares, E.T. Bradlow, C. Terwiesch, and D. Corsten. 2010. "Structural Estimation of the Effect of Out-of-Stocks." *Management Science* 56:1180-1197.

National Marine Fisheries Service. *Fisheries Economics of the United States 2012*. (NOAA Technical Memorandum NMFS-F/SPO-137, February 2014).

National Park Service Campground Reservations – Yosemite National Park; Campgrounds – Grand Canyon National Park; Fees & Reservations – Yosemite National Park, http://www.nps.gov/yose/planyourvisit/camping.htm; http://www.nps.gov/grca/planyourvisit/campgrounds.htm; http://www.nps.gov/yose/planyourvisit/feesandreservations.htm

Shaw, W.D., and M.T. Ozog. 1999. "Modeling Overnight Recreation Trip Choice: Application of a Repeated Nested Multinomial Logit Model." *Environmental and Resource Economics* 13:397-414.

Chapter 2

An Analysis of the Cost of an Undergraduate Degree and the Incentives of the State, the University, and the Student

2.1 Chapter Abstract

To expand undergraduate enrollments or to make decisions regarding rule changes for degrees, administrators need information on how much expansions and contractions in each department cost. This paper presents several methods of accounting for per-credit hour cost across departments. Using internal data from UCSD, we find that most social sciences are relatively cheap and engineering is relatively expensive.

This paper then simulates the university's allocation of funding to undergraduate departments and the student response. We find that a university with

static undergraduate fund-per-student preferences will allocate funds-per-student away from departments with large number of students to discourage them from majoring in those departments and instead majoring in a less-filled field. Using data from UCSD, we show that departments with large numbers of graduates are cheaper per degree, have higher modified student-to-faculty ratios, and graduate *sooner* than their colleagues in a different program at the university.

2.2 Introduction

2.2.1 Instruction-Related Cost per Degree

Decades of higher-than CPI education inflation in higher education have led administrators and policy makers to look inward at cost. For bachelors degrees, which are the focus of much of the policy debate on the cost of education, we can measure many aspects of instruction-related costs. The first is instruction-related cost. The other two other large categories would include are largely fixed costs: first, the cost of running the university, such as building maintenance, libraries, police, and utilities. The second is contract and grant related research. Our goal in this study is to determine how much money it costs *to add students* to the university; thus, we are concerned with instruction related costs.

Few studies attempt to deconstruct cost on a department basis (an attempt is found in Johnson (2009)). However, policy makers have discussed or hypothesized about the cost of different departments. These statements have often generated controversy, as did one made by former University of California

President James Yudolf:

> Many if our, if I can put it this way, businesses, are in good shape. We're doing very well there. Our hospitals are full, our medical businesses, our medical research, the patient care. So, we have this core problem: Who is going to pay the salary of the English department? We have to have it. Who's going to pay it in sociology, in the humanities? And that's where we're running into trouble (Michels 2009).

Yudolf may not have been engendering any sympathy from the humanities. He mentions a system, the Responsibly Controlled Management System (RCM), which showed that the humanities were among the most efficient users of university money. Watson implies in the article that the RCM was mysteriously abandoned when it found that Yudolf's "businesses" were using the system least effectively.

2.2.2 Funding

Higher education in the United States is funded at the individual and household, state, and federal levels. In 2008, the U.S. spent a combined 2.7% of its GDP on tertiary education, which was virtually unchanged since 2000 (OECD 2011). What had changed, however, was the level of student indebtedness, the direct cost of education to the students, and the late-2000's recession. Student debt in the first quarter of 2007 amounted to $363 billion; seven years later, this had ballooned to $920 billion (NY Fed).

Part of the problem in public education is state funding; state funding for the University of California fell from an inflation-adjusted per-student outlay of

$16,430 in 1990-91, to $8,220 in 2010-11 (UC Office of the President 2011). Increasing cost and decreasing state funding has led universities to raise tuition. Tuition increases, however, leverage students' future earnings and are unsustainable in the long-run. As an example, likely in response to worse hiring outcomes and higher law school tuition, those taking an LSAT dropped dramatically from 2009-10 to 2011-12 – 25%, or around 42,000 test takers (Segal 2012).

This paper will model resources allocated to students within a university. Students themselves are both a resource and a cost. Bound and Turner (2007) find that states that have abnormally large college cohorts in comparison to other states that year graduate *less* students. Further, large universities experience very little outflow to other schools; more commonly, students change majors. A university can be seen to be a customer-locked institution (Arcidiacono 2004).

2.2.2.1 Funding: Differential Tuition

States, however, also face large budget shortfalls and are questioning the wisdom of funding higher education – and what to fund. Some majors take longer to complete (Babcock and Marks 2011), and some majors earn more than others as well, leading to *potentially* higher net tax revenues. California has provided subsidies on a per-student basis, but other states are exploring providing *more* funding for students who major in STEM fields (science, technology, engineering, and mathematics). Florida considered stopping increases in STEM fields while allowing non-STEM tuition to increase (Webley 2013). This is clearly controversial.

Differential tuition has only recently been in the direction of encouraging enrollment in more expensive STEM majors (Webley 2013, Carter and Curry 2011). This is because the actual delivery system of educating these majors is (perhaps) higher. However, it is not clear that this policy is rational for the state that is funding public institutions. Particularly in the light of same per-student allocations, there is little, besides salary, in the university-student relationship to encourage study of high paying majors. Could it be rational for the state to fund STEM majors at a higher rate?

Recent work on differential tuition has shown that students are price conscious of tuition; however, the elasticity of this response in respect to enrollment ratios is controversial. Carter and Curry (2011) attribute this to the setup of various studies. In their analysis, they look at an individual's choice to major in a particular major instead of a cross-school analysis.

2.3 Measuring the Cost

In the first part of this study, we quantify the resources needed to train additional undergraduate students in each particular field. Our hope here is to provide a summary statistic that encompasses small (but not *infinitely small*) desires to increase undergraduate enrollments in particular departments, holding the essential departmental characteristics intact. Here, we assume the additional students in these departments will provide little additional cost to the broadly cost-distributed university goods, such as libraries, gyms, roads, etc.

One particularly challenging problem is that of the marginal student. We are not calculating the cost of the marginal student; her contribution to a large department is essentially zero, but not quite. The theoretical underpinnings of adding one additional student to a department are outlined in Hoenack, et.al. (1986). In the case of dropping an additional student in a lecture, the only marginal cost is the cost of meeting with that particular student, grading that particular student's exams, and any other minor administrative cost which is directly attributed to that student (such as the small burden of record keeping, etc.). This marginal cost is clearly small.

Continuing with Hoenack, et.al., once lectures need to be increased, however, the university can operate through several channels. They can have instructors repeat a course twice, which raises the cost, but not quite at the level of adding a completely new instructor. They can lower the marginal cost by ensuring that the faculty member teaches a subject they enjoy teaching.

The university can also raise teaching loads and decrease research and other activities. This essentially is a work increase and, equivalently, a pay decrease, as faculty will still have to be competitive in the same fields they were before the increase. As noted in Nelson and Hevert (1992), and applicable to our study.

> It would be inappropriate to assume that one could reduce the marginal costs to their allocated level by cutting faculty salaries or increasing teaching loads in proportion to the percentage of time allocated to research. Many faculty view the ability to do research as partial compensation for relatively low academic salaries: salaries would thus have to increase to attract sufficient numbers of faculty to positions with higher teaching loads.

Our analysis works at a more extensive margin. In this paper, we assume that the university wants to expand a department to accommodate a large number of students – but not so large that the university needs to add cost-distributed resources. An assumption that makes it possible to calculate this cost is that departments could scale up to teach a new block of students at its current average cost.

2.3.1 Cost Simulation

We create a simulation to explain how undergraduate spending on *instruction*. As mentioned, the university is a customer-locked institution (Arcidiacono 2004). In this simulation, we look at the attractiveness of majors, we lock students in the university, and we look at the post-admitted behavior of students after the university assigns funding.

In terms of undergraduate education, assume that the university is only concerned about the funds it allocates per student to each department; these funds per student are a measure of the amount of faculty that the department can hire per student, the quality of instruction it provides, etc. In this scenario, a dynamic relationship exists between the students and the university. Because the university allocates funds based on students, the students are actually interpreted by the university as a "cost." The university can manipulate this cost by increasing or decreasing the funds and making different majors more or less attractive.

The results show that higher salaries and fun majors actually lead to *less* funds per student, not more. Students pack these majors, and the university finds less costly majors to increase the quality and prestige of the undergraduate program at the university. Furthermore, it costs more to lure students away from majors with high salary or high utility.

2.3.1.1 Students

Students file into majors based on salary, funds-per-student, and likeability of majors. Denote students $s = 1, \ldots, S$ as choosing a major from a variety of majors, $m = 1, \ldots, M$. Denote year 1 as the year in which students enter, and all actors discount years 25 and further infinitely. Each major comes with an average time to graduation, G_m. The tuition payment per year is p_{tm}. Denote γ_m as the exogenous non-salary appeal of each major. Also denote the exogenous salary over the career years $G_m + 1$ to 24 as w_{tm}. Each student has an exogenous discount rate of r_s, and $\beta_{r_s} = \frac{1}{1+r_s}$. The university provides funds f_m to each major, and the number of students in each major is denoted S_m. Each major has an exogenous inefficiency factor of σ_m. This σ_m maps funds-to-students to effective quality of instruction. For instance, a major may have a high cost to teach – this implies a large σ_m. The quality of instruction is denoted by $\rho_m = \frac{f_m}{\sigma_m S_m}$. Here, ρ_m is a modified funds-per student ratio, which tells us the amount of instructional funds allocated to a student – but adjusted for cost of instruction. The mean-adjusted salary of students is:

$$w_m = \frac{1}{\bar{w}} \left[\left(\sum_{t=G_m+1}^{24} \beta_{r_s}^t w_{tm} \right) - \left(\sum_{t=1}^{G_m} \beta_{r_s}^t p_{tm} \right) \right] \tag{2.1}$$

Here, $\bar{w}$ is the average discounted income stream. The *proportion of* *students* in each major i given a funding allocation is given by:

$$S_i|\mathbf{f} = \frac{S}{\sum_{m=1}^{M}\left[\frac{\gamma_m}{\gamma_i}\right]^{\frac{b_1}{1+b_2}}\left[\frac{w_m}{w_i}\right]^{\frac{b_3}{1+b_2}}\left[\frac{\sigma_m}{\sigma_i}\right]^{-\frac{b_2}{1+b_2}}\left[\frac{f_m}{f_i}\right]^{\frac{b_2}{1+b_2}}} \tag{2.2}$$

We derive equation 2.2 in section 2.6.2.3. This proportion is similar to a multinomial logit framework, with the caveat that we assume that the proportions are given as in this equation instead of idiosyncratic. The equivalent utility function for each student is $b_1 \log \gamma_m + b_2 \log \rho_m + b_3 \log w_m$. Here, b_1, b_2, and b_3 are exogenous.

2.3.1.2 The University

We assume that a benevolent administrator will want to allocate cost-adjusted per-student funds relatively equally across departments so that the quality of instruction is as equal as possible across the university, subject to perhaps a known, exogenous administration preference parameter built into her utility function. The university administrator decides to spend $\xi < 1$ of its funds $f_1 + \ldots + f_M = F$ on undergraduate instruction, and $1-\xi$ on other utility enhancing services, such as administration, consistent with Cobb-Douglas utility. For now, assume that F is given. The total quality of undergraduate departments is given by:

$$Q(\rho) = \left(\sum_{m=1}^{M} \alpha_i \left[\frac{f_i}{\sigma_i (S_i|\mathbf{f})}\right]^r\right)^{\frac{1}{r}} \tag{2.3}$$
$$= \frac{F}{S}\left(\sum_{m=1}^{M} \alpha_i \left[\frac{\frac{f_i}{F}}{\sigma_i \frac{(S_i|\mathbf{f})}{S}}\right]^r\right)^{\frac{1}{r}}$$
$$= \frac{F}{S}\left(\sum_{m=1}^{M} \alpha_i \left[\frac{\widetilde{f}_i}{\sigma_i \left(\widetilde{S}_i|\mathbf{f}\right)}\right]^r\right)^{\frac{1}{r}}$$

Here, $\widetilde{f}_i$ and $\widetilde{S}_i$ are the fractions of funds and students (respectively) in each major. The outer $1/r$ is not relevant to the maximization. The price of Q can be manipulated through student filing and the fraction of funds devoted to each major. The price of Q can be given by:

$$F = Q \frac{S}{\left(\sum_{m=1}^{M} \alpha_i \left[\frac{f_i}{\sigma_i (S_i|\mathbf{f})}\right]^r\right)^{\frac{1}{r}}}; p_Q = \frac{S}{\left(\sum_{m=1}^{M} \alpha_i \left[\frac{f_i}{\sigma_i (S_i|\mathbf{f})}\right]^r\right)^{\frac{1}{r}}} \tag{2.4}$$

Additionally, we start with the assumption that there is no tuition or state-funding differential. This implies that the only relevant part of the maximization problem is maximizing $Q(\rho)$, and because F and S are endogenous, only maximizing the utility over the fractional students and funds. We thus have a bounded problem and will perform a maximization search.

Let $M = 3$, $\mathbf{b} = (1, 0.5, 3)$, and $r = 0.25$. Let total funds for undergraduates equal \$500 million, and let the number of undergraduates equal 30,000 (\$16,667 per degree). For simplicity, call the majors economics, engineering, and psychology (in that order, $m = 1,2,3$). We will vary $\boldsymbol{\alpha}$, $\boldsymbol{\gamma}$, and $\mathbf{w}$ to show how differently the university responds. This procedure is run using the Nedler-Mead simplex algorithm for local minima, with a grid search of initial points (0.01-0.97 per major). The local maximum are compared, and the global maximum is selected. We ignore behavior at the far extremes where one major is incredibly unattractive to students such that it is not a viable major.

Assume $\sum \alpha_i = 1$ and $\bar{w} = 1$. Here, letting $\sum \alpha_i = 1$ will allow the utility to be a standard weighted average, and allowing the average income stream, $\bar{w} = 1$, will make interpretation easier as well. Here is a table of initial values; we will systematically vary these throught the simulation:

	$\boldsymbol{\alpha}$	$\boldsymbol{\sigma}$	$\boldsymbol{\gamma}$	$\mathbf{w}$
Economics	0.33	1.5	0.33	1.0
Engineering	0.50	3.0	0.17	1.2
Psychology	0.17	1.0	0.50	0.8

- University Preference

First, lets vary $\boldsymbol{\alpha}$. We set $\alpha_3 = 1/6$, and we vary α_1 from 0.001 to 0.83. For engineering, α_2 is simply $1 - \alpha_1 - \alpha_3$. Firstly, as α_1 increases, funding in engineering drops and funding in economics increases. Funding in psychology is not constant, even though α_3 is constant.

As funds are shifted into economics, more students want to major in economics, eroding some of the increased funds. The modified funds per student ratio, ρ_1, also increases. As α_2 dips, ρ_2 goes down as well.

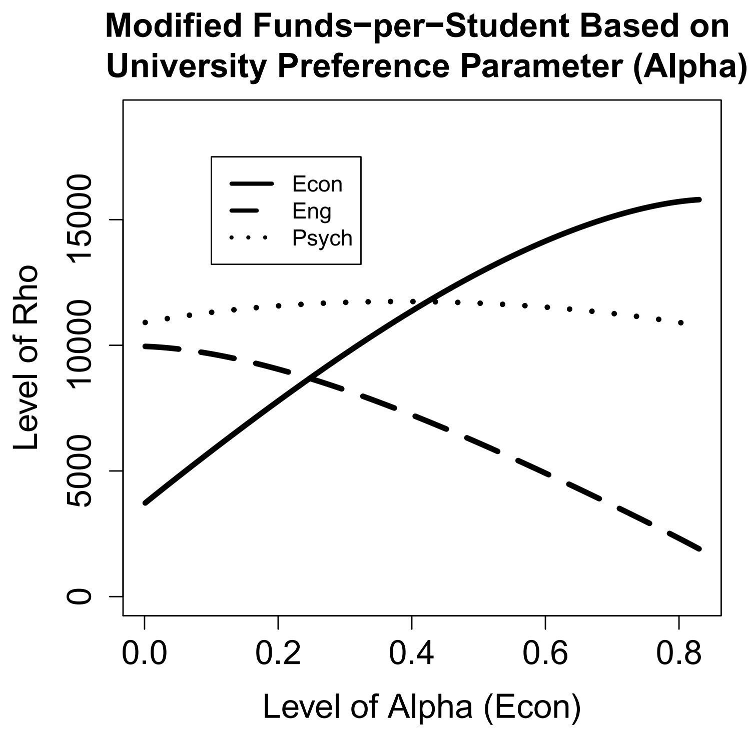


Figure 2.1: The University Preference Parameter and Modified Funds-Per-Student. The university's department preference parameter on economics, α_1, is varied from 0.001 to 0.83, and the engineering parameter, α_2, changes in tandem so that $\alpha_1 + \alpha_2 + \alpha_3 = 1$ with fixed $\alpha_3 = \frac{1}{6}$. In the picture, we see the effective funds per student in economics, $\rho_i = \frac{f_i}{S_i \sigma_i}$, increases in tandem with its relative preference parameter. This happens even given given the larger number of students in economics (see Figure 2.2).

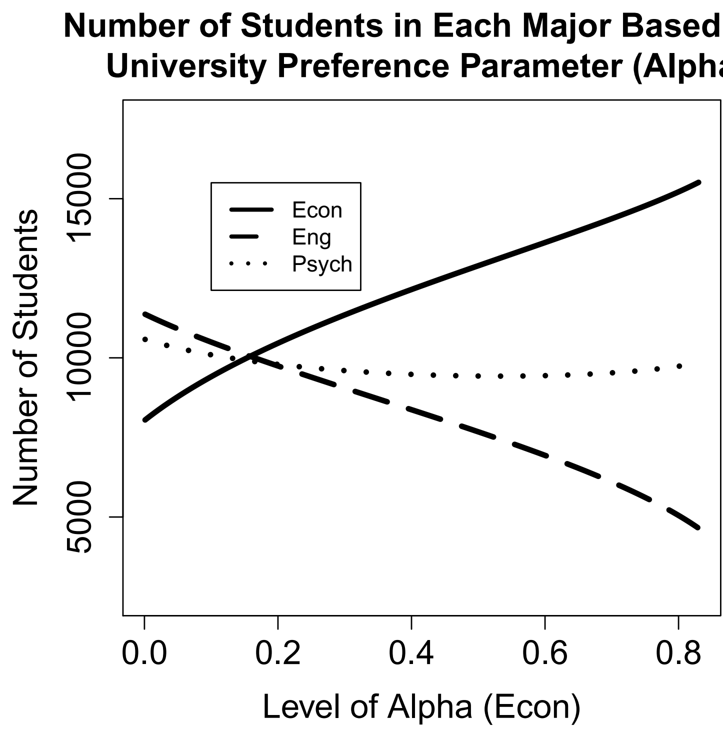


Figure 2.2: The University Preference Parameter and Student Major Choices. The university's department preference parameter on economics, α_1, is changed from 0.001 to 0.83, and the engineering parameter, α_2, is also changed so that $\alpha_1 + \alpha_2 + \alpha_3 = 1$ and $\alpha_3 = \frac{1}{6}$. In the picture, we see the number of students in economics increases as the university begins increasing effective per-student funds into this department.

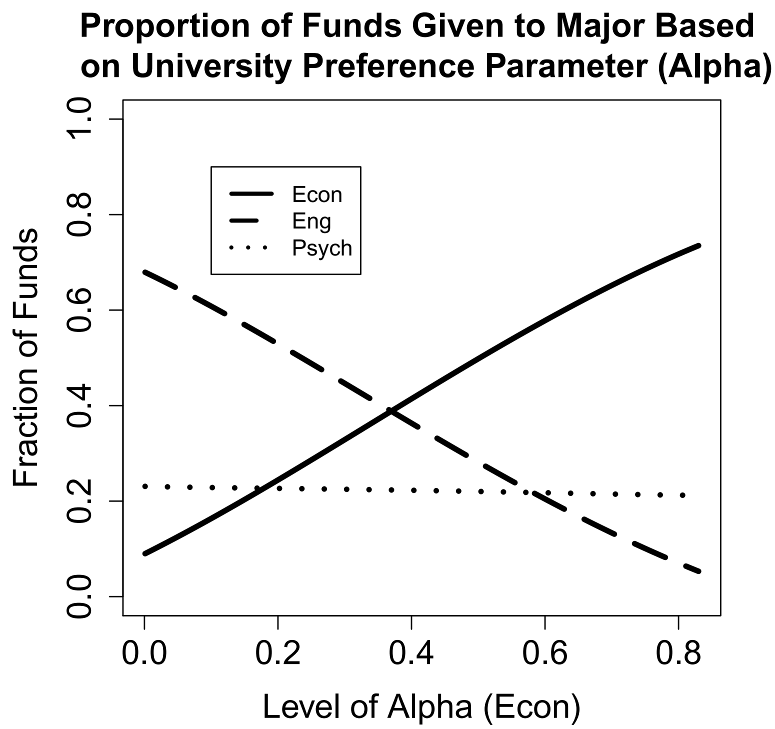


Figure 2.3: The University Preference Parameter and Funds Given to Each Department. The university's department preference parameter on economics, α_1, is changed from 0.001 to 0.83, and the engineering parameter, α_2, is also changed so that $\alpha_1 + \alpha_2 + \alpha_3 = 1$ and $\alpha_3 = \frac{1}{6}$. In the picture, we see, unsurprisingly, the funds in engineering drop and the funds for economics rise.

- Cost of Education Parameter

Next, we vary the cost of delivering effective undergraduate economics education. In this simulation, we vary σ_1, but we do not alter σ_2 or σ_3. As the cost increases, the university decides to change the funding proportions *out* of economics and into the other two majors. In response, students leave economics, and, in this simulation, not enough students leave to increase the modified funds-per-student ratio. The increased cost also has a negative effect on engineering and psychology; increased costs are bad for all majors.

One important note should be made in this section: Often, high costs are associated with high salaries for students after school. This is an important point. Most of the discussion around higher paid professors are around cost – but this simulation shows that this association both results in fewer majors through the cost channel and more majors through the salary channel. The end result will depend on the sum of these effects (and hence underlying parameters). But methods that universities use to control costs could result in students shifting out of high-paying departments.

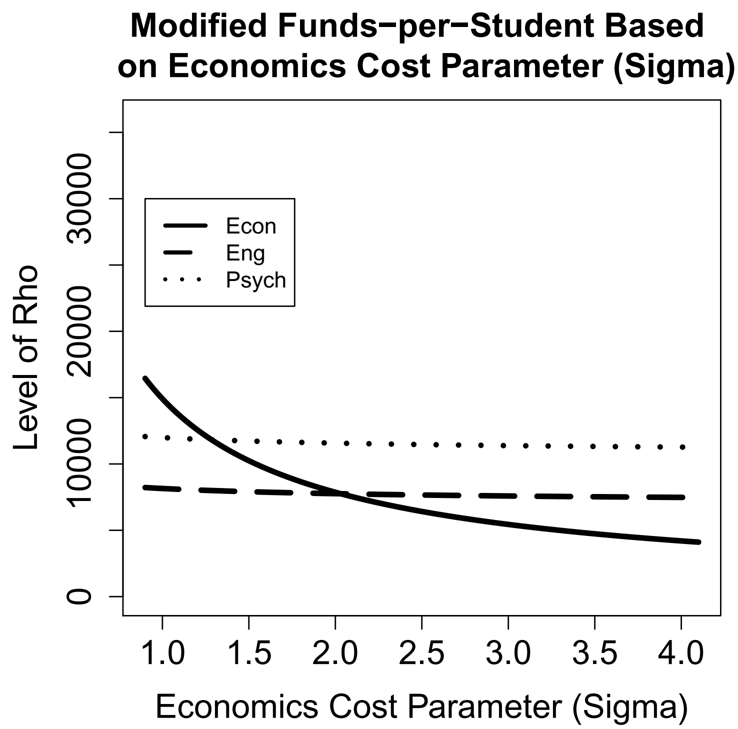


Figure 2.4: Modified Funds-per-Student Based on Economics Cost Parameter (Sigma). As σ_1 (economics) increases, the effective funds per students drop – in all majors but in particular, in economics.

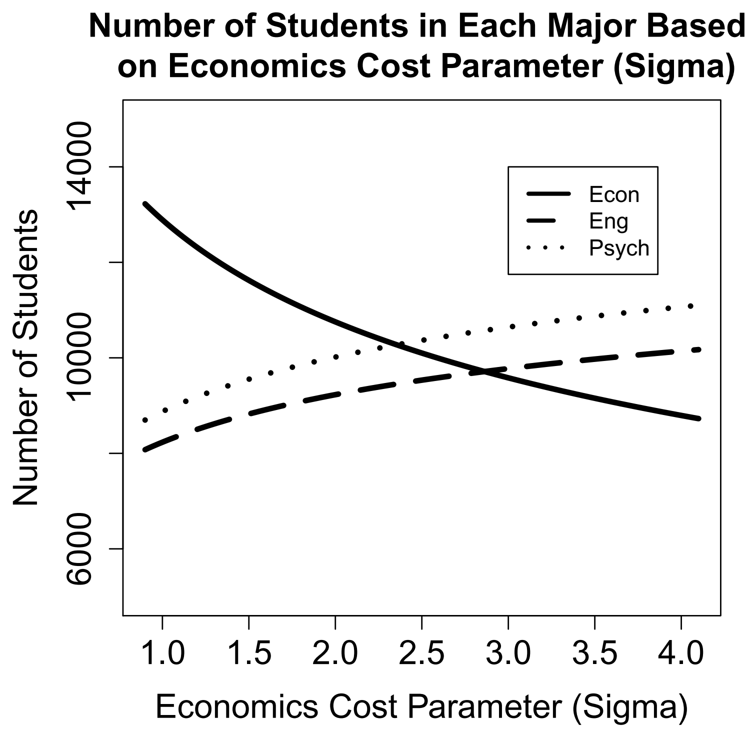


Figure 2.5: Number of Students in Each Major Based on Economics Cost Parameter (Sigma). As σ_1 (economics) increases, the effective funds per students drop – in all majors but in particular, in economics. This leads students who are on the margin of majoring in economics to leave economics and join the other majors.

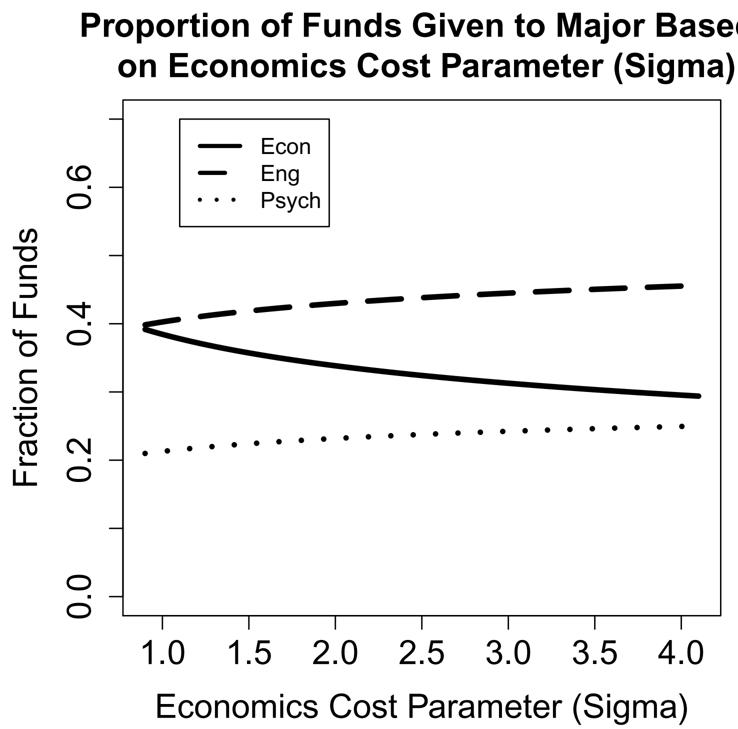


Figure 2.6: Number of Students in Each Major Based on Economics Cost Parameter (Sigma). As σ_1 (economics) increases, the effective funds per students drop – in all majors but in particular, in economics. In our simulation, the university also drops funds to economics.

- Student Preference Parameter

Next, we vary γ_1. We keep $\gamma_3 = 0.5$, but we vary $\gamma_2 = 1 - \gamma_1 - \gamma_3$. At extremely low γ_1, the university gets a great deal for students who actually do major in economics. For low funding, the university awards a few very high quality degrees to students who really like the field relative to their colleagues in engineering and psychology. On the other hand, having university preferences so out of line with student preferences may not be desirable.

After an initial drop in funding as γ_1 increases, increased students in economics drive total funding of the department higher. Still, it is not enough to increase the total funding per student. Students liking economics have major impacts on ρ_2 and ρ_3.

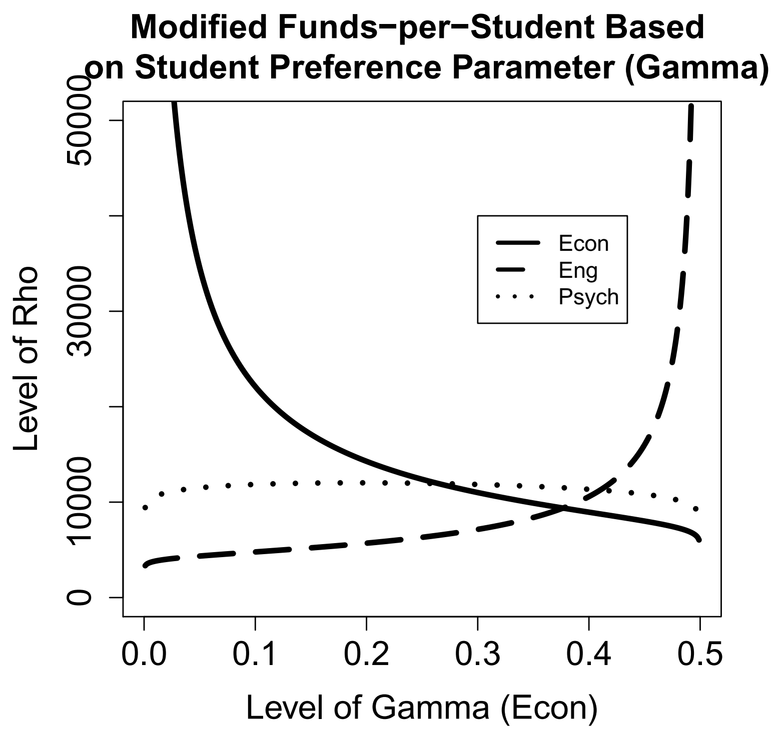


Figure 2.7: Modified Funds-per-Student Based on Student Preference Parameter (Gamma). In this simulation, we vary γ_1 (student preference parameter on economics) so that $\gamma_3 = 0.5$ and $\gamma_2 = 1 - \gamma_1 - \gamma_3$. Economics (major 1) experiences an influx of majors, making the funds-per-student more expensive for that major. This leads to decreased ρ_1, as the university tries to lower its cost for funds-per-student by making the other majors more attractive.

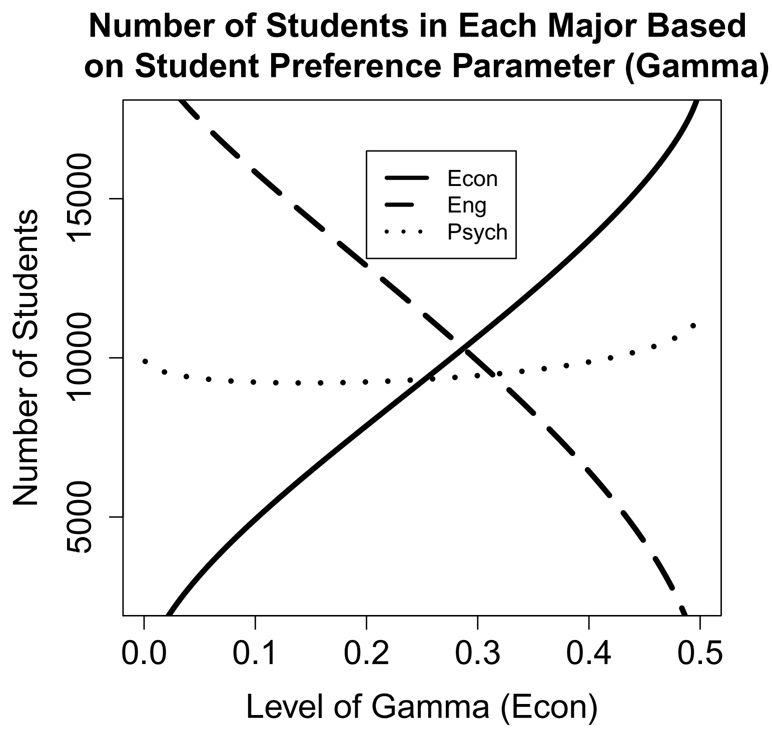


Figure 2.8: Number of Students in Each Major Based on Student Preference Parameter (Gamma). In this simulation, we vary γ_1 (student preference parameter on economics) so that $\gamma_3 = 0.5$ and $\gamma_2 = 1 - \gamma_1 - \gamma_3$. Economics (major 1) experiences an influx of majors, making the funds-per-student more expensive for that major. This leads to decreased ρ_1 and increased S_1, as the university tries to lower its cost for funds-per-student by making the other majors more attractive.

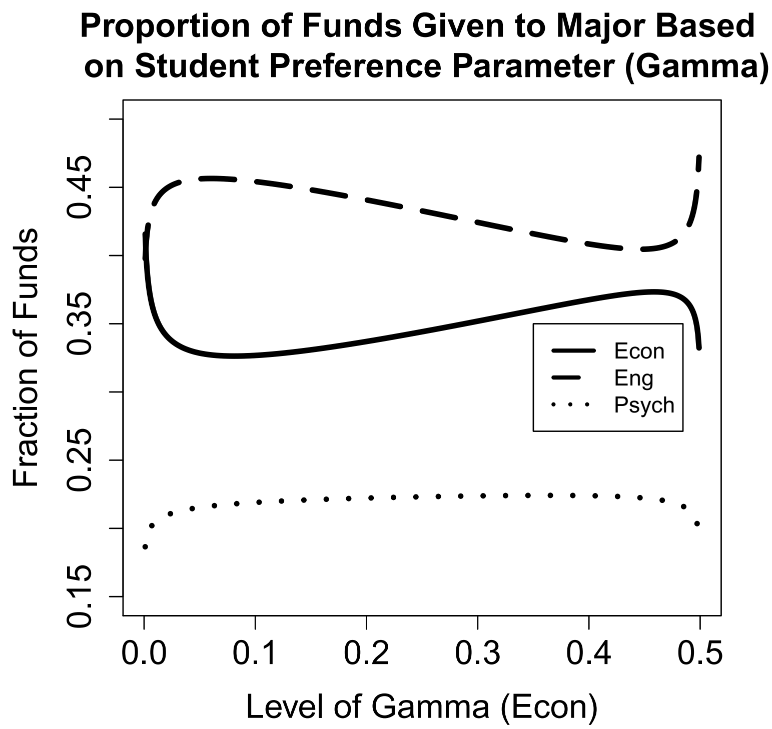


Figure 2.9: Proportion of Funds Given to Major Based on Student Preference Parameter (Gamma). In this simulation, we vary γ_1 (student preference parameter on economics) so that $\gamma_3 = 0.5$ and $\gamma_2 = 1 - \gamma_1 - \gamma_3$. Economics (major 1) experiences an influx of majors, making the funds-per-student more expensive for that major. The university, at very low levels of γ_1, first lowers funding as students start to come in. Over more reasonable relative γ_1's, the funds increase, but not enough to make up for the students in the major (Figure 2.8).

- Salary Parameter

Similarly, we see what increased salary does in terms of attracting students to major in economics. As we increase w_1, more students flow into the major. The dynamics of this simulation are similar to γ_1, as the parametrization is the same with different parameters. In this simulation, we set $w_3 = 0.8$ and $w_2 = 3 - w_1 - w_3$.

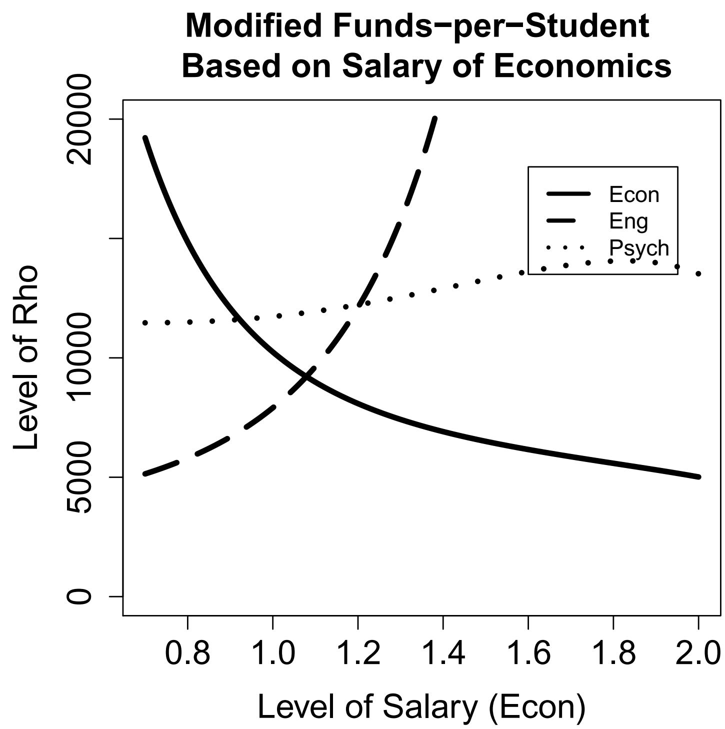


Figure 2.10: Modified Funds-per-Student Based on Salary of Economics (Salary). As we increase the salary of economics and decrease the salary of engineering, the modified funds-per-student (ρ) increases in engineering and decreases in economics. This is even with more total funds entering in economics (see Figure 2.12.)

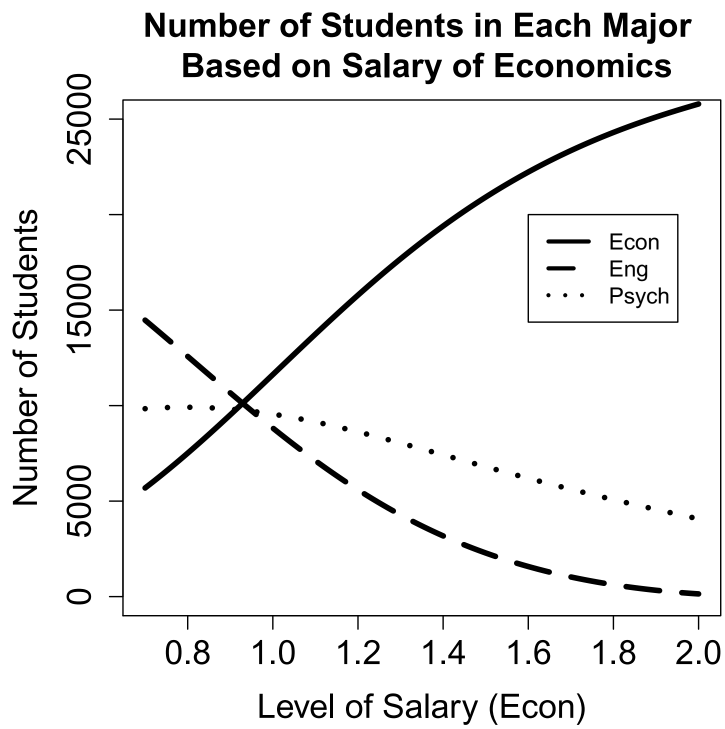


Figure 2.11: Number of Students in Each Major Based on Salary of Economics (Salary). As we increase the salary of economics and decrease the salary of engineering, more students enter economics – even students who once majored in psychology.

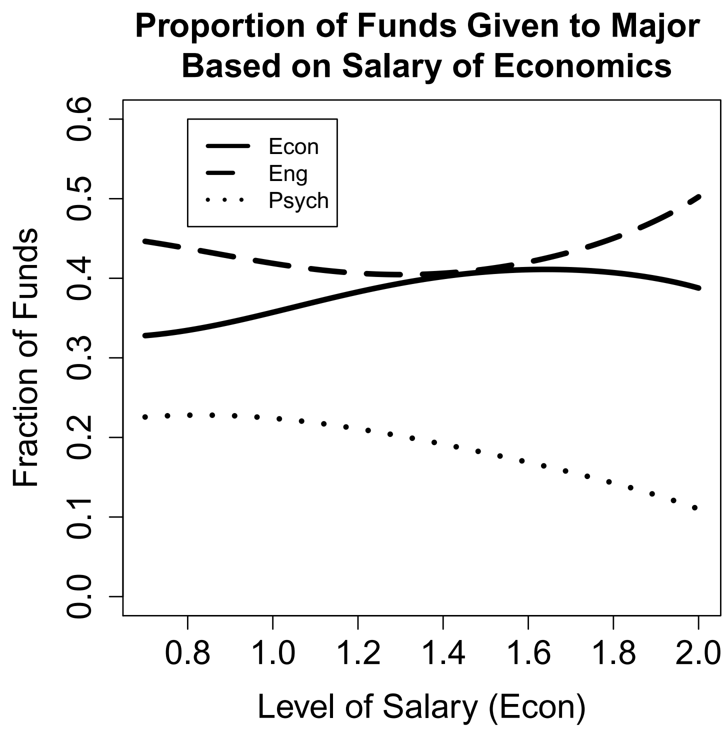


Figure 2.12: Proportion of Funds Given to Major Based on Salary of Economics (Salary). As we increase the salary of economics and decrease the salary of engineering, the impact on the actual proportions of funds in each department is ambiguous.

2.3.1.3 Differential Payments

To evaluate the impact of differential tuition, we need to look at two factors. Overall, we find that the results of the simulation are ambiguous. This may seem counter-intuitive; most economists view differential tuition in the prism of (a) higher tuition leads to less students in the major and (b) higher payments will let the university provide more resources to allow more students into the major.

However, in the context of our model, higher tuition payments lead to an incentive for the university to offer higher ρ_i to the department to attract more students. This may outweigh the loss in w_i. Thus, these differential payments are parameter dependent and ambiguous.

We have already shown what happens in our simulation with changed w_i under particular parameters. However, we have not shown what happens to the university with an influx of variable money. To do this, we run a separate simulation where the state differentially funds engineering majors. We will show that the university devotes more resources to engineering, leading to an increased enrollment in engineering.

To simulate the impact differential state funding would have on the university, we must first look at the maximization problem of the university. The university picks both the price of Q and the resulting amount of funds it receives from the state (under the prediction that it can forecast student enrollment). Summarizing, the university finds:

$$
\begin{array}{rcl}
\max\limits_{\mathbf{f}} & & Q\,(\rho)^{\zeta}\,O^{1-\zeta} \\
s.t.: & & \\
p_O O + p_Q Q & = & S(B + b\widetilde{S}_2) \\
F & \equiv & S(B + b\widetilde{S}_2) \\
p_Q & = & S/\left[\sum_{i=1}^{M} \alpha_i \left(\dfrac{f_i}{\sigma_i\,(S_i|\mathbf{f})}\right)^{r}\right]^{\frac{1}{r}} \\
S_i|\mathbf{f} & = & \dfrac{S}{\sum_{m=1}^{M}\left[\frac{\gamma_m}{\gamma_i}\right]^{\frac{b_1}{1+b_2}}\left[\frac{w_m}{w_i}\right]^{\frac{b_3}{1+b_2}}\left[\frac{\sigma_m}{\sigma_i}\right]^{-\frac{b_2}{1+b_2}}\left[\frac{f_m}{f_i}\right]^{\frac{b_2}{1+b_2}}}
\end{array}
\qquad (2.5)
$$

List of variables and initial parameters:

$$M = 3, \mathbf{b} = (1, 0.5, 3), \text{ and } r = 0.25$$

		α	σ	γ	w
(1)	Economics	0.33	1.5	0.33	1.0
(2)	Engineering	0.50	3.0	0.17	1.2
(3)	Psychology	0.17	1.0	0.50	0.80

Parameter	Initial Value	Description
B	\$16,667	Base payment, regardless of major.
b	(will vary)	Additional payment per student enrolled in engineering.
$\mathbf{f} = (f_1, f_2, f_3)$		Funds devoted to each department.
$\tilde{\mathbf{f}} = (\widetilde{f_1}, \widetilde{f_2}, \widetilde{f_3})$		Fraction of funds devoted to each department.
O		Amount of spending not related to $Q(\rho)$.
p_O	\$50,000	Price of O.
p_Q		Price of Q.
S	30,000	Number of students.
S_i		Number of students in each major i.
$\widetilde{S_i}$		Fraction of students in each major i.
ζ	2/3	Proportion of funds devoted to $Q(\rho)$; which is also the Cobb-Douglas parameter.

This maximization is simulated and proceeds the following way:

1. A previous utility, the current maximum, is stored. If this is the first round, set to zero.

2. An initial funding vector, $\tilde{\mathbf{f}} = (\widetilde{f}_1, \widetilde{f}_2, \widetilde{f}_3)$, is selected.

3. $\widetilde{S}_2$ is determined from the funding vector, $\tilde{\mathbf{f}}$.

4. Price p_Q is determined from the initial vector.

5. F is found.

6. 2/3 of F is devoted to Q, 1/3 to P. This is based on prices $p_O = \$50,000$, and p_Q from step 4.

7. The utility function is calculated. If higher than it was in step 1, this is the new calculation.

The results of the simulation show that as a state-funded subsidy of engineering occurs, funds increase and more students major in engineering. The modified funds-per-student improves in every major, but it grows fastest in engineering. An important point: per-student allocation increase not just in engineering – some of the subsidy is going to educate students in other majors. Further, by assumption, some of the money is going to other (O) parts of the university.

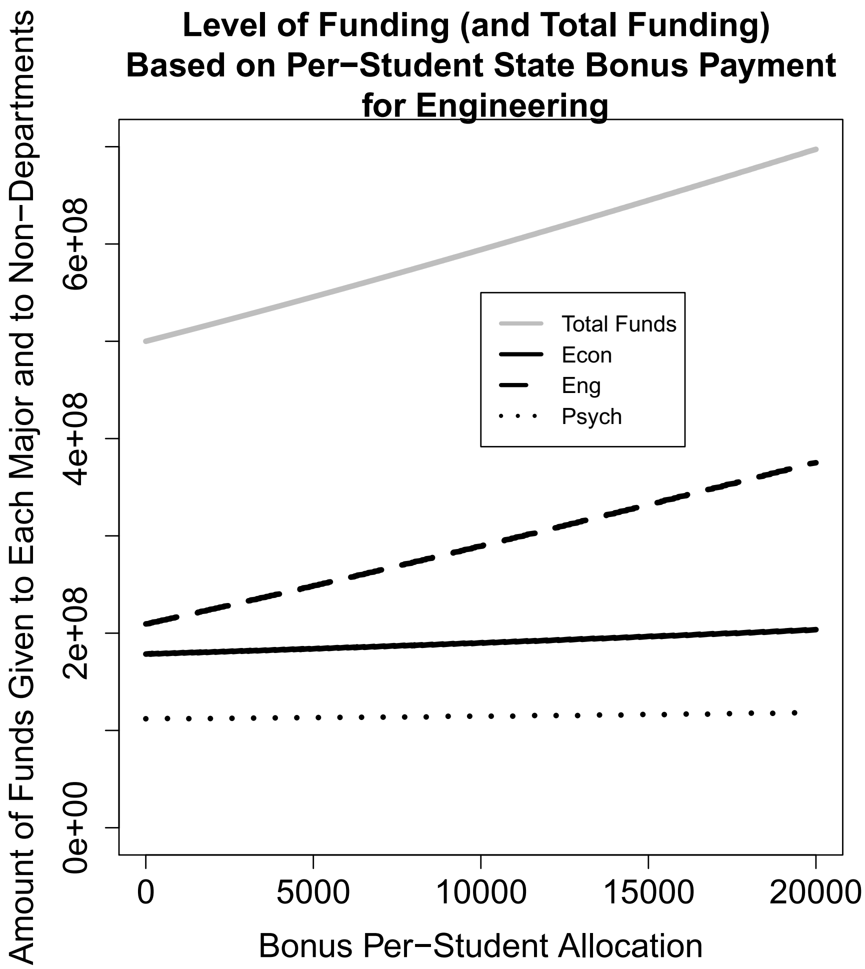


Figure 2.13: Simulation: Level of Funding and Total Funding Based on Per-Student State Bonus Payment for Engineering. As the state increases b, the per-engineering differential subsidy, total funds increase, and the university decides to grow engineering spending faster than other departments.

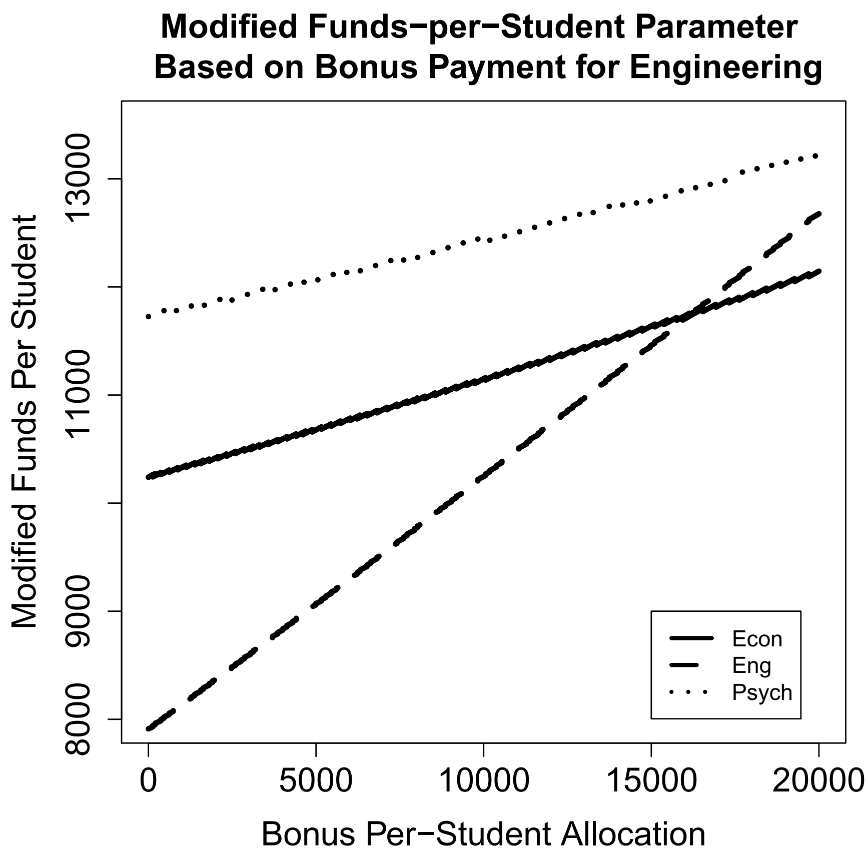


Figure 2.14: Simulation: Level of Funding and Total Funding Based on Per-Student State Bonus Payment for Engineering. As the state increases b, effective funds per student increase in *all* fields, not just engineering – although the differential is much larger in engineering.

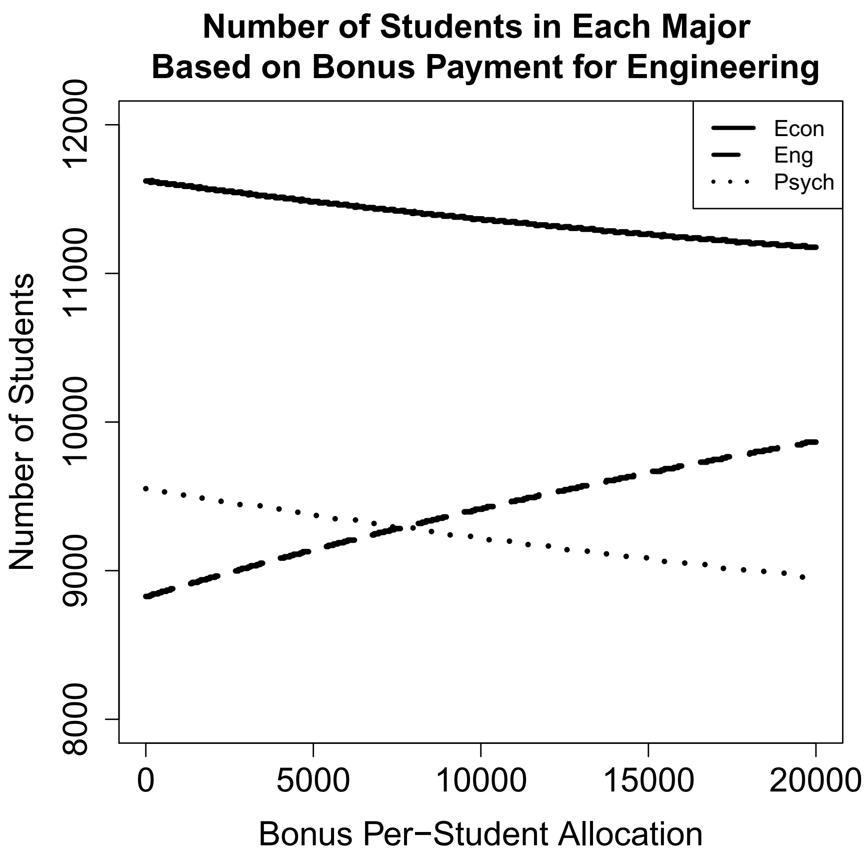


Figure 2.15: Simulation: Number of Students in Each Major Based on Bonus Payment for Engineering. As the state increases b, the per-engineering differential subsidy, more students fill in engineering as ρ_2 increases more differentially than other fields.

2.4 Education Funding Research

How does our simulation explain what is happening with actual data? Many of the impacts of increasing tuition, differential tuition, major choice, school choice, and departmental funding are already well-researched. Hoxby

(1997) looks at increasing tuition throughout the latter half of the 20th century and concludes that increased competition led to higher price tags, better matched students, higher quality education, lower variance in student abilities at schools, higher variance between schools, and increased diversity in geography at schools. She made no decomposition into effects by departments.

Very few studies attempt to dissect cost into departments. One recent attempt to do this is Johnson (2009). Johnson looks at five different ways to compute cost of a bachelors degree. The first estimates the cost of the degree if a student follows the prescribed catalog and does not fail any courses. The second looks at the actual classes that students in each major take. The third ("Full Cost Contribution") includes students who fail out of the program or waste time taking classes not attributable to their degree and transfer to a different (often, easier) field. The fourth takes IPEDS data on institutional costs and does a large regression. The final looks at the sticker price for the student. Estimates do not change between the various calculations, with engineering uniformly higher than other fields in virtually all of the calculations. The Full Cost Attribution is notable because of students in the State University System of Florida who end up in Leisure Studies, only 9% started off wishing to finish in that field, dramatically decreasing the cost of the degree in the Full Cost Attribution calculation. Johnson's method finds Florida's costs per bachelors degree of $26,485 for the Catalog Cost, $31,764 for the Transcript Cost Method, and $37,757 for the Full Cost Attribution.

The question though is what to include in the calculation. What is the relevant margin? In the Johnson paper, notably, capital costs are excluded. In the IPEDS analysis, costs include funds from "contracts, grants, endowment income and gifts," which are not included in the other cost analyses, which include, "direct and indirect" costs, and not "auxiliary" activities, such as housing.

The Johnson paper is a framework for a working paper by Romano, Losinger, and Millard, who look at the cost of a community college degree. Surprisingly, the more expensive community college degrees are very expensive, even compared to four year degrees. At the upstate New York community college they looked at, Broome Community College, the Full Cost Transcript method yielded $47,968, for the Dental Hygiene Degree. Clearly, the college is losing money on these degrees; they are being subsidized by lower-cost liberal arts degrees. Another working paper by Romano and Djajalaksana actually finds it is *cheaper* per full-time equivalent to educate students at a masters-level university than it is to educate them at a community college.

Part of what may be driving low-cost masters-level university teaching is economies of scale and scope. Readily available college-level data proliferated a large number of studies on the marginal cost of activities on campus. Nelson and Hevert (1992) find that economies of scale occur if colleges decide to increase class size. It also finds laboratory courses are associated with higher cost. Dundar and Lewis (1995) are also concerned with economies of scale and scope. In the process, they discover that social science courses have the lowest cost and engineering the highest. They find economies of scale and scope in at the departmental level that differ between types of departments (i.e., social sciences) but not within type; they also control for quality using departmental rankings. de Groot, et.al. (1991) also finds economies of scale for U.S. universities in 1983. In contrast to previous studies, Fu, et.al. (2011) finds that Taiwanese universities are too large and are experiencing *diseconomies* of scale.

Another way to compare departments is to use a multidimensional microeconomic analysis, such as data envelopment analysis. Kao and Hung (2006) does this for a Taiwanese university, and a working paper by Halkos, Tzeremes, and Kourtzidis does this for a Greek university. One fatal downfall of this approach when comparing departments is that outputs are *different* across majors. While scientists may take pride in publishing papers, art faculty may both publish papers and put on exhibits. Since all departments could have different outputs, the efficiency scores are near one, particularly at a school with few academic departments.

2.5 Data Analysis

2.5.1 UCSD

The University of California, San Diego, is a highly-ranked "very high research activity" public university in La Jolla, CA – a northern outlying neighborhood of San Diego. There are over 23,000 students, multiple graduate program, and a medical school (*U.S. News and World Report* 2015 and Carnegie Classification of Institutions for Higher Education). UCSD is a residential university with six colleges[1] – the academic "home" of students. These colleges determine the general education requirements for the students. Any student in any college can major in any of the departments on campus, as long as they meet the requirements for that department. UCSD has strict guidelines[2] for students who want to transfer colleges, and this happens infrequently. Colleges house their own freshman writing programs, which are hybrid courses where colleges introduce their themes and students write about them.

Academic departments are divided into divisions[3]. Divisions also have budgets, which we will include in the cost per credit hour.

2.5.2 Cost Data

We download the list of all courses taken by all students in FY2008 and FY2009 (years 2007-8 and 2008-9) from the university's database. The number of students comes from the query "Campus Classlist Statistics 3rd Week" (the drop date for classes is in the third week). Since some courses have more than one credit hour option, we also find the average credit hours taken in the course by downloading the query "Campus Classlist 3rd Week." We assign the courses to the department listed, with a few exceptions, some of which are listed below, and some of which are noted in the Appendix.

Each course has both a department code and a subject code. In the administration of the university, the subject code is uniquely in a particular department.

[1]Thurgood Marshall, Earl Warren, Eleanor Roosevelt, Revelle, Muir, Sixth
[2]See "The College System: FAQ" for more details on these requirements.
[3]Undergraduate majors are in Art, Biology, Social Sciences, Engineering, and Science

The theater department has a code “THEA,” and the Dance and Movement subject within theater has a subject code “TDMV.” However, some instructors teach outside of their department code (i.e., cross-listed courses), so in our calculations, some subject codes may span multiple departments.

Most undergraduate independent and lab courses are not included in the *initial* department code calculation, including many practicum courses. These are, however, included when we find the cost for each *subject code*; some subject codes are clearly only for independent study purposes and thus have cost $0. *Department code calculations do not reflect “independent” courses – these are assigned $0, but when we aggregate to subject code, these $0’s are reinserted to lower the cost per credit hour for the final cost of the degree and to reflect actual cost.* We relegate the technical description of this calculation to the appendix. The calculation is done in this way because any undergraduate receiving credit for laboratory experience or independent research is contributing to the research goals of the university, which is a worthy goal but not what we are trying to calculate here. However, the calculation of the subject code will reflect the basket of research and non-research courses in that particular subject code and will be a good comparison. Furthermore, we can take this average to student-by-student data containing lower and upper division hours and subject code to compute the cost of a degree: Our data for this purpose has total hours by subject codes *including and not separating independent study courses*.

We find cross-listed courses based on similarities on the course schedule, and we assign these courses to the department of the faculty which taught the course as they are listed in the *UCSD General Catalog*. If the faculty home is not listed, we assign it to the listed department as a last resort.

Weighted Penner-ratios are a UCSD measure for student-to-faculty ratios and are adjusted for whether instructional hours are lower division undergraduate, upper division undergraduate, early graduate, or late doctoral. The factors for the weighting of the hours are, respectively, 1, 1.5, 2.5 × (15/12), and 3.5 × (15/12). The latter two have an adjustment for the fact that a full-time graduate student is considered 12 hours instead of 15. Since this is the university's way to adjust for the difficulty of and issues related to teaching the course, and we do not have finer data (such as total amount of time preparing for each type of course), we defer to the university approach to adjust core units. As a sidenote, the Penner-ratio, after all adjustments, is divided by a campus-wide average, so that departments with an average student-to-faculty ratio will have a ratio of 1.0.

2.5.3 Cost Calculation

We have four methods to compute these costs, which are highly correlated (see Table 2.3). The data is from UCSD Academic Affairs Resource Profiles for FY2008 and FY2009, the Office of Graduate Studies, and UCSD's Blink System. Except for Winter 2009 tuition[4], all calculations are inflated to FY2009 dollars[5]. There were a few additional quid-pro-quos to these data, which we relegate to the appendix.

[4]Tuition for each quarter is nearly the same, payments are available, so this is nearly equivalent to the FY2009 calculation.

[5]We define FY2009 dollars as the average CPI-U over months July 2008-June 2009.

We relegate the formal formulation of cost to the appendix, section 2.6.2.2. There are several variables added to all of the cost calculations. These are:

1. Budgeted support funds.
2. Faculty salaries.
3. Lecturer salaries.
4. Teaching assistant salaries.
5. Tutor and reader salaries.
6. Diversity awards.
7. Block grant awards.
8. OGS non-specified awards.
9. Teaching assistant tuition waivers.

Four cost measures are computed. Measure one includes no space. The first three have various amounts of space included in addition to the costs above. Measure two includes the following space at $36/year[6]:

1. Office space allocated to department.
2. Classroom space allocated.
3. Teaching labs.

Measure three includes the three space requirements measures above and the following space measures at $36/year:

[6]Christopher Beliare from the real estate firm Newmark Grubb Knight Frank assisted us in lease data for the UCSD area.

1. Assembly space.
2. Research space.
3. Other space.

The fourth measure uses the same space as the second measure, but it does not do the Penner adjustment explained in Subsection 2.5.2.

First, a cost per credit hour is found per department by dividing the cost measure by the hours awarded in the department code in FY2008 and FY2009 (summer excluded). As described earlier in Section 2.5.2, independent courses are excluded and valued at $0 in the department code run, but will be bundled appropriately at the subject code level. We do this for all of the departments and separately for divisions. We then sum the money used in each subject code by computing the hours awarded times the sum of division cost plus department costs. We then divide by the hours awarded by subject code (thus independent courses are awarded some monies at the end). We then get a cost per credit hour for each subject code. All cost per credit hours are in the Appendix, Tables 2.1 and 2.2.

2.5.4 Cost at the University of California, San Diego

To determine the cost of an undergraduate degree, and to assign benefits to the state, the university, and the student, we first use the cost per credit hour calculations from section 2.5.3. Denote the set of courses each student S takes as $\mathbf{k}^S$, which is a vector of courses (k). The subject code of each course is $\kappa(k)$, the number of hours for that course is h_k, and the Penner parameter of the course is P_k.

The Penner parameter is a university-assigned weight for the level of the course. Upper division courses, in this calculation are considered harder to teach and more costly per hour than lower division courses. The Penner parameters used in this calculation are 1 [Lower Division Undergraduate], 1.5 [Upper Division Graduate], 2.5 × (15/12) [Lower Graduate], and 3.5 × (15/12) [Upper Graduate].

The cost of the degree $\left(c^{\kappa(k)}\right)$ is given by $\sum_{k \in \mathbf{k}^S} c^{\kappa(k)} h_k P_k$. The cost per credit hour, by subject code, are listed in Table 2.2.

We begin an analysis of the cost of each degree at UCSD in the spirit of Johnson (2009). In our analysis, we look at the courses that students *take* and the degrees with which they graduate. Say a student fails out of engineering and enters economics. This throws the cost of the now non-contributing engineering courses into the economics major. We trade this bias for another bias – one where only the required courses are included. A cost measure which only looks at the required courses does not allow us to compare majors where students are likely to take non-required courses, for any particular reason.

Our student-level degree awarded data tracks non-transfer students who entered in year 2006 as freshmen, and graduated by Spring 2012. We see dropouts, and we consider those not earning their degrees at the end of this period as not earning a degree. We have lower division, upper division, and graduate hours by subject code. We do not use AP credit in the cost calculations, as these are not costs to the university.

We compute costs for summer courses as if they were during-the-year courses separately and report these. The default measure presented in the regressions in this report includes the summer-session hours. The summer session has a fundamentally different cost structure and is run by UCSD extension, not the school itself. Although not relevant to the cost calculation – the funding for the session is also different (tuition is assigned per credit hour and the funding mechanism between the university is also different). However, one can argue that the summer session utilizes the same type of instruction (with perhaps a higher lecturer percentage), the same buildings, and other resources – all for required courses which would have appeared in the degree anyways.

We do not have data on inter-school transfer hours, which are generally small at UCSD.

Each degree has a particular code, and we combine some degrees and treat them as one degree in our analysis. For instance, an econ-coded Joint Math-Econ degree is the same as a math-coded Math-Econ degree; we bundle several literature majors, etc. A listing of these majors that are bundled as one are in the appendix, Section 2.6.2.1. In aggregate analysis, some information is only available for departments; for instance, salary per faculty FTE will not be available for programs without faculty. For the most part, programs do not have faculty independent of departments. Much of the analysis is thus restricted to degrees in departments. The number of observations in the analysis is in each table.

A listing of the degrees and the costs are available in the appendix, Section 2.9. Biology at UCSD is an enormous major, in fact, taking its own division, and has many majors in this table, all of them fairly low-cost. Also of low-cost are many of the social science majors. Humanities are not on the whole cheap; there are few of these majors,.

2.5.4.1 Explaining Cost

Next, we use a regression framework to look at the role of factors in explaining cost. The dependent variable in these regressions is the log of costs. Independent variables include (a) the log of the salary per FTE faculty member in the department, (b) the log of the students graduating in the department in the dataset, (c) the log of the adjusted Penner Ratio [an adjusted student-faculty ratio; different from the Penner parameter[7]], (d) the log of the indirect funds per FTE faculty measure, (e) the log of the office space per FTE faculty member, and (f) the log of the number of hours[8].

Since the definition of cost is functionally dependent in a non-linear way with adjusted hours, we also take the cost and subtract off Average Cost per Adjusted Hour × Adjusted Hours. This helps wash out the hours portion of the discrepancy and then the regressand only reflects additional cost above and beyond the hours. We regress this adjusted cost on the non-log measures in the

[7] The Penner ratio is an adjusted student-faculty ratio. We will use the term Penner ratio to contrast it with what we term the Penner parameter, which was discussed earlier in the paper. The Penner parameter is actually used in the calculation of the adjusted Penner ratio.

[8] We adjust logs by a trivial fraction of a dollar to avoid zeros and small, negative numbers. While this is a fairly controversial procedure, we also do an adjusted-mean procedure.

previous paragraph. This should create confidence in the log regression if the signs are similar.

Additionally, we have indicators for a double major, one Bachelors of Science, and colleges. “Transfer” students are students who transfer from one college to another. UCSD is a residential college with six colleges[9] – the academic “home” of students. These colleges determine the general education requirements for the students. Any student in any college can major in any of the departments on campus, as long as they meet the requirements for that department. UCSD has strict guidelines for students who want to transfer colleges, and this happens rarely.

Information on the data is relegated to the appendix.

The cost-regression results are consistent with the presented model and are presented in Tables 2.4 to 2.6. The student-faculty ratio, the salary, and the number of students are the only department-level significant variables when we account for hours. The coefficient on adjusted hours is not surprising; the cost is a function of hours. This paper does find a (very small) increasing return to scale – but the coefficient, while significant, is economically small. A doubling of students would yield a <4% increase in degree price. Notably, most of the returns to scale that much of the education literature finds may be occurring through lower-quality education: once we control for the Penner ratio, the cost savings elasticities in the hours regressions go from -0.07 to -0.03.

If we do not control for hours, double majors are clearly more expensive; Bachelors in Science Degrees do not appear significantly more expensive, but

[9]Thurgood Marshall, Earl Warren, Eleanor Roosevelt, Revelle, Muir, Sixth

the signs are mostly positive and non-trivially small. Once we control for hours, however, this washes away; double majors are actually *cheaper* controlling for hours.

Removing the Penner ratio from the regression yields expected results. The percent taught by faculty, which is negatively correlated with the Penner (see Table 2.8), become significant. The returns to scale increase significantly. More importantly, we must include the Penner to even see the impact on FTE faculty salary. This is the most surprising result; our simulation predicts a negative coefficient: That if you do not take into account faculty per student, you should see a lower funds-per-student in higher salary departments. This in fact is true. A regression of log cost of degree on log salary (and other binary variables) is negative, not positive, and fairly large (-20.0%; see regression (14) in Table 2.5). However, it is not significant at the 10% level. Higher paid faculty in expensive departments teach more students to more than make up for their higher salaries.

Next, let's look at time-to-completion, shown in Table 2.7. While the elasticity is small, higher Penners are associated with lower time-to-degree. This has the implication that higher Penner departments are also better from a time perspective. However, we must be careful in assigning too much to the faculty-student ratio. The effects are small, and the explanatory power of virtually all relevant variables is incredibly small.

2.5.5 Conclusions

Universities and students face substantially different incentives in procurement of bachelor's degrees. The university, if state funds are not allocated towards majors differentially and tuition is equal, will face incentives to fund popular majors at lower levels relative to unpopular majors. This is because marginal students are a cost to the university. On the other hand, the state wants students to major in hard, heavy-return majors. Students want to balance their future potential salary against how difficult or enjoyable the major is.

Looking at UCSD, what we have shown in the model simulation seems to be what is actually happening. The major, relevant factor in the analysis is the modified student-to-faculty ratio. The paper shows that much of the so-called returns to scale in education, at least at UCSD, is actually a reduction in faculty time with students. Surprisingly, faculty salary across departments does not impact the cost of education unless we account for the amount of per-capita students the faculty member must bear; higher paid faculty deal with more students. Indeed, a lone regression on faculty salary and other binary parameters shows that higher paid fields lead to less costly degrees.

2.6 Appendix, Education Paper

2.6.1 Tables

Table 2.1: Cost per Credit Hour (Department Code). This is the listing of cost per credit hours for departments and programs as described in the text. Programs are highlighted in grey, and writing programs are highlighted in black. The table is sorted by Measure II and does not include division costs.

Department or Program	Measure I	II	III	IV
Basic Writing Program Admin	$12.89	$20.57	$20.57	$20.57
Making of the Modern World	$60.09	$63.48	$63.48	$63.60
Culture, Art, Technology	$67.05	$74.54	$74.85	$83.28
Dimensions of Culture	$79.09	$86.53	$86.53	$87.03
Revelle Humanities	$84.25	$92.68	$92.68	$93.91
Muir College Writing Program	$95.49	$107.25	$107.25	$111.89
Warren College Writing Program	$99.02	$115.83	$115.83	$115.83
Science, Technology, Public Affairs	$40.83	$40.83	$40.83	$46.30
Critical Gender Studies	$41.44	$41.44	$41.44	$52.96
International Studies Program	$52.80	$52.80	$52.80	$79.19
Urban Studies & Planning	$49.69	$56.08	$56.08	$78.23
Third World Studies	$55.38	$59.04	$59.04	$59.71
Earth Systems	$65.94	$70.65	$70.65	$98.41
Human Development Program	$59.87	$79.34	$79.34	$106.63
Linguistics Language Program	$86.19	$93.68	$93.68	$98.66
Academic Internship Program	$98.22	$135.64	$135.64	$203.47
Psychology	$69.93	$79.17	$93.39	$115.79
History, CAESER	$79.49	$85.82	$85.91	$121.24
Economics	$80.64	$86.02	$86.23	$126.08
Biology	$87.18	$103.95	$132.94	$166.39
Chemistry	$84.14	$108.08	$139.01	$166.27
Political Science	$102.92	$109.24	$109.24	$163.27
Educational Studies	$96.14	$110.35	$110.42	$275.86
Bioengineering	$88.21	$115.55	$142.76	$293.31
Communications	$104.53	$115.66	$115.83	$181.55
Mathematics	$107.67	$118.98	$119.55	$146.61
Sociology	$120.24	$130.75	$130.75	$204.51
Cognitive Science	$111.55	$131.95	$152.75	$193.21
Ethnic Studies	$125.22	$137.90	$137.90	$204.85
Mechanical & Aeronautical Engineerir	$123.67	$143.73	$159.98	$266.26
Structural Engineering	$120.75	$144.64	$188.81	$289.81
Computer & Science Engineering	$129.69	$159.75	$175.98	$326.19
Visual Arts	$131.54	$161.27	$196.05	$232.65
Anthropology	$148.51	$165.26	$178.61	$287.44
Music	$132.75	$167.88	$183.59	$230.82
Literature	$155.62	$170.91	$171.35	$246.72
Linguistics (not FLP)	$146.14	$172.93	$192.58	$265.22
Electrical & Computer Engineering	$149.36	$178.29	$216.39	$461.49
Theater	$143.94	$190.76	$220.86	$293.87
Physics	$163.52	$200.01	$239.55	$260.53
Philosophy	$198.75	$216.80	$216.80	$287.29
Rady School of Management	$197.24	$246.93	$249.64	$646.12

Table 2.2: Cost per Credit Hour (Subject Code). [includes division cost]

	Subject Code	Measure I	Measure II	Measure III	Measure IV
Academic Internship Program	AIP	$ 98.22	$ 135.64	$ 135.64	$ 203.47
Anthropological Archeology	ANAR	$ 151.49	$ 168.54	$ 181.90	$ 292.41
Anthro/Biological Anthropology	ANBI	$ 148.12	$ 165.11	$ 179.18	$ 285.77
Anthropology/Sociocultural	ANSC	$ 151.49	$ 168.54	$ 181.90	$ 292.41
Anthropology	ANTH	$ 146.46	$ 162.94	$ 175.86	$ 284.82
Bioengineering	BENG	$ 98.83	$ 129.15	$ 155.87	$ 319.79
Biology/Biochemistry	BIBC	$ 91.37	$ 108.46	$ 137.35	$ 173.54
Biol/Genetics,Cellular&Develop	BICD	$ 91.25	$ 108.32	$ 137.32	$ 173.39
Biol/Ecology, Behavior, & Evol	BIEB	$ 91.25	$ 108.32	$ 137.32	$ 173.39
Biology/Lower Division	BILD	$ 91.11	$ 108.16	$ 137.11	$ 173.13
Biology/Molec Biol, Microbiol	BIMM	$ 91.31	$ 108.41	$ 137.39	$ 173.62
Biology/Animal Physiol&Neurosc	BIPN	$ 91.25	$ 108.32	$ 137.32	$ 173.39
Biology/Special Studies	BISP	$ 14.25	$ 16.91	$ 21.44	$ 27.07
Culture, Art, and Technology	CAT	$ 118.34	$ 142.94	$ 145.48	$ 159.87
Chemical Engineering	CENG	$ 136.88	$ 161.00	$ 177.79	$ 303.66
Critical Gender Studies	CGS	$ 47.69	$ 48.63	$ 48.64	$ 63.38
Chemistry and Biochemistry	CHEM	$ 86.99	$ 111.20	$ 141.89	$ 170.62
Chinese Studies	CHIN	$ 82.74	$ 89.69	$ 89.78	$ 126.74
Classical Studies (Hon Thesis)	CLAS	$ -	$ -	$ -	$ -
Communication and Culture	COCU	$ 107.50	$ 118.94	$ 119.12	$ 186.52
Communication/General	COGN	$ 103.56	$ 114.58	$ 114.75	$ 181.36
Cognitive Science	COGS	$ 111.63	$ 131.81	$ 152.09	$ 194.06
Communication/Human Info Proc	COHI	$ 107.50	$ 118.94	$ 119.12	$ 186.52
Communication Media Methods	COMT	$ 90.26	$ 99.86	$ 100.01	$ 156.60
Communication as Social Force	COSF	$ 107.50	$ 118.94	$ 119.12	$ 186.52
Computer Science & Engineering	CSE	$ 142.12	$ 175.95	$ 192.65	$ 362.12
Dimensions of Culture	DOC	$ 141.69	$ 165.69	$ 165.69	$ 167.96
Education Abroad Program	EAP	$ 110.63	$ 129.01	$ 142.71	$ 195.91
Electrical & Computer Engineer	ECE	$ 162.14	$ 194.96	$ 233.20	$ 495.73
Economics	ECON	$ 83.40	$ 89.09	$ 89.31	$ 130.73
Education Studies	EDS	$ 98.46	$ 112.18	$ 112.60	$ 260.31
Engineering	ENG	$ 110.63	$ 129.01	$ 142.71	$ 195.91
Environmental Studies	ENVR	$ 120.14	$ 142.62	$ 148.01	$ 161.61
Eleanor Roosevelt College	ERC	$ 77.07	$ 89.88	$ 99.42	$ 128.23
Environmental Systems	ESYS	$ 62.05	$ 68.39	$ 70.72	$ 98.24
Ethnic Studies	ETHN	$ 128.10	$ 142.23	$ 143.12	$ 209.94
Exchange Programs	EXPR	$ 110.63	$ 129.01	$ 142.71	$ 195.91
Film Studies	FILM	$ 120.37	$ 142.90	$ 148.30	$ 161.89
Family and Preventive Medicine	FPM	$ -	$ -	$ -	$ -
Human Development Program	HDP	$ 59.60	$ 77.18	$ 77.20	$ 106.15
History of Africa	HIAF	$ 82.80	$ 89.76	$ 89.85	$ 126.81
History of East Asia	HIEA	$ 82.76	$ 89.71	$ 89.80	$ 126.74
History of Europe	HIEU	$ 82.76	$ 89.71	$ 89.80	$ 126.74
History of Latin America	HILA	$ 82.80	$ 89.76	$ 89.85	$ 126.81
History, Lower Division	HILD	$ 82.80	$ 89.76	$ 89.85	$ 126.81

Table 2.2. Cost per Credit Hour (Subject Code), *continued.*

	Subject Code	Measure I	Measure II	Measure III	Measure IV
History of the Near East	HINE	$ 82.57	$ 89.50	$ 89.60	$ 126.45
History of Science	HISC	$ 82.69	$ 89.64	$ 89.73	$ 126.64
History Topics	HITO	$ 69.40	$ 75.23	$ 75.31	$ 107.00
History of the United States	HIUS	$ 82.98	$ 90.31	$ 90.40	$ 127.04
Human Rights	HMNR	$ 132.17	$ 146.38	$ 156.01	$ 245.84
Humanities	HUM	$ 157.84	$ 183.60	$ 184.67	$ 186.03
Computing and the Arts	ICAM	$ 129.11	$ 157.92	$ 184.61	$ 226.25
International Studies	INTL	$ 55.05	$ 55.35	$ 55.36	$ 83.08
Japanese Studies	JAPN	$ 82.64	$ 89.59	$ 89.68	$ 126.64
Judaic Studies	JUDA	$ 80.64	$ 87.41	$ 87.51	$ 124.31
Latin American Studies	LATI	$ 107.10	$ 124.91	$ 138.17	$ 191.57
Law and Society	LAWS	$ 110.63	$ 129.01	$ 142.71	$ 195.91
Linguistics/Arabic	LIAB	$ 89.16	$ 96.96	$ 96.97	$ 103.63
Linguistics - Directed Study	LIDS	$ 89.16	$ 96.96	$ 96.97	$ 103.63
Linguistics/French	LIFR	$ 89.16	$ 96.96	$ 96.97	$ 103.63
Linguistics/German	LIGM	$ 89.16	$ 96.96	$ 96.97	$ 103.63
Linguistics/General	LIGN	$ 146.44	$ 172.93	$ 192.07	$ 265.75
Linguistics/Heritage Languages	LIHL	$ 89.16	$ 96.96	$ 96.97	$ 103.63
Linguistics/Italian	LIIT	$ 89.16	$ 96.96	$ 96.97	$ 103.63
Linguistics/Portuguese	LIPO	$ 89.16	$ 96.96	$ 96.97	$ 103.63
Linguistics/Amer Sign Language	LISL	$ 89.16	$ 96.96	$ 96.97	$ 103.63
Linguistics/Spanish	LISP	$ 89.16	$ 96.96	$ 96.97	$ 103.63
Literature of Africa	LTAF	$ 158.94	$ 174.85	$ 175.29	$ 252.28
Literature of the Americas	LTAM	$ 158.94	$ 174.85	$ 175.29	$ 252.28
Literature of China	LTCH	$ 158.94	$ 174.85	$ 175.29	$ 252.28
Literature/Cultural Studies	LTCS	$ 158.46	$ 174.33	$ 174.77	$ 251.63
East Asian Literature	LTEA	$ 158.35	$ 174.20	$ 174.64	$ 251.35
Literatures in English	LTEN	$ 157.45	$ 173.45	$ 174.06	$ 250.47
Literature/European & Eurasian	LTEU	$ 136.98	$ 150.37	$ 150.72	$ 217.07
Literature/French	LTFR	$ 158.94	$ 174.85	$ 175.29	$ 252.28
Literature/Greek	LTGK	$ 131.14	$ 143.83	$ 144.15	$ 217.95
Literature/German	LTGM	$ 121.74	$ 133.28	$ 133.55	$ 203.47
Literature/Italian	LTIT	$ 125.34	$ 137.30	$ 137.59	$ 208.71
Literature/Korean	LTKO	$ 158.94	$ 174.85	$ 175.29	$ 252.28
Literature/Latin	LTLA	$ 86.36	$ 93.74	$ 93.85	$ 131.27
Literature / Portuguese	LTPR	$ 158.94	$ 174.85	$ 175.29	$ 252.28
Literature/Russian	LTRU	$ 82.56	$ 89.49	$ 89.59	$ 126.51
Literature/Spanish	LTSP	$ 158.43	$ 174.29	$ 174.72	$ 251.71
Literature/Theory	LTTH	$ 100.26	$ 109.27	$ 109.44	$ 155.58
Literatures of the World	LTWL	$ 146.73	$ 161.21	$ 161.60	$ 233.57
Literature/Writing	LTWR	$ 150.80	$ 165.90	$ 166.32	$ 241.22
Mechanical & Aerospace Engin	MAE	$ 137.73	$ 162.09	$ 179.64	$ 305.78
Mathematics	MATH	$ 111.11	$ 122.87	$ 123.43	$ 151.98
Muir College Writing Program	MCWP	$ 174.41	$ 208.71	$ 214.11	$ 220.81
Medicine	MED	$ -	$ -	$ -	$ -

Table 2.2. Cost per Credit Hour (Subject Code), *continued.*

	Subject Code	Measure I	Measure II	Measure III	Measure IV
Rady School of Management	MGT	$ 197.24	$ 246.93	$ 249.64	$ 646.12
Making of the Modern World	MMW	$ 96.35	$ 123.34	$ 126.14	$ 123.62
Muir College	MUIR	$ 102.76	$ 119.84	$ 132.57	$ 181.17
Music	MUS	$ 134.06	$ 168.85	$ 184.04	$ 233.47
Neurosciences	NEU	$ -	$ -	$ -	$ -
Ophthalmology	OPTH	$ -	$ -	$ -	$ -
Orthopaedics	ORTH	$ -	$ -	$ -	$ -
Pathology	PATH	$ -	$ -	$ -	$ -
Pediatrics	PEDS	$ -	$ -	$ -	$ -
Pharmacology	PHAR	$ -	$ -	$ -	$ -
Philosophy	PHIL	$ 172.89	$ 188.58	$ 188.60	$ 258.42
Physics	PHYS	$ 166.74	$ 203.62	$ 243.09	$ 265.68
Political Science	POLI	$ 103.34	$ 110.13	$ 110.15	$ 164.40
Psychiatry	PSY	$ -	$ -	$ -	$ -
Psychology	PSYC	$ 69.77	$ 78.90	$ 92.51	$ 115.96
Radiology	RAD	$ -	$ -	$ -	$ -
Religion, Study of	RELI	$ 99.18	$ 112.97	$ 121.23	$ 169.96
Revelle College	REV	$ 107.93	$ 125.87	$ 139.23	$ 191.13
Reproductive Medicine	RMED	$ -	$ -	$ -	$ -
San Diego Community College	SDCC	$ 16.20	$ 24.51	$ 24.51	$ 26.13
Structural Engineering	SE	$ 135.13	$ 162.96	$ 205.76	$ 327.70
Scripps Inst of Oceanography	SIO	$ 108.25	$ 126.24	$ 139.64	$ 192.74
Soc/Theory & Methods	SOCA	$ 119.85	$ 130.34	$ 130.36	$ 202.59
Soc/Cult, Lang, & Soc Interact	SOCB	$ 124.43	$ 136.48	$ 137.86	$ 215.45
Soc/Soc Organiz & Institutions	SOCC	$ 123.02	$ 133.85	$ 133.86	$ 209.20
Soc/Comparative & Historical	SOCD	$ 116.10	$ 126.24	$ 126.27	$ 194.93
Soc/Ind Research & Honors Prog	SOCE	$ 61.98	$ 67.42	$ 67.42	$ 105.37
Soc/Lower Division	SOCL	$ 123.21	$ 134.03	$ 134.04	$ 209.48
Sci, Technology&Public Affairs	STPA	$ 186.05	$ 223.92	$ 229.17	$ 231.09
Surgery	SURG	$ -	$ -	$ -	$ -
Theatre / Acting	TDAC	$ 146.82	$ 194.13	$ 224.15	$ 298.72
Dance/Choreography	TDCH	$ 147.25	$ 194.70	$ 224.80	$ 299.44
Theatre / Design	TDDE	$ 146.65	$ 193.91	$ 223.89	$ 298.46
Theatre / Directing&Stage Mgmt	TDDR	$ 139.35	$ 184.25	$ 212.74	$ 283.37
Theatre / General	TDGE	$ 145.87	$ 192.88	$ 222.70	$ 297.43
Dance/History	TDHD	$ 97.36	$ 113.79	$ 120.94	$ 166.50
Theatre / History & Theory	TDHT	$ 131.83	$ 169.59	$ 192.51	$ 266.50
Dance/Movement	TDMV	$ 147.25	$ 194.70	$ 224.80	$ 299.44
Dance/Performance	TDPF	$ 147.25	$ 194.70	$ 224.80	$ 299.44
Theatre Dance/Practicum	TDPR	$ 140.67	$ 186.00	$ 214.76	$ 289.61
Theatre / Playwriting	TDPW	$ 144.56	$ 191.14	$ 220.70	$ 295.16
Dance/Theory	TDTR	$ 147.25	$ 194.70	$ 224.80	$ 299.44
Thurgood Marshall College	TMC	$ 109.32	$ 127.49	$ 141.02	$ 193.98
Third World Studies	TWS	$ 58.70	$ 62.98	$ 62.98	$ 65.27
Urban Studies & Planning	USP	$ 61.70	$ 68.87	$ 68.88	$ 97.41

Table 2.2. Cost per Credit Hour (Subject Code), *continued.*

	Subject Code	Measure I	Measure II	Measure III	Measure IV
Visual Arts	VIS	$ 129.91	$ 158.28	$ 190.14	$ 229.61
Warren College	WARR	$ 100.07	$ 116.71	$ 129.10	$ 181.34
Warren College Writing Program	WCWP	$ 244.24	$ 298.92	$ 304.17	$ 300.62

Table 2.3: Correlation Between Measures.

Department	I	II	III	IV
I	1.0000			
II	0.9867	1.0000		
III	0.9509	0.9786	1.0000	
IV	0.8432	0.8698	0.8463	1.0000

Subject	I	II	III	IV
I	1.0000			
II	0.9875	1.0000		
III	0.9573	0.9864	1.0000	
IV	0.8773	0.8989	0.9095	1.0000

Table 2.4: Regression of Log Cost, Including Adjusted Hours as a Variable.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Log(Adjusted Penner Ratio)	-0.407***	-0.460***		-0.484***	-0.487***		
	(0.0582)	(0.0503)		(0.0426)	(0.0420)		
Log(Indirect Funds per FTE)	0.00475*	0.00442	-0.00305	0.00490*	0.00594***	-0.0110*	
	(0.00267)	(0.00259)	(0.00390)	(0.00248)	(0.00184)	(0.00564)	
Log(Office Space per FTE)	0.0130	0.0315	0.0125	0.0255			
	(0.0418)	(0.0436)	(0.0542)	(0.0393)			
Log(Percent Taught by Faculty)	0.0388	0.0360	0.220***				
	(0.0503)	(0.0523)	(0.0627)				
Log(Salary Per FTE)	0.268***	0.306***	0.0989	0.301***	0.280***	-0.0698	-0.0363
	(0.0658)	(0.0723)	(0.0867)	(0.0685)	(0.0668)	(0.180)	(0.0765)
Log(Adjusted Hours)	1.011***	1.016***	1.014***	1.018***	1.023***	1.097***	1.019***
	(0.0254)	(0.0244)	(0.0353)	(0.0250)	(0.0241)	(0.0530)	(0.0405)
Log(Students Graduating in Dept)	-0.0168*		-0.0567***				-0.0769***
	(0.00875)		(0.0106)				(0.0116)
Double Major	-0.0570***	-0.0584***	-0.0483***	-0.0588***	-0.0591***	-0.0498***	-0.0479***
	(0.00788)	(0.00675)	(0.0118)	(0.00687)	(0.00700)	(0.0134)	(0.0142)
One B.S.	-0.000355	-0.0177	0.0519	-0.0160	-0.00931	0.0593	0.0585**
	(0.0184)	(0.0190)	(0.0313)	(0.0189)	(0.0176)	(0.0414)	(0.0269)
Muir College	0.0287***	0.0267***	0.0370***	0.0259***	0.0273***	0.0299***	0.0311***
(ERC is the base.)	(0.00486)	(0.00490)	(0.00679)	(0.00485)	(0.00454)	(0.00658)	(0.00662)
Revelle College	0.0364***	0.0347***	0.0405***	0.0342***	0.0361***	0.0318***	0.0317***
	(0.00680)	(0.00692)	(0.00786)	(0.00704)	(0.00660)	(0.00780)	(0.00940)
Sixth College	0.0499***	0.0482***	0.0655***	0.0473***	0.0482***	0.0719***	0.0643***
	(0.00421)	(0.00417)	(0.0101)	(0.00429)	(0.00425)	(0.0138)	(0.00937)
Thurgood Marshall College	0.0279***	0.0270***	0.0322***	0.0266***	0.0277***	0.0297***	0.0274***
	(0.00538)	(0.00556)	(0.00643)	(0.00560)	(0.00504)	(0.00769)	(0.00696)
Transfer	0.0505***	0.0480***	0.0671***	0.0484***	0.0485***	0.0864***	0.0792***
	(0.0113)	(0.0112)	(0.0140)	(0.0112)	(0.0110)	(0.0236)	(0.0164)
Warren College	0.0574***	0.0557***	0.0715***	0.0558***	0.0578***	0.0927***	0.0750***
	(0.00753)	(0.00745)	(0.0118)	(0.00761)	(0.00804)	(0.0186)	(0.0147)
Constant	1.785*	1.143	3.955***	1.213	1.582**	5.151**	5.530***
	(0.919)	(0.903)	(1.166)	(0.830)	(0.711)	(2.061)	(0.918)
Observations	3,616	3,616	3,616	3,616	3,616	3,616	3,616
R-squared	0.890	0.887	0.840	0.886	0.886	0.707	0.802

STANDARD ERRORS CLUSTERED BY DEPARTMENT.

*** P<0.01, ** P<0.05, * P<0.1

Table 2.5: Regression of Log Cost without Hours.

	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
Log(Adjusted Penner Ratio)	-0.417***	-0.514***		-0.565***	-0.582***			
	(0.0862)	(0.0858)		(0.0881)	(0.0827)			
Log(Indirect Funds per FTE)	0.00392	0.00331	-0.00407	0.00436	0.00946**	-0.0108		
	(0.00393)	(0.00391)	(0.00490)	(0.00404)	(0.00388)	(0.00754)		
Log(Office Space per FTE)	0.102	0.136**	0.102	0.124**				
	(0.0595)	(0.0623)	(0.0733)	(0.0561)				
Log(Percent Taught by Faculty)	0.0821	0.0774	0.268***					
	(0.0757)	(0.0848)	(0.0863)					
Log(Salary Per FTE)	0.368***	0.436***	0.195	0.426***	0.325**	-0.0947	-0.202	-0.0143
	(0.108)	(0.125)	(0.121)	(0.119)	(0.127)	(0.237)	(0.149)	(0.107)
Log(Students Graduating in Dept)	-0.0303**		-0.0712***					-0.0998***
	(0.0115)		(0.0138)					(0.0150)
Double Major	0.102***	0.101***	0.112***	0.101***	0.103***	0.129***	0.126***	0.116***
	(0.0194)	(0.0182)	(0.0225)	(0.0181)	(0.0188)	(0.0264)	(0.0257)	(0.0241)
One B.S.	0.0565	0.0256	0.110**	0.0293	0.0639	0.153**	0.0946	0.163***
	(0.0367)	(0.0348)	(0.0513)	(0.0355)	(0.0398)	(0.0661)	(0.0604)	(0.0368)
Muir College	0.0131	0.00932	0.0216	0.00771	0.0138	0.0158	0.0113	0.0202
(ERC is the base.)	(0.0143)	(0.0145)	(0.0165)	(0.0145)	(0.0159)	(0.0170)	(0.0174)	(0.0169)
Revelle College	0.0218	0.0185	0.0259	0.0174	0.0266	0.0205	0.0115	0.0244
	(0.0182)	(0.0187)	(0.0189)	(0.0187)	(0.0214)	(0.0203)	(0.0205)	(0.0206)
Sixth College	0.0763***	0.0735***	0.0924***	0.0717***	0.0770***	0.108***	0.109***	0.0948***
	(0.0158)	(0.0161)	(0.0208)	(0.0158)	(0.0163)	(0.0261)	(0.0262)	(0.0206)
Thurgood Marshall College	0.0282	0.0265	0.0326	0.0256	0.0309	0.0336	0.0292	0.0318
	(0.0193)	(0.0196)	(0.0204)	(0.0196)	(0.0203)	(0.0223)	(0.0220)	(0.0203)
Transfer	0.0835*	0.0793*	0.101**	0.0802*	0.0814*	0.130**	0.133***	0.116**
	(0.0437)	(0.0436)	(0.0448)	(0.0434)	(0.0442)	(0.0475)	(0.0465)	(0.0459)
Warren College	0.0722***	0.0693***	0.0867***	0.0696***	0.0795***	0.124***	0.121***	0.0989***
	(0.0116)	(0.0112)	(0.0130)	(0.0113)	(0.0149)	(0.0181)	(0.0168)	(0.0189)
Constant	5.665***	4.536***	7.899***	4.697***	6.613***	11.38***	12.57***	10.91***
	(1.359)	(1.510)	(1.556)	(1.395)	(1.456)	(2.726)	(1.734)	(1.222)
Observations	3,616	3,616	3,616	3,616	3,616	3,616	3,616	3,616
R-squared	0.428	0.419	0.375	0.416	0.400	0.141	0.116	0.325

STANDARD ERRORS CLUSTERED BY DEPARTMENT.

*** P<0.01, ** P<0.05, * P<0.1

Table 2.6: Regression of Mean-Hour Adjusted Cost on Variables.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Adjusted Penner Ratio	-8,326***	-8,169***		-9,286***	-10,085***		
	(1,663)	(1,800)		(1,589)	(1,629)		
Indirect Research Funds ($ per FTE)	0.00904	-0.0230*	-0.0382	-0.0199	-0.00161	-0.0390*	
	(0.0211)	(0.0118)	(0.0336)	(0.0119)	(0.00739)	(0.0203)	
Office Space	-0.976	4.478*	4.976	4.023			
(Assigned Square ft. per FTE)	(3.522)	(2.362)	(5.608)	(2.365)			
Percentage Taught by Faculty	2,003	3,873	14,293***				
	(3,631)	(3,370)	(3,703)				
Salary per FTE	0.0792***	0.0698**	-0.0160	0.0749**	0.0876***	-0.0734**	-0.0411
	(0.0256)	(0.0294)	(0.0279)	(0.0298)	(0.0286)	(0.0279)	(0.0308)
Number of Students in Dept.	-3.158**		-2.528				
	(1.405)		(2.393)				
Double Major	-1,594***	-1,637***	-1,219***	-1,667***	-1,667***	-965.6**	-1,092**
	(255.1)	(253.6)	(324.0)	(254.8)	(257.7)	(427.8)	(435.0)
One B.S.	1,130	871.7	2,608*	788.3	635.4	3,396	261.1
	(936.9)	(936.3)	(1,316)	(928.1)	(787.2)	(2,002)	(1,571)
Muir	1,095***	1,053***	1,257***	998.3***	1,021***	1,075***	794.7***
(ERC is omitted)	(194.2)	(199.3)	(226.2)	(190.3)	(180.6)	(241.7)	(196.8)
Revelle	1,387***	1,420***	1,519***	1,370***	1,378***	1,331***	697.0***
	(254.7)	(258.2)	(266.3)	(256.8)	(244.3)	(262.2)	(222.7)
Sixth	1,843***	1,864***	2,313***	1,805***	1,835***	2,584***	2,533***
	(256.8)	(261.1)	(406.4)	(253.6)	(262.0)	(511.8)	(576.1)
Thurgood Marshall	1,020***	1,046***	1,130***	1,014***	1,034***	1,093***	814.9**
	(177.4)	(181.4)	(222.8)	(179.8)	(173.3)	(297.7)	(301.4)
Transfer	1,409***	1,544***	1,971***	1,536***	1,524***	2,737***	2,865***
	(395.4)	(392.3)	(455.9)	(394.9)	(389.1)	(674.7)	(908.0)
Warren	1,878***	1,916***	2,408***	1,896***	1,981***	3,275***	2,902***
	(287.7)	(287.2)	(440.6)	(288.0)	(300.5)	(625.2)	(574.0)
Constant	2,282	-997.5	-8,806**	2,430	3,929*	7,318**	3,359
	(4,023)	(3,841)	(4,254)	(2,548)	(2,033)	(3,439)	(3,646)
Observations	3,616	3,616	3,616	3,616	3,616	3,616	3,616
R-squared	0.667	0.653	0.522	0.647	0.631	0.208	0.091

Standard errors are clustered by department.

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 2.7: Regression of Log Time in Years.

	(TIME 1)	(TIME 2)	(TIME 3)	(TIME 4)
Log(Adjusted Penner Ratio)	-0.0493	-0.0554**	-0.0413**	
	(0.0289)	(0.0261)	(0.0175)	
Log(Indirect Funds per FTE)	0.000643			
	(0.00198)			
Log(Office Space per FTE)	0.0365			
	(0.0287)			
Log(Percent Taught by Faculty)	0.00633	-0.00295		
	(0.0274)	(0.0299)		
Log(Salary Per FTE)	0.0573	0.0373		-0.0199
	(0.0372)	(0.0403)		(0.0331)
Log(Students Graduating in Dept)	-0.00365			
	(0.00428)			
Double Major	0.0412***	0.0430***	0.0434***	0.0450***
	(0.00983)	(0.00985)	(0.00965)	(0.00983)
One B.S.	0.0173	0.0369**	0.0372**	0.0349**
	(0.0204)	(0.0139)	(0.0146)	(0.0136)
COLLEGE FIXED EFFECTS	YES	YES	YES	YES
Constant	0.566	0.997**	1.428***	1.644***
	(0.532)	(0.467)	(0.0116)	(0.388)
Observations	3,596	3,596	3,596	3,596
R-squared	0.038	0.031	0.030	0.023

Standard errors clustered by department.

*** p<0.01, ** p<0.05, * p<0.1

Table 2.8: Regression of Log Penner Ratio on Other Variables.

	(1)	(2)
Log(Percent Taught by Faculty)		-0.898***
		(0.150)
Log(Students Graduating in Department)	0.174***	
	(0.0477)	
Constant	-0.575**	-0.178*
	(0.237)	(0.0952)
Observations	3,616	3,616
R-squared	0.477	0.520

Standard errors are clustered by department. *** p<0.01, ** p<0.05, * p<0.1

Table 2.9: Cost per Degree.

Major(s)	N	Cost (Academic Year)	Cost (Summer)	Full Cost	Time to Degree (yrs)
Did not graduate.	635	$16,448.77	$1,199.78	$17,648.55	2.81
History	45	$23,874.82	$819.04	$24,693.86	3.92
International Studies-Econ	53	$23,380.92	$1,544.56	$24,925.48	4.07
Joint Math-Econ	17	$24,371.09	$1,130.17	$25,501.27	3.87
International Studies-History	17	$24,001.92	$1,929.34	$25,931.26	4.25
Urban Studies and Planning	31	$24,179.09	$1,835.81	$26,014.90	4.25
Psychology	207	$24,719.81	$1,566.09	$26,285.90	4.30
Psychology	83	$24,696.48	$1,735.07	$26,431.54	4.11
International Studies-Poli Sci	63	$25,475.86	$1,088.10	$26,563.96	4.15
Human Development	90	$25,813.61	$1,265.64	$27,079.25	4.12
Economics	216	$25,385.83	$1,817.36	$27,203.20	4.20
Management Science	213	$25,650.79	$1,613.07	$27,263.87	4.19
Political Science and Specialties	256	$26,122.42	$1,274.13	$27,396.55	4.07
Microbiology	9	$26,269.37	$1,492.50	$27,761.87	4.14
History & Political Science and Specialties	7	$26,004.42	$1,850.23	$27,854.65	4.25
Economics & Political Science and Specialties	15	$25,614.44	$2,589.98	$28,204.42	4.27
International Studies-Sociol	31	$27,213.99	$1,104.28	$28,318.27	4.27
International Studies-Linguist	5	$27,153.27	$1,328.75	$28,482.02	4.20
Molecular Biology	40	$26,943.91	$1,547.18	$28,491.09	4.21
Human Biology	305	$27,039.56	$1,682.87	$28,722.44	4.23
Mathematics (Applied)	10	$27,828.17	$1,195.19	$29,023.36	3.98
Environmental Systems and Specialties	5	$27,935.87	$1,227.98	$29,163.86	4.70
General Biology	252	$27,223.66	$2,057.29	$29,280.96	4.35
Communications	148	$27,932.40	$1,368.70	$29,301.11	4.08
Mathematics	14	$27,852.94	$1,596.73	$29,449.67	3.96
Communications & Political Science and Specialties	7	$27,998.80	$1,508.89	$29,507.68	4.07
Physiology & Neuroscience	114	$27,762.95	$1,756.51	$29,519.46	4.27
Biochemistry and Cell Biology	257	$27,642.97	$2,049.69	$29,692.66	4.29
Ecology, Behavior & Evolution	29	$28,600.88	$1,294.61	$29,895.49	4.44
Biochemistry/Chemistry	45	$28,073.75	$1,954.66	$30,028.41	4.21
Environmental Systems and Specialties	40	$28,234.96	$1,957.20	$30,192.16	4.35
Sociology and Specialties	58	$29,117.89	$1,226.02	$30,343.91	4.25
Pharmacological Chemistry	52	$28,837.81	$2,244.80	$31,082.60	4.39
Chemistry	20	$30,494.16	$1,225.09	$31,719.25	4.39
International Studies-Anthro	7	$29,122.59	$2,673.74	$31,796.33	4.43
Political Science and Specialties & Sociology and Specialties	8	$30,778.45	$1,866.02	$32,644.47	4.19
Anthropology and Specialties	23	$31,124.18	$1,689.37	$32,813.55	4.13
Cognitive Science and Specialties	80	$30,943.50	$1,909.21	$32,852.71	4.35
Joint Math-Econ	9	$29,670.18	$3,590.83	$33,261.01	4.94
Probability & Statistics	7	$30,705.91	$2,918.23	$33,624.14	4.25
Visual Arts (Media)	37	$32,367.08	$1,281.13	$33,648.21	4.18
Ethnic Studies	28	$32,034.24	$1,935.84	$33,970.08	4.74
Literature (French, Spanish, English, World), Lit/Writing, or Lit Cultural	74	$32,604.37	$1,519.13	$34,123.50	4.18
Linguistics and Specialties, except Language Studies	13	$33,856.66	$941.75	$34,798.42	4.23
Visual Arts (Studio)	11	$32,399.66	$2,511.31	$34,910.97	4.57
Bioengineering and Specialties	114	$33,549.85	$1,553.45	$35,103.30	4.31
Visual Arts(Art Hist/Criticsm)	15	$34,021.72	$1,340.12	$35,361.84	4.37
Cognitive Science and Specialties	10	$33,112.10	$2,520.93	$35,633.03	4.56
Philosophy	16	$34,706.30	$2,341.40	$37,047.70	4.55
Interdisc Computing & the Arts	16	$35,893.48	$1,518.77	$37,412.24	4.52
Mathematics	8	$36,592.74	$1,864.12	$38,456.86	4.88
Chemical Engineering	40	$36,273.87	$2,338.92	$38,612.79	4.39
Computer Science	91	$36,719.21	$2,043.44	$38,762.65	4.46
Interdisc Computing & the Arts	14	$38,091.79	$823.80	$38,915.59	4.45
Theatre	16	$37,984.48	$1,392.05	$39,376.52	4.34
Structural Engineering	72	$39,483.88	$2,268.06	$41,751.94	4.60
Environmental Engineering	9	$37,309.38	$4,515.03	$41,824.41	4.56
Physics and Specialties	16	$41,146.91	$917.95	$42,064.86	4.05
Aerospace Engineering	55	$39,382.63	$3,015.68	$42,398.31	4.41
Mechanical Engineering	126	$39,127.46	$3,292.05	$42,419.51	4.65
Computer Engineering	14	$41,698.22	$1,058.53	$42,756.76	4.38
Electrical Engineering	59	$41,756.24	$1,514.85	$43,271.09	4.53
Music	8	$42,221.32	$1,872.81	$44,094.13	4.50

2.6.2 Technical Appendix

2.6.2.1 Bundled Majors

This listing reflects majors that we bundled into one major. Some majors have the same title but are under different codes – for instance, some Math-Econ majors are classified under math, whereas others are classified under Economics. Additionally, some specialized majors have very small enrollments – instead of showing the actual cost of a small number of students, these students were included along with other majors in their department.

1. Anthropology, Anthropology(Conc in Bio Anth), Anthropology(Conc in Archaeol), Anth (Conc Sociocultural Anth)
2. Bioengineering, Bioengineering: Pre-Medical, Bioengineering (Biotechnology), Bioengineering: Bioinformatics
3. Cognitive Science, Cogn Sci w/Specializ Human Cog, Cogn Sci w/Specializ Neurosci, Cogn Sci w/Spec Hum Comp Inter, Cogn Sci w/Spec Clin Asp Cogn
4. Two differently coded Computer Engineering Degrees
5. Joint Major Mathematics & Econ (Home: Econ Dept), Joint Major Mathematics & Econ (Home: Math Dept)
6. Environ Sys (Earth Sciences), Environ Sys (Ecol,Behav&Evol), Environ Sys(Environ Chemistry), Environ Sys (Environ Policy)
7. Linguistics, Linguistics(Spec Lang&Society), Linguistics(Spec Cogn & Lang)
8. French Literature, Spanish Literature, Literature/Writing, Literatures in English, Literatures of the World, Literature/Cultural Studies
9. Political Science, Political Sci/Amer Politics, Political Sci/Compar Politics, Political Sci/Intntl Relations, Political Sci/Political Theory, Political Sci/Public Law, Political Sci/Public Policy
10. Two differently coded Communication Degrees

11. Physics-Biophysics, Physics, Physics w/Specializ Mtrls Phys, Phys w/Spec Computational Phys, Physics w/Specializ Astrophys
12. Sociology, Sociology-American Studies, Sociology-Culture/Communic, Sociology-Economy and Society, Sociology-International Stu, Sociology-Law and Society, Sociology-Social Inequality
13. Two differently coded Computer Science Degrees

2.6.2.2 Technical Description of Cost Calculation

In order to compute the cost, we use the following variables:

$1(k)$	Identity function equal to 1 if course k is *not* a graduate teaching course or an undergraduate independent study course.
$1(k,m)$	Identity function equal to 1 if course k is taught by department m.
BG_m	Block grant awards for department m (FY2008-9).
C_m^q	Budget of department or division m, measure q.
$c_{m,q}$	Cost per credit hour for major m, measure q.
$c^{\kappa,q}$	Cost per credit hour for subject code κ, measure q.
${}_{d,q}c$	Cost per credit hour for division d, measure q.
$D(m)$	Division of major m.
DIV_m	Diversity awards for graduates in department m.
F_m^1	Office space allocated to department m (assigned ft^2 FY2008-9).
F_m^2	Classroom space allocated to department m (assigned ft^2 FY2008-9).
F_m^3	Teaching labs allocated to department m (assigned ft^2 FY2008-9).
F_m^4	Assembly space allocated to department m (assigned ft^2 FY2008-9).
F_m^5	Research space allocated to department m (assigned ft^2 FY2008-9).
F_m^6	Other space allocated to department m (assigned ft^2 FY2008-9).

h_k	Non-adjusted average credit hour taken for course k.
k	Course
$\mathbf{k}_\kappa$	Set of courses in subject code κ.
$\mathbf{k}_m$	Set of courses in major m.
κ	Subject code
$\boldsymbol{\kappa}_m$	Set of subject codes in department m.
LEC_m	Lecturer salaries in department m.
m	Department
$M(k)$	Function returning departmental home of course k.
OGS_m	OGS non-specified awards for department m.
P_k	Penner measure for course k.
S_k	Number of students in a course k.
SF_m	Budgeted support funds for department m (FY2008-9).
SAL_m	Faculty salaries for department m (FY2008-9).
TA_m	Teaching assistant salaries for department m (FY2008-9).
TAT_m	Teaching assistant tuition waivers for department m (FY2008-9).
TU_m	Tutor and reader salaries (FY2008-9).

The cost per credit hour calculation is given by:

$$C_m^1 = SF_m + SAL_m + LEC_m + TA_m + TU_m + DIV_m + BG_m + OGS_m + TAT_m \quad (2.6)$$

$$C_m^2 = C_m^4 = C_m^1 + 36\sum_{t=1}^{3} F_m^t$$

$$C_m^3 = C_m^1 + 36\sum_{t=1}^{6} F_m^t$$

$$c_{m,q} = \begin{cases} q \neq 4 & \dfrac{C_m^q}{\displaystyle\sum_{\kappa\in\boldsymbol{\kappa}_m}\sum_{k\in\mathbf{k}_\kappa} S_k h_k P_k \times 1(k,m)1(k)} \\ q = 4 & \dfrac{C_m^q}{\displaystyle\sum_{\kappa\in\boldsymbol{\kappa}_m}\sum_{k\in\mathbf{k}_\kappa} S_k h_k \times 1(k,m)1(k)} \end{cases} \tag{2.7}$$

$${}_{d,q}c = \begin{cases} q \neq 4 & \dfrac{C_d^q}{\displaystyle\sum_{\kappa\in\boldsymbol{\kappa}_d}\sum_{k\in\mathbf{k}_\kappa} S_k h_k P_k \times 1(k,d)1(k)} \\ q = 4 & \dfrac{C_d^q}{\displaystyle\sum_{\kappa\in\boldsymbol{\kappa}_d}\sum_{k\in\mathbf{k}_\kappa} S_k h_k \times 1(k,d)1(k)} \end{cases}$$

$$c^{\kappa,m} = \begin{cases} q \neq 4 & \dfrac{\displaystyle\sum_{k\in\mathbf{k}_\kappa} 1(k)\times S_k h_k P_k \left(c_{M(k),q} + {}_{d(k),q}c\right)}{\displaystyle\sum_{k\in\mathbf{k}_\kappa} S_k h_k P_k} \\ q = 4 & \dfrac{\displaystyle\sum_{k\in\mathbf{k}_\kappa} 1(k)\times S_k h_k \left(c_{M(k),q} + {}_{d(k),q}c\right)}{\displaystyle\sum_{k\in\mathbf{k}_\kappa} S_k h_k} \end{cases}$$

2.6.2.3 Student Utility Problem

We take the ratio of students as implied by a logit utility function as given instead of idiosyncratic. The utility function is $U_{ms} = b_1 \log \widetilde{\gamma}_m + b_2 \log \rho_m + b_3 \log \frac{w_m}{\bar{w}} + \epsilon_{ms}$. The ratio of the students is given by:

$$
\begin{aligned}
\frac{S_i|\mathbf{f}}{S_j|\mathbf{f}} &= \frac{\exp\left(b_1 \log \tilde{\gamma}_i + b_3 \log \frac{w_i}{\bar{w}}\right) \exp\left(b_2 \log \rho_i\right)}{\exp\left(b_1 \log \tilde{\gamma}_j + b_3 \log \frac{w_j}{\bar{w}}\right) \exp\left(b_2 \log \rho_j\right)} \\
&= \exp\left(b_1 \left[\log \widetilde{\gamma}_i - \log \widetilde{\gamma}_j\right]\right) \exp\left(b_2 \left[\log \rho_i - \log \rho_j\right]\right) \\
&\times \exp\left(b_3 \left[\log \frac{w_i}{\bar{w}} - \log \frac{w_j}{\bar{w}}\right]\right) \\
&= \left[\frac{\widetilde{\gamma}_i}{\widetilde{\gamma}_j}\right]^{b_1} \left[\frac{\frac{w_i}{\bar{w}}}{\frac{w_j}{\bar{w}}}\right]^{b_3} \left[\frac{\rho_i}{\rho_j}\right]^{b_2} \\
&= \left[\frac{\widetilde{\gamma}_i}{\widetilde{\gamma}_j}\right]^{b_1} \left[\frac{w_i}{w_j}\right]^{b_3} \left[\frac{\frac{f_i}{\sigma_i S_i}}{\frac{f_j}{\sigma_j S_j}}\right]^{b_2} \\
&= \left[\frac{\widetilde{\gamma}_i}{\widetilde{\gamma}_j}\right]^{b_1} \left[\frac{w_i}{w_j}\right]^{b_3} \left[\frac{f_i}{f_j} \frac{\sigma_j S_j}{\sigma_i S_i}\right]^{b_2} \\
&= \left[\frac{\widetilde{\gamma}_i}{\widetilde{\gamma}_j}\right]^{b_1} \left[\frac{w_i}{w_j}\right]^{b_3} \left[\frac{f_i}{f_j}\right]^{b_2} \left[\frac{\sigma_i}{\sigma_j}\right]^{-b_2} \left[\frac{S_i}{S_j}\right]^{-b_2} \\
\left[\frac{S_i|\mathbf{f}}{S_j|\mathbf{f}}\right]^{1+b_2} &= \left[\frac{\widetilde{\gamma}_i}{\widetilde{\gamma}_j}\right]^{b_1} \left[\frac{w_i}{w_j}\right]^{b_3} \left[\frac{f_i}{f_j}\right]^{b_2} \left[\frac{\sigma_i}{\sigma_j}\right]^{-b_2} \\
\frac{S_i|\mathbf{f}}{S_j|\mathbf{f}} &= \left[\frac{\widetilde{\gamma}_i}{\widetilde{\gamma}_j}\right]^{\frac{b_1}{1+b_2}} \left[\frac{w_i}{w_j}\right]^{\frac{b_3}{1+b_2}} \left[\frac{f_i}{f_j}\right]^{\frac{b_2}{1+b_2}} \left[\frac{\sigma_i}{\sigma_j}\right]^{-\frac{b_2}{1+b_2}}
\end{aligned} \tag{2.8}
$$

2.6.2.4 Some Additional Remarks Concerning Treatment

Cost calculation: TAT_m in Section 2.6.2.2 should be assigned to the department offering the stipend, but the data feed is based on where the student is located. For instance, if a student in Economics is a teaching assistant for Culture, Art, and Technology, we see lines related to ECON, but we only see the line for CAT where the CAT program budget specifically funds the student. Many of the other lines are from general funds, even though the CAT program has initiated these expenses. To assign costs to the proper department, we take this group of perhaps one or more students and assign the tuition and fee waiver to the departments in proportion to the instructional stipend paid by each department. We do not include the stipend in TAT_m, as this is included in TU_m and/or TA_m.

Student-level data for TAT_m is unavailable for FY2008[10]; for FY2008, we regress (without a constant) the FY2009 stipend on FY2009 TA salaries, and then we use the inflation-adjusted TA salary for FY2008 to estimate the FY2009 tuition waiver.

For SAL_m, we were unable to obtain non-salary benefits package summaries for departments at UCSD until late in the research. In actuality, these costs are assigned department-to-department by a formula which has little relevance to the benefits actually received or to the burden implied by any particular faculty or staff member.

[10]We received this data, but the student-level data for FY2008 does not match the summary statistics for FY2008 by a large margin, whereas a check for a few departments on the FY2009 data against summary statistics match within rounding, so we are certain FY2009 is accurate.

Variables BG_m, DIV_m, and OGS_m come from a data feed from the Office of Graduate Studies. Block grants are a department-by-department sum of money allocated by OGS; for more information on BG_m, see Arovas, et.al. (2010). The Material Science and Bioinformatics programs receive an allocation of BG_m, DIV_m, and OGS_m; we reassign these costs across relevant departments based on the course makeup of the graduate students in those departments.

On occasion, the department m is ambiguous. Firstly, the History Department is combined with the CAESER Program. This is because the CAESER Program, which includes majors such as Classical Studies, Russian & Soviet Studies, and others, is administered by the History Department. Secondly, we have separated out the Linguistics Department, which is a true academic department, from the Linguistics Language Program, which teaches undergraduate foreign language courses. While they are administered in the same department, they have separate budgets, and so many students take courses in the language program that it is worth separating. Next, the Nanoengineering department is created in the middle of the dataset. Thus, we combine it with a similar department that taught many of the courses prior to its founding, the Mechanical and Aerospace Engineering Department.

We do not have data for some departments. For instance, the Scripps Institute of Oceanography is a part of UCSD but is assigned a completely different budget process. These department codes, usually located in specialty departments, are assigned the average cost per credit hour[11].

[11]Department codes ERTH, LAWS, LATI, RELI, SOE, SIO, and UNAF. Code ERTH is different than ESYS; we do have data for ESYS. We use average adjusted under-

The $36 figure comes from lease information on office and research space for similar buildings in the nearby Torrey Pines area in San Diego; this information was provided by Newark Grubb Night Frank.

Provost and division calculations are meant to include administrative costs, but often include courses taught by those provosts and divisions as well. We do have information to properly separate several of these departments (i.e., writing programs, Muir Interdisciplinary and Critical Gender Studies), but for some specific courses, there is no information. Unfortunately, this biases both courses taught by those departments *and* the divisional administrative cost. In a degree aggregation, these costs will aggregate properly as long as the student taking these college-specific courses is in the college. Secondly, some division-unidentifiable courses were assigned the average divisional cost.

Information on Right-Hand Side Variables: Most right-hand side data is from Academic Affairs' Resource Profiles and averaged across two years (we do not average the averages; we take the numerator as the sum of the figures and the denominator as the sum of the faculty over both years). However, the percentage of courses taught by faculty is taken from "Teaching Statistics for the UCSD General Campus Academic Year 2007-2008, Excluding Summer Sessions." Since many programs have 0 FTE Faculty Members, many students graduating in these programs are not included in the regression. Mechanical Engineering and Nanoengineering are combined into one department. History is combined with CAESER for the cost-per credit hour calculation (and for lo-

---

graduate hours to weight these parameters, except in measure 4, which we use average unadjusted undergraduate hours.

cation of degree to determine department size), but for most right-hand side variables, we use only History funds. This is because there are no FTE faculty in CAESER, so adding extra office space in the numerator for a large, linked, but technically different, operation would make the combined History-CAESER department a strange outlier. A similar rationale works for Linguistics and the Linguistics Language Program.

A note on inflation: All data from FY2008 is inflated by 1.0140 (using an average of July-June CPI for both years). We refer to this figure as FY2009 dollars.

2.7 Acknowledgments for "An Analysis of the Cost of an Undergraduate Degree and the Incentives of the State, the University, and the Student"

This paper is co-authored with Richard Carson and Melissa Famulari. I would like to thank Melissa Famulari both for her assistance in getting the education chapter completed and for her expertise as a teaching mentor. I would like to thank Devaney Kerr, Financial Manager in the Mechanical & Aerospace Engineering Department, and Sue King, Chief Administrative Officer in the Economics Department, with help procuring and understanding the administrative data. Further, I would like to thank Christine Hurley and Kirk Belles. Christopher Beliare from the real estate firm Newmark Grubb Knight Frank assisted me in locating lease data for the UCSD area. Further, input from department chair Valerie Ramey was invaluable.

2.8 Bibliography

Arcidiacono, P. 2004. "Ability Sorting and the Returns to College Major." *Journal of Econometrics* 121:343-375.

Arovas, D.P., J.D. Cohen, and V.A. Ramey. 2010. "UCSD Graduate Council - Report of Block Grant Allocation Process Subcommittee." La Jolla, CA.

The Association for Legal Career Professionals. 2012. "Employment for The Class of 2011 – Selected Findings."

Babcock, P., and M. Marks. 2011. "The Falling time Cost of College: Evidence from Half a Century of Time-Use Data." *Review of Economics and Statistics* 93:468-78.

Bound, J., and S. Turner. 2007. "Cohort Crowding: How Resources Affect Collegiate Attainment." *Journal of Public Economics* 91:877-899.

Carter, R.E., and D.J. Curry. 2011. "Using Student-Choice Behavior to Estimate Tuition Elasticity in Higher Education." *Journal of Marketing Management* 27:1186-1207.

Carnegie Classification of Institutions of Higher Education, the. "University of California-San Diego."

de Groot, H., W.W. McMahon, and J.F. Volkwein. 1991. "The Cost Structure of American Research Universities." *Review of Economics and Statistics* 73:424-31.

Dundar, H., and D.R. Lewis. 1995. "Departmental Productivity in American Universities: Economies of Scale and Scope." *Economics of Education Review* 14:119-144.

Federal Reserve Bank of New York. "Student Loan Debt History." http://www.newyorkfed.org/studentloandebt/

Fu, T., C.J. Huang, and Y. Yang. 2011. "Quality and Economies of Scale in Higher Education: A Semiparametric Smooth Coefficient Estimation." *Contemporary Economic Policy* 29:138-49.

Halkos, G.E., N.G. Tzeremes, and S.A. Kourtzidis "A DEA Approach for Measuring University Departments' Efficiency." Working Paper.

Hoenack, S.A., W.C. Weiler, R.D. Goodman, and D.J. Pierro. 1986. "The Marginal Costs of Instruction." *Research in Higher Education* 24:335-415.

Hoxby, C.M. 1997. "How the Changing Market Structure of U.S. Higher Education Explains College Tuition."

Johnson, N. 2009. *What Does a College Degree Cost?* Delta Cost Project.

Kao, C., and H. Hung. 2006. "Efficiency Analysis of University Departments: An Empirical Study." *Omega* 36:653-64.

Michels, S. 2009. "In California, Budget Cuts and Higher-Priced Education," Nov. 20.

Nelson, R., and K.T. Hevert. 1992. "Effect of Class Size on Economies of Scale and Marginal Costs in Higher Education." *Applied Economics* 24:473-82.

Organization for Economic Co-operation and Development. 2011. *Education at a Glance*, OECD Publishing.

Romano, R.M., and Y.M. Djajalaksana. 2010. "Using the Community College to Control College Costs: How Much Cheaper is It?" Cornell University School of Industrial and Labor Relations Working Papers.

Romano, R.M., R. Losinger, and T. Millard. 2010. "Measuring the Cost of a College Degree: A Case Study of a SUNY Community College." Cornell University School of Industrial and Labor Relations Working Papers.

Segal, D. 2012. "For 2nd Year, a Sharp Drop in Law School Entrance Tests." *The New York Times*.

"Teaching Statistics for the UCSD General Campus Academic Year 2007-2008, Excluding Summer Sessions"

UC Office of the President. 2011 "The UC Budget: Myths and Facts."

—. 2009. "The UC Budget: Myths and Facts."

University of California, San Diego. 2007. *UCSD General Catalog 2007-2008*. La Jolla, CA.

—. 2008. *UCSD General Catalog, 2008-2009*. La Jolla, CA.

—. "The College System: FAQ." http://admissions.ucsd.edu/colleges/about/faq.html.

U.S. News and World Report College Profiles. 2015. "University of California – San Diego."

Watson, R.N. 2010. "The Humanities Really Do Produce a Profit." *The Chronicle of Higher Education*.

Webley, K. 2013. "College Costs: Would Tuition Discounts Get More Students to Major in Science?" *Time*.

Chapter 3

Pollution Whack-a-Mole: Ambient Acetaldehyde and the Introduction of E-10 Gasoline in the Northeast

3.1 Chapter Abstract

This paper uses a complicated set of phase-ins and phase-outs of oxygenated motor fuel in the Northeast to determine whether E-10 ethanol-enhanced fuel contributes to acetaldehyde air pollution over the pre-ethanol methyl tertiary-buthyl ether (MTBE) fuel. Oil companies phased out MTBE because of groundwater pollution concerns, and now E-10 is the standard fuel in EPA reformulated gas areas. Using a difference-in-difference approach, I find a large percentage increase in acetaldehyde pollution is associated with the switch from MTBE to E-10. Using EPA carcinogenic estimation techniques, I find that the cost of this

increase in acetaldehyde pollution is around $3 million annually for the New York City Metropolitan area. This smaller cost estimate comes from a pollution increase that – while large in percentage terms – is small in level terms.

3.2 Introduction

The make-up of motor vehicle fuel impacts the air we breathe. Increasingly complicated phase-ins and phase-outs of gasoline oxygenate requirements occurred from 1973-2006. In this paper, I explain how one particular phase-out and phase-in can be used to measure the impact of E-10 gasoline – on one particular air pollutant, acetaldehyde – in one particular region, the Northeast United States. Because of the complexity of the regulations, I begin the paper with an overview of the regulatory environment followed by a brief overview of acetaldehyde.

This paper agrees with scientific papers which find small, positive increases in acetaldehyde from E-10. This study does find large percentage impacts, however, as prior acetaldehyde pollution is low in this region. Here, I use a changing regulatory environment and monitor data to compare E-10 gasoline to MTBE-enhanced gasoline. Previous scientific work on the problem has not utilized social counterfactuals; I will use a control group of states as a counterfactual. Additionally, this paper will also compute approximate acetaldehyde pollution costs for a large city. Since I only look narrowly at this one type of pollution, I cannot make a larger determination about air pollution from E-10; however, few economic papers have considered novel air pollutants from ethanol-enhanced motor fuels.

3.3 Regulatory Framework

The United States phased out lead gasoline beginning in 1973, leading to, "one of the great environmental achievements of all time," preventing large amounts of lead poisoning (U.S. EPA 1996). Lead was an octane enhancer; octane helps prevent engine knocking. Oil companies needed a substitute to keep octane levels high, so they began adding methyl tertiary-buthyl ether (MTBE) (U.S. EPA [7]).

In 1990, the Clean Air Act Amendments (CAAA90) compelled oil companies to add even higher amounts of MTBE. In 1996, however, Santa Monica, CA, discovered MTBE leaked out of underground fuel tanks and polluted groundwater. This lead many states to ban MTBE, and the industry phased out MTBE in 2006. In the process, the industry switched to E-10, a 10%-ethanol enhanced gasoline, which also met reformulated gasoline requirements (U.S. EPA [3]; U.S. EPA [7]).

The phase-in and phase-out of MTBE occurred in several stages. While MTBE was used in much of the country, in actuality, several additives were available. After the CAAA90, in the Midwest, oil companies used ethanol, and elsewhere, they used MTBE (U.S. EPA [7]; U.S. EPA [9]). Ethanol has one particular disadvantage that caused these separate markets – it is not easily mixed into gasoline and must be added close to sale (U.S. EIA 2006).

After the Santa Monica water pollution discovery, states moved to ban

MTBE. Connecticut and New York did this in 2004, so they began receiving E-10 while the rest of the Northeast continued receiving MTBE-enhanced gasoline (U.S. EIA 2003; U.S. EIA 2006; U.S. EIA Office of Oil and Gas 2003; U.S. EPA [9]). This changed in 2006 when oil companies moved rapidly to rid the system of MTBE, fearing pollution liabilities. This was realized in 2013 when a New Hampshire jury fined Exxon Mobil $236 million for MTBE-related pollution (Tuohy 2013, U.S. EIA 2006; U.S. EPA [9]).

3.4 A Description of Acetaldehyde

Acetaldehyde (CH_3CHO) is a “probable human carcinogen” that causes skin, eye, and lung irritation (U.S. EPA 2000). Scientific models predict large increases in this substance when ethanol is burned. One of these studies, Jacobson (2007), found a 2000% increase in acetaldehyde pollution in Los Angeles in 2020 if the city switched from a baseline gasoline to E-85.

In contrast, none of the papers on E-10 find these large increases. A public review draft on the California transition to ethanol predicted only small increases in acetaldehyde over non-ethanol fuels (Allen et al. 1999). Anderson, Lanning and Wilkes (1997) used an ARIMA model and found no impact on acetaldehyde when Denver, CO, switched from MTBE to E-10.

Several papers look at acetaldehyde pollution in Brazil, which has high ethanol consumption. Goldemberg, Coelho and Guardabassi (2008) look at the transition to ethanol and do not find a concerning level of acetaldehyde pollution in the São Paulo region. An earlier paper, however, Grosjean, Miguel and

Tavares (1990) finds high levels of acetaldehyde in the same region.

Acetaldehyde has a very fickle atmospheric residence time. During the day, it is relatively short. In St. Louis on a clear July day, acetaldehyde has a 3 hour residence time; in New York, it is 5 hours. On a cloudy or rainy July day, this ups to 6 hours in St. Louis and 11 hours in New York. However, this rapidly increases to 170 hours (St. Louis) and 40 hours (New York) at night on a clear July day. On a clear January night, it has a 3000 hour residence time in St. Louis and New York (U.S. EPA Technical Support Branch 1993). Thus, in the summer, there is a short residence time during the day and a long one in the evening.

3.5 Environmental Economics Research on Pollution

Environmental economists have utilized monitor data and economic tools to answer regulatory and economic questions. These studies have looked at a variety of air pollution topics – the impact of total suspended particulates on infant mortality (Chay and Greenstone 2003), whether the Clean Air Act and Amendments had an impact on SO_2 levels (Greenstone 2004), and even whether agricultural workers in California's Central Valley are less productive when there are high levels of ground-level ozone (Graff Zivin and Neidell 2012).

More specifically to this paper, gasoline and driver regulations have been studied extensively. High levels of air pollution in Mexico City led to the city passing *Hoy No Circula*, a policy which required drivers to avoid using their cars a particular day of the week based on their license plates. Davis (2008) finds that drivers utilized different cars and taxis to get around the regulation, and no criterion pollutant in the study went down. Chakravorty, Nauges and Thomas (2008) finds that market segmentation in the United States increases cost. Finally, this paper takes one approach used in Auffhammer and Kellogg (2011). In this paper, the authors find that gasoline regulations in the United States have not largely lowered ozone levels with the exception of regulations in California.

3.6 Natural Experiment and Data

Connecticut and New York phased out MTBE in 2004, so they began receiving E-10 while the rest of the Northeast continued receiving MTBE-enhanced gasoline (U.S. EIA 2003; U.S. EIA 2006; U.S. EPA [9]). I use EPA's reformulated gas survey from 2004-2006 to generate levels of ethanol in the gasoline by metropolitan area (U.S. EPA [9]), and I match monitors from EPA's AQS Datamart (U.S. EPA [2]) to metropolitan areas using an online lookup tool (Silver Biology) and metro data from the U.S. Government Accountability Office (2004). From both an internet archive of the survey explanation (U.S. EPA [6]) and personal communication (Lenski), the gasoline survey reflects the gas sold in each metro area. Figure 3.1 shows the percentage ethanol in the gasoline in each metro area.

The survey and the report from the EPA (U.S. EPA [6]) indicate that

MTBE was transitioned during the Winter 2006 driving season. From the survey, all of the MTBE was out by Summer 2006 and was perfectly substituted to ethanol. So, while there was some variation from 2004-2006, all areas received treatment in 2006. Because of the transition, I focus on summer gasoline. Further, in 2005, Hurricane Katrina resulted in a waiver of summer gasoline requirements (Kumins and Bamberger 2005), so I drop all observations after August 22 (Dyre 2005; U.S. EPA [5]).

I must focus on the Northeast for another particular reason in this setup. No other area of the country is free from ethanol plants, which are likely sources of acetaldehyde pollution. The EPA is monitoring acetaldehyde, for instance, in Lynn County, IA (Kintz, Lundberg, and Dodge 2011). Figure 3.2 shows the ethanol plants that were operated according to a 2006 snapshot of Ethanol Producer. As expected, the Midwest is awash in ethanol production, which increased in the 2006 season (Renewable Fuels Association). California's RFG surveys are not available for a portion of the study (U.S. EPA [9]), and other areas of the country pose other problems, not least of which is the fact that these are completely different air spaces.

Other data used in this analysis includes annual per-capita gross metropolitan product from the U.S. Bureau of Economic Analysis (2013) in chained 2005 dollars. Monthly miles traveled were downloaded by state from the U.S. National Highway Safety Administration through Pro Quest Statistical Datasets, and these were divided by the estimated state population from the United States Census Bureau through Pro Quest Statistical Datasets. Metro area populations were from the U.S. Census Bureau. MSA's were determined from a listing mapping counties to MSA (U.S. GAO). RFG counties were from a listing archive from the EPA (U.S. EPA [8]). I exclude 24 observations from a monitor in rural Essex County, NY, near Whiteface Mountain[1](Foy 1994; U.S. Code of Federal Regulations 2003, 40 §80.70).

Acetaldehyde monitors used in the report are shown in Figure 3.3. To be included in the analysis, the monitor must have had at least one sample before and after the main ethanol transition in 2006. Acetaldehyde monitors were matched to weather monitors through a canned distance matching algorithm. First, I downloaded a set of monitors from the National Oceanic and Atmospheric Administration National Climate Data Center database. However, the closest monitor often did not have the requisite weather variables. Thus, while I found the closest monitor from this database for each variable, I also downloaded weather monitor data for nearby airports. Since the monitors are in locations that are highly urban, there are airport weather stations sufficiently close to the acetaldehyde monitors (the maximum distance from the algorithm is 27.6 miles; the mean is 10.5 miles). Because of the reliability of airport monitors is excellent, I will use this data for the exposition in this paper. Using the alternative weather variables do not change the results substantially, and using them also requires a complicated algorithm and assumptions about the time of day highs occurred.

[1]From Paul Foy of Albany, NY's Daily Gazette, December 7, 1994: "It is one of the odd mandates of the state's clean-air program that only reformulated gasoline can be sold above 4,500 feet on Whiteface mountain.

"There are no gas stations on Whiteface Mountain."

Monitors reported either every 24 hours or every 3 hours. AQS Datamart has collocated monitors in the same location on occasion, per personal communication with the EPA (Mangus). If the “Data Source Reference ID” was different, I considered this a separate monitor for the purposes of the analysis. For the purposes of matching to weather data only, the date was moved back one day if the monitor was a three hour and began at 5 A.M. or earlier. The date was moved forward one day if the monitor was a 24 hour reporting monitor beginning 1:01 P.M. or after; in this case, the majority of the day was actually the next day. This was only for purposes of matching these to weather variables; for the main analysis, the actual day was used.

3.7 Data

A plot of median acetaldehyde measures are shown in Figures 3.4 and 3.5. These are arranged by state – except for New York and New Jersey, where the New York City metropolitan area is separated from the rest of the state (these are separated because the New York metropolitan area starts out with around half E-10, half MTBE in 2004).

Figure 3.4 shows that New Jersey is a major outlier, even ignoring the unusually high acetaldehyde readings in the New Jersey suburbs of Philadelphia in 2004. Figure 3.5 excludes New Jersey. I run the model with both New Jersey and without it. Notably, from Figure 3.5, the controls, New York and Connecticut start out above other states and trend downward in 2006. Some states trend upward, such as Virginia. The monitors are reporting very low levels of acetaldehyde, in the 0-3 ppbC range.

3.8 Specification

I run a difference-in-difference setup for all of the monitors, only the 24 hour monitors, and only the three hour monitors. Additionally, I estimate the model with a weighting scheme to try to control for over-sampling of some areas. The long-form specification is as follows:

$$A_{imt} = 10\delta E_{mt} + T_t + I_i + R_{mt} + \boldsymbol{\beta}'\mathbf{Weather}_{it} + \boldsymbol{\gamma}'\mathbf{Regional}_{mt} + \varepsilon_{imt} \tag{3.1}$$

Here, A_{imt} is the level of acetaldehyde in ppbC for monitor i, metro area m, and time t, E_{mt} is the amount of ethanol in the gasoline as a decimal. Here, δ is the impact of E-10. T_t is a time dummy, either annual or monthly. I_i is a monitor fixed effect. $\mathbf{Weather}_{it}$ is a set of weather variables from airport monitors, and $\mathbf{Regional}_{mt}$ is a set of controls. ε_{imt} is error.

Table 3.1 shows this regression on levels. In columns (5) and (6), the impact of E-10 is a 1.03 ppbC increase in acetaldehyde levels. Including New Jersey seems to increase the δ coefficient, from columns (1)-(4). I report both robust standard errors and standard errors clustered by metropolitan area, which is the level of analysis across many of the explanatory variables.

In specification (7) in table 1, I exclude controls and the ethanol coefficient. I then run the regression to see which years had the highest acetaldehyde levels. From the regression, weather and monitor-controlled acetaldehyde levels were trending downwards in 2004-2005, but they spiked in 2006 (the absorbed year).

Results for other demographic variables in Regionalmt are not reported in Table 3.1. They are reported for log-log regressions, which I will discuss later; however, the small variation in regional variables between years do not absorb much of the variation and lead to numerical issues. The monitor fixed effects over-fit the model to accommodate the new, numerically unidentified variable. If the monitors are in a fixed spot and these values do not change substantially over the three year window, then monitor fixed effects will absorb much of the variation of the effect of $\mathbf{Regional}_{mt}$. Let ρ_i denote the approximate regional values for monitor i. Then, if $\mathbf{Regional}_{mt}$ does not change substantially over time:

$$\begin{aligned} \boldsymbol{\rho}_i &\approx \mathbf{Regional}_{m1} \approx \cdots \approx \mathbf{Regional}_{mN} \\ A_{imt} &\approx 10\delta E_{mt} + T_t + I_i + R_{mt} + \boldsymbol{\beta}'\mathbf{Weather}_{it} + \boldsymbol{\gamma}'\boldsymbol{\rho}_i + \varepsilon_{imt} \qquad (3.2) \\ A_{imt} &\approx 10\delta E_{mt} + T_t + R_{mt} + (I_i + \boldsymbol{\gamma}'\boldsymbol{\rho}_i) + \boldsymbol{\beta}'\mathbf{Weather}_{it} + \varepsilon_{imt} \end{aligned}$$

Adding metro population and GDP to the regression changes the estimate on δ to 0.946 (a difference of 0.09 ppbC), but it also changes the monitor fixed effects to unreasonable values ranging from -2.6 to 90.9 ppbC. Since I am not interested in $\boldsymbol{\gamma}$ per se, allowing monitor fixed effects to take care of this is fine.

Tables 3.2 to 3.5 use log specifications instead of levels. A total of 138 (out of 4,574) data points read 0 ppbC acetaldehyde (none in New Jersey). In this case, I specified two different modifications. The first is simply adding 0.001

pbbC acetaldehyde to all observations before logging, and the other is setting the reading equal to log(max(pbbC acetaldehyde,0.1)). Specification (8) in Table 3.2 shows an estimate of exp(0.641), or a near doubling of acetaldehyde pollution under E-10. Notably, this agrees with the level specification, as acetaldehyde pollution remains low in both regressions. It does, however, increase substantially in percentage terms.

Table 3.3 is slightly different. Here, I have substituted T_t with a time trend, ϑt. Notably, day-to-day, acetaldehyde seems to be going down, but when ethanol is introduced, it rebounds. However, this effect is not statistically significant when clustered by metro area. In Table 3.4, I drop monitor fixed effects to try to identify regional variables. However, I get little fit from the regression, and the impact does not change substantially.

Lastly, in Table 3.5, I add weights to the regression to balance oversampling in some regions. Let Z equal the number of counties in the analysis. Let N_{CO} equal the observations in the analysis for a particular county (CO), and let N equal the total number of observations. Then, the weight is[2]:

$$w_{CO} = \frac{1/Z}{N_{CO}/N} \tag{3.3}$$

The numbers change slightly, but they still indicate a large positive percentage change but small level change in acetaldehyde pollution.

[2] I consulted http://www.atlas.illinois.edu/support/stats/resources/spss/create-post-stratification-weights-for-survery-analysis.pdf for a description of how to form weights from survey data.

3.9 Robustness Checks

The regression on levels suffers from having values near zero. It is not possible to have negative values. Additionally, some monitors have detection limits as high as 0.6 ppbC acetaldehyde. The regression on logs somewhat ameliorates this; however, as a robustness check, I also run a Poisson regression. Here, I "count" the number of units of 0.6 ppbC acetaldehyde; the dependent variable is $\lfloor A_{imt}/0.6 \rfloor$. While this is clearly inferior to the log regression in that continuity is lost, it has a few advantages – numbers below any monitor's detection limit are bundled together, and there is no probability of values less than 0.

Table 3.6 lists two Poisson regressions, one with robust standard errors and one with clustered by metro area. The estimate for the ethanol variable is 0.581, with a corresponding IRR of 1.79, indicating a near 80% increase in acetaldehyde pollution with E-10. Thus, the findings are robust to the MDL.

Next, define the following:

$$\Theta_m = E_{m(t \in \text{year } 2006)} - \frac{E_{m(t \in \text{year } 2004)} + E_{m(t \in \text{year } 2005)}}{2} \tag{3.4}$$

Here, Θ_m is an intensity of treatment measure. Higher values of Θ_m indicate larger values of ethanol change from 2004-2005 to 2006. To test whether the treatment areas are different from non-treatment areas, I also run the following regression.

$$A_{imt} = \begin{aligned} &10\omega_1\Theta_m + 10\omega_2\Theta_m 1\left(t \in \text{year } 2006\right) + R_{mt} + \theta t + \text{Month}_t + \\ &\boldsymbol{\beta}'\textbf{Weather}, \textbf{Miles}_{mt} + \epsilon_{imt} \end{aligned} \tag{3.5}$$

In equation 3.5, I cannot identify monitor fixed effects because metro area m contains many monitors. Thus, a regression of Θ_m on I_i yields an R^2 of 1. I am interested in both ω_1 and ω_2. If ω_1 is statistically significant, then the change in ethanol is correlated with acetaldehyde measures. Now, ω_2 is the intensity-controlled measure.

I do find that ω_1 is statistically significant and negative. Here, I am taking the conservative approach to not reject ω_1 of using robust (as opposed to metro-area clustered standard errors). However, even including ω_1 in the regression, I find ω_2 is statistically significant, even using metro-area clustering, which is now the conservative choice. The value for ω_2 is 0.641, which is 62.1% of estimate Table 3.1, specification (5).

Since equation 3.5 cannot identify monitor fixed effects, the new difference-in-difference regression does not control for monitors – it's not possible to know whether the actual answer is 0.641 or 1.033 ppbC acetaldehyde. However, as I mention in the conclusion section, both answers are incredibly small in comparison to the amount of damage MTBE causes to the water table.

3.10 Cancer Risk & Conclusions

In the specification in Table 3.1, Specification (5), and Table 3.7, Specification (7), I find that E-10 likely increased by around 1.033 ppbC and 0.641 ppbC, respectively. This translates to 0.516 ppbV acetaldehyde and 0.321 ppbV acetaldehyde (Holland 2001). From the EPA's approximation of 1 ppm acetaldehyde = 1.8 mg / m^3 and U.S. EPA 2000; Satterfield 2004, 0.516 ppbV = 0.000516 ppbV = 0.000929 mg / m^3 and 0.321 ppbV = 0.000321 ppbV = 0.000577 mg / m^3. The EPA estimates that the risk of developing cancer over a lifetime is equal to this final figure divided by 500 (U.S. EPA 2000). This is equal to 1.86×10^{-6} in the first case, and it's equal to 1.15×10^{-6} in case 2. Using 78.54 years as life expectancy (World Bank 2010), a metro of 19 million (like New York (U.S. Census Bureau)), this would equal one cancer every 2.2 years in the first case – and one cancer every 3.58 years in the second case. An upper bound assuming mortality for each cancer and a Department of Transportation Value of Statistical Life of $9.1 million (Trottenberg and Rivkin 2013), this policy costs $4.09 million annually in the first case – and $2.54 million annually in the second case. If the U.S. switched to E-10 and faced similar impacts to the urban Northeast environment, assuming 310 million people, the annual cost is $66.8 million in the first case and $41.5 million annually in the second case.

While I have compared pollution to MTBE here, it's difficult to make a general conclusion here about the use of E-10 because it is difficult both to determine which gasoline should be the comparison. Firstly, it is possible to produce high-quality gasoline without oxygenates. In 2004, oil companies provided

California-standard gasoline without oxygenates to the non-EPA RFG-required San Francisco Bay Area (Fong et.al. 2005).

Secondly, if MTBE-enhanced gasoline is in actuality the next best gasoline for comparison, E-10 may be the better additive. There are other unknowns in MTBE use, but even the knowns indicate extreme cost. It is still uncertain whether MTBE is carcinogenic. According to the EPA, "...the data support the conclusion that MTBE is a potential human carcinogen at high doses" (U.S. EPA 2012). This study does not look at MTBE groundwater pollution and its carcinogenic impact. Further, MTBE groundwater pollution is very costly, but the cost estimates very substantially. The American Water Works Association (2005) estimates the costs could range from $4 billion to $85 billion. Based on the assessment here, even assuming groundwater pollution in their current locations, and switching E-10 to the entire country, $4 billion would pay for 59.9 non-discounted payments of $66.8 million. Additionally, the point estimates used in this calculation may suffer from large pharmacokinetic variances not available because of lack of variance data on point estimates (Rogers, et.al. 2011).

While I emphasize that the impact of E-10 on acetaldehyde pollution appears small, the U.S. has made a decision to rid one pollution at the expense of another. This was true when the U.S. ridded itself of lead gasoline for MTBE-enhanced gasoline. While both of this switch and the latter switch to ethanol may have been better for the environment at the time, policy makers should be aware of the tradeoffs and the new consequences of different fuel additives. Additionally, economic non-pollution factors are also a consideration. And, from

Chakravorty, Nauges, and Thomas (2008), the U.S. has severe market segmentation in gasoline. All of these factors need to be considered in the decision about gasoline in the near-future.

3.11 Figures

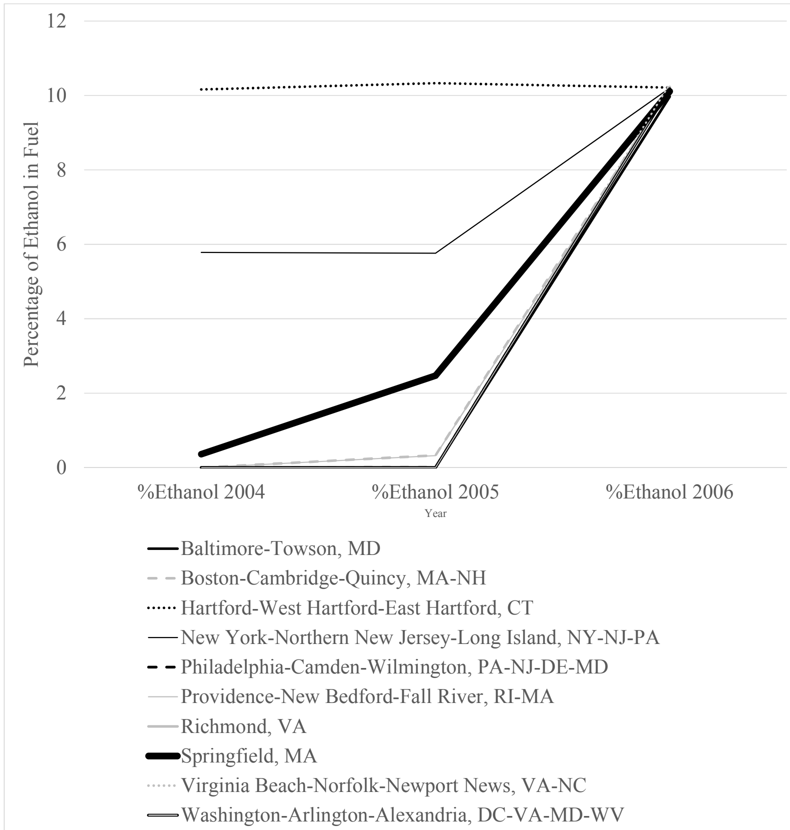


Figure 3.1: Percentage of Ethanol Present in Fuel by Metro Area. The difference-in-difference setup in this paper relies on a differential ethanol fuel regulatory regime. This data comes from (U.S. EPA [9]), and from personal communication and the EPA (Lenski 2013; U.S. EPA [6]), the surveys are representative of fuel sold in the region. Hartford, CT, is an obvious control, but other metro areas also had some ethanol content in their gasoline before 2006. The super thick line in the center is Springfield, MA.

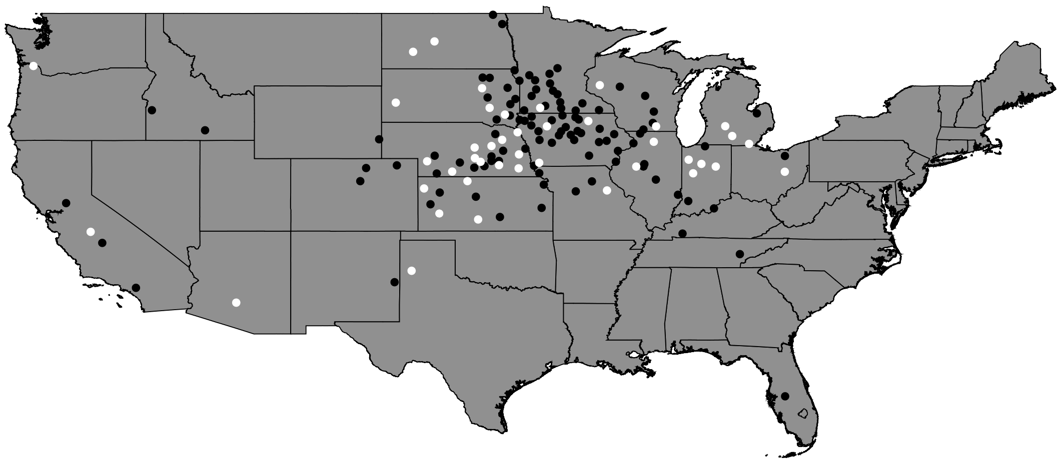


Figure 3.2: U.S. Ethanol Plants in 2006. This is a snapshot of ethanol plants (black dot) and ethanol plants under construction (white dot) in 2006. The data comes from Ethanol Producer magazine, and dots indicate the city where the plant was located, not the plant itself – cities were matched to coordinates with a matching routine in R (Loecher 2013) with Google Maps (2015). Ethanol plants are concentrated in the Midwest, and ethanol production has increased throughout the 2000's (Renewable Fuels Association). Under reasonable assumptions, this would lead to an increase in acetaldehyde, and an attenuation of the treatment effect, if I used the Midwest as a control.

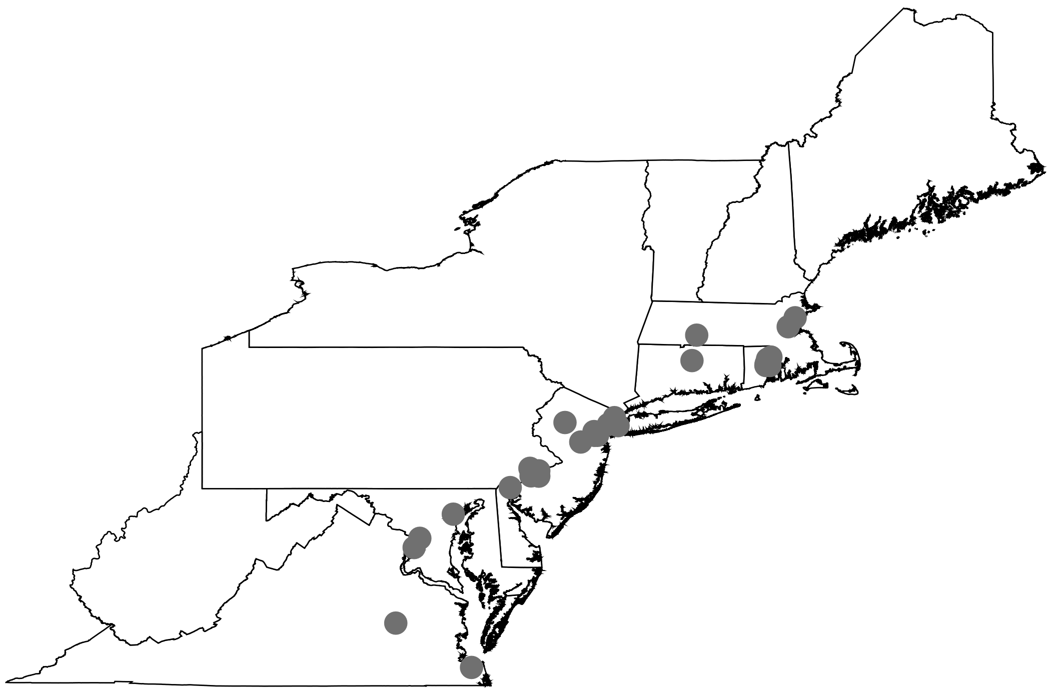


Figure 3.3: Acetaldehyde Monitors. This is a map of the EPA acetaldehyde monitors used in the report. All of the monitors are in urban areas. An interactive Google Map of the monitors is available on the author's website.

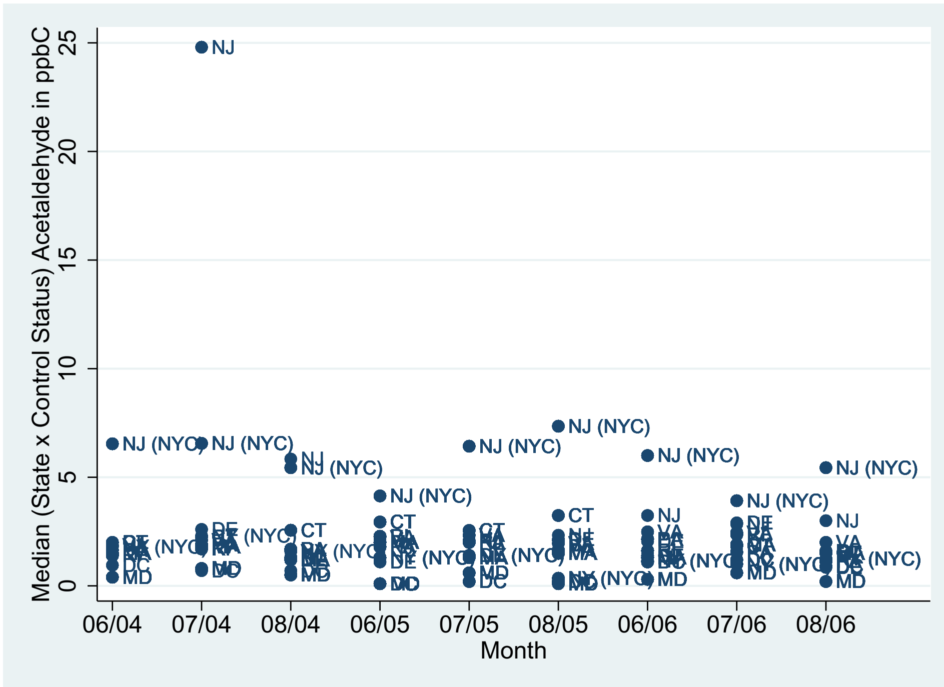


Figure 3.4: Median Acetaldehyde Measures in Each State. This scatterplot shows the median acetaldehyde monitor reading in ppbC for each state – with the New York metro area separated from other parts of New York and New Jersey. Ignoring the extreme outlier in the New Jersey suburbs of Philadelphia, acetaldehyde readings are really low. However, New Jersey, including the New Jersey suburbs of New York, appears to be an outlier. Figure 3.5 shows the same plot with New Jersey excluded.

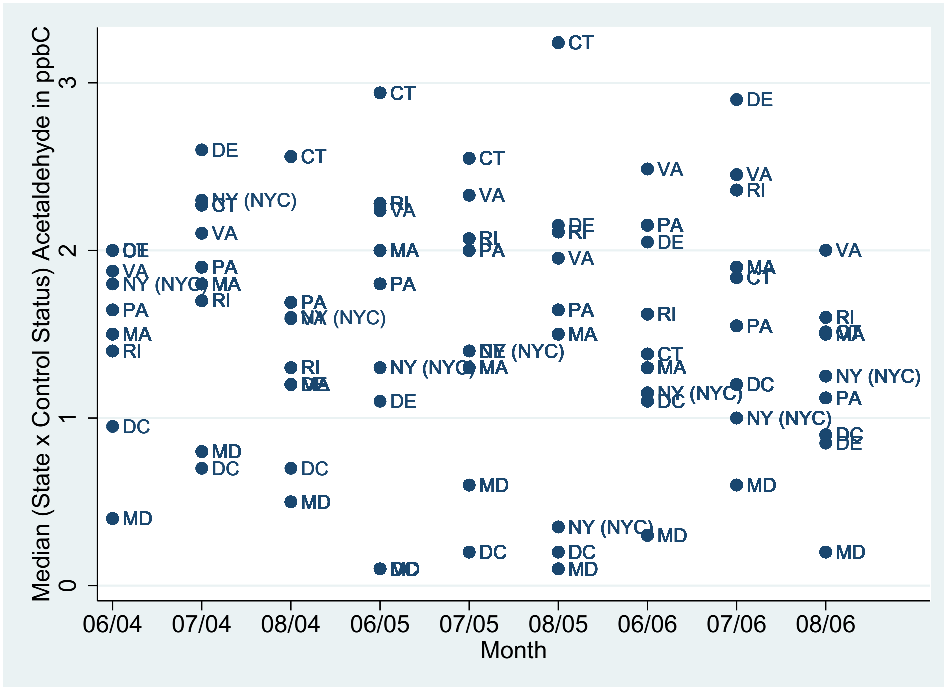


Figure 3.5: Median Acetaldehyde Measures Outside of New Jersey. This scatterplot, unlike Figure 3.4, excludes New Jersey and shows the median acetaldehyde monitor reading in ppbC for each state. Acetaldehyde readings are very low, and, as a group, there is no discernible trend among the treatment group. However, New York and Connecticut trend downward in the period.

3.11.1 Tables

Table 3.1: Regression on Levels. From specification (5), E-10 likely adds 1.03 ppbC acetaldehyde in the atmosphere for the typical urban area studied during the summer. Hour × Duration Bins are fixed effects where the monitors are separated into 6 bins based on the time of day and the duration of the monitor. Including monitors in New Jersey raised the coefficient between (1)-(2) and (3)-(4), but New Jersey is an outlier, as described in the text. Specification (7) excludes the ethanol coefficient and the control states. Notably, 2006 had the regression-controlled highest level of acetaldehyde. Hour × Duration Bins are 3-hour: (a) 12:00-18:59:59, (b) 19:00-22:59:59, (c) 23:00-5:59:59, (d) 6:00-11:59:59, 24-hour: (e) 0:00 or 23:00, and (f) 12:00.

VARIABLES	(1) ppbC Acetaldehyde	(2) ppbC Acetaldehyde	(3) ppbC Acetaldehyde	(4) ppbC Acetaldehyde
	Robust	*Clustered - Metro*	*Robust*	*Clustered - Metro*
	All Data	*All Data*	*NJ excluded*	*NJ excluded*
10% Ethanol = 1	1.466***	1.466*	0.844***	0.844**
	(0.140)	(0.697)	(0.0677)	(0.336)
Year Dummy (2004)	1.320***	1.320	0.584***	0.584
	(0.161)	(0.762)	(0.0771)	(0.362)
Year Dummy (2005)	1.201***	1.201*	0.626***	0.626*
	(0.141)	(0.603)	(0.0737)	(0.279)
Constant	0.489***	0.489	1.044***	1.044**
	(0.148)	(0.746)	(0.0788)	(0.453)
Observations	4,574	4,574	4,431	4,431
R-squared	0.032	0.032	0.024	0.024

*** p<0.01, ** p<0.05, * p<0.1

Table 3.1. Regression on Levels, *continued.*

VARIABLES	(5) ppbC Acetaldehyde	(6) ppbC Acetaldehyde	(7) ppbC Acetaldehyde
	Robust	*Clustered - Metro*	*Robust*
	NJ excluded	*NJ Excluded*	*NJ, NY, CT excluded*
10% Ethanol = 1	1.033***	1.033***	
	(0.0979)	(0.159)	
Year Dummy (2004)			-0.236***
			(0.0495)
Year Dummy (2005)			-0.262***
			(0.0477)
1 = 062004	1.339***	1.339***	
	(0.120)	(0.146)	
1 = 072004	1.118***	1.118***	
	(0.107)	(0.0871)	
1 = 082004	1.106***	1.106***	
	(0.111)	(0.139)	
1 = 062005	1.408***	1.408***	
	(0.103)	(0.183)	
1 = 072005	0.942***	0.942***	
	(0.102)	(0.131)	
1 = 082005	0.651***	0.651***	
	(0.104)	(0.106)	
1 = 062004	0.815***	0.815***	
	(0.105)	(0.219)	
1 = 072004	0.212***	0.212	
	(0.0788)	(0.135)	
Airport Daily Maximum Temperature	0.0114***	0.0114***	0.00945***
(Tenths Degrees Celcius)	(0.000468)	(0.00157)	(0.000435)
Airport Daily Average Wind Speed	-0.0177***	-0.0177**	-0.0156***
(Tenths m/s)	(0.00184)	(0.00584)	(0.00186)
Per Capita Miles Traveled	-0.139	-0.139	-3.268***
(1000s)	(0.574)	(1.602)	(0.558)
MONITOR FIXED EFFECTS	YES	YES	YES
HOUR X DURATION BINS	YES	YES	YES
Constant	-0.334	-0.334	0.290
	(0.694)	(0.955)	(0.327)
Observations	4,431	4,431	3,945
R-squared	0.325	0.325	0.276

*** p<0.01, ** p<0.05, * p<0.1

Table 3.2: Log Levels. Two separate log formulations (to deal with zero readings) were regressed on policy variables. In specification (8), E-10 nearly doubled at exp(.641), and in specification (9), the result was exp(.553). Since the average acetaldehyde levels in the region were low, these results are consistent with the level specifications. [log(ppbC Acetaldehyde + 0.001) is LOG M-I, log(max(ppbC Acetaldehyde, 0.1)) is LOG M-II]

VARIABLES	(8) LOG M-I	(9) LOG M-II	(10) LOG M-1	(11) LOG M-II
	Robust	*Robust*	*Clustered - Metro*	*Clustered - Metro*
	NJ excluded	*NJ excluded*	*NJ excluded*	*NJ excluded*
10% Ethanol = 1	0.641***	0.553***	0.641*	0.553**
	(0.0880)	(0.0588)	(0.312)	(0.229)
1 = 062004	0.862***	0.752***	0.862**	0.752***
	(0.112)	(0.0721)	(0.307)	(0.207)
1 = 072004	0.785***	0.699***	0.785**	0.699***
	(0.110)	(0.0693)	(0.311)	(0.182)
1 = 082004	0.883***	0.677***	0.883**	0.677***
	(0.102)	(0.0680)	(0.330)	(0.198)
1 = 062005	0.759***	0.701***	0.759***	0.701***
	(0.0912)	(0.0627)	(0.232)	(0.125)
1 = 072005	0.650***	0.508***	0.650**	0.508***
	(0.0971)	(0.0650)	(0.244)	(0.127)
1 = 082005	0.373***	0.328***	0.373	0.328**
	(0.0991)	(0.0670)	(0.205)	(0.118)
1 = 062004	0.321***	0.378***	0.321	0.378**
	(0.0955)	(0.0578)	(0.192)	(0.162)
1 = 072004	0.0951	0.132***	0.0951	0.132**
	(0.0793)	(0.0500)	(0.119)	(0.0547)
Log(Airport Daily Maximum Temperature)	1.684***	1.634***	1.684***	1.634***
(Tenths Degrees Celcius)	(0.119)	(0.0724)	(0.249)	(0.222)
Log(Airport Daily Average Wind Speed)	-0.249***	-0.224***	-0.249*	-0.224*
(Tenths m/s)	(0.0823)	(0.0447)	(0.129)	(0.105)
Log(Per Capita Miles Traveled)	-0.140	0.192	-0.140	0.192
(1000s)	(0.361)	(0.233)	(0.804)	(0.554)
MONITOR FIXED EFFECTS	YES	YES	YES	YES
HOUR X DURATION BINS	YES	YES	YES	YES
Constant	-6.979***	-7.220***	-6.979***	-7.220***
	(0.805)	(0.508)	(1.440)	(1.321)
Observations	4,431	4,431	4,431	4,431
R-squared	0.362	0.438	0.362	0.438

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 3.3: Log Levels with Other Variables. Specifications (12)-(15) show a time trend control along with the ethanol coefficient and other variables of interest. There is not enough power to detect a difference between E-10 and MTBE fuel in this specification; however, results are similar to those found in tables 3.1 and 3.2. [log(ppbC Acetaldehyde + 0.001) is LOG M-I, log(max(ppbC Acetaldehyde, 0.1)) is LOG M-II]

VARIABLES	(12) LOG M-I	(13) LOG M-1	(14) LOG M-II	(15) LOG M-II
*** p<0.01, ** p<0.05, * p<0.1	*Robust* *NJ Excluded*	*Robust* *NJ, NY, CT Excluded*	*Robust* *NJ Excluded*	*Robust* *NJ, NY, CT Excluded*
10% Ethanol = 1	0.445***	0.475***	0.424***	0.441***
	(0.0758)	(0.0963)	(0.0442)	(0.0545)
Day Time Trend (Each Day = +1)	-0.000721***	-0.000766***	-0.000587***	-0.000621***
	(0.000118)	(0.000162)	(6.43e-05)	(8.62e-05)
Log(Airport Daily Maximum Temperature)	1.513***	1.624***	1.454***	1.529***
(Tenths Degrees Celcius)	(0.113)	(0.121)	(0.0693)	(0.0730)
Log(Airport Daily Average Wind Speed)	-0.207***	-0.236***	-0.181***	-0.196***
(Tenths m/s)	(0.0789)	(0.0881)	(0.0436)	(0.0481)
Log(Per Capita Miles Traveled)	-0.714***	-0.893***	-0.610***	-0.699***
(1000s)	(0.267)	(0.265)	(0.199)	(0.202)
MONITOR FIXED EFFECTS	YES	YES	YES	YES
HOUR X DURATION BINS	YES	YES	YES	YES
Constant	6.378***	2.764	3.792***	0.855
	(1.884)	(2.537)	(1.063)	(1.398)
Observations	4,431	3,945	4,431	3,945
R-squared	0.357	0.347	0.426	0.414

VARIABLES	(16) LOG M-I	(17) LOG M-1	(18) LOG M-II	(19) LOG M-II
*** p<0.01, ** p<0.05, * p<0.1	*Clustered - Metro* *NJ Excluded*	*Clustered - Metro* *NJ, NY, CT Excluded*	*Clustered - Metro* *NJ Excluded*	*Clustered - Metro* *NJ, NY, CT Excluded*
10% Ethanol = 1	0.445	0.475	0.424	0.441
	(0.270)	(0.354)	(0.237)	(0.299)
Day Time Trend (Each Day = +1)	-0.000721	-0.000766	-0.000587*	-0.000621
	(0.000405)	(0.000579)	(0.000288)	(0.000421)
Log(Airport Daily Maximum Temperature)	1.513***	1.624***	1.454***	1.529***
(Tenths Degrees Celcius)	(0.210)	(0.193)	(0.193)	(0.178)
Log(Airport Daily Average Wind Speed)	-0.207*	-0.236*	-0.181*	-0.196
(Tenths m/s)	(0.106)	(0.115)	(0.0948)	(0.104)
Log(Per Capita Miles Traveled)	-0.714	-0.893	-0.610	-0.699
(1000s)	(0.800)	(0.828)	(0.750)	(0.821)
MONITOR FIXED EFFECTS	YES	YES	YES	YES
HOUR X DURATION BINS	YES	YES	YES	YES
Constant	6.378	2.764	3.792	0.855
	(6.528)	(9.201)	(4.933)	(7.132)
Observations	4,431	3,945	4,431	3,945
R-squared	0.357	0.347	0.426	0.414

Table 3.4: Log Levels with Other Variables. I exclude monitor fixed effects to check the variables which do not change enough within the time period to identify. Here, GDP has a positive and significant coefficient. The values are similar to previous specifications in Tables 3.1 to 3.3; however, the policy variable is not significant with clustered standard errors. [log(ppbC Acetaldehyde + 0.001) is LOG M-I, log(max(ppbC Acetaldehyde, 0.1)) is LOG M-II]

VARIABLES	(20) LOG M-I	(21) LOG M-II	(22) LOG M-I	(23) LOG M-II
	Robust *NJ Excluded*	*Robust* *NJ Excluded*	*Clustered - Metro* *NJ Excluded*	*Clustered - Metro* *NJ Excluded*
10% Ethanol = 1	0.764***	0.554***	0.764	0.554
	(0.0958)	(0.0662)	(0.676)	(0.483)
Year Dummy (2004)	0.925***	0.612***	0.925	0.612
	(0.103)	(0.0663)	(0.735)	(0.469)
Year Dummy (2005)	0.681***	0.414***	0.681	0.414
	(0.0843)	(0.0589)	(0.565)	(0.365)
Log(Airport Daily Maximum Temperature)	0.919***	0.894***	0.919**	0.894**
(Tenths Degrees Celcius)	(0.126)	(0.0806)	(0.381)	(0.298)
Log(Airport Daily Average Wind Speed)	0.804***	0.492***	0.804	0.492
(Tenths m/s)	(0.0911)	(0.0505)	(0.493)	(0.298)
Log(Per Capita Miles Traveled)	-0.772***	0.423***	-0.772	0.423
(1000s)	(0.263)	(0.162)	(2.611)	(1.754)
Log(Real GDP)	0.428***	0.247***	0.428	0.247
Millions of Chained 2005 $	(0.119)	(0.0845)	(0.813)	(0.635)
Log(Metro Population)	-0.827***	-0.496***	-0.827	-0.496
	(0.168)	(0.116)	(1.042)	(0.775)
HOUR X DURATION FIXED EFFECTS	YES	YES	YES	YES
Constant	-1.780	-2.434**	-1.780	-2.434
	(1.529)	(0.980)	(8.093)	(5.212)
Observations	4,431	4,431	4,431	4,431
R-squared	0.066	0.091	0.066	0.091

*** p<0.01, ** p<0.05, * p<0.1

Table 3.5: Weighted Regressions. I weigh observations as described in the text in order to control for over-sampling of particular counties. The results remain consistent with the previous analysis. Additionally, I weigh and run regressions on only 3- and only 24-hour monitors. [log(ppbC Acetaldehyde + 0.001) is LOG M-I, log(max(ppbC Acetaldehyde, 0.1)) is LOG M-II]

VARIABLES	(24) LOG M-I	(25) LOG M-I	(26) LOG M-II	(27) LOG M-II
	Robust *NJ Excluded* *County Weighted*	*Clustered - Metro* *NJ Excluded* *County Weighted*	*Robust* *NJ Excluded* *County Weighted*	*Clustered - Metro* *NJ Excluded* *County Weighted*
10% Ethanol = 1	1.173***	1.173*	0.768***	0.768**
	(0.241)	(0.524)	(0.159)	(0.267)
1 = 062004	1.004***	1.004*	0.729***	0.729**
	(0.231)	(0.458)	(0.170)	(0.238)
1 = 072004	1.092***	1.092**	0.759***	0.759***
	(0.240)	(0.460)	(0.173)	(0.206)
1 = 082004	1.047***	1.047**	0.657***	0.657***
	(0.236)	(0.443)	(0.171)	(0.192)
1 = 062005	0.762***	0.762*	0.587***	0.587**
	(0.227)	(0.397)	(0.165)	(0.183)
1 = 072005	0.950***	0.950**	0.599***	0.599***
	(0.273)	(0.368)	(0.192)	(0.143)
1 = 082005	0.716***	0.716*	0.368**	0.368*
	(0.253)	(0.355)	(0.182)	(0.177)
1 = 062004	0.0866	0.0866	0.225**	0.225
	(0.164)	(0.249)	(0.102)	(0.159)
1 = 072004	-0.379**	-0.379	-0.0426	-0.0426
	(0.177)	(0.530)	(0.104)	(0.237)
Log(Airport Daily Maximum Temperature)	0.630***	0.630	0.757***	0.757
(Tenths Degrees Celcius)	(0.244)	(0.691)	(0.169)	(0.544)
Log(Airport Daily Average Wind Speed)	-0.126	-0.126	-0.211***	-0.211***
(Tenths m/s)	(0.130)	(0.0874)	(0.0793)	(0.0618)
Log(Per Capita Miles Traveled)	-0.274	-0.274	0.108	0.108
(1000s)	(0.491)	(1.312)	(0.372)	(0.865)
MONITOR FIXED EFFECTS	YES	YES	YES	YES
HOUR X DURATION FIXED EFFECTS	YES	YES	YES	YES
Constant	-4.288***	-4.288	-3.948***	-3.948
	(1.488)	(3.558)	(1.022)	(3.138)
Observations	4,431	4,431	4,431	4,431
R-squared	0.406	0.406	0.474	0.474

*** p<0.01, ** p<0.05, * p<0.1

Table 3.5. Weighted Regressions, *continued.*

VARIABLES	(28) LOG M-I	(29) LOG M-I	(30) LOG M-II	(31) LOG M-II
*** p<0.01, ** p<0.05, * p<0.1	*Robust*	*Clustered - Metro*	*Robust*	*Clustered - Metro*
	NJ Excluded	*NJ Excluded*	*NJ Excluded*	*NJ Excluded*
	3 Hour Data	*3 Hour Data*	*3 Hour Data*	*3 Hour Data*
	County Weighted	*County Weighted*	*County Weighted*	*County Weighted*
10% Ethanol = 1	0.882***	0.882*	0.829***	0.829*
	(0.0862)	(0.385)	(0.0591)	(0.392)
Year Dummy (2004)	0.766***	0.766**	0.669***	0.669**
	(0.0800)	(0.301)	(0.0511)	(0.237)
Year Dummy (2005)	0.355***	0.355**	0.354***	0.354***
	(0.0505)	(0.129)	(0.0422)	(0.0934)
Log(Airport Daily Maximum Temperature)	1.529***	1.529***	1.436***	1.436***
(Tenths Degrees Celcius)	(0.174)	(0.273)	(0.0951)	(0.198)
Log(Airport Daily Average Wind Speed)	-0.0592	-0.0592	-0.0591	-0.0591
(Tenths m/s)	(0.118)	(0.113)	(0.0596)	(0.103)
Log(Per Capita Miles Traveled)	-1.322***	-1.322	-0.871***	-0.871
(1000s)	(0.495)	(1.343)	(0.307)	(1.087)
MONITOR FIXED EFFECTS	YES	YES	YES	YES
HOUR X DURATION FIXED EFFECTS	YES	YES	YES	YES
Constant	-8.972***	-8.972***	-8.286***	-8.286***
	(1.110)	(1.852)	(0.619)	(1.284)
Observations	3,462	3,462	3,462	3,462
R-squared	0.355	0.355	0.461	0.461

VARIABLES	(32) LOG M-I	(33) LOG M-I	(34) LOG M-II	(35) LOG M-II
*** p<0.01, ** p<0.05, * p<0.1	*Robust*	*Clustered - Metro*	*Robust*	*Clustered - Metro*
	NJ Excluded	*NJ Excluded*	*NJ Excluded*	*NJ Excluded*
	24 Hour Data	*24 Hour Data*	*24 Hour Data*	*24 Hour Data*
	County Weighted	*County Weighted*	*County Weighted*	*County Weighted*
10% Ethanol = 1	1.101**	1.101**	0.518	0.518**
	(0.551)	(0.341)	(0.340)	(0.211)
Year Dummy (2004)	1.506***	1.506***	0.529	0.529**
	(0.545)	(0.264)	(0.342)	(0.210)
Year Dummy (2005)	1.284**	1.284***	0.485	0.485**
	(0.519)	(0.175)	(0.326)	(0.148)
Log(Airport Daily Maximum Temperature)	0.530	0.530	0.575***	0.575
(Tenths Degrees Celcius)	(0.325)	(0.808)	(0.186)	(0.572)
Log(Airport Daily Average Wind Speed)	-0.166	-0.166	-0.192*	-0.192**
(Tenths m/s)	(0.201)	(0.114)	(0.104)	(0.0705)
Log(Per Capita Miles Traveled)	-0.687	-0.687	-0.822*	-0.822
(1000s)	(0.697)	(1.507)	(0.433)	(0.857)
MONITOR FIXED EFFECTS	YES	YES	YES	YES
HOUR X DURATION FIXED EFFECTS	YES	YES	YES	YES
Constant	-3.475*	-3.475	-2.857**	-2.857
	(1.900)	(4.745)	(1.166)	(3.550)
Observations	969	969	969	969
R-squared	0.489	0.489	0.568	0.568

Table 3.6: Poisson Regression. As a robustness check, I run a Poisson regression to see whether specifying the measure values as floor(ppbC acetaldehyde / 0.6) impacts the analysis. This is done to check the results robustness to zeros in the regression and values below the highest mdl (0.6 ppbC). The results remain similar to the original analysis.

VARIABLES	(1) Acetaldehyde Count	(2) Acetaldehyde Count
	floor(ppbC acetaldehyde / 0.6)	*floor(ppbC acetaldehyde / 0.6)*
	NJ Excluded; Robust	*NJ Excluded; Clustered - Metro*
10% Ethanol = 1.0	0.581***	0.581***
	(0.0608)	(0.109)
Airport Daily Max Temp	0.00785***	0.00785***
(Tenths Degrees Celcius)	(0.000302)	(0.000668)
Airport Daily Average Wind Speed	-0.0136***	-0.0136***
(Tenths m/s)	(0.00118)	(0.00314)
Per Capita Miles Traveled	-0.137	-0.137
(1000s)	(0.343)	(1.047)
1 = 062004	0.761***	0.761***
	(0.0764)	(0.138)
1 = 072004	0.656***	0.656***
	(0.0689)	(0.0902)
1 = 082004	0.644***	0.644***
	(0.0734)	(0.111)
1 = 062005	0.824***	0.824***
	(0.0652)	(0.0948)
1 = 072005	0.524***	0.524***
	(0.0665)	(0.0898)
1 = 082005	0.332***	0.332***
	(0.0686)	(0.0696)
1 = 062006	0.482***	0.482***
	(0.0653)	(0.117)
1 = 072006	0.127**	0.127*
	(0.0514)	(0.0716)
HOUR X DURATION FIXED EFFECTS	YES	YES
MONITOR FIXED EFFECTS	YES	YES
Constant	-0.472	-0.472
	(0.334)	(0.744)
Observations	4,431	4,431

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 3.7: Controlling the Controlls. To test whether or not the controls are different from other monitors, I run specification (3.5). Indeed, the monitors that have highest ethanol increases are different than the controls. This explains part of the level-specification ethanol coefficient. However, the results are still robust to this specification, and E-10 seems to be increasing acetaldehyde pollution by 0.641 ppbC in the northeast, according to specification (7).

VARIABLES	(1) ppbC Acetaldehyde	(2) ppbC Acetaldehyde	(3) ppbC Acetaldehyde	(4) ppbC Acetaldehyde
*** p<0.01, ** p<0.05, * p<0.1	*Clustered - Metro*	*Clustered - Metro*	*Clustered - Metro*	*Robust*
	NJ Excluded	*NJ Excluded*	*NJ Excluded*	*NJ Excluded*
10ΔEthanol X 1(Year = 2006)	0.264**	0.398	0.398	0.622***
	(0.110)	(0.281)	(0.281)	(0.0778)
10ΔEthanol	-0.439	-0.490	-0.490	-0.533***
	(0.396)	(0.448)	(0.448)	(0.0622)
Day Time Trend		-0.000262	-0.000262	-0.000742***
		(0.000486)	(0.000486)	(9.88e-05)
Airport Daily Maximum Temperature				0.00811***
(Tenths Degrees Celcius)				(0.000467)
Airport Daily Average Wind Speed				0.000558
(Tenths m/s)				(0.00171)
Per Capita Miles Traveled				0.818***
(1000s)				(0.238)
MONTH OF YEAR FIXED EFFECT	NO	YES	YES	YES
HOUR X DURATION FIXED EFFECT	NO	NO	NO	NO
Constant	2.116***	6.329	6.329	11.74***
	(0.331)	(8.230)	(8.230)	(1.627)
Observations	4,431	4,431	4,431	4,431
R-squared	0.013	0.019	0.019	0.084

VARIABLES	(5) ppbC Acetaldehyde	(6) ppbC Acetaldehyde	(7) ppbC Acetaldehyde	(8) ppbC Acetaldehyde
*** p<0.01, ** p<0.05, * p<0.1	*Clustered - Metro*	*Robust*	*Clustered - Metro*	*Clustered - Metro*
	NJ Excluded	*NJ Excluded*	*NJ Excluded*	*All Data*
10ΔEthanol X 1(Year = 2006)	0.622**	0.641***	0.641***	0.853**
	(0.199)	(0.0773)	(0.191)	(0.312)
10ΔEthanol	-0.533	-0.499***	-0.499	-1.143**
	(0.495)	(0.0653)	(0.502)	(0.504)
Day Time Trend	-0.000742**	-0.000790***	-0.000790**	-0.00143*
	(0.000317)	(9.81e-05)	(0.000300)	(0.000668)
Airport Daily Maximum Temperature	0.00811***	0.00812***	0.00812***	0.00944***
(Tenths Degrees Celcius)	(0.00186)	(0.000466)	(0.00183)	(0.00196)
Airport Daily Average Wind Speed	0.000558	0.000556	0.000556	0.00282
(Tenths m/s)	(0.00993)	(0.00169)	(0.00950)	(0.00946)
Per Capita Miles Traveled	0.818	0.698***	0.698	-0.818
(1000s)	(2.550)	(0.244)	(2.617)	(2.689)
MONTH OF YEAR FIXED EFFECT	YES	YES	YES	YES
HOUR X DURATION FIXED EFFECT	NO	YES	YES	YES
Constant	11.74**	12.14***	12.14**	24.24*
	(4.756)	(1.609)	(4.660)	(11.11)
Observations	4,431	4,431	4,431	4,574
R-squared	0.084	0.096	0.096	0.097

3.12 Acknowledgments for "Pollution Whack-a-Mole: Ambient Acetaldehyde and the Introduction of E-10 Gasoline in the Northeast"

This paper is available in different form on AgEcon Search and was presented at the Agricultural and Applied Economics Association Meetings in Minneapolis, 2014. I would like to thank several NOAA NCDC and EPA AQS Datamart personnel for extensive help with downloading and interpreting data. Feedback from Julie Cullen was invaluable. I am also grateful to the Center for Environmental Economics for feedback; in particular, my adviser, Richard Carson.

3.13 Bibliography

Allen, P., R. Bradley, B.E. Croes, J. DaMassa, R. Effa, M. Fuentes, A. Hebert, L. Dongmin, R. Vincent, L. Woodhouse, and E. Yang. 1999. "Analysis of the Air Quality Impacts of the Use of Ethanol in Gasoline." Sacramento, CA: California Air Resources Board.

American Water Works Association. 2005. "A Review of Cost Estimates of MTBE Contamination of Public Wells."

Anderson, L.G., J.A. Lanning, and E. Wilkes. 1997. "Effects of Using Oxygenated Fuels on Carbon Monoxide, Formaldehyde and Acetaldehyde Concentrations in Denver, Toronto." Air & Waste Management Association's 90th Annual Meeting & Exhibition.

Auffhammer, M., and R. Kellogg. 2011. "Clearing the Air? The Effects of Gasoline Content Regulation on Air Quality." *American Economic Review* 101:2687-2722.

Chakravorty, U., C. Nauges, and A. Thomas. 2008. "Clean Air regulation and heterogeneity in US gasoline prices." *Journal of Environmental Economics and Management* 55:106-22.

Chay, K.Y., and M. Greenstone. 2003. "The Impact of Air Pollution on Infant Mortality: Evidence from Geographic Variation in Pollution Shocks Induced by a Recession." *Quarterly Journal of Economics* 118:1121-67.

Davis, L.W. 2008. "The Effect of Driving Restrictions on Air Quality in Mexico City." *Journal of Political Economy* 116:38-81.

Drye, W. 2005. "Hurricane Katrina: The Essential Timeline." *National Geographic*, September 14.

Ethanol Producer. "Plant List," http://web.archive.org/web/20060615010103/http://www.ethanolproducer.com/plant-list.jsp?country=USA; and http://web.archive.org/web/20060827035003/http://www.ethanolproducer.com/plant-list.jsp?view=construction&country=USA

Fong, D., C. Mizutani, R. Shapiro, and S.W. Matthews. 2005. "Options to Reduce Petroleum Fuel Use." California Energy Commission Report CEC-600-2005-024.

Foy, P. 1994. "Gas Plan Irks Local Officials." Albany, NY: *The Daily Gazette*.

Goldemberg, J., S.T. Coelho, and P. Guardabassi. 2008. "The Sustainability of Ethanol Production from Sugarcane." *Energy Policy* 36:2086-97.

Google Maps.

Graff Zivin, J.S., and M.J. Neidell. 2012. "The Impact of Pollution on Worker Productivity." *American Economic Review* 102:3652-73.

Greenstone, M. 2004. "Did the Clean Air Act Cause the Remarkable Decline in Sulfur Dioxide Concentrations?" *Journal of Environmental Economics and Management* 47:585-611.

Grosjean, D., A.H. Miguel, and T.M. Tavares. 2007. "Urban Air Pollution in Brazil; Acetaldehyde and Other Carbonyls." *Atmospheric Science B* 24:101-6.

Holland, J.C. 2001. "Volatile Organic Compounds (VOC) Audit" [NPAP-SOP-017].

Illinois Atlas. "Post-stratification weights." http://www.atlas.illinois.edu/support/stats/resources/spss/create-post-stratification-weights-for-survery-analysis.pdf

Jacobson, M.Z. 2007. "Effects of Ethanol (E85) versus Gasoline Vehicles on Cancer and Mortality in the United States." *Environmental Science and Technology* 41:4150-7.

Kintz, W.R., K. Lundberg, and S. Dodge. 2011. "Enhanced Monitoring of Acetaldehyde in Linn County, Iowa," http://www.epa.gov/ttnamti1/files/ambient/airtox/2011workshop/day3commWandaKintzCarbonyls.pdf.

Kumins, L., and R. Bamberger. 2005. "Oil and Gas Disruption from Hurricanes Katrina and Rita," RL33124. Congressional Research Service.

Lenski, F. (Personal communication.)

Loecher, M. 2013. Overlays on Google map tiles in R, RgoogleMaps-package: RgoogleMaps.

Mangus, N. (Personal Communication)

Renewable Fuels Association. "Statistics," http://www.ethanolrfa.org/pages/statistics.

Rogers, J., J. Wilbur, S. Cole, P. W. Bernhardt, J. L. Bupp, M. J. Lennon, N. Langholz, and C. P. Steiner. 2011. "Quantifying Uncertainty in Predictions of Hepatic Clearance." *Statistics in Biopharmaceutical Research* 3:515-25.

Satterfield, Z. "What Does ppm or ppb Mean?" *On Tap Q & A*. http://www.nesc.wvu.edu/ndwc/articles/ot/fa04/q&a.pdf.

Silver Biology County Lookup via Google Maps, http://labs.silverbiology.com/countylookup/.

Trottenberg, P., and R.S. Rivikin. 2013. "Memorandum to Secretarial Officers Modal Administrators: Guidance on Treatment of the Economic Value of a Statistical Life (VSL) in U.S. Department of Transportation Analyses."

Tuohy, L. 2003. "Exxon Mobil Must Pay $236m in NH Pollution Case." *Associated Press*.

U.S. Bureau of Economic Analysis. "Per Capita GDP by Metropolitan Area (Chained 2005 Dollars)." 2013.

U.S. Census Bureau. "Annual Estimates of the Population of Metropolitan and Micropolitan Statistical Areas: April 1, 2000 to July 1, 2009."

—. "Population Estimates of the United States."

U.S. Code of Federal Regulations. 2003. "Protection of Environment, 40 Parts 72 to 80 (para. 80.70)."

U.S. Energy Information Administration. 2006. "Eliminating MTBE in Gasoline in 2006."

—. 2003. "Status and Impact of MTBE Bans."

U.S. Energy Information Administration, Office of Oil and Gas. 2003. "Preparations for Meeting New York and Connecticut MTBE Bans."

U.S. Environmental Protection Agency [1], 2000. "Acetaldehyde Hazard Summary." Available at http://www.epa.gov/ttnatw01/hlthef/acetalde.html.

— [2]. AQS Datamart. Available at http://www.epa.gov/ttn/airs/aqsdatamart/.

— [3]. 2012. "Drinking Water [MTBE]." Available at http://www.epa.gov/mtbe/water.htm.

— [4], 1996. "EPA Takes Final Step in Phaseout of Leaded Gasoline."

— [5]. "Fuel Waiver Response to 2005 Hurricanes." Available at http://www.epa.gov/compliance/katrina/waiver/.

— [6]. "Information on Reformulated Gasoline (RFG) Properties and Emissions Performance by Area and Season – Methodology and Explanation." Live and Internet Archived. Available at http://www.epa.gov/otaq/fuels/gasolinefuels/rfg/properf/perfmeth.htm, http://web.archive.org/web/20050921013044/http://www.epa.gov/otaq/regs/fuels/rfg/properf/perfmeth.htm, http://web.archive.org/web/20021213103539/http://www.epa.gov/otaq/regs/fuels/rfg/properf/perfmeth.htm, http://web.archive.org/web/20061007061614/http://www.epa.gov/otaq/regs/fuels/rfg/properf/perfmeth.htm.

— [7]. "MTBE in Fuels." Available at http://www.epa.gov/mtbe/gas.htm.

— [8], 2004, 2005, and 2007. "Reformulated Gasoline Where You Live." Available through the Internet Archive at http://web.archive.org/web/20050920142033/http://www.epa.gov/otaq/rfg/whereyoulive.htm, http://web.archive.org/web/20060925232122/http://www.epa.gov/otaq/rfg/whereyoulive.htm, and http://web.archive.org/web/20080709040328/http://www.epa.gov/otaq/rfg/whereyoulive.htm.

— [9]. "RFG Survey Data." Available at http://www.epa.gov/otaq/fuels/rfgsurvey.htm.

U.S. Environmental Protection Agency Technical Support Branch, Emission Planning and Strategies Division, Office of Mobile Resources, and Office of Air and Radiation. 1993. *Motor Vehicle-Related Toxics Study*.

U.S. Government Accountability Office. 2004. "Metropolitan Statistical Areas: New Standards and their Impact on Selected Federal Programs," GAO-04-758.

U.S. National Highway Traffic Safety Administration. "Miles Traveled."

U.S. National Oceanic and Atmospheric Administration National Climate Data Center Weather Data.

World Bank. 2010. "Life Expectancy at Birth, Total (Years)."