Title
Essays in the Economics of a Decarbonized Transportation Sector

Author
Blundell, Marshall

Publication Date
2023

Peer reviewed|Thesis/dissertation

Essays in the Economics of a Decarbonized Transportation Sector

by

Marshall Blundell

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Agricultural and Resource Economics

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor James Sallee, Chair
Professor Severin Borenstein
Professor Maximilian Auffhammer

Spring 2023

Essays in the Economics of a Decarbonized Transportation Sector

Abstract

Essays in the Economics of a Decarbonized Transportation Sector

by

Marshall Blundell

Doctor of Philosophy in Agricultural and Resource Economics

University of California, Berkeley

Professor James Sallee, Chair

The transportation sector is the largest contributor to anthropogenic U.S. greenhouse gas (GHG) emissions. If the U.S. is to address climate change, a key mechanism will be through a transition to a decarbonized transportation sector. Given environmental externalities, it is clear that the private market will produce too little low-carbon transportation infrastructure. Policy-makers are attempting to overcome this scale problem by encouraging the build out of alternative transportation infrastructure. In this dissertation, I examine existing clean infrastructure policies — the spatial efficiency of infrastructure installed to-date, their effect on adoption of the new transportation mode, and their effect on other transportation modes.

In the first chapter of this dissertation, I study the location of electric vehicle charging stations. Absent policy, it is well understood that EV charging network size may be inefficient because of pollution and network externalities. This paper argues that there will also be spatial inefficiencies in the location of electric vehicle fast charging stations. I empirically test for the presence of a spatial inefficiency in the free market by comparing the location decisions of a vertically integrated firm that sells vehicles and provides a charging network to those of a non-vertically integrated charging network. I define a metric, enabled e-miles, that captures whether a charging station mitigates consumers' range anxiety. I then combine estimates from a spatial demand model and a simulation of charging behavior to compare charging demand and enabled e-miles at charging stations in California. I find strong evidence in favor of a spatial inefficiency in the location decisions of non-vertically integrated charging providers. These findings show that EV policy should not be location-neutral, but should consider spatial inefficiencies as well as network size.

In the second chapter of this dissertation, I examine whether the availability of home- and workplace-charging infrastructure for renters has a detectable effect on EV adoption. Policies supporting charging infrastructure in workplaces and multi-unit dwellings (MUDs) are nascent. I aim to address two questions: (i) are such policies effective in increasing investment in charging stations, and (ii) whether they are effective in promoting EV adoption. I

provide the first causal evidence on whether charging infrastructure represents a barrier to adoption for MUD occupants. I find that the subsidies increased the number of charging stations in a census block group (CBG) by .01. Against the sample average of 0.053, this represents a 20% increase. I did not detect any effect of charging stations on EV adoption.

In the third chapter of this dissertation, I explore the installation of a bike sharing system in New York City and its interaction with existing vehicle infrastructure. Bike sharing is one form of shared mobility service. These increasingly ubiquitous services provide shared use of a vehicle, bicycle, scooter, or other mode of transportation. They can simultaneously complement and conflict with existing transportation infrastructure. New York City's Citibike is the largest bike-sharing system in the country and among its goals is the relief of traffic congestion. I estimate the causal effect of Citibike on historical street speeds on Manhattan avenues. I employ Google Maps data to chart the routes between Citibike docks. I match these routes to rider counts and to novel data on traffic speeds at a 10-meter spatial resolution. This allows us to exploit variation in treatment intensity along and across avenues in fine resolution. I control for bike lanes, and other changes in street conditions over time. I find that the Citibike system decreased speeds on avenues in Manhattan. Overall, I estimate that the system decreased speed by 2.3% on average. At the maximum observed system utilization, travel time increased by 9.6%.

Together, these chapters employ travel data, spatial variation and tools, causal identification, and a combination of data and economic reasoning to draw policy relevant insights about how to decarbonize the transportation sector. A theme throughout is that the efficient deployment of infrastructure supporting low carbon transportation should take into account the location, scale, and utilization of existing infrastructure, be it the installed base of charging stations, or alternative transportation options.

To Amy.

Contents

Contents **ii**

List of Figures **iv**

List of Tables **vi**

1 Spatial Inefficiencies in Electric Vehicle Fast Charging **1**
1.1 Introduction . . . 1
1.2 Data . . . 6
1.3 A Spatial Demand Model . . . 8
1.4 A Simulation of Charging Behavior . . . 11
1.5 A Test for Spatial Inefficiencies . . . 16
1.6 Conclusion . . . 29

2 Subsidizing Electric Vehicle Charging Stations for Renters **30**
2.1 Data . . . 32
2.2 Policy Background . . . 36
2.3 Empirical Strategy . . . 38
2.4 Results . . . 40
2.5 Conclusion . . . 49

3 Bike Sharing and Traffic Congestion: A Case Study of Citibike in New York City **51**
3.1 Introduction . . . 51
3.2 Background: Citibike Bike-Sharing System . . . 53
3.3 Data . . . 55
3.4 Empirical Strategy . . . 60
3.5 Results . . . 65
3.6 Conclusion . . . 75

Bibliography **77**

A Appendix for Chapter 1 **82**

A.1 Alternative Normalization 82

B Appendix for Chapter 3 **85**
B.1 Space-Mean Speed 85
B.2 Back-of-the-envelope decomposition of speed change 86

List of Figures

1.1 Schematic Presentation of Spatial Demand Model 10
1.2 Distribution of Annual Charging Demand per Charger 21
1.3 Average Annual E-miles Versus Range by Year 22
1.4 Average Annual Minutes of Fast Charging Versus Range by Year 23
1.5 Annual Charging Quantity Versus Annual E-miles Enabled at 2019 Charging Stations 24
1.6 Distribution of Annual E-miles Enabled and Annual Charging Demand for Tesla and Non-Tesla Charging Stations 25
1.7 E-miles Enabled at 2019 Charging Station Locations, by Network 26
1.8 Demand at 2019 Charging Station Locations, by Network 27

2.1 Percent of Vehicles Registered at Addresses with more than Four Vehicles, for Bay Area Census Block Groups, 2018. 34
2.2 Number of Charging Stations by Type for California, 2015–2018. 35
2.3 California Census Tracts by Disadvantaged Community (DAC) Status 37
2.4 Event Study Plot for Impact of Subsidy on Charging Stations 43

3.1 Total Weekly Citibike Rides in Manhattan, June 2013 to July 2015 54
3.2 Kernel Density Plot of Time of Day of Citibike Rides in Manhattan, June 2013 to July 2015 55
3.3 Kernel Density Plot of Duration of Citibike Rides in Manhattan, June 2013 to July 2015 56
3.4 Total Monthly Citibike Rides in Manhattan, June 2013 to July 2015 61
3.5 Frequency of Citibike Rides on Road Segments in Manhattan, June 2014 63
3.6 Monthly Time Series of Travel Time in Log Average Seconds per Meter Above versus Below 59th Street, January 2009 to July 2015 64
3.7 Historical Street-Level Traffic Speed on East Side Avenues, Pre- and Post-Citibike 67
3.8 Historical Street-Level Traffic Speed on West Side Avenues, Pre- and Post-Citibike 68
3.9 Time Series of Monthly Uber Pickups in Manhattan 74

A.1 Annual Charging Quantity Versus Annual E-miles Enabled at 2019 Charging Stations 83

A.2 Distribution of Annual E-miles Enabled and Annual Charging Demand for Tesla and Non-Tesla Charging Stations 84

List of Tables

1.1 Demand Model Estimates 12
1.2 Counts of Vehicles, Charging Stations, and Chargers by Standard for 2019 . . . 17
1.3 Prices ($/kWh) by Network 19
1.4 Price Regression Results 28

2.1 PG&E Charging Station Subsidy Amount by Segment and Location 38
2.2 First Stage Impact of Subsidy on Charging Stations: Comparison across Charging Station Type 42
2.3 First Stage Impact of Subsidy on Charging Stations: Alternative Specifications . 45
2.4 First Stage Impact of Subsidy on Charging Stations: Alternative Samples 47
2.5 Instrumental Variables Results for Effect of Charging Stations on EV Registrations 49

3.1 Bike Share Usage Statistics by City 56
3.2 Seconds-per-meter for within-avenue taxi trips, January 2009 to August 2015 . . 59
3.3 Regression Sample Summary Statistics 66
3.4 Main Difference-in-Differences Results 69
3.5 Manhattan Commuter Mode Shares for 2012, 2014 71
3.6 Robustness and Specification Check Results 73

Acknowledgments

First and foremost, I am indebted to my advisor, Jim Sallee, who has provided much needed guidance, support, and enthusiasm for my work. He helped me get back on the rails after the world, and I, seemed to have shut down. He taught me to ask, and tackle, questions I didn't immediately know how to answer. He would not only ask, "What do you need from me?" but also made me feel as though I could answer honestly. I am grateful to Severin Borenstein for kindly delivered, thoughtful, and constructive comments, and for teaching me that I can ask questions and find the data later. I thank Max Auffhammer for his guidance during my qualifying exam, and for all his feedback, and good humor.

I owe much to faculty and staff at the Department of Agricultural and Resource Economics, and the Energy Institute at Haas. I'm extremely grateful to Michael Anderson for so much guidance on my third chapter, as well as to Thibault Fally and Marco Gonzalez-Navarro for helpful comments. I thank Carmen Karahalios and Elisabeth Sadoulet for gently corralling me to candidacy. At the Energy Institute at Haas, I am grateful to Lucas Davis for his comments and enthusiasm, and for continually reaching out to catch up on my research. I thank Cristina Bentley for everything she does and for putting up with my poor email response times! For so many helpful comments and enlightening discussions I thank Karen Notsund, Andrew Campbell, Meredith Fowlie, Reed Walker, Joseph Shapiro, and Catherine Wolfram. I thank ERE seminar participants and organizers. I am grateful for all the other interactions that I have learned from.

I must acknowledge the numerous other educators and colleagues who helped pave the way. My high school economics teacher was Geoff Riley, and I do not think I would be here without him. For encouragement, guidance, and inspiration as a wayward undergraduate, I thank Frank Wolak. I also thank Gopi Goda for seeing my potential. For support, direction, and valuable insight into the energy and utilities industry I thank Steve George, Jon Cook, and all my other colleagues at Freeman, Sullivan and Company, and Nexant.

In Matt Tarduno, Jenya Kahn-Lang, and Jesse Buchsbaum, I found wonderful classmates and friends. My other fellow graduate students and Energy Institute GSRs also provided many insightful comments, and unwavering camaraderie. I am forever grateful to Kevin and Yuen for their companionship during the restrictions imposed during the COVID-19 pandemic, and for their friendship ever since. I also thank the other residents of University Village, whom I will miss dearly, for their neighborliness. For helping me find some balance I thank my climbing partners, including but not limited to: Yuen, Leo, Braden, and Kris. I'm grateful to all my other friends for believing in me, and for providing some perspective. I would also like to thank my parents for all their love and support, even though my Dad thought this only took three years.

At home, the cats Matcha and Emily provided unconditional emotional support and keyboard-warming services. Coco the puppy provided a welcome daily escape through walks at Berkeley Marina and Albany Bulb. Last but not least, I lovingly thank my fiancée Amy, who was there throughout to help me overcome my lows and celebrate my highs.

Chapter 1

Spatial Inefficiencies in Electric Vehicle Fast Charging

1.1 Introduction

The transportation sector is the largest contributor to anthropogenic U.S. greenhouse gas (GHG) emissions. If the U.S. is to address climate change, a key mechanism will be through electrification of the transportation sector. This is most likely to occur through the adoption of electric vehicles (EVs). Adoption to date, however, has been slow: in 2018 EVs had a 2% market share nationwide. Widely cited barriers to EV adoption include range anxiety, cost, and the availability of public charging infrastructure.

In this paper, I define a metric, enabled e-miles, that captures whether a charging station mitigates consumers' range anxiety. I then combine estimates from a spatial discrete choice demand model and a simulation of charging behavior to compare charging demand and enabled e-miles at charging stations in California.

A particularly challenging question is how to efficiently subsidize public charging. Given environmental externalities, it is clear that the private market will produce too little charging so long as the EV market is inefficiently small. But the two-sided nature of charging–EV adoption encourages charging build out, and vice versa–makes the optimal policy more complicated (Rochet and Tirole, 2006; Springel, 2019). With these issues in mind, the existing literature has considered how policy can improve on the private market allocations considering the scale of charging infrastructure. Less attention has been paid to whether the private market will create spatial inefficiencies in where chargers are located. Following the logic of Hotelling's seminal paper on spatial inefficiency (Hotelling, 1929), business locations in space can be inefficient due to business stealing effects and travel costs. In this setting, location decisions can also be socially inefficient because the benefits of vehicle adoption spurred by a charging station in a new location mostly accrue to competing charging networks. This would not be the case were there a monopoly charging station network, and would also not be the case for a charging network with a proprietary charging standard.

To examine spatial inefficiencies in this setting, I study the location of charging stations and ask if there is a role for policy. I empirically test for the presence of a spatial inefficiency in the free market by comparing the location decisions of a vertically integrated firm that sells vehicles and provides a charging network to those of non-vertically integrated charging networks. The vertically integrated network has features in common with a social planner. In particular, its proprietary charging standard means it should internalize any business stealing effect. Furthermore, it will internalize the benefit that a more comprehensive charging network has on vehicle adoption, and in turn, future charging station demand.

Range anxiety, which is a driver's fear that a vehicle has insufficient battery capacity to reach their intended destination, is related to the availability of public charging infrastructure. Public charging infrastructure is a remedy to range anxiety in that it allows more destinations to be reached with a given state of battery charge. For each driver in a typical year, there exists a set of trips that cannot be completed in an EV of a given battery capacity given the existing charging network. If a charging station is located appropriately, it may reduce the size of a driver's set of unachievable trips. I define a charging station's *e-miles enabled* as the total mileage of the previously unachievable trips that are now possible as a charging station location is added. E-miles enabled at potential charging station location depends on the location of other charging stations in the network. In this setting, e-miles enabled is a major component of consumer value from a charging station that is heterogeneous over space. Policy makers are attempting to foster electrification by encouraging the build-out of a more comprehensive charging network, but many open questions remain about the role of charging stations and the ideal design of policies to subsidize charging stations[1].

My test comprises two questions: A. How does profit correlate with e-miles enabled at a marginal charging station? B. Do we observe the vertically integrated network entering in locations with higher e-miles enabled? The magnitude of the correlation between profit and e-miles is of interest because if the correlation were strong, then firms ignoring e-miles would nevertheless locate efficiently. Thus a necessary condition for spatial inefficiencies to occur is that the correlation is weak. Furthermore, if locations with relatively high e-miles enabled are relatively unprofitable we could expect fewer charging stations in those locations. Question B examines a manufacturer, Tesla, that puts more weight on e-miles enabled from charging station locations. This occurs by virtue of its vertical integration, manufacturing vehicles as well as charging stations which it installs and manages. A finding that the vertically integrated network enters locations with higher incremental e-miles would be evidence in favor of a spatial inefficiency in free entry. In theory, how profits compare between a vertically integrated and non-vertically integrated provider is unclear. In the absence of any business stealing effects, we might expect revenues to be lower if incremental e-miles is imperfectly correlated with revenue over space. However, business stealing among competing networks should lead to excess entry and lower revenue for non-vertically integrated providers.

My paper addresses these questions in four sections. Section I discusses the data used

[1] *www.whitehouse.gov/briefing-room/statements-releases/2021/12/13/fact-sheet-the-biden-harris-electric-vehicle-charging-action-plan/*

for estimation, which includes a novel dataset of "check-ins" to charging stations. Section II presents an innovative model of spatial demand that allows me to predict charging quantity at a charging station at an arbitrary location, given an existing network. Section III presents a simulation of charging behavior used to estimate additional e-miles enabled by a charging station at a point in space, given an existing network of charging stations. Section IV combines these methods to empirically test for the presence of spatial inefficiencies in the fast charging market in California.

I estimate a model of spatial demand based on the existing network of charging stations in California in 2019. For my application, the model must, for any set of charging station locations, be able to predict demand for individuals' charging over space. I construct a measure of latent demand for refueling over space by using cell phone data to count annual visitors to gas stations in California that are over 100 miles from home. The restriction to gas station visits over 100 miles from home aims to model the behavior of EV drivers, the vast majority of whom have access to home charging. I then use a discrete choice framework that takes into account distance and charging station characteristics to map refueling demand to charging station demand using data on visits to charging stations. I use this model to run counterfactual scenarios to compare demand for a charging station network with and without each charging station.

A fundamental feature of fast charging networks is that the e-miles enabled of a particular location depends on the location of all other charging stations in the network. I estimate the e-miles enabled of a marginal charging station using a simulation approach. I simulate charging behavior using a large dataset of daily travel diaries, assuming the trips were taken in a representative EV. Specifically, I construct a recursive EV route-planning function that I apply to 114,455 trips from the 2017 National Household Transportation Survey California Add-on. I use this function to estimate, for any set of charging stations, the e-miles enabled by a marginal charging station in that set. With the simulation architecture in place, this estimation is simply the weighted sum of miles traveled with a network including a charging station, minus the weighted sum of miles traveled with a network excluding a charging station. I then use existing estimates of the value of an electric vehicle's range to value the incremental e-miles enabled by a charging station.

I combine my estimates from my spatial demand model and charging simulation, to test for the presence of a spatial inefficiency. I compare charging demand, and incremental e-miles enabled for the marginal charging station in the 2019 network in California. I make this comparison in a counterfactual scenario with equivalent EVs per charging station across charging standards. I first show that, for a marginal charging station, demand is negatively correlated with e-miles enabled over space. This suggests that locations with relatively high consumer value are relatively unprofitable.

A station may be able to maintain higher profits with lower quantities if it can charge higher prices, or lower costs. If so, then quantities are an incomplete measure of profitability. While I do not observe prices and costs at all charging stations so cannot estimate profit directly, the empirical evidence suggests that demand is the main driver of profit in this setting. In particular, the majority of charging networks in this setting have a uniform price

across charging stations. Results from regressing price on demand and competition variables show little correlation, which is consistent with networks not setting prices in a way to extract rents at remote locations. Identifying the correct explanation for uniform pricing within a network is beyond the scope of this paper. In an emerging market, a likely explanation is that varying prices across charging stations would lead to negative reactions from consumers and would in turn lower demand for that network's charging stations. DellaVigna and Gentzkow, 2017 identified a similar mechanism as playing a role in uniform pricing at U.S. retail chains.

I then compare demand and e-miles enabled between vertically integrated and non-vertically integrated charging providers. The average charging stations in the vertically integrated Tesla network enables 700.6k total annual e-miles, whereas the average charging stations at other networks enables 55.5k total annual e-miles. This result is consistent with Tesla entering locations with a higher consumer value, and is strong evidence in favor of a spatial inefficiency. Tesla's primary concern in developing a charging network is to sell vehicles. To that end, they face a stronger incentive to design a network which maximizes consumer than charging networks that do not sell vehicles. Moreover, other charging networks that do not sell vehicles share a charging standard with competitors. Therefore they only receive a fraction of the benefit of vehicle adoption spurred by their investment. I find average annual demand at Tesla chargers is slightly lower than at chargers in other networks. Lower demand per charger at Tesla chargers suggests that the effect of entering locations with a high consumer value that tend to have lower demand outweighs the business stealing effect that increases crowding and reduces demand at other networks. These findings suggest that leaving location decisions to non-vertically integrated, charging station providers may result in sub-optimal location choice. EV policy should not be location-neutral, but should consider spatial inefficiencies as well as network size.

Under the bipartisan infrastructure law, states are eligible for $5 billion in funding for charging stations under the National Electric Vehicle Infrastructure (NEVI) Formula Program.[2] To receive funding, eligible stations must be no more than 50 miles apart and must be installed on interstates and major highways that are designated as Alternative Fuel Corridors. A maximum distance between stations, and a requirement that stations are installed along high-demand corridors, shows that policy makers recognize the importance of a comprehensive network of charging stations to vehicle adoption. However, without more structured guidance on location, a roll-out could still be inefficient. In California's plan[3], corridors are segmented and vendors submit proposals to install charging stations on segments. Segments with lower demand may be undersupplied with charging stations under such a scheme. Nevertheless, the structure of the NEVI program and state plans shows that location is a concern for policy makers. This work represents a formal justification for such a location-based policy.

This paper makes three main contributions. The first is to a growing literature on

[2] *www.fhwa.dot.gov/bipartisan-infrastructure-law/evs_5year_nevi_funding_by_state.cfm*

[3] *dot.ca.gov/-/media/dot-media/programs/sustainability/documents/nevi/2022-ca-nevi-deployment-plan-a11y.pdf*

charging stations and the electric vehicle market. J. Li, 2019 and Springel, 2019 both model electric vehicle purchases and charging station investment as a two-sided market. J. Li, 2019 examines the ambiguous impact of mandating compatibility standards on market outcomes and welfare. J. Li, 2019 models charging station investment in markets and along city corridors. Springel, 2019 investigates optimal subsidy structure and models charging station entry in a market as a function of the installed base of EVs in that market. This work contributes in several dimensions. First, I highlight the potential divergence in consumer value and producer value of charging stations across space. Even under a flat subsidy set optimally, entry across or within a market or city corridor may be socially inefficient. Next, my work is the first to consider how incentives facing charging stations differ according to whether they are owned and operated by a vehicle manufacturer. This can exacerbate spatial inefficiencies because the surplus from vehicle purchases does not accrue to the charging station entrant. Dorsey, Langer, and McRae, 2022 estimate an imperfect information discrete choice model of drivers' gas station refueling preferences. They find that investment in charging speed lowers drivers' refueling time more than investment in additional charging stations. This paper is the first to estimate charging station demand using empirical data from installed charging stations.

This work also contributes to a literature on the general phenomenon whereby profits are poorly correlated with total surplus from a firm's operations, and inefficient outcomes result. This phenomenon was first investigated in the context of monopolistic competition and optimal product variety (Dixit and Stiglitz, 1977; Koenker and Perry, 1981; Mankiw and Whinston, 1986; Spence, 1976). My setting, in which companies choose to enter different markets (location) with different profits, is similar to Borenstein, 1988. Borenstein, 1988 shows that the difference between private and social incentives in a competitive permit allocation process can be large. S. Li, 2017 empirically quantifies the welfare consequences of the two mechanisms for vehicle licenses in China, a context with large negative externalities. The analysis shows that different allocation mechanisms lead to dramatic differences in social welfare. This paper contributes by examining the correlation between profits and total surplus in a two-sided markets context, where some firms are vertically integrated.

Finally, this work contributes to a literature examining ownership structure and product differentiation. Existing studies have examined the effect of horizontal integration on product variety. Berry and Waldfogel, 2001 study the effect of mergers on variety in radio broadcasting. They find that increased concentration from mergers increased program variety relative to the number of stations in the market. Sweeting, 2010 also studies the effect of mergers and business stealing in radio broadcasting. The authors find that when firms buy competing stations, aggregate variety does not increase, and the gains in market share come at the expense of other stations in the same format. Seim and Waldfogel, 2013 examine the entry decisions of Pennsylvania's state liquor retailing monopoly, with an emphasis on location choice. The authors estimate a spatial model of demand for liquor and compare market outcomes and locations under the status quo of a state monopoly to the profit- and welfare-maximizing locations. My demand model is similar to that in Seim and Waldfogel, 2013, but I relax the assumption that each consumer visits the nearest location.

This study also contributes by examining the effect of *vertical integration* on spatial product differentiation in a novel two-sided markets setting where a vertically integrated firm has characteristics in common with a social planner.

1.2 Data

This study draws on three main data sources. For my charging simulation, I use the California Add-On for the 2017 National Household Transportation Survey (NHTS). My demand model draws on "check-ins" to charging stations sources from a large online platform, and travel demand data in the form of visits to gas stations in California. The travel demand data includes counts of monthly visitors to gas stations in California, broken out by the visitors' home census block group (CBG).

Geocoded Trips Data

I use the California Add-On for the 2017 National Household Transportation Survey (NHTS). These data comprise a detailed travel diary for a selected day of the year for 23,405 households in California. The geocoded version that I employ includes the latitude and longitude of the start and end points of each trip and includes whether the trip started or ended at home.

I restrict the sample to trips taken by vehicle (car, SUV, pickup, or van), where a member of the respondent household was the driver, and to trips originating or ending in California. I also remove 216 "loop" trips that have the same origin and destination location. This leaves 114,455 trips taken by 20,868 households. I define a trip chain as all trips in a day taken by a household's vehicle. The data comprise 29,644 trip chains.

These data do not include detailed information on the route taken. I use a routing API, HERE Matrix Routing API, to calculate detailed routes for each trip.[4] The API computes energy consumption of a typical electric vehicle on the route. The consumption model takes into account elevation change, traffic speed, and other factors.[5] I supply the departure time of the trip and the routing model uses historical traffic information to estimate consumption.[6] The charging station data I use only contains charging stations for the state of California, so I truncate trips that span the boundary of the state.

Charging Stations Data

My charging stations data are drawn from a comprehensive dataset of locations of public EV charging stations in California. The dataset includes "check-ins" by users, which I employ as a proxy for demand. The data were obtained from a mobile application and website that allows EV owners to locate charging stations. The data include charging stations for

[4] *developer.here.com/documentation/matrix-routing-api/dev_guide/index.html*

[5] *developer.here.com/documentation/routing-api/dev_guide/topics/use-cases/ev-routing.html*

[6] Heating and air conditioning use, which affects energy use, is not included in the model.

each of the three fast charging standards: Supercharger, which is Tesla's charging stations; CHAdeMO, which is used by Nissan and the Kia Soul MK1; and CCS, which is used by all other manufacturers.

I observe charging stations installed up to February 2020. For each charging station, I observe the installation date, the network, the number of chargers of each standard, and the total number of check-ins since installation. I also observe more detailed information on the 50 most recent check-ins, including username, date, and vehicle type. For some charging stations, I interpolate total check-ins since the installation date to find the check-ins for my study period, February 2019 to February 2020. This interpolation is necessary for charging stations that were installed before February 2019 and that have all 50 of the most recent check-ins occur after February 2019.

A potential source of bias in these data is that the platform is primarily used to find charging stations and plan trips. That same function is provided by software in Tesla vehicles, which route users to Superchargers when charging is necessary on a trip. While some vehicles with the CCS and CHAdeMO standard include similar software, many do not. This means Tesla users, and therefore check-ins to Superchargers are likely to be underrepresented. To address this issue, I perform separate normalizations of check-ins to the population of registered vehicles of each standard in the state of California in 2019. To do so, I use detailed data on the 50 most recent check-ins to estimate the total number of vehicles of each standard represented in the data. For users who report their vehicle type, I compute check-ins per vehicle for vehicles of each standard. I then divide total annual check-ins of each standard by the appropriate vehicles per check-in, to estimate number of vehicles of each standard represented in the data. This method also accounts for a potential source of bias in that some standards such as Tesla tend to have larger ranges and so check-in less frequently. Finally, using statewide data on the number of registered EVs of each charging standard, I normalize check-ins at charging stations of each standard according to the ratio of population vehicles to vehicles represented among platform users.

In addition to a charging station's location, and check-ins, I obtain its name, a description, a description of the price for charging, and a categorical variable that encodes places of interest, such as a school, or workplace, or apartment building, where the charging station is located.

Travel Demand Data

I use a demand model that predicts charging demand from consumer travel demand over space. To model latent travel demand over space, I first discretize consumer locations to the level of a gas station location. I then construct an annual count of visitors to gas stations in California that are over 100 miles from home using cell phone data from the data company Safegraph. My data ranges from January 2019 to December 2019. The data records, for every gas station, monthly counts of visitors from every CBG to yield a monthly panel at the gas station, home CBG level. I restrict to home CBG-gas station combinations that are over 100 miles apart in great-circle distance. To create a more representative sample, for

each CBG, I create a weight equal to the population as measured in the ACS divided by the number of devices in the Safegraph sample. Since the travel demand of EV owners, who tend to be wealthy, might differ from that of non-EV owners, I weight visitors from each CBG by the proportion of electric vehicles of a particular charging standard in each home CBG. I apply this weight and aggregate to the gas-station level to yield total visits in 2019 that are over 100 miles from home.

Other

I use vehicle registrations data from the California DMV. The data provide counts of total vehicle registrations by census block group (CBG), by year. Within a CBG-year, the data are further broken out in a fuel-technology subcategory that includes battery electric vehicle (BEV), fuel-cell electric vehicle (FCEV), internal combustion engine (ICE), and plug-in-hybrid electric vehicle (PHEV). I use fuel-technology variable to categorize vehicles as EV or non-EV.

I use county-level estimates of registered EVs by charging standard from California Air Resources Board. Charging standards include CCS, CHAdeMO, and Tesla.

1.3 A Spatial Demand Model

For this application, a demand model must, for any set of charging station locations, be able to predict charging demand from travel demand over space. I use a discrete choice framework to model consumers' decision to visit a charger and estimate its parameters based on the existing network of charging stations in 2019.

I discretize consumer locations and model latent charging demand over space at the level of a gas station location. I take the number of potential consumers at a location to be observed visits to gas stations, considering only visits that are over 100 miles from home. Since the travel demand of EV owners, who tend to be wealthy, might differ from that of non-EV owners, I weight visitors from each CBG by the proportion of electric vehicles of a particular charging standard in each home CBG. My demand function then relates charging station demand, derived from the check-ins data, to the share of latent demand at each charging station that is within 15 miles of a gas station. The share of latent demand obtained by each charging station is a function of distance of a charging station from a gas station, and other charging station characteristics such as the number of chargers at a station.

Formally, let C_{sg} be the set of charging stations of standard s such that gas station g is within 15 miles of great-circle distance. By restricting the choice set at a gas station location to charging stations within 15 miles I assume consumers would not travel more than 15 miles out of the way to use a charging station. This means that travel demand at a gas station is never allocated to a charger more than 15 miles away.

Consumer i at gas station g's conditional indirect utility from traveling to charging station c of standard s to charge is

$$V_{igcs} = X'_{gc}\beta_x - \beta_d d_{gc} + \epsilon_{igcs}, \tag{1.1}$$

where X_{gc} is a vector of charging station characteristics, and d_{gc} is the great-circle distance from charging station c to gas station g. I assume ϵ_{igcs} is distributed extreme value. I assume charging stations only compete with charging stations of the same standard. This yields logit purchase probabilities π for charger c of standard s at a gas station location g:

$$\pi_{gcs} = \frac{\exp\left(X'_{gcs}\beta_x - \beta_d d_{gcs}\right)}{1 + \sum_{c \in C_{sg}} \exp\left(X'_{gcs}\beta_x - \beta_d d_{gcs}\right)}. \tag{1.2}$$

Aggregate demand for charging from gas station g, $\hat{Q}_{gcs}$, and at charging station c, $\hat{Q}_c$ are given by

$$\hat{Q}_{gcs} = \pi_{gcs} M_{gs} \tag{1.3}$$

, and

$$\hat{Q}_{cs} = \sum_{g \in G_{cs}} \frac{\hat{Q}_{gcs}}{M_{cs}} M_{cs} = \pi_{cs} M_{cs} \tag{1.4}$$

, where M_{gs} denotes the number of potential consumers at gas station location g with standard s and $M_{cs} = \sum_{g \in G_{cs}} M_{gs}$. M_{gs} is the number of consumers visiting gas station g over 100 miles from home, weighted by the proportion of EV registrations in their home CBGs of a particular standard.

Estimation

I estimate the demand model using maximum likelihood. Parameter estimates maximize likelihood of observing check-ins to charger c, Q_c, given data on characteristics and distance from the charging station of the gas stations making up the charging station's catchment area. The log-likelihood function is given by

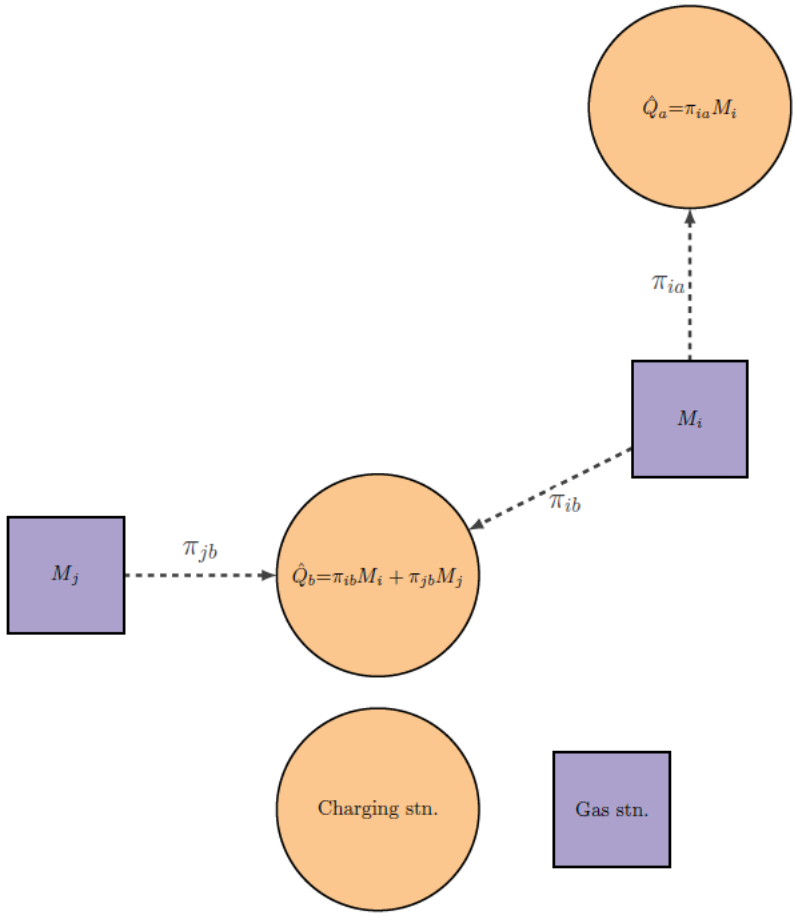
$$lnL = -\sum_{c=1}^{C} \sum_{s \in \{Tesla, Combo\}} \left(Q_{cs} \ln(\pi_{cs}) + (M_{cs} - Q_{cs}) \ln(1 - \pi_{cs})\right), \tag{1.5}$$

where Q_{cs} is check-ins at a charging station of each standard, scaled up using the population of registered vehicles of each standard, as discussed in the data section.

Model Estimates

I estimate the model using data at the charging station, gas station level for 2019. I restrict the choice set for each gas station to charging stations that are within 15 miles, measured

Figure 1.1: Schematic Presentation of Spatial Demand Model



Note: This figure shows how gas station demand is allocated to charging stations. Charging stations are shown as orange circles and gas stations as violet squares. For simplicity, I omit the standard subscript s. Both charging stations a and b are within 15 miles of gas station i and so are included in the choice set for that gas station, but only the lower charging station b is within 15 miles of gas station j. M_g, the number of potential consumers at each gas station, is allocated to charging stations using estimated choice probabilities π_{gc} to compute $\hat{Q}_c$, aggregate demand for charging at each charging station. In estimation, aggregate demand for each charging station is used to compute charging station market shares $\pi_c = \frac{\hat{Q}_c}{M_c}$. Market size for charging station c is $M_c = \sum_{g \in G_c} M_g$ where M_g is the number of consumers visiting gas station g over 100 miles from home, and G_c is the set of gas stations within 15 miles of charging station c.

in great-circle distance. I include the great-circle distance between each gas station and charging station in my model to capture the effect of distance on demand.[7]

Table 1.1 shows the coefficients from my estimated demand function. Column 1 reports

[7]It was cost prohibitive to include measure driving time or driving distance for each 702 x 10,359 gas station, charging station combination, or even for just the 291,690 combinations in my analysis dataset that are within 15 miles in great-circle distance.

estimates including only distance as a variable. Column 2 includes controls for the number of chargers of each standard at a charging station. Tesla charging stations, or Superchargers, only ever include Tesla chargers, while other charging stations can include chargers for both CCS and CHAdeMO standards. Column 3, my preferred specification, includes binary variables encoding the network, for which the base category is unnetworked charging stations. The coefficient on distance is then identified by within-network variation in distance and demand.

The estimated coefficient on distance is stable across specifications. The average marginal effect of a one mile increase in distance based on column (3) is -0.021.

Figure 1.2 shows the distributions of annual quantity per charger that I observe in the data, alongside predictions from the demand model. To account for variation the in number of chargers at charging stations, I display quantities per charger at each charging station, weighted by the number of chargers at each station. Predictions from the model fit the data closely. The average annual predicted quantity is 229.2 per charger, and the average annual quantity in the data is 212.3. While I normalize quantities to match the number of registered EVs of each charging standard, I cannot account for the fact that an average user does "check-in" every time they charge. An implication of this is that annual charging quantities that I observe are almost certainly lower than the true annual quantities.

1.4 A Simulation of Charging Behavior

A fundamental feature of fast charging networks is that the consumer value of a particular location depends on the location of all other charging stations in the network. A simple example is that the value of adding an additional charging station to a location depends on whether the location is already occupied by an existing charging station. Another example is an additional charging station in a remote location that is only valuable with sufficient charging stations in less remote areas such that the remote location is reachable. This feature presents multiple difficulties in conceiving of an experiment to estimate the value of charging stations.

A simple experiment where California is divided in two and charging stations added to one half would not yield causal estimates because vehicle purchasers state wide value the entire charging network and so potential outcomes are affected by other units' treatment. In an ideal experiment, there would be multiple states of the world, where in each state a charging station was randomly assigned to some point in space, with the existing network unchanged. With data on vehicle demand, electric vehicle sales in different states of the world could be compared to estimate the consumer value of charging stations over space, given the existing network. The empirical intractability of this problem lends itself to simulation. I construct a route-planning function that I apply to 114,455 trips from the 2017 NHTS California Add-on. I use this function to estimate, for any set of charging stations, the incremental e-miles enabled by a marginal charging station in that set. With the simulation architecture in place, this estimation is simply the weighted sum of miles traveled with a network including

Table 1.1: Demand Model Estimates

	(1)	(2)	(3)
Distance (miles)	-1.061^{***}	-1.112^{***}	-1.101^{***}
	(0.234)	(0.230)	(0.233)
Tesla chargers		0.041^{**}	0.001
		(0.015)	(0.022)
CHAdeMO chargers		0.003	0.020
		(0.067)	(0.043)
CCS chargers		−0.020	0.026
		(0.018)	(0.020)
ChargePoint			−0.103
			(0.401)
Electrify America			−0.500
			(0.401)
EV Connect			−0.144
			(0.621)
EVgo			−0.309
			(0.376)
Greenlots			−0.064
			(0.502)
Supercharger			0.911
			(0.567)

Note: $^{*}p<0.05$; $^{**}p<0.01$; $^{***}p<0.001$. Results based on 291,690 charging station-gas station level observations, for 702 charging stations. The dependent variable is annual check-ins at a charging station of each standard, scaled up using the population of registered vehicles of each standard, as discussed in the data section. Standard errors are block bootstrapped at the charging station level (250 replications). Omitted category for network fixed effects is unnetworked charging stations.

a charging station, minus the weighted sum of miles traveled with a network excluding a charging station. I then use existing estimates of the value of an electric vehicle's range to value these miles.

A Recursive Charging Function

Many electric vehicles are equipped with software that helps a driver find charging stations and plan when to charge to reach their destination. Otherwise, there exist apps, such as A Better Route Planner, that provide the same service.[8] I construct a charging route planner that operates as a simplified version of such software.

The function operates on a trip chain, which is the set of trips that a household's vehicle makes in a day. I define a trip chain $T \in \mathcal{T}$ as the tuple $T := (R, E, PD, H)$, where R is a vector of routes in a day completed by a household's vehicle, E is a vector of trip energy consumption values in kWh, PD is the vector of destination parting durations, and H is a binary vector encoding whether the destination is the household's home. The function takes into account battery parameters, which are a vector $B := (B_{max}, B_{min}, B_s)$, where B_{max} is capacity in kWh, B_{min} is the minimum battery state in kWh, and B_s is the state of charge in kWh at the time of the function call. I set B_{max} to 75 kWh in my main specification, which is commensurate with the usable battery capacity of a long range Tesla Model 3 in my study period.[9] I set B_{min} to be 10% of B_{max}. All trip chains begin with a full charge i.e. $B_s = B_{max}$. The set of available charging stations $C \in \mathcal{C}$ is a set of points with longitude and latitude. The output of the function is in $\mathcal{M}$, the space of mileage (completed and remaining) and minutes charging variables, where $M := (M_c, M_r, minutes)$. Formally, the function $Charge : \mathcal{T} \times \mathcal{B} \times \mathcal{C} \times \mathcal{M} \to \mathcal{M}$ is recursively defined as follows,

$$
\begin{aligned}
&Charge \begin{pmatrix} T \\ B \\ C \\ M \end{pmatrix} \\
&= \begin{cases}
(M_c + \sum_j miles(R_j), 0, minutes), & \text{if } \sum_j E_j \leq B_s - B_{min} \\
Charge \begin{pmatrix} T_{-1} \\ (B_{max}, B_{min}, B_s - E_1) \\ (M_c + miles(R_1), M_r - miles(R_1), minutes) \end{pmatrix}, & \text{if } E_1 \leq B_s - B_{min} \\
Charge \begin{pmatrix} T' \\ (B_{max}, B_{min}, B'_s) \\ (M'_c, M'_r, minutes') \end{pmatrix}, & \text{if fast charger available} \\
(M_c, \sum_j miles(R_j), minutes), & \text{otherwise.}
\end{cases}
\end{aligned}
\tag{1.6}
$$

[8] *www.abetterrouteplanner.com*

[9] *www.electrek.co/2017/08/08/tesla-model-3-battery-packs-50-kwh-75-kwh-elon-musk/*

The first conditional statement considers whether the sum of consumption of all trips in the chain is less than the battery state minus the minimum tolerance: $\sum_j E_j \leq B_s - B_{min}$. In that case, the entire chain can be completed and the mileage is updated with the miles of all trips in the chain ($M_c + \sum_j miles(R_j)$), miles remaining is set to 0, and $minutes$ which is minutes charged, is unchanged.

The second conditional statement considers whether the consumption in the first trip in the chain is less than the battery state minus the minimum tolerance: $E_1 \leq B_s - B_{min}$. In that case the trip can be completed, and the trip is removed from the chain (T_{-1}), the battery state is updated to $B_s - E_1$, and the miles completed and miles remaining are updated with the trip's mileage.

Finally, the third conditional statement examines whether a fast charging station is available. In each call to the function, the set of potential charging stations C is projected onto the trip route R_1, and those within a mile of a route are available to use. When multiple charging stations are available along a route, the function selects the furthest such that battery state remains above the minimum upon arrival. When a charging station is used, the route is split at that charging station and updated to T', the battery capacity is updated to $B'_s = .8B_{max}$. Electric vehicles have a "charging curve" which describes the relationship between charging speed and battery capacity. Charging speeds slow considerably above 80% capacity, so I model a user as only being willing to charge to 80%. Before being called again, the function updates mileage to M'_c by adding the distance of the route to the charger to mileage covered so far, subtracts that distance from mileage remaining, and updates minutes charged. If a charging station is not available, the chain cannot be completed and the function returns miles completed up to the current trip, the total miles remaining, and the minutes charged.

The recursive charging algorithm is "greedy" in that it operates on each trip in the chain without looking ahead to the next trip. This assumption is consistent with the behavior of a myopic driver who does not know, or does not think about, their next trip.[10] A simple refinement for future work might allow a driver to estimate the state of charge they finish the current trip with, and look ahead to the next trip to see if they can complete the trip, and if not, charge at the last available charging station in the current trip. I allow a driver to charge at home in a trip chain whenever the destination is their home, where I assume there is a level 2 charging station that charges at 6.6 kW.

[10] An unrealistic result that can arise from the assumption that drivers are myopic is that adding a charging station can actually *reduce* e-miles enabled in some trip chains. This occurs when, under the addition of a charging station, a driver finishes a non-terminal trip in a chain with a lower state of battery than if the charging station had not been added. This might occur, for example because the driver charges once in the middle of the trip at the new charging station, rather than at existing charging stations at the beginning and at the end of the trip. Finishing the non-terminal trip with a lower state of charge might make it harder to complete the subsequent trip and therefore e-miles might be lowered for a trip chain.

Estimating Incremental E-miles of Charging Stations

To estimate annual incremental e-miles for a set of charging stations C, I create subsets of C corresponding to charging stations of each standard: $\{C_s : s \in Tesla, CHAdeMO, CCS\}$. The subsets are overlapping, since many charging stations accommodate both CHAdeMO and CCS standards. I compute e-miles for trip chain T^j for charging station $c_i \in C_s$, by applying the $Charge$ function with and without c_i:

$$\begin{aligned} & e-miles_i \\ & \quad = 365 \sum_s \sum_j \left(Charge \begin{pmatrix} T^j \\ B \\ C_s \\ M \end{pmatrix} - Charge \begin{pmatrix} T^j \\ B \\ C_{s,-i} \\ M \end{pmatrix} \right) W_j P(EV,S)_j \\ & \quad = 365 \sum_s \sum_j \left(M_{s,j} - M_{s,j,-i} \right) W_j P(EV,S)_j \end{aligned} \tag{1.7}$$

I compute the difference in miles completed of each trip chain with and without the charging station, of a particular standard, $M_{s,j}$, and $M_{s,j,-i}$, and take the weighted sum over all trip chains. W_j are the NHTS sample weights. $P(EV,S)$ is the probability of an EV of a particular standard being observed in the census block group (CBG) of that sample household. I construct $P(EV,S)_j$ by merging the household's CBG with the proportion of passenger vehicles in a CBG in 2019 in CA that are EVs. I then multiply by the proportion of registered EVs in each county in 2019 of each standard to yield the probability of an EV of a particular standard in each CBG.

For a small number of trip chains, adding a charging station reduces e-miles traveled. This result is an artifact of the "greedy" behavior of the $Charge$ function. This occurs in the function because adding a charging station results in a the driver charging further along their route, meaning they can complete more trips in a chain, but the battery becomes depleted on a trip without a charging station. If the charging station had not been added, the battery would have been depleted on an earlier trip in the chain that had a charging station available. To correct this, I restrict the trip chain-level difference in miles completed under an additional charging station to be weakly positive.

Simulation Results

To evaluate my simulation methodology, I simulate charging behavior under the expansion of the fast charging network from 2013 to 2020. I use a selection of vehicle range parameters as an input to the simulation. For the charging network, I use the observed network installed in February of each year. I weight the output using NHTS 2017 sample weights multiplied by the 2019 proportion of vehicles of each standard at the CBG, which yields weighted average across charging standards in 2019.

Figure 1.3 presents simulation results showing the relationship between e-miles and an EV's range under the expansion of the fast charging network from 2013 to 2020. For a given year's charging network, as a vehicle's range increases, average annual e-miles increases. In 2013, when the charging network is incipient, average annual e-miles are still over 10,000 for all vehicle ranges. Because most trips occur close to home, and I assume individuals leave home at the start of a trip chain with a full battery, the majority of annual mileage can be completed with an EV with relatively low range. As one would expect, as the charging network expands, for a given range, more e-miles can be completed. The year-on-year increase in e-miles is larger for vehicles with a lower range. In 2019, the effect of range on e-miles is fairly linear, whereas in earlier years the relationship is concave.

Figure 1.4 presents simulation results showing the relationship between annual minutes spent fast charging and an EV's range under the expansion of the fast charging network from 2013 to 2020. For a given year's charging network, increases in range reduce the average annual minutes spent fast charging. As the charging network expands, some trips that were not possible under a small charging network become possible and are taken under an expanded charging network. This effect increases average annual time spent charging as a network expands. The year-on-year increase in minutes spent charging is larger for vehicles with a lower range. The same long trip might be possible under an expanded network for a vehicle of 175 mile range and 300 mile range, but the 175 mile range vehicle will have to charge multiple times.

1.5 A Test for Spatial Inefficiencies

In this section I combine estimates from my choice model and my simulation in a simple test for spatial inefficiencies in fast charging location decisions. I address two questions: A. How does profit correlate with consumer value at marginal charging stations over space? B. Do we observe the vertically integrated network entering in locations with higher consumer value?

How does Profit Correlate with Consumer Value across Space?

Theory tells us that in general, profits do not capture net surplus. I use my simulation methodology to estimate e-miles enabled, which is the ability of a charging station to alleviate range anxiety by enabling trips at marginal charging stations. E-miles enabled is a major component of consumer value that is heterogeneous over space in this setting. I use my demand model to estimate the net change in a network's demand when a charging station location is added to a network, and I show that demand is the main driver of profit in this setting. If the correlation between e-miles enabled and profit across space was strong, this would be evidence against the possibility of spatial inefficiency. Unless business stealing effects were large, we would expect networks to enter locations with high profitability, which would also have high enabled e-miles. However, if the correlation was weak, we should expect

spatial inefficiencies to be present, since entry to a marginal location that is profitable would not necessarily mean that location had a high value of e-miles enabled.

A challenge in comparing charging stations across space is that different charging standards have a different installed base of electric vehicles. Table 1.2 shows counts of vehicles, charging stations, and chargers by standard for 2019. Firstly, note that the number of registered Tesla vehicles is almost double the number of registered vehicles of the other standards combined. Secondly, note that while Tesla has more chargers than other standards, Tesla still has more vehicles per charger, at 83.52, than the other standards, which have 38.16 vehicles per charger, when combined. An implication of this disparity is that a comparison of charging demand at Tesla chargers versus chargers of different standards, suffers from the fact that there are many more registered Tesla EVs compared to EVs of other standards, even when normalized on a per charger basis. All else equal, charging demand at Tesla charging stations will by higher simply because the installed base of EVs is higher. Similarly, weighting trip chain level output of my simulation by the installed base of vehicles of each standard will lead to more annual e-miles enabled at Tesla charging stations simply due to the number of registered Tesla EVs.

Table 1.2: Counts of Vehicles, Charging Stations, and Chargers by Standard for 2019

Standard	Vehicles	Charging stns.	Chargers	Vehicles per charging stn.	Vehicles per charger
Tesla	168,551	143	2,018	1,178.7	83.5
CCS	67,449	512	1,490	131.7	45.3
CHAdeMO	27,298	557	993	49	27.5
CCS/CHAdeMO	94,747	559	2,483	169.5	38.2

Note: Vehicle counts were obtained from the California Department of Motor Vehicles (CBG-level counts) and California Air Resources Board (county-level counts by standard). Counts of charging stations and chargers were obtained from a mobile application and website that allows EV owners to locate charging stations.

To compare charging quantity and e-miles enabled across space in a way that addresses the different vehicles per charger of different charging standards, I increase the installed base of registered EVs of CCS and CHAdeMO standards such that the vehicles per charger is the same as Tesla. In my demand model, this results in a higher value of M_{gs}, the number of potential consumers at gas station location g with standard $s \in \{CCS, CHAdeMO\}$. In my simulation method, inflating the installed base of CCS and CHAdeMO EVs results in a higher value of $P(EV, S)$, the probability of an EV of a particular standard being observed in a CBG of a household, which is the weight applied to results in equation (7). Appendix A.1 presents results from an alternative normalization, which rolls the development of the CCS and CHAdeMO networks back to 2015, while maintaining the 2019 Tesla network,

so as to have an equivalent number of vehicles per charger. The results of the alternative normalization are very similar to those I present here.

Figure 1.5 compares annual charging demand, and annual e-miles enabled for marginal charging station locations in California. Each data point is a charging station location. The size of each point represents the number of chargers at the charging station. In the top left quadrant are locations with high demand and low e-miles enabled. The bottom right quadrant shows locations with low demand and high e-miles enabled. There are no locations with high values for both e-miles enabled and demand. The blue line is a linear fit showing the negative relationship between e-miles enabled and demand across locations. Locations with a high demand tend to be smaller in terms of chargers per location and locations with high e-miles tend to be larger. Networks that put some weight on consumer value and install in locations with a high value for e-miles enabled should also have a incentive to reduce crowding at those locations by installing more chargers. Overall, negative correlation between demand and e-miles enabled over space is evidence that spatial inefficiencies could be present. If the top right quadrant of this graph were populated, and there existed locations with both high annual e-miles enabled and high demand, there would be less reason to expect locations with high consumer value to be absent from the market.

In figure 1.5 I compared annual charging *demand* with e-miles enabled, while I sought to compare *profitability*. A station may be able to maintain higher profits with lower quantities if it can charge higher prices, or lower costs. If so, then quantities are an incomplete measure of profitability. While I do not observe prices and costs at all charging stations so cannot estimate profit directly, the empirical evidence suggests that demand is the main driver of profit in this setting. Table 1.3 summarizes charging station prices in $/kWh across networks.[11] EVgo, the largest charging network in terms of number of charging stations, has four price zones in California: Bay Area, Los Angeles and Inland Empire, San Diego, and greater California. Overall, prices are very similar with a standard deviation of just $0.02 across charging stations. The next two networks by size, Tesla's Supercharger network, and Electrify America, both have a uniform price statewide. In Appendix 2, I present results from regression charging station price on demand and competition variables. There is little correlation between price and measures of competition and demand, which suggests networks are not setting prices in a way to extract rents at remote locations.

Do Vertically Integrated Networks Enter Locations with Higher Consumer Value?

To answer question B, I examine a manufacturer, Tesla, that is vertically integrated. Tesla manufactures and sells vehicles in addition to installing and managing a network of fast charging stations with a proprietary standard that can only be used by Tesla vehicles. While

[11] In the price data that I observe, charging stations price per minute, per session, per kWh or a combination of the three. I convert per minute and per session prices to per kWh prices by assuming a user charges 25 kWh at a rate of 50 kW.

Table 1.3: Prices ($/kWh) by Network

Network	N	Mean	SD	min	max
EVgo	278	0.30	0.02	0.28	0.34
Supercharger	143	0.31	0	0.31	0.31
Electrify America	89	0.31	0	0.31	0.31
ChargePoint	68	0.28	0.12	0	0.50
Greenlots	31	0.25	0.19	0	0.60
None	19	0.30	0.26	0	0.59
Nissan	14	0	0	0	0
EV Connect	10	0.23	0.19	0	0.49
Other	4	0.22	0.22	0	0.49

Note: Data were obtained from a mobile application and website that allows EV owners to locate charging stations. I convert per minute and per session prices to per kWh prices by assuming a user charges 25 kWh at a rate of 50 kW.

an ideal test would compare charging station locations under free entry and a social planner, my test is meaningful because Tesla shares features in common with a social planner. Firstly, it should put some weight on e-miles enabled at charging locations, because consumer demand for vehicles is a function of the quality of a charging network, and the e-miles enabled by a charging network is a major component of consumer value at a charging station location. Secondly, the proprietary standard means EVs from other manufacturers cannot use Tesla charging stations, so Tesla does not steal business by locating near other charging stations. Tesla can only steal business from its own charging stations, which is an effect that it should internalize.[12] A finding that the vertically integrated network enters locations with higher incremental e-miles would therefore be evidence in favor of a spatial inefficiency in free entry. In theory, how profits compare between a vertically integrated and non-vertically integrated provider is unclear. In the absence of any business stealing effects, we might expect revenues to be lower if incremental e-miles is imperfectly correlated with revenue over space. However, business stealing among competing networks should lead to excess entry and lower demand for non-vertically integrated providers.

Figure 1.6 compares distributions of annual e-miles enabled and annual charging demand for Tesla and non-Tesla charging stations. The left panel presents the annual e-miles enabled for Tesla and non-Tesla charging stations. Observations are weighted by the number of chargers at a station. The average charging stations in the vertically integrated Tesla network enables 700.6k total annual e-miles, whereas the average charging stations at other networks

[12] An adapter allowing Tesla model S and X drivers to use CHAdeMO charging stations has been available since 2013 for $450. It became available to model 3 drivers in mid-2019. An adapter allowing Tesla vehicles to use charging stations from the CCS standard became available in 2022. There is no adapter allowing other drivers to use Tesla charging stations.

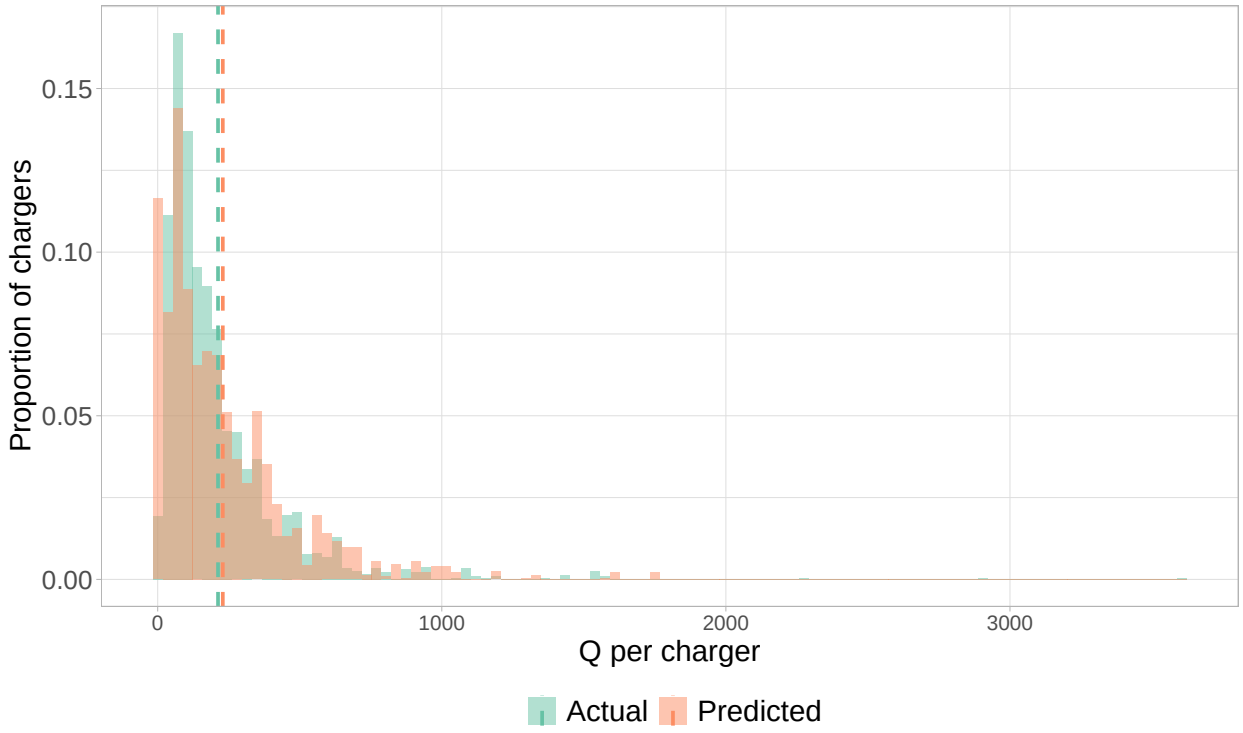
enables 55.5k total annual e-miles. The difference in means is statistically significant at the 1% level. This result is consistent with Tesla entering locations with a higher consumer value, and is strong evidence in favor of a spatial inefficiency in free market entry.

The right panel of figure 1.6 compares annual charging demand for Tesla and non-Tesla charging stations. I find average annual demand at Tesla chargers is slightly lower than at chargers in other networks. Lower demand per charger at Tesla chargers suggests that the effect of entering locations with higher e-miles enabled that tend to have lower demand outweighs the business stealing effect that increases crowding and reduces demand at other networks. These findings suggest that leaving location decisions to the free market, comprised mostly of non-vertically integrated charging station providers, may result in sub-optimal location choice.

Figure 1.7 presents a map comparing e-miles enabled at 2019 charging station locations for Tesla and non-Tesla charging stations. Charging stations in the top quartile of overall e-miles enabled (weighted by chargers per station) are shown in teal, with bottom three quartiles shown in orange. Overall, most of the charging stations are relatively close to the San Francisco Bay Area and Greater Los Angeles. Tesla, shown in the left-hand panel, has a larger share of stations in more rural areas than non-Tesla networks, whose stations are densely clustered near metropolitan areas. Related, is that stations in the top quartile of e-miles enabled tend to be in relatively rural areas, and therefore stations in the top quartile of e-miles enabled make up a larger share of Tesla stations.

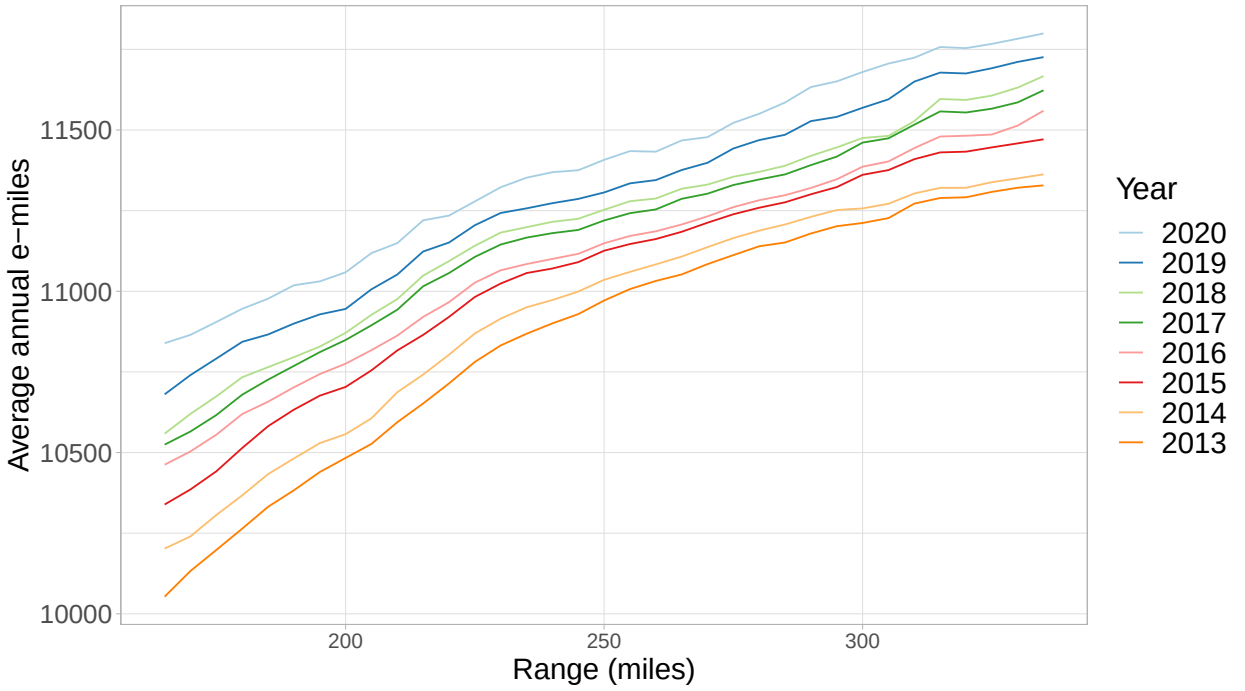
Figure 1.8 presents a map comparing demand at 2019 charging station locations for Tesla and non-Tesla charging stations. Charging stations in the top quartile of overall demand per charger enabled (weighted by chargers per station) are shown in teal, with bottom three quartiles shown in orange. In contrast with 1.7, which showed e-miles enabled, stations in the top quartile of demand tend to be close to metropolitan areas, rather than in more rural areas. This holds true for both Tesla and non-Tesla stations. However, stations in the top quartile of demand tend to belong to non-Tesla networks, shown in the map on the right-hand panel.

Figure 1.2: Distribution of Annual Charging Demand per Charger



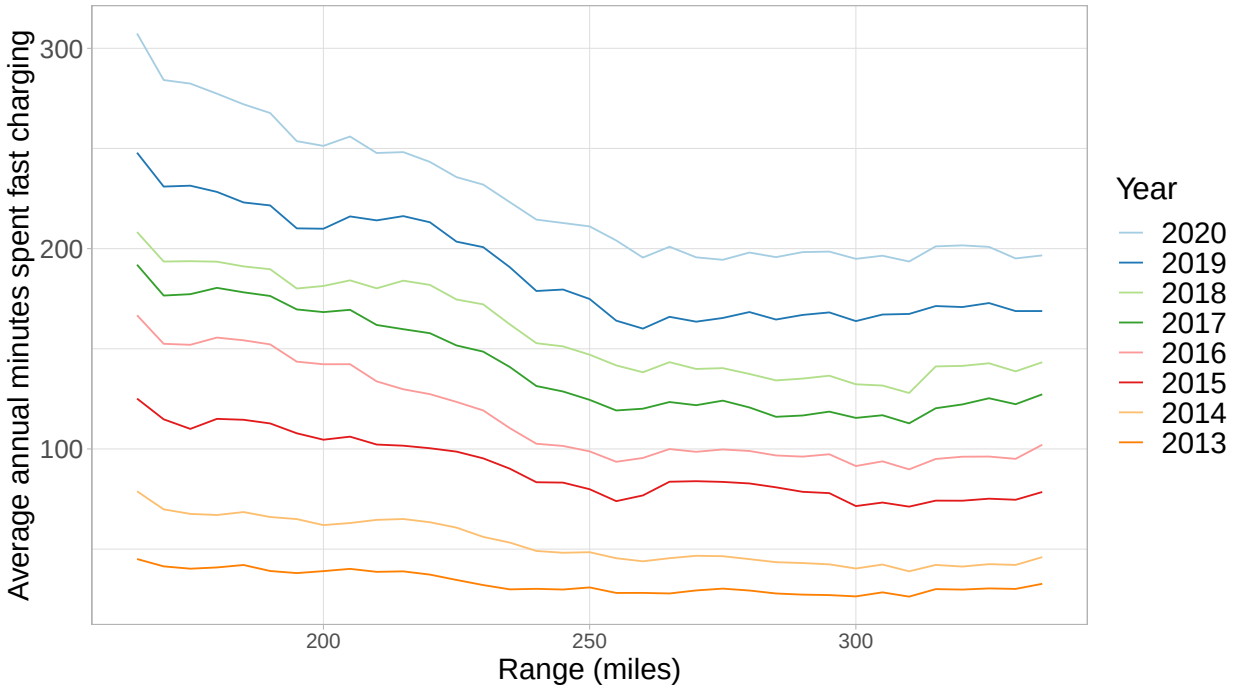
Note: This figure shows the distribution of annual charging demand per charger. Values are at the charging station level (N = 702), weighted by the number of chargers at each station. Actual values that I observe in the data are shown in teal, and predicted values from the demand model are shown in orange. The dashed vertical lines show the mean, which is 212.3 for actual values, and 229.2 for predicted values.

Figure 1.3: Average Annual E-miles Versus Range by Year



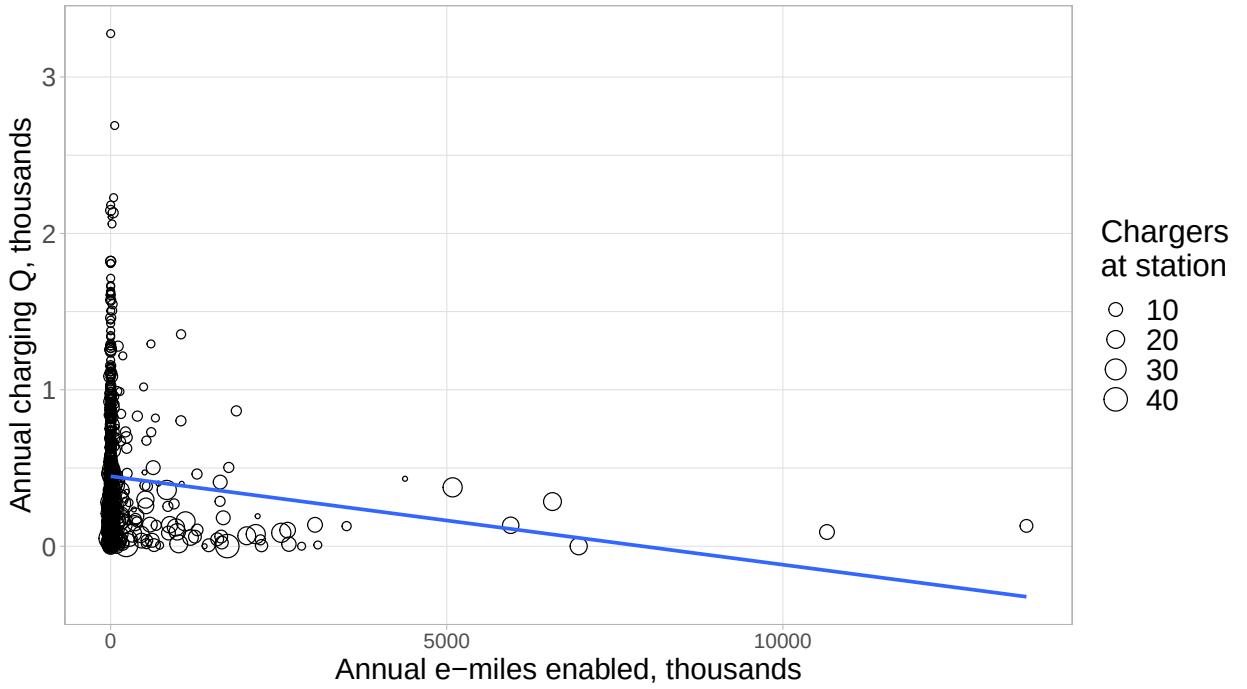
Note: This figure shows simulated estimates of the relationship between average annual e-miles and an EV's range under the expansion of the fast charging network from 2013 to 2020. Average annual e-miles is total household mileage of observed NHTS trips that are possible to complete in an electric vehicle using home charging and fast charging stations, averaged over households in the 2017 NHTS sample.

Figure 1.4: Average Annual Minutes of Fast Charging Versus Range by Year



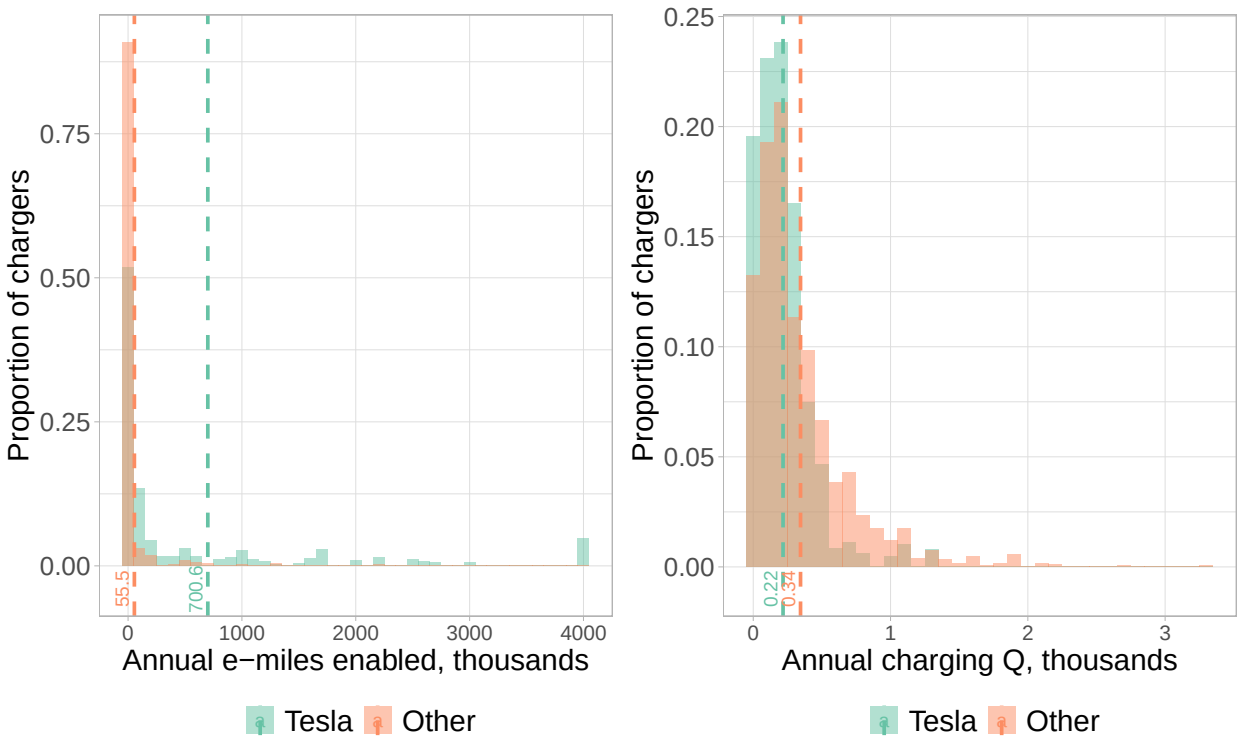
Note: This figure shows simulated estimates of the relationship between average annual minutes spent fast charging and an EV's range under the expansion of the fast charging network from 2013 to 2020. Average annual minutes spent fast charging is defined as total household minutes spent fast charging to complete observed trips, averaged over households in the NHTS 2017 sample.

Figure 1.5: Annual Charging Quantity Versus Annual E-miles Enabled at 2019 Charging Stations



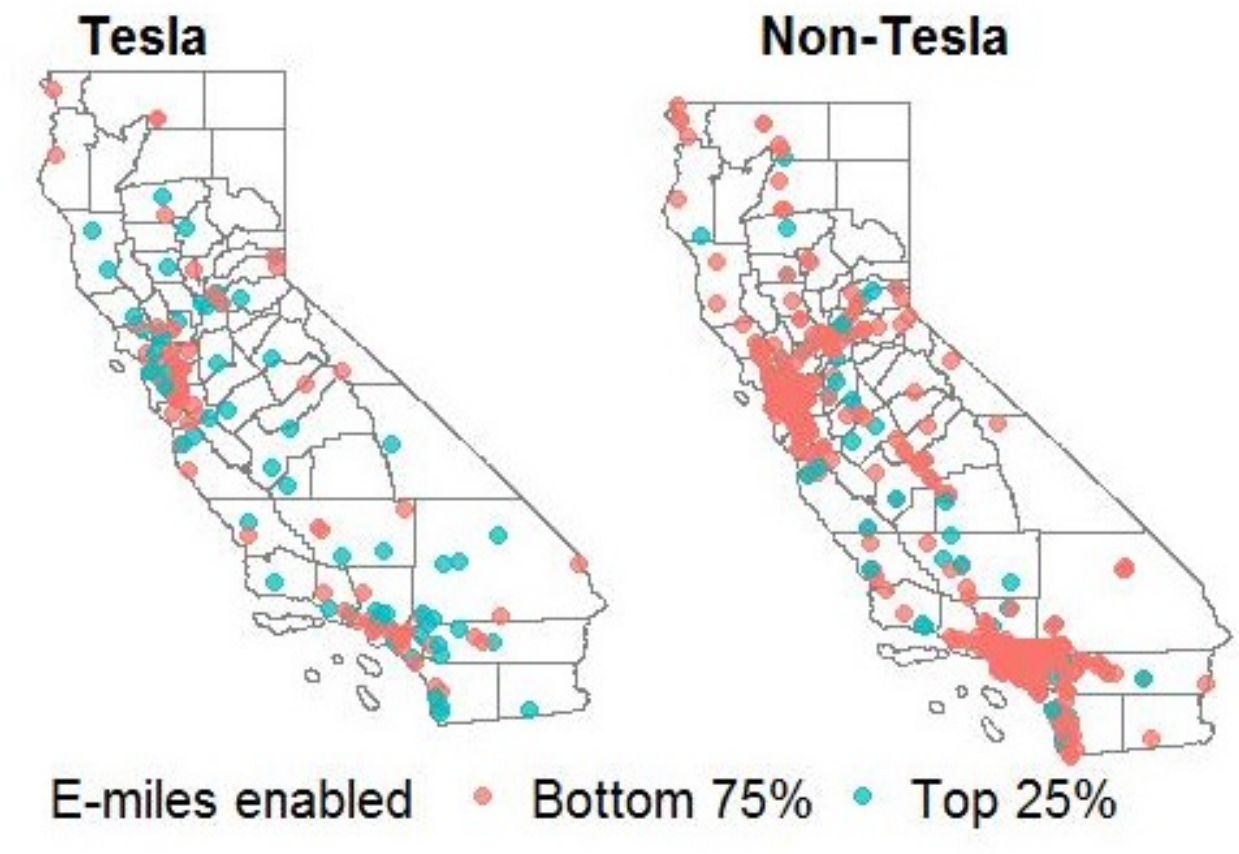
Note: N = 702. This figure shows annual e-miles enabled and annual charging demand for marginal charging station locations in California. Each data point is a charging station location. The size of each point represents the number of chargers at the charging station.

Figure 1.6: Distribution of Annual E-miles Enabled and Annual Charging Demand for Tesla and Non-Tesla Charging Stations



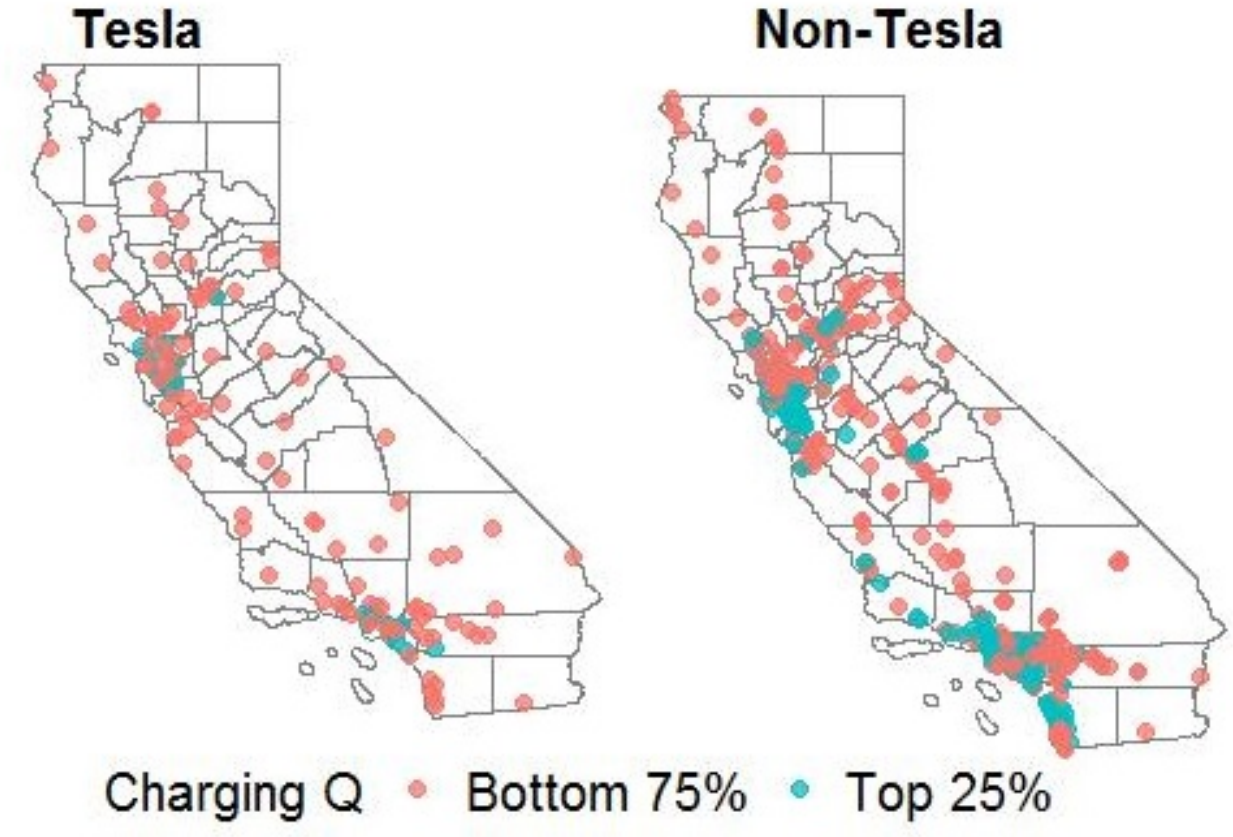
Note: This figure shows distributions of annual e-miles enabled and annual charging demand for Tesla and non-Tesla chargers. Observations are at the charging station level, weighted by the number of chargers at a station. The left panel presents the annual e-miles enabled, and the right panel presents annual charging demand. The vertical dashed lines show average values for Tesla and non-Tesla chargers.

Figure 1.7: E-miles Enabled at 2019 Charging Station Locations, by Network



Note: This figure shows the location of Tesla and non-Tesla charging stations in California, categorized by whether they fall in the top quartile for estimated e-miles enabled. E-miles enabled is the estimated total mileage of the previously unachievable trips that become possible as a charging station location is added to a network.

Figure 1.8: Demand at 2019 Charging Station Locations, by Network



Note: This figure shows the location of Tesla and non-Tesla charging stations in California, categorized by whether they fall in the top quartile for charging station demand. Charging station demand is the net change in a network's demand when a new charging station is added to a network.

Table 1.4: Price Regression Results

	Dependent variable:			
	Dollars per kWh			
	(1)	(2)	(3)	(4)
log(Annual charging Q)	−0.005	−0.0003	−0.002	0.001
	(0.003)	(0.001)	(0.003)	(0.002)
Stations within 10 miles	−0.001	−0.001	−0.001	−0.0004
	(0.0004)	(0.0004)	(0.0004)	(0.0003)
Intercept	0.326***	0.310***		
	(0.012)	(0.006)		
Fixed effects?	No	No	Network	Network
Weights?	No	Chargers per station	No	Chargers per station
Observations	656	656	656	656
R^2	0.029	0.022	0.309	0.292
Adjusted R^2	0.026	0.019	0.295	0.278
Residual Std. Error	0.089 (df = 653)	0.147 (df = 653)	0.075 (df = 642)	0.126 (df = 642)

Note: *p<0.05; **p<0.01; ***p<0.001. The mean of the dependent variable is 0.291, unweighted, and 0.299, weighted. Each regression is at the charging station level. Standard errors are clustered at the network level.

1.6 Conclusion

Electrification of the transportation sector is necessary if the U.S. is to address climate change. This is most likely to occur through the adoption of electric vehicles (EVs). Adoption to date, however, has been slow. Widely cited barriers to EV adoption include range anxiety, cost, and the availability of public charging infrastructure.

Range anxiety, which is a driver's fear that a vehicle has insufficient battery capacity to reach their intended destination, is related to the availability of public charging infrastructure. If a charging station is located appropriately, it may reduce the size of a driver's set of unachievable trips, and thereby alleviate range anxiety. Policy makers are attempting to foster electrification by encouraging the build-out of a more comprehensive charging network, but many open questions remain about the role of charging stations and the ideal design of policies to subsidize charging stations.

In this paper I examine the location decisions of firms providing electric vehicle fast charging stations. I empirically test for the presence of a spatial inefficiency in the location decisions by comparing locations of vertically integrated firm that sells vehicles and provides a charging network to those of a non-vertically integrated charging network. I define a metric at the charging station station level, *e-miles enabled*, as the total mileage of the previously unachievable trips that are now possible as a charging station location is added. E-miles enabled at potential charging station location depends on the location of other charging stations in the network. In this setting, e-miles enabled is a major component of consumer value from a charging station that is heterogeneous over space. I combine estimates from a spatial demand model and a simulation of charging behavior to compare charging demand and enabled e-miles at charging stations in California. I find strong evidence in favor of a spatial inefficiency in the location decisions of non-vertically integrated charging providers. These findings suggest that leaving location decisions to non-vertically integrated, charging station provider may result in sub-optimal location choice. EV policy should not be location-neutral, but should consider spatial inefficiencies as well as network size. A location-based subsidy that varies with a location's e-miles enabled should be considered.

Chapter 2

Subsidizing Electric Vehicle Charging Stations for Renters

As discussed in the introduction to Chapter 1, the transportation sector is the largest contributor to anthropogenic U.S. greenhouse gas (GHG) emissions. If the U.S. is to address climate change, a key mechanism will be through electrification of the transportation sector. This must occur primarily through the adoption of electric vehicles (EVs). Adoption to date has been slow: In 2018 EVs had a market share of about 2% of nationwide sales. Widely cited barriers to EV adoption include range anxiety, cost, and availability of public charging infrastructure. The availability of home- and workplace-charging infrastructure for renters has received little academic attention. According to the National Household Transportation Survey, in 2017, 34% of vehicles in California, and 24% nationally, are owned by renters. Davis, 2019 finds that homeowners are three times more likely than renters to own an EV, even after controlling for income and other factors. Policies supporting charging infrastructure in workplaces and multi-unit dwellings (MUDs) are nascent. In this paper, I aim to address two questions: (i) are such policies effective in increasing investment in charging stations, and (ii) whether they are effective in promoting EV adoption. This work represents the first causal evidence on whether charging infrastructure represents a barrier to adoption for MUD occupants.

The problem of estimating the impact of charging infrastructure in MUDs and workplace on EV adoption is well-suited to the use of two-sided market framework. In the two-sided market framework introduced by Rochet and Tirole, 2006 and Armstrong, 2006, the benefits for one group, EV owners, from joining a platform, depend on the size of the other group, EV charging operators, that joins a platform. Given the variation I observe in the data, I cannot implement such a two-sided market framework. I will report on instrumental variables estimates of the effect of an additional MUD or workplace charging station on EV registrations. The instrumental variable exploits spatial and temporal variation in subsidies that reduce the cost of installing charging stations in MUDs and workplaces. Treatment timing differs across the three largest investor-owned utilities (IOUs) in California, with subsidized charging stations installed in Southern California Edison (SCE) and San Diego

Gas & Electric (SDG&E) territories in 2017, and in Pacific Gas & Electric (PG&E) territory in 2018, the final year of the sample. Areas of California served by other utilities did not receive subsidies in my sample period. The first-stage estimates are identified using variation in the number of charging stations in untreated vs treated CBGs within counties that span IOU boundaries, as subsidies are introduced. Using vehicle registration counts at the census block group-year level that I obtained from California Air Resources Board (CARB), I estimate the effect of an additional MUD or workplace charging station on the number of registered EVs.

As it stands, I do not detect a statistically significant effect of MUD or workplace charging stations on EV registrations. My coefficient of interest is 0.379, indicating that an additional MUD or workplace charging station in a census block group (CBG) yields a 38% increase in registrations. However, the standard error is 1.131, so I cannot rule out extremely large or even negative effects. This analysis would possibly be improved by the use of monthly data. This would add precision, as well as allow more precise encoding of the month of year that subsidies were introduced. Furthermore, my estimation sample spans 2015 through 2018, and the subsidies were available from 2017 to 2019 and likely have a lagged effect. The inclusion of additional years of data, not yet available at the time of writing, would allow one to estimate the effect at the program's full deployment.

While the primary goal of the study was to execute the two-stage design described above, the first-stage results on take-up of subsidies for MUD and workplace charging stations may itself be a finding of importance for the design of subsidies. I find that the subsidies increased the number of charging stations in a census block group (CBG) by .01. Against the sample average of 0.053, this represents a 20% increase. Transfers to inframarginal recipients of subsidies are a problem in new technologies with externalities because they induce a welfare loss (Boomhower and Davis, 2014). In this setting, as is often the case in energy-related investments, there are multiple distortions. Charging stations offer external benefits through both positive environmental externalities, network externalities, and landlord-tenant problems may also mean investment falls below the socially optimal level. Furthermore, Borenstein and Davis, 2016 showed that across a broad range of "clean energy" investment tax credits, take-up was strongly correlated with income. The effect was most extreme for electric vehicle tax credits whereby the top quintile claimed 90% of the credits.

In this setting policy makers appear to have attempted to overcome some of theses issues by targeting subsidies to disadvantaged communities (DACs).[1] The subsidy amount to locations in disadvantaged communities is twice that to locations elsewhere. PG&E's program has a target of 15% of subsidized ports in disadvantaged communities. Nevertheless, my first stage results of the effect of charging subsidies on MUD and workplace charging stations are monotonically increasing in income quartile. Alternative samples for disadvantaged and non-

[1]Disadvantaged communities (DACs) are census tracts in California identified by CalEPA as suffering most from a combination of economic, health, and environmental burdens. The tracts are identified using an analytic tool, CalEnviroScreen, which scores tracts based on pollution and economic factors. Tracts above the 95th percentile are defined as DACs.

disadvantaged communities show the effect is larger in the non-disadvantaged communities that received the smaller subsidy amount.

This study is the first to yield any causal estimates of adding charging stations to MUDs and workplaces on EV adoption. I will make a meaningful contribution to the nascent literature examining network externalities and subsidy structure in the EV and charging infrastructure markets. J. Li, 2019 studies compatibility and investment in the fast-charging station industry. Corts, 2010 studies the effect of the installed base of flexible fuel vehicles on the deployment of fueling stations. S. Li et al., 2017 model the development of the EV sector in the US as a two-sided market (comprised of *public* charging infrastructure and vehicles) with network externalities. Springel, 2019, models the EV sector in Norway as a two-sided market with network externalities by using a BLP-type framework. My innovation is to focus on a new barrier to adoption: home and workplace charging station availability to renters. I will also exploit rich subsidy variation, and vehicle registration data at fine geographic resolution in order to yield credible causal estimates.

This research also contributes to a literature studying environmental policy and the market for automobiles and distributional considerations. Muehlegger and Rapson, 2018 exploit quasi-experimental variation in EV subsidies in California to study EV demand. Allcott, C. Knittel, and Taubinsky, 2015; Borenstein and Davis, 2016 find that take-up of subsidies for energy efficient products is strongly correlated with income. I contribute by estimating heterogeneous effects by income, and as well as effects for disadvantaged communities (DACs).

My work makes contribution to a literature on landlord-tenant problems related to energy efficiency and environmentally friendly goods.[2] Gerarden, Newell, and Stavins, 2015 examine the energy efficiency gap and potential explanations found in previous literature, including market failures, behavioral anomalies, and model and measurement errors. Davis, 2019 finds that homeowners are three times more likely than renters to own an EV. Davis argues that this is a version of the landlord-tenant problem.

2.1 Data

This study draws on two main data sources. The first is vehicle registrations collected from the California Department of Motor Vehicles. These data were aggregated and made available to researchers by the California Air Resources Board. The second is a comprehensive dataset of locations of public and private EV charging stations in California. I used a web-scraping program to obtain the data directly from PlugShare, a mobile application and website that allows EV owners to locate charging stations.

[2] A landlord-tenant problem is a form of principal-agent problem whereby renters do not want to invest in property they do not own, and landlords may not make the investments on renters' behalf. It is unclear whether this represents a market failure in the case of EV charging infrastructure.

Vehicle Registrations Data

I use vehicle registrations data from the California DMV. The data provide counts of total vehicle registrations by census block group (CBG), by year. Within a CBG-year, the data are further broken out in a fuel-technology subcategory that includes battery electric vehicle (BEV), fuel-cell electric vehicle (FCEV), internal combustion engine (ICE), and plug-in-hybrid electric vehicle (PHEV). I use fuel-technology variable to categorize vehicles as EV or non-EV. The data are also aggregated by the number of vehicles registered at the same address, including categories for one, two, three and more than four vehicles. I use the number of vehicles at the address to construct multiple-vehicle (more than 4 registered) and non-multiple-vehicle counts of registrations. Multiple-vehicle addresses represent 25% of EV registrations in the sample. Figure 2.1 shows percent of vehicles in a CBG registered at multi-vehicle addresses. Most notable are the bright green CBGs around major airports, where car rental companies are located. Because multi-vehicle addresses appear to be commercial addresses, I exclude multi-vehicle registrations from the analysis.

Charging Stations Data

PlugShare is a mobile application and website that allows EV owners to locate charging stations. Using a web-scraping program, I obtained a comprehensive dataset of locations of public and private EV charging stations in California. A notable feature of the PlugShare data is that they feature restricted station, which are not available to the general public, but might be located at private workplaces, or apartment buildings. I use these data to construct a variable encoding where MUD and workplace charging stations have been installed, which is the dependent variable in my first-stage.

The vast majority of EV charging stations belong to networks, in which case the station is connected to online management tools that facilitate pricing, and real-time availability monitoring. My understanding from conversations with industry experts is that networks add their locations to PlugShare as new stations are available. Non-networked stations must be added by PlugShare users or the station's owner. I construct a panel of charging station locations using the date the charging station was first added to PlugShare. While it is difficult to verify whether charging stations are added in a timely manner, many stations appear to be added before they are open for use, and are listed as "coming soon." This suggests that networks and users make charging station locations available promptly, as new locations are installed.[3]

In addition to a charging station's location, I obtain its name, a description, and a categorical variable that encodes places of interest, such as a school, or workplace, or apartment building, where the charging station is located. I use a combination of the above variables to

[3]If a charging station is removed during the sample period, it will not appear in my panel, since I can only obtain data for currently installed charging stations. While it's possible to remove charging stations, the occurrences should be rare since charging stations have a high upfront fixed cost of installation and low variable costs of operation.

Figure 2.1: Percent of Vehicles Registered at Addresses with more than Four Vehicles, for Bay Area Census Block Groups, 2018.

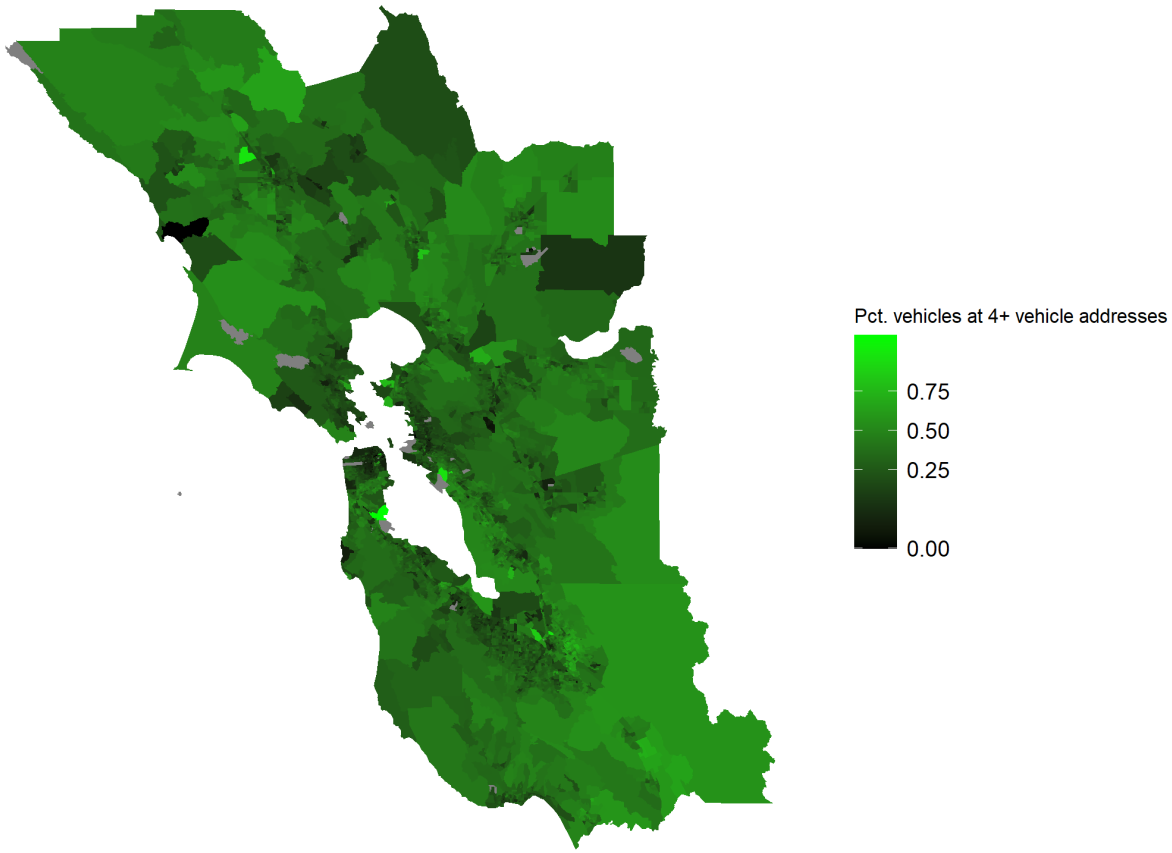
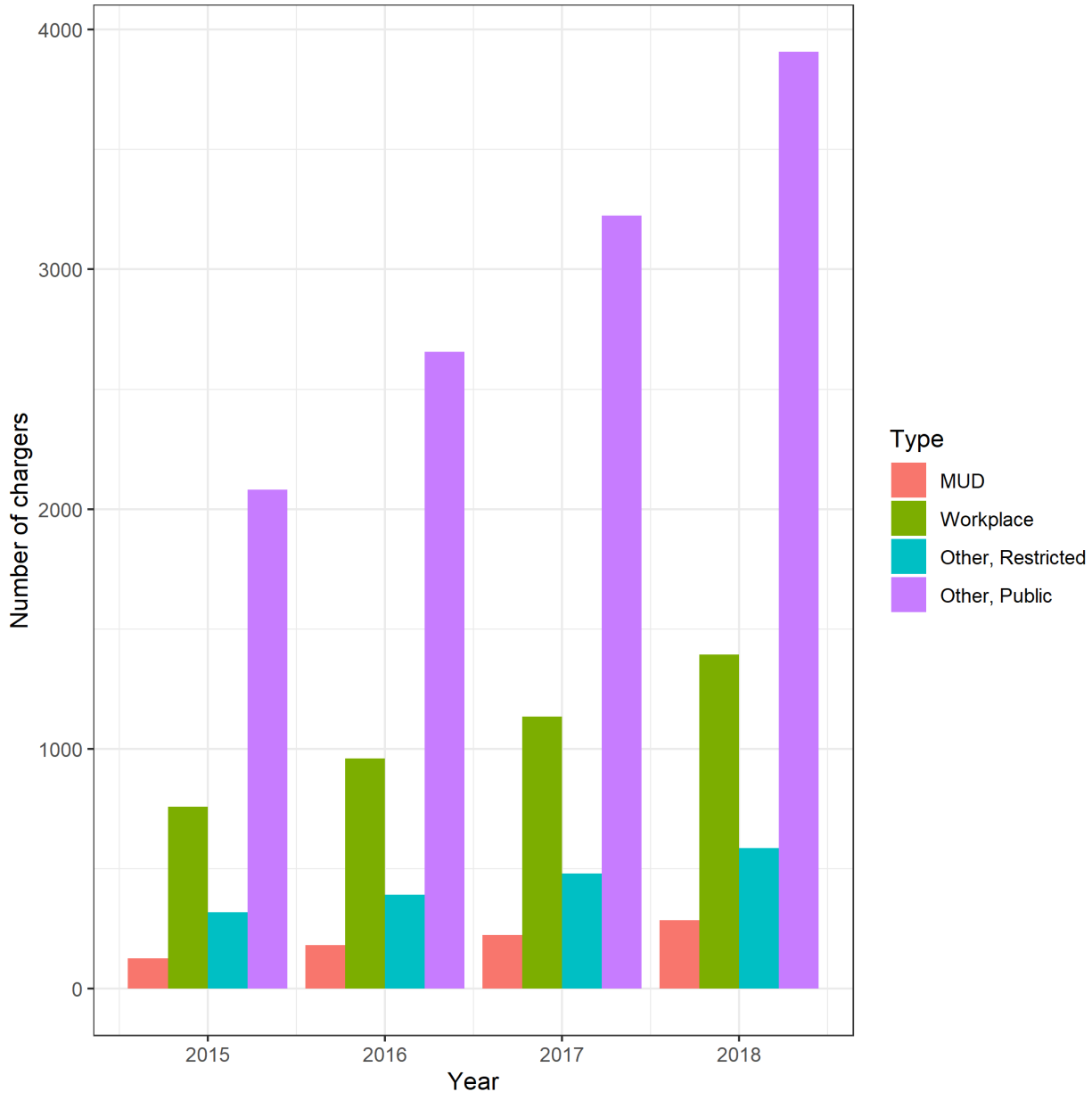


Figure 2.2: Number of Charging Stations by Type for California, 2015–2018.



identify charging stations as workplace charging stations, MUD charging stations, or public charging stations. Figure 2.2 shows the number of charging stations of each type for the sample period from 2015 to 2018. There are about 4,000 public charging stations in 2018, 1,400 workplace charging stations, and about 250 MUD charging stations. While MUD charging stations are few in number, the market has grown rapidly: there were just 105 MUD charging stations in 2015.

2.2 Policy Background

PG&E, SCE and SDG&E are currently subsidizing investment in charging stations. They will install infrastructure to support up to 12,500 charging stations with a total combined budget of $197 million. The cost of a charging station installation is comprised of the electric vehicle supply equipment (EVSE) as well as the cost of site infrastructure from the distribution line to the parking space. The EVSE is the equipment from the wall-plug to the car that controls charging. The infrastructure includes distribution infrastructure, panels, conduit and wiring. The cost of EVSE is about $2,300 per port and comprises about 30% of the total cost of installation for an average charging station, with the other 70% being made up by infrastructure costs. The subsidy is for the full value of infrastructure, which the IOU will own, plus 25 to 100% of the EVSE cost. The IOUs proposed these incentives which were ultimately approved by the CPUC.[4]

While the primary objective is to increase adoption of EVs, regulators also aim to improve air quality and economic conditions in disadvantaged communities (DACs). DACs are census tracts in California identified by CalEPA as suffering most from a combination of economic, health, and environmental burdens. The tracts are identified using an analytic tool, CalEnviroScreen, which scores tracts based on pollution and economic factors. Tracts above the 95th percentile are defined as DACs. The investment subsidies are higher for workplaces and MUDs in DACs.

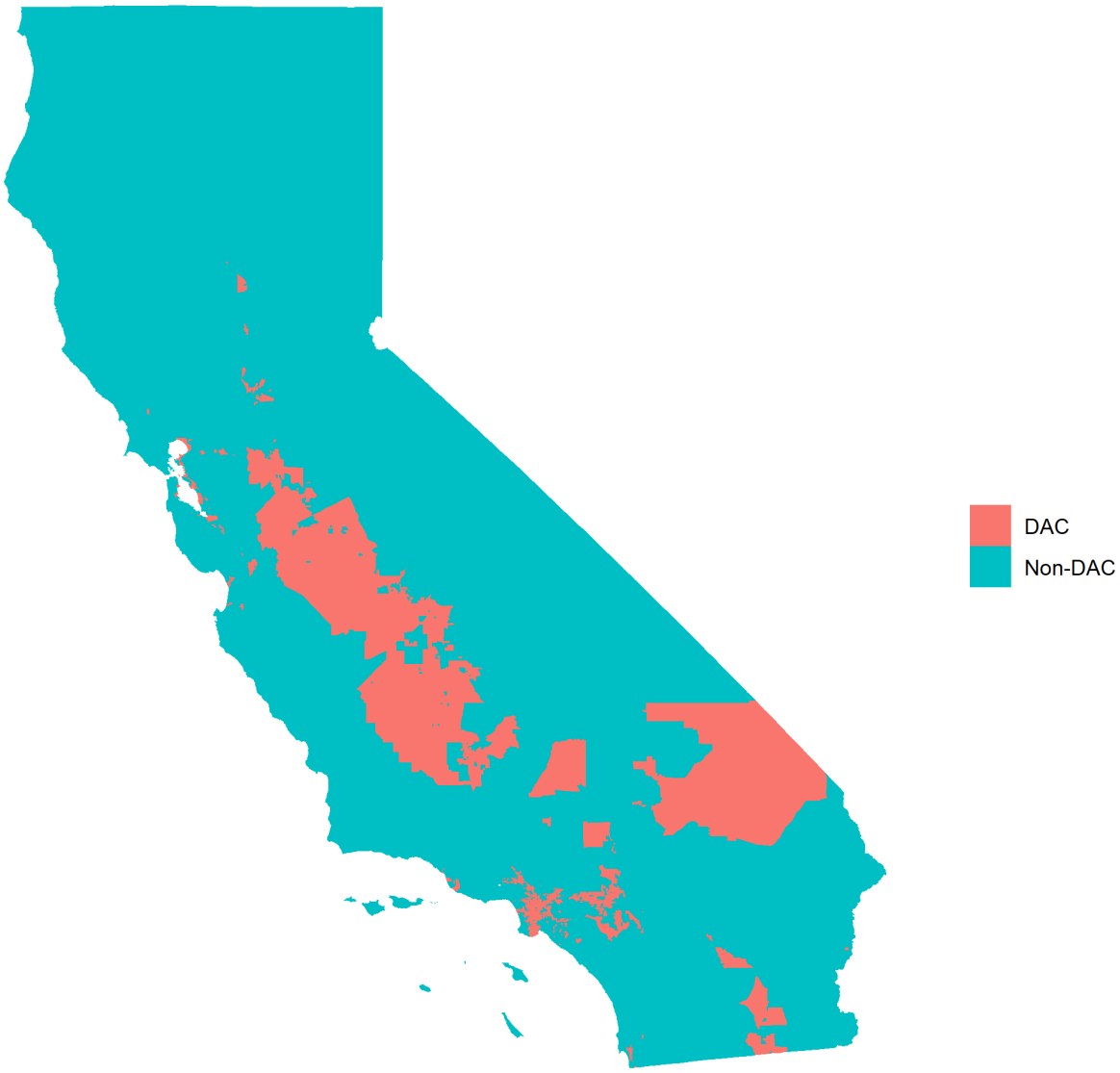
PG&E's EV Charge Network program is an investment subsidy. Its goal is to incentivize installation of EV charging stations at multi-unit dwellings and workplaces with more than 10 parking spaces. The program includes targets under-served communities through explicit targets of 15% of program charging stations deployed in Disadvantaged Communities (DACs), and 20% deployed in MUDs. The program was approved by the CPUC on December 15, 2016, and officially launched in January, 2018. The program's duration is 5 years and the budget is $130 million.

For all program participants, PG&E pays for infrastructure from the pole (distribution line) to the parking space. The customer then pays for the charging station hardware that is subsidized with a rebate. The amount of the subsidy varies according to whether the customer is a MUD or workplace, and whether the customer is located in a DAC. Table 2.1 shows the amount of subsidy the customer receives for charging station hardware based on their segment (MUD or workplace) and location (DAC or non-DAC). Additionally, certain customers may opt to allow PG&E to own the hardware, or own the hardware themselves. While the table shows subsidy amounts for PG&E, the subsidy amounts are very similar in structure and amount at the other IOUs.

Customers that are MUDs, or workplaces in DACs may opt for the PG&E to own the EVSE. (Workplaces outside of DACs are not eligible for this ownership option.) This lowers the overall cost. Multi-unit dwellings in DACs pay no upfront EVSE costs, and multi-unit dwellings outside of DACs pay only a participation payment of $1,150 per port. Workplaces

[4] docs.cpuc.ca.gov/PublishedDocs/Published/G000/M157/K835/157835660.PDF

Figure 2.3: California Census Tracts by Disadvantaged Community (DAC) Status



Note: Figure shows California census tracts with disadvantaged community (DAC) status. DACs are census tracts identified by CalEPA as suffering most from a combination of economic, health, and environmental burdens. The tracts are identified using an analytic tool, CalEnviroScreen, which scores tracts based on pollution and economic factors. Tracts above the 95th percentile are defined as DACs.

Table 2.1: PG&E Charging Station Subsidy Amount by Segment and Location

Segment	Location	Rebate ($ per port)
Multi-unit dwelling	DAC	2,300
	Non-DAC	1,150
Workplace	DAC	1,150
	Non-DAC	575

in DACs also only pay a participation payment of $1,150 per port.

SCE's Charge Ready is an investment subsidy for EVSE. Similar to PG&E's program, SCE installs and maintains the infrastructure from the distribution line to the parking spot at no cost to the participant, and provides a subsidy for charging hardware purchased by the participant. Non-residential customers are eligible. The budget is $22 million, and the subsidy amount varies by whether the customer is in a DAC, a multi-unit dwelling, or other eligible facility. The program launched in January 2017, and as of December 2018, 1,063 charging ports at 71 sites had been built.[5]

SDG&E's Power Your Drive is also an investment subsidy for charging station installation in MUDs and workplaces. The structure is similar to PG&E and SCE's programs. The program's budget is $45 million and installation of charging stations started in June 2017. Similar to the other programs, the generosity of the subsidy varies by whether the location is in a DAC or not, as well as whether it is a MUD or workplace. Participation is free to locations in DACs, and participants in other locations pay a one time payment equal to $630 per port for businesses and $235 per port for MUDs.

2.3 Empirical Strategy

First Stage

An ideal experiment would randomly assign CBGs in California to treatment and control groups. The treatment group would receive charging station subsidies, and the control group would not. Due to random assignment, treatment assignment is independent of potential outcomes of treated and untreated units. Simply taking the difference between the average number of charging stations in the subsidized and unsubsidized groups would yield an estimate of the average treatment effect of subsidies on the number of charging stations in California CBGs. In the non-experimental setting that I analyze, treatment assignment is not randomly assigned, but units select into treatment based on whether they are in an IOU.

My empirical strategy exploits the geographic and temporal variation in deployment of charging station subsidies in a difference-in-differences design. Estimates are identified using variation in the number of charging stations in unsubsidized versus subsidized CBGs within counties that span IOU boundaries, as subsidies are introduced.

[5] *www.sce.com/business/electric-cars/Charge-Ready*

I estimate a panel data difference-in-differences specification with variation in treatment timing:

$$Charging\ stns._{bt} = \beta_{DID} Subsidy\ treatment_{bt} + x_{bt} + \gamma_b + \delta_{ct} + \varepsilon_{bt} \tag{2.1}$$

where $Charging\ stn.s_{bt}$ is the number of charging stations in a CBG (b) at the end of calendar year (t) from 2015 to 2018. $Subsidy\ treatment_{bt}$ is a binary variable encoded as one after 2017 for SCE and SDG&E, and after 2018 for PG&E, and zero otherwise.[6] x_{bt} includes controls for any factors that may affect the number of charging stations in a way that varies over time and geography. I include fixed effects γ_b at the level of an CBG, which absorb time-invariant features of an area, as well as yearly time effects δ_{ct} at the county level, which absorb the variation in number of charging stations over time that is common across CBGs within a county. Relative to a specification with common time effects, which would require that parallel trends hold between arbitrary CBGs in California, county-level time effects relax this assumption, requiring that parallel trends hold between CBGs within a county. However, the more palatable parallel trends assumption comes as the expense of external validity. The difference-in-differences coefficient β_{DID} is estimated from the difference within each county between the number of charging stations in subsidized and unsubsidized CBGs, pre- versus post-subsidy roll-out. The treatment effect is therefore estimated using variation within counties that span IOU boundaries, rather than the entire IOU territory.

I identify the causal effect of subsidies on charging stations in equation 2.1 if the parallel trends assumption holds across CBGs within a county – that is if the number of charging stations in subsidized versus unsubsidized CBGs within counties that span IOU boundaries is not trending apart for reasons other than the subsidies themselves. A potential threat to identification would be local shocks to charging infrastructure correlated with subsidies. An example is incentives that local governments provide to charging stations. These subsidies would have to be administered the sub-county-level and be correlated with the introduction of IOU subsidies in order to invalidate the identifying assumption of parallel trends.

EV Demand

I estimate the following equation:

$$arsinh(EV_{bt}) = \beta_1 MUD/workplace\ charging\ stns._{bt} + x_{bt} + \gamma_b + \delta_{ct} + \varepsilon_{bt} \tag{2.2}$$

where $arsinh(EV_{bt})$ is the inverse hyperbolic sine of the number of registered EVs at the level of a CBG (b) and year(t) from 2015 to 2018. $MUD/workplace\ charging\ stns._{bt}$ denotes the total number of MUD and workplace charging stations in a CBG at the end of a calendar year. The fixed effects are the same as in the first-stage specification above. I include fixed effects γ_b at the level of an CBG, which absorb time-invariant features of an area, as well as yearly time effects δ_{ct} at the county level, which absorb the variation in number of EVs over

[6] I categorize a small number of census block groups that overlap utility boundaries as treated if they overlap a boundary between treated and untreated IOU territories in a year. This is a conservative approach as it will attenuate any estimates of a positive treatment effect.

time that is common across CBGs within a county. x_{bt} includes control variables such as the number of public charging stations. The specification suffers from omitted variable bias if, for example, EV incentives that vary within a county are also correlated with changes in the number of MUD/workplace charging stations. I am collecting information on local EV incentives that I plan to incorporate into the above equation. ε_{bt} represents unobserved time-varying local shocks to EV demand. Demand shocks that result from a location-specific model preference, such as a strong preference for Tesla model 3 in San Francisco, are absorbed by fixed effects if the demand shock is county-specific.

The above specification still suffers from endogeneity due to simultaneity that arises due to unobserved time-varying and CBG-specific demand shocks that affect the number of charging stations through investment decisions. I hope to address this through the use of IV. I document my instrumental variable and first-stage analysis above. A valid IV must be correlated with the independent variable of interest, and uncorrelated with other factors that affect EV demand. A potential concern is endogeneity arising if the IOU EV incentives are a response to local unobserved demand shocks (S. Li et al., 2017). In this setting this is a particular concern as the subsidy generosity was higher in disadvantaged communities, possibly as a response negative demand shocks that affect both EVs. To overcome this, I do not use subsidy generosity as a instrument, I use a binary variable for if a CBG was subsidized or not. Furthermore, I also only use variation within counties that span IOU boundaries. Demand shocks are more likely to be common across CBGs within counties than they are across arbitrary CBGs in CA.

2.4 Results

In this section I first report results for the first-stage estimates of the effect of subsidies on number of charging stations. Then, I report OLS and IV estimates for the effect of MUD and workplace charging stations on EV registrations.

First Stage Results

Table 2.2 reports results for the effect of the subsidy on number of charging stations. The dependent variable in all regressions is the number of charging stations of a particular type, measured at the level of a census block group-year. Each column shows estimates for a different categorization of charging station type (MUD and workplace, MUD, workplace, public). The independent variable in all regressions is *Subsidy treatment*, and is the standard independent variable in a difference-in-differences specification with variation in treatment timing. Treatment timing differs across the three IOUs, with subsidized charging stations installed in SCE and SDG&E territories in 2017, and in PG&E territory in 2018, the final year of the sample.

The coefficient on *Subsidy treatment* in column 1 is statistically significant at the 1% level. It represents the average causal effect of subsidies on the number of MUD and workplace

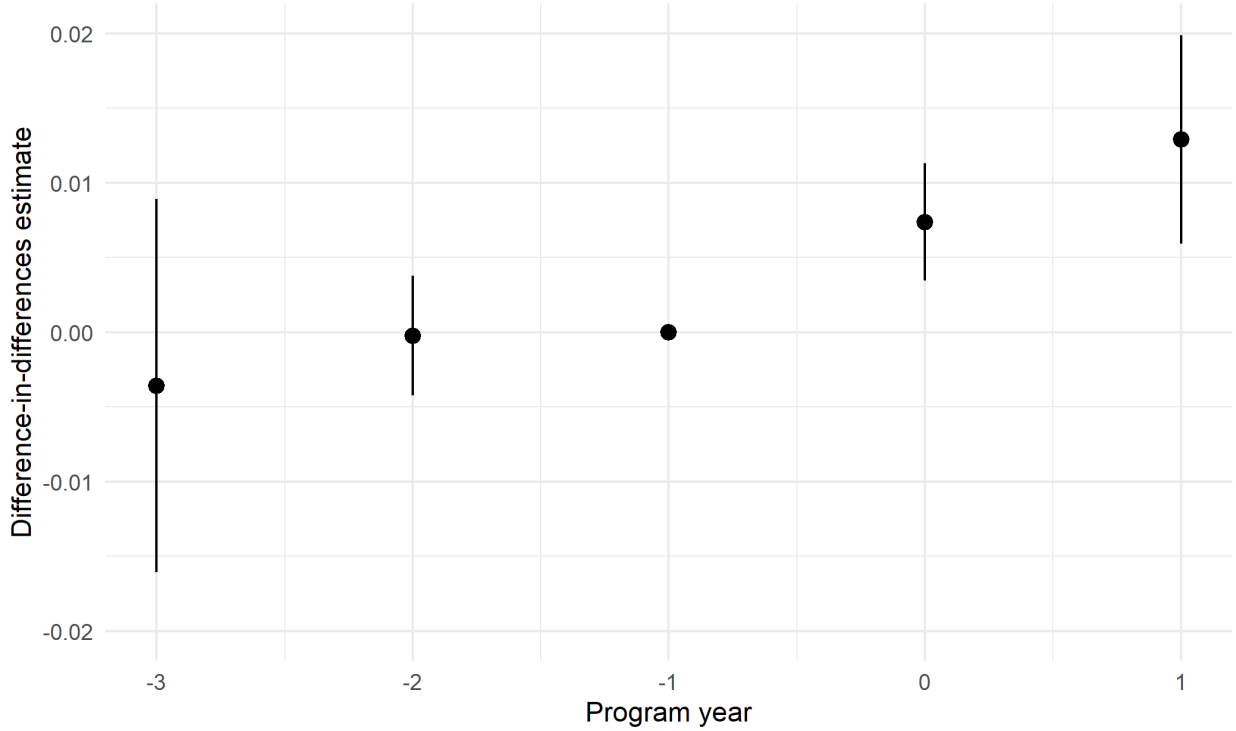
charging stations in treated CBGs within counties that span IOU boundaries. I find that the subsidies increased the number of charging stations in a CBG by .01. Against the sample average of 0.053, this represents a 20% increase. Columns 2 and 3 decompose this effect into the effect on MUD and workplace charging stations. In column 2, the effect on MUD charging stations is not statistically significant at the 10% level, while the effect on workplace charging stations is approximately equal to that in column 1. The effect of treatment on MUD charging stations is imprecisely estimated, and we cannot rule out treatment effects as large as 37.5% in 95% two-sided confidence interval. Column 4 examines the effect of the subsides on *public* charging stations, which serves as a check on the exclusion restriction in consideration of the treatment as an instrument for charging station deployment. The coefficient on the treatment variable is not statistically significant at the 10% level.

Table 2.2: First Stage Impact of Subsidy on Charging Stations: Comparison across Charging Station Type

	Dependent variable:			
	N. MUD/workplace charging stations	N. MUD charging stations	N. workplace charging stations	N. public charging stations
	(1)	(2)	(3)	(4)
Subsidy treatment (binary)	0.010***	0.001	0.009***	−0.006
	(0.003)	(0.001)	(0.002)	(0.006)
CBG FE	X	X	X	X
County-year FE	X	X	X	X
Mean of dep.var	0.053	0.008	0.046	0.155
F-stat	12.1	0.22	14.3	1.12

Note: *p<0.1; **p<0.05; ***p<0.01. N = 92,140. An observation is a block group-year combination. Standard errors clustered at the block group level.

Figure 2.4: Event Study Plot for Impact of Subsidy on Charging Stations



Note: Dependent variable is number of MUD and workplace charging stations. 95% point-wise confidence intervals are for standard errors clustered at the CBG-level. The baseline period is one year before the subsidies were introduced in each IOU, and program year is labeled relative to the first year of the subsidy program, defined as 0.

In order to estimate a causal effect, I require identical trends in subsidized and unsubsidized CBGs within a county. Figure 2.4 shows estimates and 95% confidence intervals from an event-study specification, which serves as a visual check on trends in pre-treatment periods. The dependent variable is number of MUD and workplace charging stations. The baseline period is normalized to one year before the subsidies were introduced in each IOU, and program year is labeled relative to the first year of the subsidy program, defined as 0. There are no pre-trends, as annual coefficients are roughly zero prior to implementation of the subsidies. The effect of subsidies increases over time and is largest in the final period.

To further assess my claim of having estimated the causal effect of subsidies on charging stations, I report estimates from alternative specifications in 2.3. Column 1 presents estimates from a Poisson regression, rather than OLS, which is statistically significant at the 5% level and represents a 14.3% increase in charging stations. Column 2 presents an estimate with the number of total number of chargers across charging stations as dependent variable.[7] The treatment effect of 0.071 represents a 26% increase. That this is larger than the 20% in-

[7]In this paper I refer to a plug on a charging station as a charger, though it is sometimes referred to as

crease in our baseline specification is consistent with the policy, since only charging stations with at least 10 connections were eligible for subsidies in many areas, which is higher than the sample average number of connections per charging station of 4.42. Column 2 presents estimates with common time effects, which, relative to the baseline specification with county-year effects, allows variation between counties over time to estimate the treatment effect. While this estimate remains significant at the 1% level, the assumption of identical trends is less palatable than in our baseline specification. It's plausible that trends in CBGs within a county are more likely to be identical than in arbitrary California CBGs. Columns 3 and 4 employ alternative clustering schemes as a check on statistical inference.[8] Column 3 presents estimates with two-way clustered errors at the CBG-year level, and column 4 presents estimates with two-way clustered errors at the county-year level. Estimates of the treatment effect remain significant at the 5% level in both cases.

a connection in the industry. Charging stations often have multiple chargers.

[8]Standard errors in my baseline model are clustered at the CBG-year level.

Table 2.3: First Stage Impact of Subsidy on Charging Stations: Alternative Specifications

	Poisson	Chargers	Across-county variation	Two-way clustered errors	Coarser two-way clustered errors
	(1)	(2)	(3)	(4)	(5)
Subsidy treatment	0.134**	0.071***	0.005***	0.010**	0.010**
	(0.056)	(0.024)	(0.001)	(0.002)	(0.002)
CBG FE	X	X	X	X	X
County-year FE	X	X		X	X
Year FE			X		
Mean of dep.var		0.272	0.053	0.053	0.053

Note: *p<0.1; **p<0.05; ***p<0.01. N = 92,140. Unless otherwise stated, the dependent variable is number of MUD and workplace charging stations. (Column 2 uses the total number of chargers across charging stations as the dependent variable.) An observation is a block group-year combination. Unless otherwise stated, standard errors clustered at the block group level.

Table 2.4 presents estimates broken out by alternative data samples. Columns 1 and 2 split the data into DAC and non-DAC census tracts. Recall that while our coefficient of interest is on a binary variable indicating whether a CBG received any subsidies, charging stations in DACs received a subsidy amount twice that of non-DAC charging stations. In spite of this, the effect for DACs is an increase of 0.007 charging stations per CBG, while the effect for non-DACs is 0.011 charging stations per CBG. The percentage effects are similar, at 17% and 19% respectively. In spite of a larger subsidy amount, both the absolute, and percentage effect of the subsidy on charging station investment is smaller in DACs. Columns 4 through 6 split the sample according to income quartile. CBGs were assigned an income quartile based on median household income from ACS 2014-2018 5-year estimates. The estimates are monotonically increasing in income quartile. Estimates in columns 3 and 4, for the lowest two income quartiles are not statistically significant at the 5% level, but represent effect sizes of 3% and 10% respectively. Columns 5 and 6, for the largest two income quartiles, represent effect sizes of 25% and 22%, respectively. Overall, effect size is more strongly correlated with income quartile than DAC-status. DAC-status is assigned at the census tract level and is based on a CalEnviroScreen score that takes into account 20 health, environmental, and socioeconomic variables. It is therefore a holistic measure of pollution-burden and pollution-vulnerability that is strongly but imperfectly correlated with income: there are 800 of the 23,200 CBGs in California fall both in a DAC and also in the top two income quartiles.

Table 2.4: First Stage Impact of Subsidy on Charging Stations: Alternative Samples

	DAC	Non-DAC	1st Income Quartile	2nd Income Quartile	3rd Income Quartile	4th Income Quartile
	(1)	(2)	(3)	(4)	(5)	(6)
Subsidy treatment	0.007*	0.011***	0.001	0.004	0.013**	0.019**
	(0.004)	(0.004)	(0.003)	(0.004)	(0.006)	(0.008)
CBG FE	X	X	X	X	X	X
County-year FE	X	X	X	X	X	X
Mean of dep.var	0.041	0.058	0.03	0.037	0.053	0.085
F-stat	3.31	9.22	0.12	0.79	5.21	6.53

Note: *p<0.1; **p<0.05; ***p<0.01. N = 92,140. Dependent variable is number of MUD and workplace charging stations. An observation is a block group-year combination. Standard errors clustered at the block group level.

EV Demand

Table 2.5 reports results for the effect of an additional MUD or workplace charging station on EV registrations. The dependent variable in all regressions is the inverse hyperbolic sine of the number of registered EVs at the CBG-year level.

Columns 1 and 2 show OLS estimates that include separate independent variables for MUD and workplace charging stations. These estimates are not causal, but serve as descriptive evidence for how changes in MUD and workplace charging stations affect EV registrations. Column 2 includes a control variable for the number of public charging stations. The effect of workplace charging stations is not statistically significant. This is unsurprising – most employees do not live in the same CBG as their workplace, so one would not expect to observe an impact of variation in workplace charging stations in a CBG on EV adoption in that same CBG. A more appropriate model would employ data on commuter flows to assign a weight to new workplace charging stations in adjacent CBGs. With CBG-level commuter flows, a workplace charging station in a CBG in downtown Berkeley, for example, would be encoded as an increase the availability of workplace charging stations in all CBGs according to the proportion of commuters to downtown Berkeley who reside in each CBG. I hope to refine this model by obtaining CBG-level commuter flows. The effect of an additional MUD charging station is statistically significant at the 5% level and corresponds to a 7% increase in the number of EVs registered. The average number of EV registrations in a CBG in the sample period is 5.5.

Column 3 includes OLS estimates for an independent variable for pooled count of MUD and workplace charging stations. This is the appropriate independent variable to use with a single instrument. Although MUD and workplace charging stations receive different subsidy amounts, the timing is identical, as is the geographic availability of subsidies. The result is that separate MUD and workplace subsidy variables are too correlated to separately identify the effects of MUD and workplace charging stations. The effect of an additional MUD/workplace charging station in my OLS specification is statistically significant and corresponds to a 3% increase in EVs registered in a CBG. The IV estimates in columns 4 and 5 do not yield statistically significant results. The first-stage F-statistic for each is 12.1 and 18.7. A problem with my empirical strategy as it stands is that the first-stage variation is in workplace charging stations rather than MUD charging stations. Recall from 2.2 that the subsidies had a large statistically significant effect on workplace charging stations, but that the effect on MUD charging stations was not statistically significant. Variation in workplace charging stations is too crudely modeled in my second stage to have a detectable effect on EV adoption. Even if workplace charging stations were randomly assigned to CBGs, one would not expect to observe an effect on EV registrations in the same CBG as the workplace charging station, because employees tend to commute from other CBGs where they reside. Researchers who examine this question in the future could refine my model with commuter flows data in order to leverage first-stage variation in workplace charging stations.

Table 2.5: Instrumental Variables Results for Effect of Charging Stations on EV Registrations

	OLS	OLS	OLS	IV	IV
	(1)	(2)	(3)	(4)	(5)
N. workplace charging stns.	0.045***	0.023			
	(0.015)	(0.016)			
N. MUD charging stns.	0.099***	0.072**			
	(0.035)	(0.036)			
N. MUD/workplace charging stns.			0.030**	0.379	0.337
			(0.015)	(1.131)	(0.942)
N public charging stns.		0.032***	0.032***		−0.067
		(0.010)	(0.010)		(0.303)
CBG FE	X	X	X	X	X
County-year FE	X	X	X	X	X
First stage F-stat				12.1	18.7

Note: *p<0.1; **p<0.05; ***p<0.01. N = 92,140. Dependent variable is inverse hyperbolic sine of EV registrations. An observation is a block group-year combination. Standard errors clustered at the block group level.

2.5 Conclusion

In this paper, I examine whether the availability of home- and workplace-charging infrastructure for renters has a detectable effect on EV adoption. I address two questions: (i) are such policies that aim increase the availability of MUD and workplace charging stations effective in increasing investment, and (ii) whether such policies are effective in promoting EV adoption. I find that the subsidies increased the number of charging stations in a census block group (CBG) by .01. Against the sample average of 0.053, this represents a 20% increase. I do not detect any effect of charging stations on EV adoption. My coefficient of interest is 0.379, indicating that an additional MUD or workplace charging station in a census block group (CBG) yields a 38% increase in registrations. However, the standard error is 1.131, so I cannot rule out extremely large or even negative effects.

Future research that examines the subsidies I study in this paper would benefit from the inclusion of additional years of data. My estimation sample spans 2015 through 2018, and the subsidies were available from 2017 to 2019 and likely have a lagged effect. The

inclusion of additional years of data, not yet available at the time of writing, would allow one to estimate the effect at the program's full deployment. The instrumental variables estimates would benefit from more refined modeling of the second stage effect of workplace charging stations on EV registrations. As it stands, the first-stage workplace effect is strong and precisely estimated. However, the inclusion of a workplace charging station in a driver's home CBG would be unlikely lead that driver to purchase an electric vehicle, as most drivers to not work in the CBG where they reside. The effect of workplace charging stations could be better modeled through the use of commuter flows data at the sub-county level using cell phone location data.

Chapter 3

Bike Sharing and Traffic Congestion: A Case Study of Citibike in New York City

3.1 Introduction

In this paper, I estimate the causal effect of the Citibike bike-sharing system on historical street-level traffic speeds in Manhattan. Bike sharing is an increasingly popular shared mobility transportation service. These services provide shared use of a vehicle, bicycle, scooter, or other mode of transportation. Shared mobility services can simultaneously complement and conflict with existing transportation infrastructure. An important way in which these services impact existing infrastructure is reallocation of finite road space to to shared vehicles and their associated infrastructure. Services that provide bicycles and scooters are often seen as purely complementary to existing transit systems. That view is at odds with the fact that road space is finite and increasingly congested in many cities. Any additional demands placed on roads should be carefully evaluated. In particular, do shared mobility services impose external costs on users of existing transportation infrastructure through increased congestion? The net effect of these services on road congestion will depend on whether the congestion mitigating effect due to substitution away from other transit that uses roads (vehicles) exceeds the congestion exacerbating effects of adding bicycles and their infrastructure (docks/storage) to the roadway.

The major contribution of this paper is to provide an estimate of the effect of a bike-sharing system on congestion in the *entire* area over which it was deployed, rather than in the vicinity of particular docks. In addition, this is the first study chart the routes between bike-sharing system docks using Google Maps bicycling directions. In doing so I obtain estimates of treatment intensity at fine spatial resolution. Another contribution is that I am the first to control for bike lanes, and other changes in street conditions over time, that could be correlated with the introduction of bike-sharing, and which could confound other

estimates. My main finding is that the Citibike system increased travel time on avenues in Manhattan. These estimates comprise combined impact of various effects of the Citibike system — Citibike users may substitute away from personal vehicles and taxis, but users also impose external cost on drivers through increasing congestion on roads. My estimates therefore represent a lower bound on the congestion externality imposed by Citibike.

My analysis uses novel data on traffic speeds at a 10-meter spatial resolution, which I estimate from 1.2 billion publicly available records of historical taxi trips. I marry this with detailed data on the Citibike system itself. I observe the locations of all docks, and records of every ride, including the start and end dock, and start and end times. To address where cyclists travel, I chart the routes between Citibike docks using Google Maps bicycling directions. Additionally, I collect and analyze other sources of urban data — in particular, I exploit data on bike lanes, and use 311 service requests related to road conditions to account for other changes in street conditions over time.

My research design exploits the sudden and localized nature of Citibike deployment as a natural experiment: Citibike launched on May 27th 2013, but only in lower Manhattan and downtown Brooklyn. This enables a comparison of street speeds above and below 59th Street, pre- and post-Citibike in a difference-in-differences specification. I find the average causal effect of the Citibike system on travel time below 59th Street is an increase of 2.3%. In a second model, I employ data on Citibike rides to measure temporal variation in treatment intensity. This variation is driven by lower ridership as the program scaled up, as well as seasonal effects — ridership is higher in summer and lower in the winter. From a panel data difference-in-differences specification that incorporates treatment intensity at the monthly level, I find that at the maximum system utilization of 987,169 monthly rides, I find travel time increased by 3.9%.

My preferred model employs a treatment intensity measure that varies temporally, at the month level, as well as spatially, at the street-level. Specifically, I divide streets into 10-meter bins. Then bike rides that crossed a 10-meter bin were measured by first mapping routes between Citibike stations using Google Maps cycling directions, then counting the rides on routes that crossed that bin. This model captures significant heterogeneity in treatment intensity both temporally as well as spatially. I find that at the maximum number of rides traversing a bin 452,056, travel time increases by 9.6%. An equivalent interpretation is that a 10% increase in utilization, of about 45,000 rides, increases travel time by 1%.

An important parameter of interest is how Citibike affected the speed of the average trip in New York City. My main results demonstrate that Citibike has slowed down trips made by car, taxi, bus. If Citibike users are substituting away from walking, then the introduction of Citibike may have sped up the average trip made by New Yorkers, in spite of the externality imposed on vehicular traffic. I perform a back-of-the-envelope calculation to evaluate whether substitution patterns that would facilitate an increase in overall trip speed are plausible. Under generous assumptions about substitution patterns (namely that all Citibike users would have walked), observed changes in the share of trips made by bike in Manhattan before and after Citibike are insufficient to have resulted in an increase in the speed of the average trip.

This study contributes to a literature examining bike-sharing systems and congestion. The costs imposed by traffic congestion are large and numerous in variety (Anderson et al., 2016; Barth and Boriboonsomsin, 2008; Currie and Walker, 2011; C. R. Knittel, Miller, and Sanders, 2016). An explicit goal of many bike-sharing systems, including Citibike, is to reduce congestion. Hamilton and Wichman, 2018 is the closest study to this one. The authors study DC's Capital Bikeshare. They match areas with docks to areas without docks using pre-treatment data. They use a panel data difference-in-differences specification and find that areas with bike-share stations (docks) experienced a reduction in congestion of 4%. Importantly, they also find that that neighboring areas experienced a larger increase in congestion, which they attribute to spillovers from drivers avoiding areas with more cyclists. I contribute by estimating the overall effect on congestion in the area in which Citibike was deployed. Furthermore, I control for changes in road use over time, such as bike lanes. Wang and Zhou, 2017 also study the effect of bike-sharing systems on congestion. They use city-level data and a difference-in-differences design. They find insufficient evidence to conclude that cities that introduced bike-sharing systems experienced a reduction in congestion as a result. Overall, prior work has not found conclusive evidence of whether bike-sharing systems increase or decrease congestion.

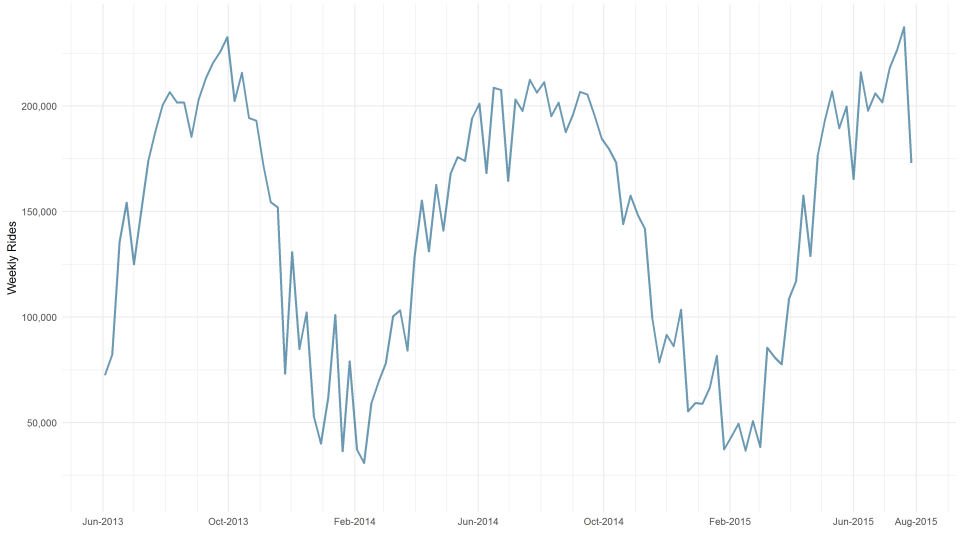
Another important literature to which I contribute is a more general literature on the determinants of road congestion. In a seminal study, Duranton and Turner, 2011 investigate the effect of lane kilometers of roads on vehicle-kilometers traveled in US cities. They find that increased provision of roads and public transit is unlikely to relieve congestion. Couture, Duranton, and Turner, 2018 also investigate the determinants of driving speed in large US cities using a structural model of demand and supply for travel. They find that centralized cities are slower, and that cities with ring roads are faster, and that a congestion tax of about 3.5 cents per kilometer yields welfare gains of about $30 billion per year. Other studies find large effects of the provision of public transit on congestion and other outcomes using strikes as a source of exogenous variation. (Anderson, 2014; Bauernschuster, Hener, and Rainer, 2017)

3.2 Background: Citibike Bike-Sharing System

The Citibike bike-sharing system makes bicycles available to the public at docks, which release the bicycle upon payment or verification of a user's subscription. Citibike's history began in September 2011 when the New York City Department of Transportation granted New York City Bike Share, a privately held company, rights to operate city's new bike-sharing system. After a public input process, the system launched on May 27, 2013 in lower Manhattan and downtown Brooklyn. Figure 3.1 shows weekly Citibike rides in Manhattan from June 2013 to July 2015. There were about 75,000 rides in the first week of June. Two months later, in early August, there were over 200,000 rides. Weekly ridership is highest in the summer at over 200,000 rides, and is lowest in the winter at fewer than 50,000 rides. In October 2014, Alta Bicycle Share took over ownership of New York City Bike Share and

announced plans to expand Citibike. The system expanded in August 2015, introducing new stations in upper Manhattan and Brooklyn. While some systems, such as DC's Capital Bikeshare, receive public funding, Citibike is privately funded.

Figure 3.1: Total Weekly Citibike Rides in Manhattan, June 2013 to July 2015



Note: Figure shows total weekly Citibike rides in Manhattan from June 2013 to July 2015. There were about 75,000 rides in the first week of June. Two months later, in early August, there were over 200,000 rides. Weekly ridership is highest in the summer at over 200,000 rides, and is lowest in the winter at fewer than 50,000 rides.

Users can subscribe to the service for a fee of $169 annually, or $14.95 per month. A subscription service grants the user unlimited rides. Non-subscribers can pay $3 for a single ride, purchase a day pass for unlimited rides in a 24 hour period, or a 3-day pass for unlimited rides in a 72 hour period. The system is designed to facilitate short, one-way trips rather than recreational trips. If subscribers keep a bicycle for more than 45 minutes at a time, they are charged $2.50 per additional 15 minutes. Non-subscribers are limited to rides of 30 minutes, and charged $4 per additional 15 minutes.

Rides by annual subscribers account for 88.5% of of all rides. Furthermore, most rides are made during rush hour, at times of peak congestion. Figure 3.2 shows the distribution of time of day for rides in Manhattan during the sample period, June 2013 to August 2015. The overall distribution, as well as those for subscribers and non-subscribers, are plotted separately. Because most rides are made by subscribers, the overall distribution is similar to that of subscribers. It is bi-modal, with most rides occurring around 9 AM and 6 PM. For non-subscribers, most rides are during the afternoon. Figure 3.3 shows the distribution of duration of rides in minutes in Manhattan during the sample period. The vast majority

of rides are under 20 minutes, and almost all are under 45 minutes. This is what one would expect given the surcharge for rides longer than 45 minutes.

Figure 3.2: Kernel Density Plot of Time of Day of Citibike Rides in Manhattan, June 2013 to July 2015

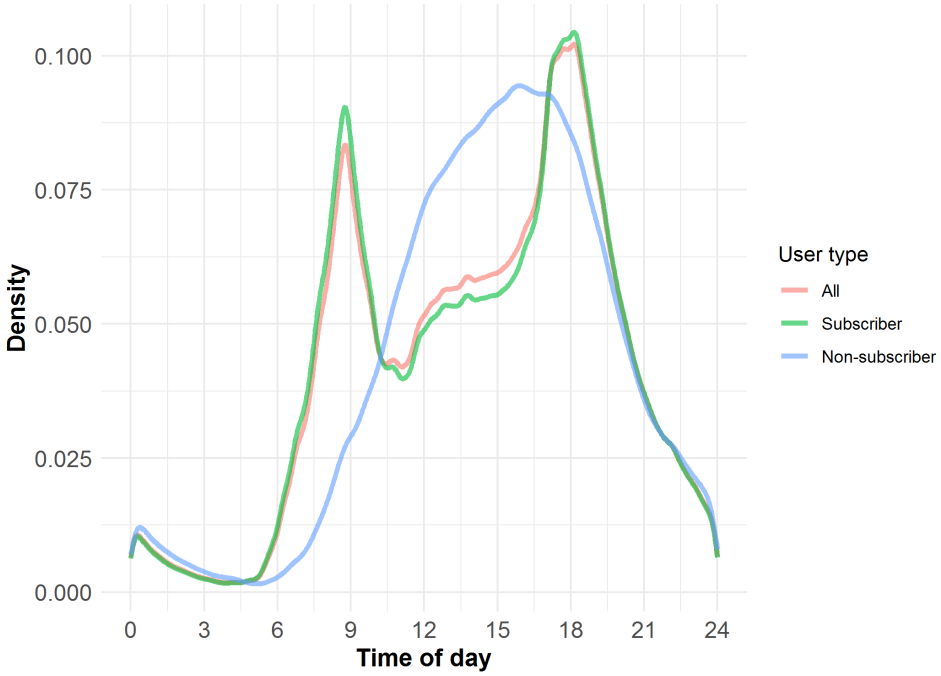
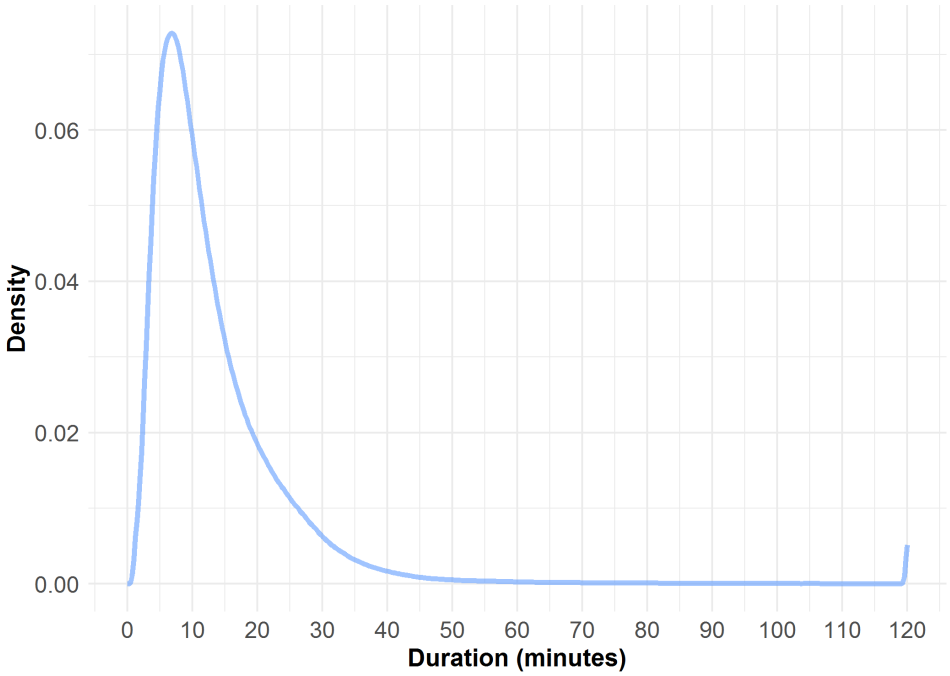


Table 3.1 shows the size of Citibike bike-sharing system compared to other bike-sharing systems in the U.S.. Citibike usage is shown for 2014, which is the last complete calendar year in the sample period. Usage for other systems is shown for 2018, the most recent complete calendar year. While Citibike is about twice as large as the Chicago and Washington D.C. systems by annual and monthly maximum rides. The number of monthly rides on the most popular ride in New York City, 11,860, is only 34% higher than the number of monthly rides on the most popular ride in Chicago, which indicates that bike-sharing penetration in Chicago is comparable to that in New York City during the sample period.

3.3 Data

The primary data for this analysis comes from yellow taxi trip records collected by the New York City Taxi and Limousine Commission (TLC). I also use bicycle trip records and bicycle dock locations collected by Citibike, the bike-sharing system serving New York City and Jersey City. Additional datasets employed include records of bike lane locations and installation dates, as well as 311 service requests pertaining to road conditions.

Figure 3.3: Kernel Density Plot of Duration of Citibike Rides in Manhattan, June 2013 to July 2015



Note: Durations above 120 minutes were truncated. The median trip duration is 10.28, the mean trip duration is 14.18.

Table 3.1: Bike Share Usage Statistics by City

City	Year	Total (annual rides)	Maximum (monthly rides)	Most popular trip (annual rides)
New York	2014	8,081,216	968,842	11,860
Chicago	2018	3,603,082	544,703	8,858
Washington, D.C.	2018	3,542,684	404,761	6,339
Boston	2018	1,767,806	242,916	5,561

Note: Table shows bike share usage for the four largest systems in the U.S.. Citibike usage is shown for 2014, which is the last complete calendar year in our sample period. Usage for other systems is shown for 2018, the most recent complete calendar year. The most popular trip is defined as the start and end dock pair between which the most rides occurred in a year.

Citibike System Data

I downloaded administrative data on historical Citibike trips. I use these data to determine areas of Manhattan that were treated, and measure treatment intensity. These data are

publicly available from the Citibike website, and I draw upon records from June 2013 to July 2015. The Citibike system launched on May 27, and I observe rides from June 1 2013 onward. July 2015 was the last month before its expansion in August 2015. The data include all historical Citibike rides with fields capturing the start and end date/time of the ride, the start and end docks, whether a rider was a subscriber or not, and the age and gender of subscribing riders. I match these data with GPS coordinates of docks obtained from Citibike's API.

I exclude a small number of records for which the ride started or ended at a Citibike depot, presumably occurring when a bicycle was serviced or delivered. I also limited to rides that started and ended in Manhattan.

Google Maps Data

I exploit Google Maps Platform's Directions API [1] to map cycling directions routes between Citibike docks. In total, I map directions between 77,000 possible combinations of the 278 docks. When combined with the Citibike system data, this allows us to count the rides traversing each road segment, assuming riders followed cycling directions from Google Maps. The routing algorithm for Google Maps is not publicly available, but anecdotal evidence suggests it gives preference to bicycling infrastructure, and is less sensitive to congestion than driving directions. That the vast majority of Citibike trips are short, less than 20 minutes in duration, suggests riders take a direct route and supports this approach.

Taxi and Limousine Commission Trip Record Data

I use administrative data on yellow taxi trips collected by the New York City Taxi and Limousine Commission. In total, these amount to 1.2 billion taxi trip records. I use these data to construct street-level historical speed. The trip records data are publicly available, and I draw upon records from January 2009 to July 2015. 2009 is the earliest date available: GPS-enabled taxi meters were rolled out in 2009 under the "Taxicab Passenger Enhancement Program." The data include all historical yellow taxi trips with fields capturing pick-up and drop-off dates/times, pick-up and drop-off GPS coordinates, trip distances, itemized fares, rate types, payment types, and driver-reported passenger counts.

I exclude a small number of records for any of the following reasons: missing values for any of pick-up and drop-off dates/times, pick-up and drop-off GPS coordinates, or trip distances; negative or zero distance value; trip duration of fewer than 10 seconds; GPS coordinates outside of New York City metropolitan area; average speed over 10% above expressway speed limit (i.e. above 77 mph); and duration over two hours. I also remove records for which the greatest circle distance is greater than ten times the recorded trip distance, since this is only possible from some kind of error in the GPS or odometer. Finally I remove trip records for which the greatest circle distance is 10 times less than recorded

[1] *developers.google.com/maps/documentation/directions/start*

trip distance, since these are either trips with circuitous routes or errors in GPS or odometer readings.

Bike Infrastructure Data

I use data on locations and installation dates of bike lanes and routes throughout New York City. The data are provided by the New York City Department of Transportation and made publicly available through the NYC Open Data Portal.[2] The dataset is at the level of a stretch of infrastructure, and variables include, start and end coordinates, installation date, and type of infrastructure. Possible infrastructure types are bike-friendly parking, curbside bike lane, protected bike lanes, standard bike lanes, sharrows[3], and signed bike routes. I restrict to bike lanes and designate curbside and standard as unprotected bike lanes.

311 Service Requests

I use 311 service request data provided by the New York City Department of Information Technology and Telecommunications and made publicly available through the NYC Open Data Portal.[4] 311 is a non-emergency phone number that people can call to find information, make complaints, or report problems like graffiti or poor street conditions. The data are at the service request level and variables include complaint type, complaint descriptor, creation date, and coordinates of location. I limit to the following types of complaints: blocked driveway — no access, blocked driveway — partial access, street condition — blocked, construction, street condition — cave-in, street condition — failed street repair, street condition — pothole, street condition — rough, pitted or cracked roads. I aggregate complaints to the 10-meter bin, month level to provide a count of the number of complaints that occurred at a 10-meter bin in a given month.

Estimation of Historical Street Speeds

To construct street-level historical speed data, I employ the method developed by Mangrum and Molnar, 2019. They use three alternative methodologies to construct measures of historical street speeds from historical records of taxi trips from the Taxi and Limousine Commission. I employ their simplest and most preferred methodology. This averages the rate of travel across all taxi trips for which one is confident that the taxi traversed a certain interval over a certain time period.

I first subset to trips that were entirely contained within a rectangular strip running along one of the north-to-south avenues of Manhattan. I use a rectangular strip to be able to capture trips that started just off an avenue on a side-street, and to accomodate any error

[2] *https://data.cityofnewyork.us/Transportation/Bicycle-Routes/7vsa-caz7*

[3] A sharrow is a street marking placed in the travel lane to indicate where people should preferably cycle

[4] *nycopendata.socrata.com/Social-Services/311-Service-Requests-from-2010-to-Present/erm2-nwe9*; *data.cityofnewyork.us/Social-Services/311-Service-Requests-for-2009/3rfa-3xsf* .

in GPS coordinates of a trip originating on an avenue. I then subset to trips whose direction of travel corresponds to the direction of travel for the avenue. Trips are classified by their avenue, direction combination, since some avenues, like Park Avenue, are bi-directional. I then calculate, for each trip, the ratio of travel time to the distance traveled, to obtain each trip's average seconds per meter. Each avenue is then subdivided into 10-meter bins, and for each, average the travel time per meter of all taxis that traversed that bin.

Table 3.2 shows summary statistics for within-avenue taxi trips, for each avenue as well as for all trips. The mean seconds per meter of all of all within-avenue taxi trips is 0.201, corresponds to a speed of 11.13 miles per hour. I do not measure speeds for 6th and 7th avenue since these stop at Central Park, which is also where the Citibike docks stop, and they therefore do not provide useful variation in my analysis.

Table 3.2: Seconds-per-meter for within-avenue taxi trips, January 2009 to August 2015

Street	N	Mean	Median	Std Dev	Min	Max
1st Ave	10,733,906	0.180	0.142	0.355	0.029	342.998
2nd Ave	14,715,273	0.219	0.176	0.367	0.029	432.475
3rd Ave	15,469,388	0.211	0.167	0.319	0.029	279.618
Lexington Ave	6,941,103	0.258	0.207	0.471	0.029	406.378
Park Ave	5,297,332	0.250	0.209	0.359	0.029	275.889
Park Ave	4,854,437	0.290	0.242	0.412	0.029	298.259
Madison Ave	7,354,580	0.245	0.202	0.437	0.029	309.444
5th Ave	10,806,262	0.291	0.236	0.394	0.029	190.575
8th Ave	13,636,670	0.281	0.217	0.574	0.029	439.932
9th Ave	10,820,429	0.242	0.194	0.408	0.029	283.346
10th Ave	3,706,137	0.206	0.157	0.501	0.029	417.562
All	104,335,517	0.241	0.191	0.420	0.029	439.932

Notes: Measured variable is seconds per meter. The mean of all within-avenue taxi trips, .201 seconds per meter, is the inverse of a speed of 11.13 miles per hour.

These data are subject to two potential types of bias that Mangrum and Molnar, 2019 identify: an aggregation bias, and a selection bias due to changes in trip selection over space and time. I replicate their discuss of these here. Consider two 10-meter bins in a relatively uncongested area in the north of the city, and assume traffic moves at the same speed along both of the bins, in the southern direction. Suppose the northernmost of these bins contains a hotel driveway, whereas its neighbor to the south does not. The bin containing the hotel originates many long trips to the relatively congested downtown area, whereas the neighboring bin originates short trips within a relatively uncongested interval. In this example, the average taken over travel time of the trips crossing each bin will be biased upwards as an estimator of the true marginal travel time of crossing each bin due to congestion that affects taxis traveling downtown from the hotel (an aggregation bias). The

bias will be larger for the bin containing the hotel since its average does not contain the shorter, local trips originating in the southern neighbor (a selection bias). Aggregation bias can vary with time as a bin may be located next to a nightclub, for example.

Mangrum and Molnar, 2019 develop alternative measures of travel time to address the issues they identify. They find their results robust across methodologies, hence I also adopt their simplest and most preferred method in this paper.

3.4 Empirical Strategy

My empirical strategy exploits the localized and sudden deployment of the Citibike system in a difference-in-differences design. Citibike was deployed on May 27, 2013 in lower Manhattan. While some docks were installed as early as April 2013, the system was not activated until the end of May, at which point all bicycles were made available for use. The boundary of the system was at 59th Street, above which there were no docks. The system expanded above 59th Street in August 2015, so I restrict the estimation sample to end in July 2015. Furthermore, I restrict the estimation sample to within 3 kilometers of 59th St. Estimates of speed in bins outside of that range become uncertain due to fewer taxi trips.

I estimate the standard panel data difference-in-differences specification in this setting:

$$s_{abt} = \beta_{DID} Post_t \cdot below\ 59th_{ab} + \beta x_{abt} + \gamma_{ab} + \delta_{at} + \varepsilon_{abt} \tag{3.1}$$

where s_{abt} is the natural log of travel time in seconds per meter at the level of an avenue-direction (a), 10-meter bin (b) and month(m) from January 2009 to July 2015. The preferred metric for measuring speed over a roadway segment in transportation engineering is space-mean speed. Space-mean speed is the harmonic mean of speeds. The outcome of interest is measured in travel time, rather than speed, so that its arithmetic mean yields the reciprocal of space-mean speed for a 10-meter bin. Positive estimates from these regressions can therefore be interpreted as increases in travel time, or equivalently, decreases in speed. More details are given in Appendix A. $Post_t$ is a binary variable encoded as one after June 2013, and zero otherwise.[5] x_{abt} includes controls for complaints related to road conditions and other factors that may affect travel time, locations of bike lanes and routes, and calendar month dummies interacted with a below 59th Street indicator, to allow for different seasonal effects above and below 59th Street. However, the above specification does not capture the initial ramp-up in Citibike usage, or heterogeneity in usage across bins. I therefore estimate two specifications which employ data on Citibike rides to measure treatment intensity.

The first such specification incorporates temporal variation in treatment intensity. This variation is driven by lower ridership as the program ramped up, as well as seasonal effects — ridership is higher in summer and lower in the winter. These effects are shown in figure 3.4.

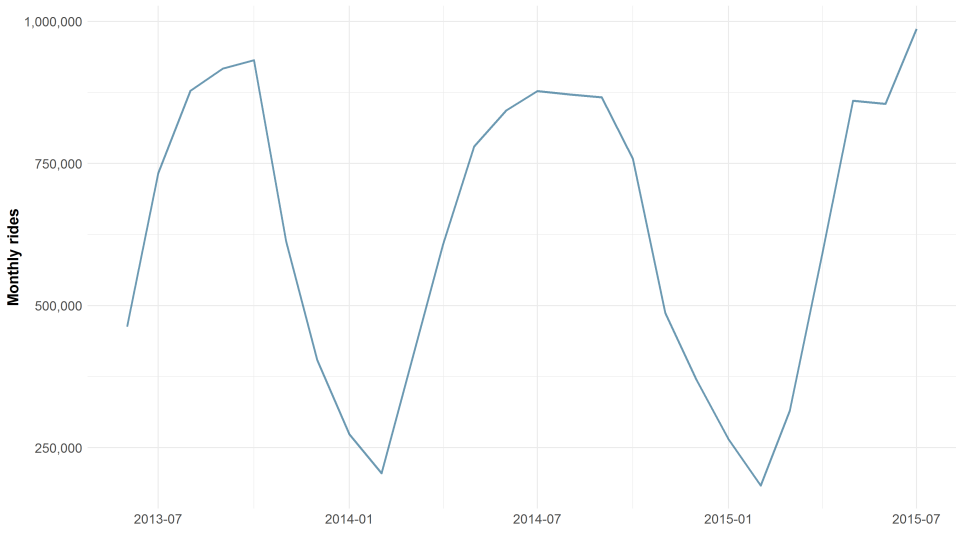
[5] I categorize May 2013 as untreated, even though Citibike launched May 27. These 5 days in May are a small enough fraction of the month, and initial ridership was sufficiently low, that this does not affect results.

Rides in June and July of 2013, soon after the system's launch, are substantially lower than in the same months in subsequent years. Furthermore, Citibike is used much less frequently in winter. In equation 3.2, the binary treatment dummy is replaced with $Monthly\ Intensity_t$, which measures intensity at the monthly level. It is defined as the ratio of monthly total rides (zero prior to June 2013) to maximum observed value of 987,169. The specification is as follows:

$$s_{abt} = \beta_{DID} Monthly\ Intensity_t \cdot below\ 59th_{ab} + \beta x_{abt} + \gamma_{ab} + \delta_{at} + \varepsilon_{abt} \tag{3.2}$$

In each of the above specifications, I employ the same fixed effects as Mangrum and Molnar, 2019. In particular, I include fixed effects γ_{ab} at the level of an avenue-direction and 10-meter bin, which absorb time-invariant features of the road, as well as monthly time effects δ_{at} at the level of an avenue-direction, which absorb the variation in speed along each avenue over time that is common across bins. The inclusion of monthly time effects that vary at the level of an avenue-direction negates the need to include weather controls; since my panel is at the monthly level, any weather variation that I might wish to control for at the monthly level is absorbed. Furthermore, to the extent that I might be concerned about weather variation across space in Manhattan, time effects at the level of an avenue-direction account for this to the extent that the variation occurs across avenues. The difference-in-differences coefficient β_{DID} is estimated from the difference within each avenue-direction between travel time north versus south of 59th Street as treatment intensity varies.

Figure 3.4: Total Monthly Citibike Rides in Manhattan, June 2013 to July 2015



My preferred specification is equation 3.3 below, which employs a treatment intensity measure at the level of a month, avenue-direction and 10-meter bin, $Bin\text{-}Month\ Intensity_{abt}$. It is defined as the ratio of bike rides traversing a bin in a month to the maximum number

of rides observed in any bin in the sample period. This measure captures significant heterogeneity in treatment intensity both temporally, as discussed above, as well as spatially. Bike rides that crossed a bin were measured by first mapping routes between Citibike stations using Google Maps cycling directions, then counting the rides on routes that crossed that bin. Spatial variation in the number of rides is significant. Note that equation 3.3 no longer includes *below 59th*$_{ab}$. Its inclusion would be redundant as the bin-month level intensity measure is always zero above 59th Street as a result of assuming riders did not take bikes above 59th Street. I discuss the plausibility of that assumption and the implications of its violation below.

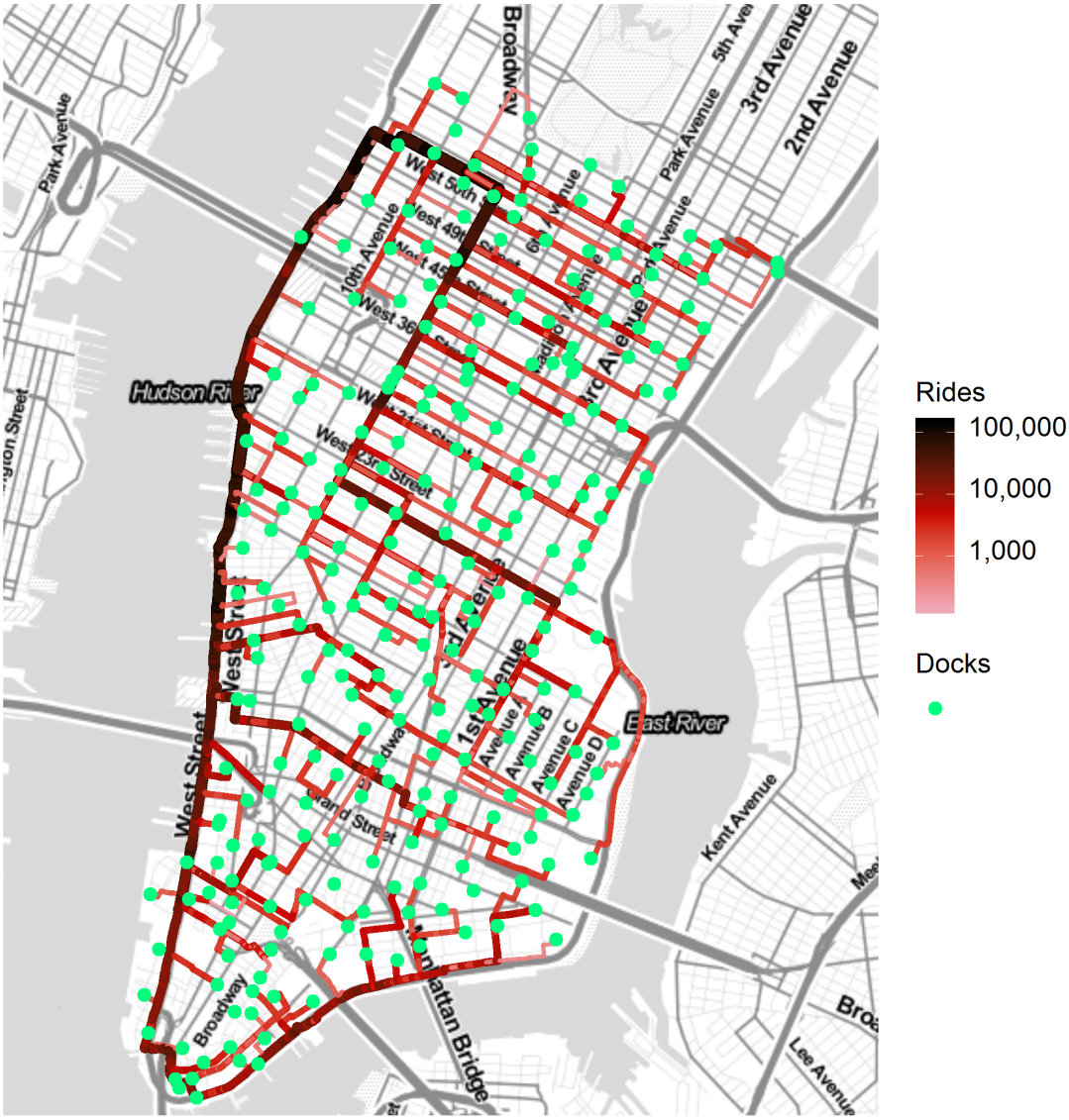
$$s_{abt} = \beta_{DID} \textit{Bin-Month Intensity}_{abt} + \beta x_{abt} + \gamma_{ab} + \delta_t + \varepsilon_{abt} \tag{3.3}$$

The specification in equation 3.3 features common time effects δ_t, which relative to the above specifications, allows variation in travel time across avenues to identify the difference-in-differences estimates. [6] This is a sensible modification, as it allows untreated bins south of 59th Street to serve as "control" areas for treated bins south of 59th Street. This mitigates any potential concern about differential trends across north-south geography as density varies. Figure 3.5 shows the identifying variation for this specification. It plots monthly rides on road segments for June 2014, a year after the system launched. The most traversed segments are shown in black, with over 100,000 rides a month. Segments traversed less often are shown in lighter shades of red. Citibike dock locations are shown as green circles. Observe that Citibike docks stop at 59th street. 8th Avenue and West Side Highway are relatively highly traversed, as are several cross-streets. There is significant variation in the frequency of rides on different road segments. Many bins that fall south of 59th Street are traversed very infrequently.

I identify the causal effect of Citibike on travel time in equations 3.1 and 3.2 if the parallel trends assumption holds - that is if the natural logarithm of travel time in treated versus untreated areas was not trending apart for reasons other than the Citibike system. I provide visual evidence in support of this assumption in figure 3.6, which shows a monthly time series of travel time in log average seconds per meter above versus below 59th Street, for the sample period of January 2009 to July 2015. Each trend has been demeaned and adjusted for seasonality to facilitate comparison. Observe that average travel time below 59th Street and average travel time above 59th street move in unison and tend to be similar and do not diverge in any systematic way prior to the treatment. For my preferred specification in equation 3.3, with treatment intensity that varies across bins, I require parallel trends across bins of different treatment intensity. This is difficult to assess visually, so I probe the robustness of the assumption by including flexible linear trends in a specification check.

[6] Common time effects absorb any weather variation that is common across bins. One might be concerned about weather shocks that are not common across bins. The available evidence suggests that weather patterns are fairly constant across space in Manhattan. Research to document the urban heat island effect (Vant-Hull et al., 2014) shows variation across Manhattan in temperature, but while these temperature differences are of up to 2 degrees Fahrenheit, such differences are not large enough to plausibly affect Citibike ridership. This is also true of variation in precipitation across space in Manhattan.

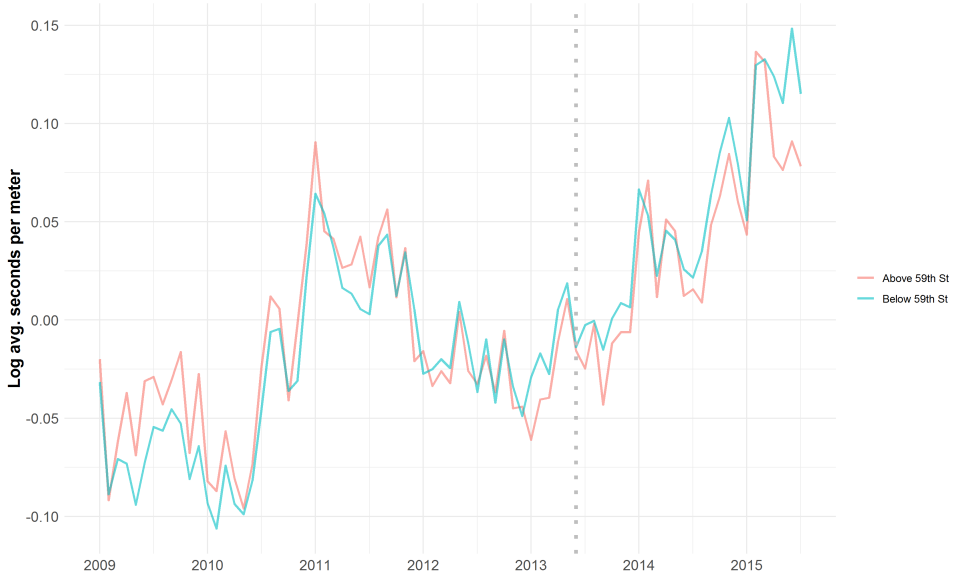
Figure 3.5: Frequency of Citibike Rides on Road Segments in Manhattan, June 2014



Note: Green circles mark Citibike dock locations. Routes assume riders follow Google Maps cycling directions.

Figure 3.6 also shows the slowdown in travel time that has occurred in Manhattan. Starting in 2013, travel time is slowing down both above and below 59th St. A potential concern might be that a third factor, other than Citibike, and correlated with the introduction of Citibike, is causing this slowdown. My experimental design ensures such a factor would not bias the estimates as long as the slowdown is at a similar rate above and below 59th St.

Figure 3.6: Monthly Time Series of Travel Time in Log Average Seconds per Meter Above versus Below 59th Street, January 2009 to July 2015



Note: Figure shows monthly time series of natural logarithm of travel time in average seconds per meter for the average bin below 59th Street and the average bin above 59th Street over time. The dotted grey line marks June 2013 when Citibike was launched. Each trend has been demeaned and adjusted for seasonality to facilitate comparison. The values shown are residuals from a regression of log average seconds per meter on a constant, indicator for below 59th Street, and calendar month dummies interacted with below 59th Street indicator.

Mangrum and Molnar, 2019 were the first to document the slowdown in Manhattan. They find that the introduction of transportation network companies (TNCs) and their subsequent growth accounts for 61.8% of the traffic slowdown in midtown Manhattan that occurs after 2013. I probe the robustness of my estimates to the introduction of TNCs below.

Other threats to identification of the causal effect of Citibike on travel time include changes in road use that affect speeds differently in treated versus untreated areas. Each baseline specification includes several control variables that account for such factors that I am able observe. First, I construct a count measure of complaints related to road conditions and other factors that may affect travel time using 311 service request data. I observe every complaint made to the 311 hot-line for the sample period, including its date, the type of complaint, and its location, where relevant. I aggregate the following types of complaint at the level of a 10-meter bin, month: blocked driveway — no access, blocked driveway — partial access, street condition — blocked, construction, street condition — cave-in, street condition — failed street repair, street condition — pothole, street condition — rough, pitted or cracked roads. I also use data on locations and installation dates of bike lanes and routes

to construct control variables for protected and unprotected bike lanes at the level of a bin, avenue-direction, and month. Additionally, I control for seasonal effects using calendar month dummies interacted with a below 59th Street indicator, to allow for different seasonal effects above and below 59th Street.

3.5 Results

Figures 3.7 and 3.8 show historical street speed for all avenues and 10-meter bins in the estimation sample. For ease of interpretation, speed is presented in miles per hour, which is the reciprocal of the dependent variable, travel time in seconds per meter, with converted units. Each figure plots speed against distance above 59th Street, and plots the average speed for the pre- and post-Citibike periods separately. On most avenues, speed increases in the northern direction, as the city becomes less dense. Over time, speeds tend to decrease both above and below 59th Street, but on many avenues the decrease is larger below 59th Street, which is the variation that identifies the treatment effect.

Table 3.3 shows summary statistics for the estimation sample, which is for the period from January 2009 to July 2015 and at the 10-meter bin, avenue-direction, month level. I include all bins within 3 kilometers of 59th Street on 1st Avenue through 5th Avenue on the East side of Manhattan and on 8th Avenue through 10th Avenue on the West side. I measure travel time in average seconds per meter. The sample average is .21 seconds per meter, which corresponds to a speed of 10.65 miles per hour. The monthly average number of Citibike rides is about 200,000, and the standard deviation of 330,000 is large relative to the mean. The bin-month Citi rides variable is right-skewed. The average number of rides traversing a bin in a month is 1,526, and the standard deviation is 14,245, while the maximum I observe is 452,065.

Main Results

Table 3.4 reports my main results. The dependent variable in all regressions is the natural log of travel time, measured in avgerage seconds per meter. Positive estimates can be interpreted as increases in travel time, or equivalently, decreases in speed. The columns show estimates for three specifications with alternative independent variables of interest. In column 1, the independent variable of interest is *Citi* x *Post*, and is the standard independent variable in a difference-in-differences specification. In column 2 the independent variable *Citi* x *Monthly Intensity* scales the dependent variable in column 1 by treatment intensity at the monthly level, normalized by maximum monthly rides observed of just under a million. In column 3 the independent variable *Citi* x *Bin-Month Intensity* measures treatment intensity at the 10-meter bin, avenue-direction, month level. This variable measures the number of Citibike rides crossing a bin in a particular direction, normalized by the maximum number of rides crossing any bin that I observe in a the sample, 452,065. The specification in column 3

Table 3.3: Regression Sample Summary Statistics

Statistic	Mean	St. Dev.	Min	Max
Seconds per meter	0.210	0.046	0.088	0.866
Monthly Citibike rides	206,733.400	329,626.600	0	987,169
Monthly intensity	0.209	0.334	0	1
Bin-month Citibike rides	1,526.092	14,245.580	0	452,065
Bin-month intensity	0.003	0.032	0	1
Protected bikelane	0.187	0.390	0	1
Unprotected bikelane	0.051	0.219	0	1
311 complaints	0.013	0.157	0	15

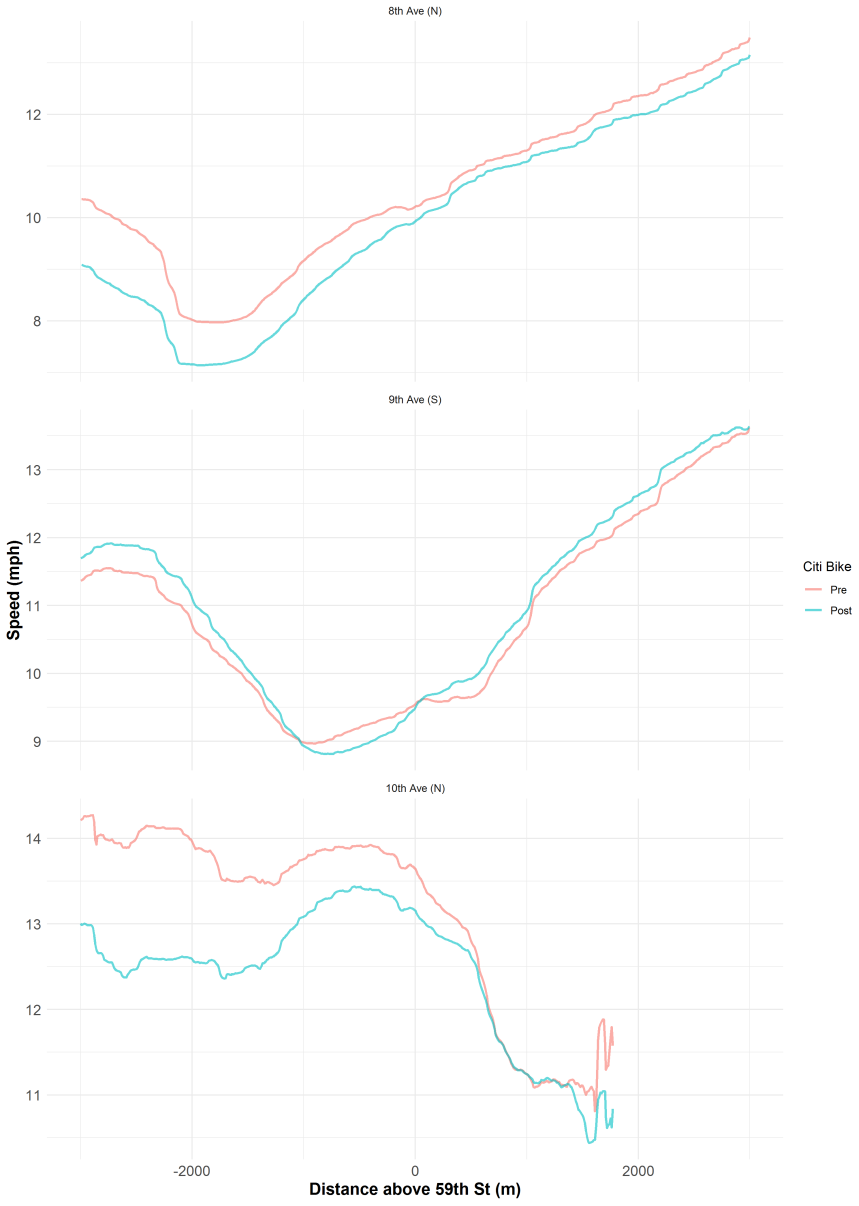
Notes: N = 510,732. Sample period is January 2009 through July 2015. Sample include all 10 meter bins within 3 kilometers of 59th St on the following avenues: 1st Ave, 2nd Ave, 3rd Ave, Lexington Ave, Park Ave, Madison Ave, 5th Ave, 8th Ave, 9th Ave, and 10th Ave. The sample average seconds per meter, .21, is the reciprocal of a speed of 10.65 miles per hour.

employ common time effects, which relative to the estimates other columns, allows variation in travel time across avenues to identify the difference-in-differences estimates.

Figure 3.7: Historical Street-Level Traffic Speed on East Side Avenues, Pre- and Post-Citibike

Note: Space-mean speed in miles per hour on selected West Side avenues in 10-meter bins, pre- and post-Citibike. The sample period is January 2009 through July 2015, and Citibike was introduced in June 2013. The mean speed of all within-avenue taxi trips is 11.13 mph.

Figure 3.8: Historical Street-Level Traffic Speed on West Side Avenues, Pre- and Post-Citibike



Note: Space-mean speed in miles per hour on West Side avenues in estimation sample, measured in 10-meter bins, pre- and post-Citibike. The sample period is January 2009 through July 2015, and Citibike was introduced in June 2013. The mean speed of all within-avenue taxi trips is 11.13 mph. 10th Avenue ends before 3000m North of 59th because it intersects with Broadway and to precisely estimate street mean speeds only taxi trips staring or ending before the intersection were used in estimates.

Table 3.4: Main Difference-in-Differences Results

	Dependent variable: log avg. seconds per meter		
	Standard DID	Monthly intensity	Bin-month intensity
	(1)	(2)	(3)
Citi x post	0.023***		
	(0.004)		
Citi x monthly intensity		0.039***	
		(0.007)	
Citi x bin-month intensity			0.096***
			(0.032)
Unprotected bikelane	−0.024***	−0.024***	−0.070***
	(0.009)	(0.009)	(0.011)
Protected bikelane	0.006	0.005	−0.044***
	(0.004)	(0.004)	(0.006)
311 complaints	0.002***	0.002***	0.002***
	(0.001)	(0.001)	(0.001)
Fixed effects?	Avenue-bin x avenue-ym	Avenue-bin x avenue-ym	Avenue-bin x year-month
Calendar month controls?	Y	Y	Y
Observations	510,732	510,732	510,732
R^2	0.968	0.968	0.939

Notes: *p<0.1; **p<0.05; ***p<0.01. Dependent variable is natural log of seconds per meter. Standard errors two-way clustered at the avenue-bin and avenue-year-month level.

In all specifications, the coefficient of interest is statistically significant at the 1% level and positive, indicating an increase in travel time (or equivalently, a decrease in speed). The coefficient on *Citi* x *Post* in column 1 represents the average causal effect of the Citibike system on travel time below 59th Street. I find that the Citibike system increased travel time by 2.3%. In column 2, the coefficient on *Citi* x *Monthly Intensity* is larger than the coefficient of interest in column 1, indicating that the effect on speed is larger in months when the Citibike system is more fully utilized. At the maximum system utilization of 987,169 monthly rides, I find travel time increased by 3.9%. In column 3, the coefficient on *Citi x Bin-Month Intensity* indicates that at the maximum number of rides traversing a bin 452,056, travel time increases by 9.6%. An equivalent interpretation is that a 10% increase in utilization, of about 45,000 rides, increases travel time by 1%. Unprotected bike lanes are associated with a decrease in travel time. Protected bike lanes are associated with a small, but statistically insignificant increase in travel time based on estimates from columns 2 and 3, but are associated with a decrease in travel time of about 4.4% in column 3. A 311 complaint related to road conditions is associated with an increase in travel time of 0.2%.

These results estimate the combined impact of various effects of the Citibike system. The finding of an increase in travel time is consistent with congestion exacerbating effects of the system outweighing any congestion mitigating effects. The only congestion mitigating effect of Citibike is substitution away from personal vehicles, taxis, and TNC services. Hypothesized congestion exacerbating effects include substitution away from walking and public transit (Campbell and Brakewood, 2017), the effect of removing parking spots for bike docks, induced demand due to the availability of a new transit option, and any service and maintenance activity of the Citibike system that occurs at the street level.

An important parameter of interest is how Citibike affected the speed of the average trip in New York City. The results presented above demonstrate that Citibike has slowed down trips made by car, taxi, TNC or bus. If Citibike users are substituting away from walking, then the introduction of Citibike may have sped up the average trip made by New Yorkers, in spite of the externality imposed on vehicular traffic. Table 3.5 shows the share of the working population residing in Manhattan that commute by different modes of transit. The estimates are derived from American Community Survey data, and are shown for 2012 and 2014, or pre- and post-Citibike. First note that a large share, about 50%, commute by subway, and the next highest share, about 20%, commute on foot. In total, only 12.6% of the Manhattan working population commute by taxi or other vehicle. This finding of limited potential for substitution away from vehicular travel is consistent with my results that Citibike increased congestion.

To estimate Citibike's effect on speed of overall trips would require estimates of substitution of Citibike riders from other modes of transit. I can however, perform a back-of-the-envelope calculation to evaluate whether substitution patterns that would facilitate an increase in overall trip speed are plausible. First I decompose the change in average trip

Table 3.5: Manhattan Commuter Mode Shares for 2012, 2014

Year	2012	2014
Working population	832836	852406
Car, truck, or van	0.088	0.082
Bus	0.081	0.075
Streetcar	0.003	0.003
Subway	0.484	0.499
Rail	0.015	0.014
Ferry	0.001	0.001
Bike	0.011	0.015
Walk	0.213	0.207
Taxi	0.038	0.037
Work from home	0.067	0.067

Notes: Table shows working population and commuter shares by mode of transport for Manhattan for 2012 and 2014. Estimates are derived from American Community Survey 1-year estimates. 2012 is pre-Citibike and 2014 is post-Citibike.

speed as follows:

$$\begin{aligned}&\text{Avg. speed}_1 - \text{Avg. speed}_0 \\ &= \sum_{m \neq bike} \left(\text{Avg. speed}_{m1} - \text{Avg. speed}_{bike1}\right) \Delta s_m + \sum_m \left(\Delta \text{Avg. speed}_m s_{m0}\right), \end{aligned} \tag{3.4}$$

where m denotes a mode of travel, $t \in \{0, 1\}$ denotes pre- and post-Citibike, and s_{mt} denotes the share of all trips made by a particular mode m at time t. Details of this decomposition can be found in Appendix B. The second term on the right-hand side of the above expression captures the externality imposed by Citibike on other modes of travel. I assume this is zero for modes other than vehicular traffic. Using the estimated decrease in speed of 0.023 and the average speed of taxis in the estimation sample of 10.619 miles per hour, the change in speed of vehicular traffic is -0.244 miles per hour. Based on the estimates from commuters in 2012, 20.7% of trips were made by bus, car, or taxi. The first term on the right-hand side is the benefit from commuters who substitute into Citibike from faster trips, in terms of speed. Since the largest difference in speed is between walking and cycling, I can calculate the percentage point change in walking that would have to occur to offset the externality of Citibike on vehicular traffic. The average Citibike speed in the sample is 8.3 miles per hour, and the average walking speed in Manhattan is 3.4 miles per hour.[7] For the externality to be offset and for the speed to increase would require a 1.468 percentage point decrease in the share of walking trips, entirely replaced by Citibike trips. I do not have estimates of the

[7] *www.richardwiseman.com/quirkology/pace_method.htm*

causal effect of Citibike, on substitution between modes of transit. However, in 2012, 1.1% of trips were made by bike, and in 2014, after Citibike's introduction, the share increased to 1.5%. This is suggestive evidence that observed changes in bike trips were insufficient to have resulted in an increase in the speed of the average trip. This is suggestive evidence that the external cost of Citibike on vehicular traffic outweighs the private benefit of the system, in terms of travel time.

Alternative Specifications and Robustness

Table 3.6 reports results from various alternative specifications and robustness checks of my preferred specification, equation 3.3. First, column 1 shows the estimates in levels, rather than logs. The estimate remains significant at the 1% level. The coefficient of 0.031 represents a percent increase in travel time of 14.7% when evaluated against the sample mean of .21 seconds per meter, which is in line with my main result. Column 2 shows estimates without covariates, to show their inclusion is not driving my main result. The estimate is consistent with my main result. Column 3 provides a check on statistical inference by two-way clustering at the avenue-direction, month of sample level. The estimate remains significant at the 1% level. Finally, column 4 provides a check against differential linear trends, for reasons other than the program's impact, that may bias my results. I include a linear trend interacted with avenue-direction and above/below 59th Street. The estimated impact decreases by only 24% after the inclusion of these flexible linear trends. To check that differential trends across north-south geography related to density or other factors are not driving my results, column 5 shows estimates that only include bins below 59th Street. The coefficient remains consistent with my main results.

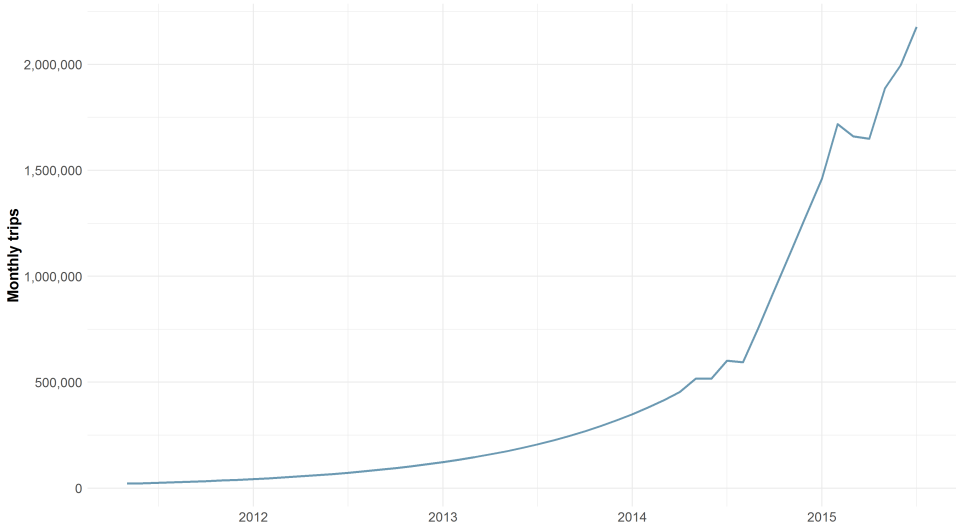
Table 3.6: Robustness and Specification Check Results

	Avg. s/m (Levels)	Log avg. s/m (No controls)	Log avg. s/m (Large clusters)	Log avg. s/m (Linear trend)	Log avg. s/m (Below)	Log avg. s/m (Linear, Uber trend)
	(1)	(2)	(3)	(4)	(5)	(6)
Citi x bin-month intensity	0.031***	0.125***	0.096***	0.073**	0.086***	0.075**
	(0.009)	(0.041)	(0.033)	(0.036)	(0.033)	(0.036)
Unprotected bikelane	−0.016***		−0.070***	−0.020**	−0.078***	−0.027***
	(0.002)		(0.009)	(0.009)	(0.012)	(0.010)
Protected bikelane	−0.012***		−0.044**	−0.027***	−0.050***	−0.024***
	(0.001)		(0.021)	(0.005)	(0.006)	(0.005)
311 complaints	0.001***		0.002*	0.002**	0.002**	0.002**
	(0.0002)		(0.001)	(0.001)	(0.001)	(0.001)
Avenue-bin x year-month FE?	Y	Y	Y	Y	Y	Y
Calendar month controls?	Y	N	Y	Y	N	Y
Observations	510,732	510,732	510,732	510,732	322,908	510,732
R^2	0.918	0.933	0.939	0.943	0.943	0.945

Notes: *p<0.1; **p<0.05; ***p<0.01. The mean of avg. second per meter in the estimtion sample is .21. Standard errors two-way clustered at the avenue-bin and avenue-year-month level, save for column 3, in which they are clustered at the avenue and year-month level. Column 4 interacts linear trends that we allow to vary above vs below 59th St on each avenue-direction. Column 5 only includes all 10 meter bins below 59th St. Column 6 interacts both linear and aggregate Uber trends that we allow to vary above vs below 59th St on each avenue-direction.

Threats to identification of the causal effect of Citibike on travel time include changes in road use that affect speeds differently in treated versus untreated areas. A major change in road use that I have not yet addressed is the introduction of TNCs such as Uber. Uber would bias my estimates to the extent that its impact is different across treated and untreated bins. Uber launched in New York City as a black car service on May 4, 2011.[8] To address this concern I follow Mangrum and Molnar, 2019 and construct an "aggregate Uber trend" from monthly Uber pickups in Manhattan. Figure 3.9 shows monthly Uber trips in Manhattan for May 2011 onward. [9] I interact this trend at the level of an avenue-direction and above/below 59th Street. Column 6 of Figure 3.6 reports estimates of my preferred specification with the joint inclusion of both the Uber trend and a linear trend. This tests the robustness of estimates to the inclusion of a trend that is not only linear, but also allowed to accelerate after May 2011 at a rate that is proportional to the growth of Uber. The estimate remains statistically significant at the 5% level and is similar in magnitude to my baseline estimate.

Figure 3.9: Time Series of Monthly Uber Pickups in Manhattan



Note: Trip-level data for Uber is available for the second and third quarters of 2014, and for 2015 onward. I extrapolate to the missing period in 2014, and to Uber's launch in May 2011 using the average growth rate of 9.09% over the period for which I have data.

One example of a potential source of bias in my estimates would be a violation of the assumption that riders do not rent bicycles to ride above 59th Street, or put another way, that the docks at 59th Street represent a boundary of the treated area. Adding cyclists

[8] https://www.nytimes.com/2011/05/04/technology/04ride.html

[9] Trip-level data for Uber is available for the second and third quarters of 2014, and for 2015 onward. I extrapolate to the missing period in 2014, and to Uber's launch in May 2011 using the average growth rate of 9.09% over the period for which I have data.

to the area north of 59th Street, the "control" area, would bias estimates downwards, as street speeds would be lower in that area than in the true counterfactual. This phenomenon does not threaten the detection of a positive affect, and only affects the magnitude of the estimate. Even then, it is sufficiently rare as to not be a concern. Some users certainly use the bicycles to go above 59th Street. In the estimation period, 69,135 rides originated from a dock on the 59th Street boundary and returned to the same dock. This represents just 0.4% of all rides, and many of these "boundary rides" could have been taken south of 59th Street, or north but remained within Central Park and not on avenues. Similarly, spatial spillovers of congestion do not threaten the detection of a positive affect. Any slow down below 59th Street that has a knock-on effect on northbound avenues will also make street speeds lower in the area above 59th Street that serves as the "control" area. This would bias my estimates downwards. Another potential concern is that Citibike riders do not follow Google Maps bicycling directions. This might be because cyclists are unable to follow a route on their phone when riding and likely deviate from these routes. First note that a rider need not consistently look at their phone when riding. Users need only map the route at the start of the ride, and because Manhattan streets form a grid, they need only memorize a few simple directions. Recently, Google Maps has incorporated real time data that shows if Citibike docks are open or closed, if the docks are accepting returns and/or offering rentals (i.e. is the dock full or not).[10] This suggests consumers have historically, and continue to use Citibike and Google Maps in an interdependent fashion. Even if Citibike users deviate from Google Maps directions, the effect would be similar to if riders went above 59th St. This phenomenon would increase congestion in areas that have a lower or no intensity, "control" areas, and thereby not threaten the detection of an effect, but merely its magnitude.

Another potential concern is endogeneity caused by simultaneity of my independent variable of interest in specifications 3.2 and 3.3 with my outcome of interest, travel time. As streets become more congested, more people who would have traveled by personal vehicle, TNC, taxi, or bus, may choose to ride Citibike. More Citi Bike riders may exacerbate congestion. This would bias my estimates upwards. While this is technically a valid concern, for any plausible elasticity of Citibike ridership with respect to congestion, the small size of the congestion effect I estimate means this source of bias would be extremely small.

3.6 Conclusion

This study estimates the effect of Citibike, New York City's bike-sharing system, on congestion along Manhattan avenues. I find that the Citibike system decreased speeds on avenues in Manhattan. Overall, I estimate that the system decreased speed by 2.3%. At the maximum number of rides traversing a 10-meter bin, 452,056, travel time increases by 9.6%. Under generous assumptions about substitution patterns (namely that all Citibike users would have walked), observed changes in the share of trips made by bike in Manhattan before and after

[10] *www.forbes.com/sites/melissakravitz/2018/04/30/google-maps-is-making-citi-bikes-so-much-easier-to-use-in-new-york-city*

Citibike are insufficient to have resulted in an increase in the speed of the average trip. This is suggestive evidence that the external cost of Citibike on vehicular traffic outweighs the private benefit of the system, in terms of travel time.

These estimates comprise the combined impact of various effects of the Citibike system. The finding of an increase in travel time is consistent with congestion exacerbating effects of the system outweighing any congestion mitigating effects. The only congestion mitigating effect of Citibike is substitution away from personal vehicles, taxis, and TNC services. Congestion exacerbating effects include substitution away from walking and public transit, the effect of removing parking spots for bike docks, induced demand due to the availability of a new transit option, and any service and maintenance activity of the Citibike system that occurs at the street level. I cannot disentangle each of these effects. These estimates therefore represent a lower bound on the congestion externality imposed by Citibike.

While bike-sharing systems may offer other benefits (Woodcock et al., 2014; Xu, 2019), this result suggests their potential to reduce congestion has been over-estimated, at least in certain settings. This should give policy makers pause. In section 2, I compare Citibike to other large bike-sharing systems in the United States. Citibike about twice as large in number of annual rides, but that the number of monthly rides on the most popular ride in New York City, 11,860, is only 34% higher than the number of monthly rides on the most popular ride in Chicago. This indicates that bike-sharing penetration in Chicago is comparable to that in New York City during the sample period. Bike-sharing systems are already widespread in the US. Nevertheless, these results should inform the analysis of whether their further expansion is welfare-improving, as well as the introduction of other shared mobility services that may conflict with existing transportation infrastructure.

Bibliography

Allcott, Hunt, Christopher Knittel, and Dmitry Taubinsky (May 2015). "Tagging and Targeting of Energy Efficiency Subsidies". In: *American Economic Review* 105.5, pp. 187–91. DOI: 10.1257/aer.p20151008. URL: http://www.aeaweb.org/articles?id=10.1257/aer.p20151008.

Anderson, Michael L. (Sept. 2014). "Subways, Strikes, and Slowdowns: The Impacts of Public Transit on Traffic Congestion". In: *American Economic Review* 104.9, pp. 2763–96. DOI: 10.1257/aer.104.9.2763. URL: http://www.aeaweb.org/articles?id=10.1257/aer.104.9.2763.

Anderson, Michael L. et al. (2016). "Superstitions, street traffic, and subjective well-being". In: *Journal of Public Economics* 142, pp. 1–10. ISSN: 0047-2727. DOI: https://doi.org/10.1016/j.jpubeco.2016.07.005. URL: http://www.sciencedirect.com/science/article/pii/S0047272716300871.

Armstrong, Mark (2006). "Competition in two-sided markets". In: *The RAND Journal of Economics* 37.3, pp. 668–691. DOI: 10.1111/j.1756-2171.2006.tb00037.x. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1756-2171.2006.tb00037.x. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1756-2171.2006.tb00037.x.

Barth, Matthew and Kanok Boriboonsomsin (2008). "Real-World Carbon Dioxide Impacts of Traffic Congestion". In: *Transportation Research Record* 2058.1, pp. 163–171. DOI: 10.3141/2058-20. eprint: https://doi.org/10.3141/2058-20. URL: https://doi.org/10.3141/2058-20.

Bauernschuster, Stefan, Timo Hener, and Helmut Rainer (Feb. 2017). "When Labor Disputes Bring Cities to a Standstill: The Impact of Public Transit Strikes on Traffic, Accidents, Air Pollution, and Health". In: *American Economic Journal: Economic Policy* 9.1, pp. 1–37. DOI: 10.1257/pol.20150414. URL: http://www.aeaweb.org/articles?id=10.1257/pol.20150414.

Berry, Steven T. and Joel Waldfogel (2001). "Do Mergers Increase Product Variety? Evidence from Radio Broadcasting". In: *Quarterly Journal of Economics* 116, pp. 1009–1025.

Boomhower, Judson and Lucas W. Davis (2014). "A credible approach for measuring inframarginal participation in energy efficiency programs". In: *Journal of Public Economics* 113, pp. 67–79. ISSN: 0047-2727. DOI: https://doi.org/10.1016/j.jpubeco.2014.03.009. URL: http://www.sciencedirect.com/science/article/pii/S0047272714000589.

Borenstein, Severin (May 1988). "On the Efficiency of Competitive Markets for Operating Licenses*". In: *The Quarterly Journal of Economics* 103.2, pp. 357–385. ISSN: 0033-5533. DOI: 10.2307/1885118. eprint: https://academic.oup.com/qje/article-pdf/103/2/357/5171078/103-2-357.pdf. URL: https://doi.org/10.2307/1885118.

Borenstein, Severin and Lucas W. Davis (2016). "The Distributional Effects of US Clean Energy Tax Credits". In: *Tax Policy and the Economy* 30.1, pp. 191–234. DOI: 10.1086/685597. eprint: https://doi.org/10.1086/685597. URL: https://doi.org/10.1086/685597.

Campbell, Kayleigh B. and Candace Brakewood (2017). "Sharing riders: How bikesharing impacts bus ridership in New York City". In: *Transportation Research Part A: Policy and Practice* 100, pp. 264–282. ISSN: 0965-8564. DOI: https://doi.org/10.1016/j.tra.2017.04.017. URL: http://www.sciencedirect.com/science/article/pii/S0965856416304967.

Corts, Kenneth S. (2010). "Building out alternative fuel retail infrastructure: Government fleet spillovers in E85". In: *Journal of Environmental Economics and Management* 59.3, pp. 219–234. ISSN: 0095-0696. DOI: https://doi.org/10.1016/j.jeem.2009.09.001. URL: http://www.sciencedirect.com/science/article/pii/S0095069609001089.

Couture, Victor, Gilles Duranton, and Matthew A. Turner (2018). "Speed". In: *The Review of Economics and Statistics* 100.4, pp. 725–739. DOI: 10.1162/rest_a_00744. eprint: https://doi.org/10.1162/rest_a_00744. URL: https://doi.org/10.1162/rest_a_00744.

Currie, Janet and Reed Walker (Jan. 2011). "Traffic Congestion and Infant Health: Evidence from E-ZPass". In: *American Economic Journal: Applied Economics* 3.1, pp. 65–90. DOI: 10.1257/app.3.1.65. URL: http://www.aeaweb.org/articles?id=10.1257/app.3.1.65.

Davis, Lucas W. (2019). "Evidence of a homeowner-renter gap for electric vehicles". In: *Applied Economics Letters* 26.11, pp. 927–932. DOI: 10.1080/13504851.2018.1523611. eprint: https://doi.org/10.1080/13504851.2018.1523611. URL: https://doi.org/10.1080/13504851.2018.1523611.

DellaVigna, Stefano and Matthew Gentzkow (Nov. 2017). *Uniform Pricing in US Retail Chains*. Working Paper 23996. National Bureau of Economic Research. DOI: 10.3386/w23996. URL: http://www.nber.org/papers/w23996.

Dixit, Avinash K. and Joseph E. Stiglitz (1977). "Monopolistic Competition and Optimum Product Diversity". In: *The American Economic Review* 67.3, pp. 297–308. ISSN: 00028282. URL: http://www.jstor.org/stable/1831401 (visited on 11/06/2022).

Dorsey, Jackson, Ashley Langer, and Shaun McRae (Mar. 2022). *Fueling Alternatives: Gas Station Choice and the Implications for Electric Charging*. Working Paper 29831. National Bureau of Economic Research. DOI: 10.3386/w29831. URL: http://www.nber.org/papers/w29831.

Duranton, Gilles and Matthew A. Turner (Oct. 2011). "The Fundamental Law of Road Congestion: Evidence from US Cities". In: *American Economic Review* 101.6, pp. 2616–52. DOI: 10.1257/aer.101.6.2616. URL: http://www.aeaweb.org/articles?id=10.1257/aer.101.6.2616.

Gerarden, Todd, Richard G. Newell, and Robert N. Stavins (May 2015). "Deconstructing the Energy-Efficiency Gap: Conceptual Frameworks and Evidence". In: *American Economic Review* 105.5, pp. 183–86. DOI: 10.1257/aer.p20151012. URL: http://www.aeaweb.org/articles?id=10.1257/aer.p20151012.

Hamilton, Timothy L. and Casey J. Wichman (2018). "Bicycle infrastructure and traffic congestion: Evidence from DC's Capital Bikeshare". In: *Journal of Environmental Economics and Management* 87, pp. 72–93. ISSN: 0095-0696. DOI: https://doi.org/10.1016/j.jeem.2017.03.007. URL: http://www.sciencedirect.com/science/article/pii/S0095069616300420.

Hotelling, H. (1929). "Stability in competition". In: *The economic journal* 39.153, pp. 41–57.

Knittel, Christopher R., Douglas L. Miller, and Nicholas J. Sanders (2016). "Caution, Drivers! Children Present: Traffic, Pollution, and Infant Health". In: *The Review of Economics and Statistics* 98.2, pp. 350–366. DOI: 10.1162/REST_a_00548. eprint: https://doi.org/10.1162/REST_a_00548. URL: https://doi.org/10.1162/REST_a_00548.

Koenker, Roger W. and Martin K. Perry (1981). "Product Differentiation, Monopolistic Competition, and Public Policy". In: *The Bell Journal of Economics* 12.1, pp. 217–231. ISSN: 0361915X. URL: http://www.jstor.org/stable/3003518 (visited on 11/06/2022).

Li, Jing (2019). "Compatability and Investment in the U.S. Electric Vehicle Market". URL: http://www.mit.edu/~lijing/documents/papers/li_evcompatibility.pdf.

Li, Shanjun (Nov. 2017). "Better Lucky Than Rich? Welfare Analysis of Automobile Licence Allocations in Beijing and Shanghai". In: *The Review of Economic Studies* 85.4, pp. 2389–2428. ISSN: 0034-6527. DOI: 10.1093/restud/rdx067. eprint: https://academic.oup.com/restud/article-pdf/85/4/2389/25848769/rdx067.pdf. URL: https://doi.org/10.1093/restud/rdx067.

Li, Shanjun et al. (2017). "The Market for Electric Vehicles: Indirect Network Effects and Policy Design". In: *Journal of the Association of Environmental and Resource Economists* 4.1, pp. 89–133. DOI: 10.1086/689702. eprint: https://doi.org/10.1086/689702. URL: https://doi.org/10.1086/689702.

Mangrum, Daniel and Alejandro Molnar (2019). "The marginal congestion of a taxi in New York City". AER revise and resubmit. URL: https://alejandromolnar.github.io/files/mm_boro.pdf.

Mankiw, N. Gregory and Michael D. Whinston (1986). "Free Entry and Social Inefficiency". In: *The RAND Journal of Economics* 17.1, pp. 48–58. ISSN: 07416261. URL: http://www.jstor.org/stable/2555627 (visited on 11/06/2022).

Muehlegger, Erich and David S. Rapson (Dec. 2018). *Subsidizing Mass Adoption of Electric Vehicles: Quasi-Experimental Evidence from California*. NBER Working Papers 25359. National Bureau of Economic Research, Inc. URL: https://ideas.repec.org/p/nbr/nberwo/25359.html.

Rochet, Jean-Charles and Jean Tirole (2006). "Two-Sided Markets: A Progress Report". In: *The RAND Journal of Economics* 37.3, pp. 645–667. ISSN: 07416261. URL: http://www.jstor.org/stable/25046265.

Seim, Katja and Joel Waldfogel (2013). "Public Monopoly and Economic Efficiency: Evidence from the Pennsylvania Liquor Control Board's Entry Decisions". In: *The American Economic Review* 103.2, pp. 831–862. ISSN: 00028282. URL: http://www.jstor.org/stable/23469684 (visited on 11/06/2022).

Spence, Michael (1976). "Product Selection, Fixed Costs, and Monopolistic Competition". In: *The Review of Economic Studies* 43.2, pp. 217–235. ISSN: 00346527, 1467937X. URL: http://www.jstor.org/stable/2297319 (visited on 11/06/2022).

Springel, Katalin (2019). "Network Externality and Subsidy Structure in Two-Sided Markets: Evidence from Electric Vehicle Incentives". URL: https://www.dropbox.com/s/ix788j0b6d24j6p/kspringel_ev.pdf?dl=0.

Sweeting, Andrew (2010). "The effects of mergers on product positioning: evidence from the music radio industry". In: *The RAND Journal of Economics* 41.2, pp. 372–397. ISSN: 07416261. URL: http://www.jstor.org/stable/40649304 (visited on 11/29/2022).

Vant-Hull, Brian et al. (2014). "FINE STRUCTURE IN MANHATTAN'S DAYTIME URBAN HEAT ISLAND: A NEW DATASET". In: *Journal of Urban and Environmental Engineering* 8.1, pp. 59–74. ISSN: 19823932. URL: http://www.jstor.org/stable/26203410.

Wang, Mingshu and Xiaolu Zhou (2017). "Bike-sharing systems and congestion: Evidence from US cities". In: *Journal of Transport Geography* 65, pp. 147–154. ISSN: 0966-6923. DOI: https://doi.org/10.1016/j.jtrangeo.2017.10.022. URL: http://www.sciencedirect.com/science/article/pii/S0966692317302715.

Woodcock, James et al. (2014). "Health effects of the London bicycle sharing system: health impact modelling study". In: *BMJ* 348. DOI: 10.1136/bmj.g425. eprint: https://www.bmj.com/content/348/bmj.g425.full.pdf. URL: https://www.bmj.com/content/348/bmj.g425.

Xu, Dafeng (2019). "Burn Calories, Not Fuel! The effects of bikeshare programs on obesity rates". In: *Transportation Research Part D: Transport and Environment* 67, pp. 89–108. ISSN: 1361-9209. DOI: https://doi.org/10.1016/j.trd.2018.11.002. URL: http://www.sciencedirect.com/science/article/pii/S1361920918304103.

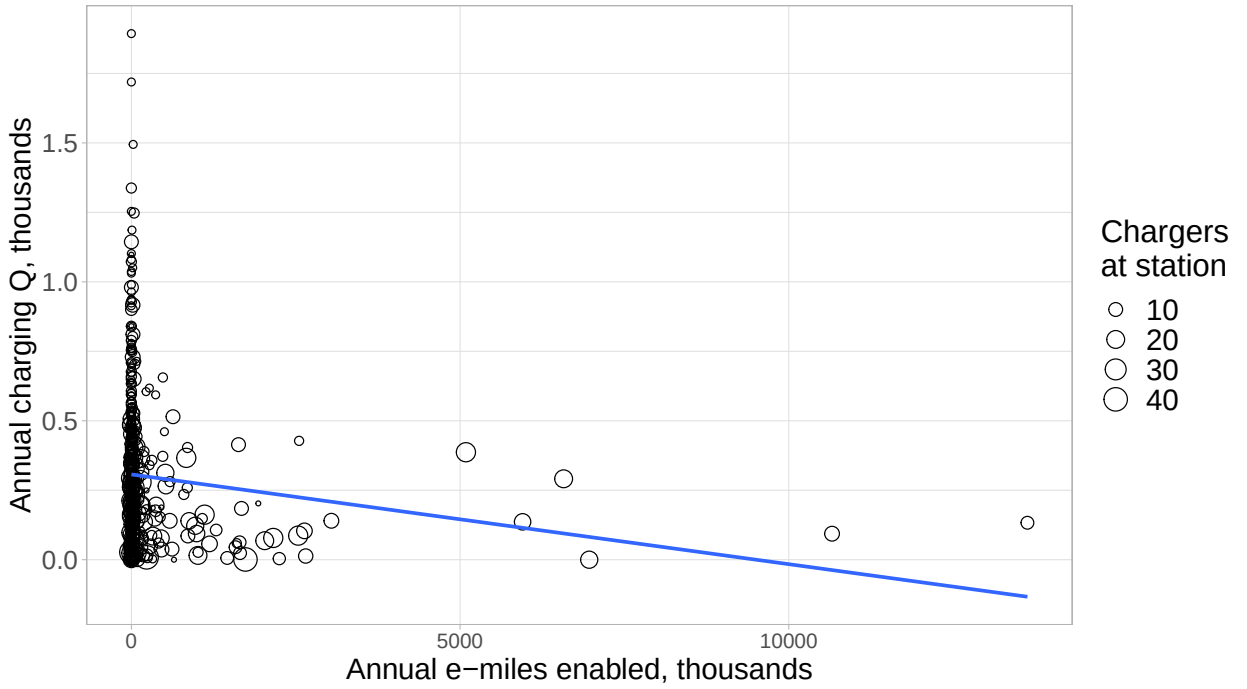
Appendix A

Appendix for Chapter 1

A.1 Alternative Normalization

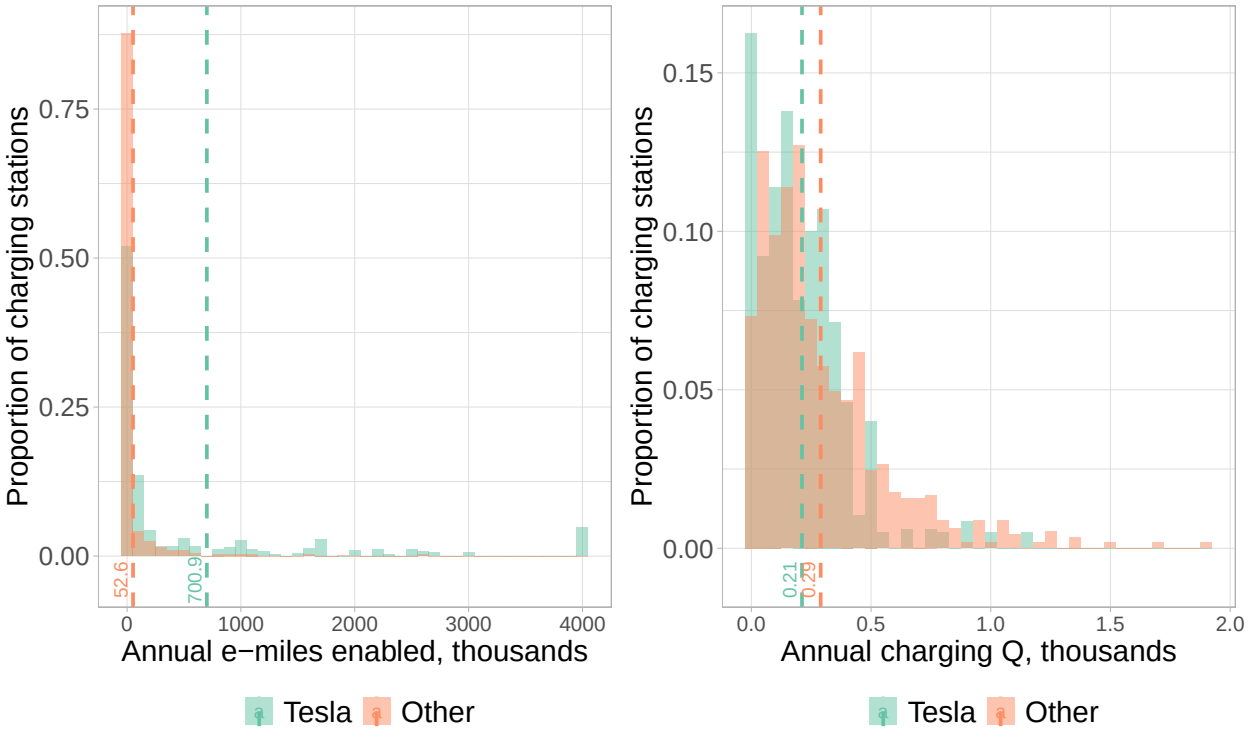
In this section I present results from an alternative normalization, which rolls the development of the CCS and CHAdeMO networks back to 2015, while maintaining the 2019 Tesla network, so as to have an equivalent number of vehicles per charger. The results of the alternative normalization are very similar to those I present in my main results.

Figure A.1: Annual Charging Quantity Versus Annual E-miles Enabled at 2019 Charging Stations



Note: N = 702. This figure shows annual e-miles enabled and annual charging demand for marginal charging station locations in California. Each data point is a charging station location. The size of each point represents the number of chargers at the charging station.

Figure A.2: Distribution of Annual E-miles Enabled and Annual Charging Demand for Tesla and Non-Tesla Charging Stations



Note: This figure shows distributions of annual e-miles enabled and annual charging demand for Tesla and non-Tesla chargers. Observations are at the charging station level, weighted by the number of chargers at a station. The left panel presents the annual e-miles enabled, and the right panel presents annual charging demand. The vertical dashed lines show average values for Tesla and non-Tesla chargers.

Appendix B

Appendix for Chapter 3

B.1 Space-Mean Speed

The Federal Highway Administration defines space-mean speed as the average speed of vehicles traveling a given segment of roadway during a specified period of time and is calculated using the average travel time and length for the roadway segment. Equation B.1 shows the formula for space-mean speed.

$$v_s = \frac{d}{\frac{\sum t_i}{n}}, \tag{B.1}$$

where d is the distance traveled, t_i is the travel time of vehicle i and n is the number of observations. For our data, we observe travel time and the length of the trip, so we convert each taxi's travel time to be average seconds per meter. This means d is one, and so $1/v_s$ yields average travel time, which is our dependent variable.

B.2 Back-of-the-envelope decomposition of speed change

Let m denote a mode of travel, let $t \in \{0, 1\}$ denote pre- and post-Citibike, and let s_{mt} denote the share of all trips made by a particular mode m at time t. We can then write

$$
\begin{aligned}
&\text{Avg. speed}_1 - \text{Avg. speed}_0 \\
&= \frac{\sum_m \text{Avg. speed}_{m1}\text{trips}_{m1}}{\sum_m \text{trips}_{m1}} - \frac{\sum_m \text{Avg. speed}_{m0}\text{trips}_{m0}}{\sum_m \text{trips}_{m0}} \\
&= \sum_m \text{Avg. speed}_{m1} s_{m1} - \sum_m \text{Avg. speed}_{m0} s_{m0} \\
&= \sum_m \text{Avg. speed}_{m1} s_{m1} - \text{Avg. speed}_{m1} s_{m0} + \text{Avg. speed}_{m1} s_{m0} - \text{Avg. speed}_{m0} s_{m0} \\
&= \sum_m \text{Avg. speed}_{m1} \Delta s_m + \Delta\text{Avg. speed}_m s_{m0} \\
&= \sum_{m \neq bike} (\text{Avg. speed}_{m1} \Delta s_m) + \text{Avg. speed}_{bike1} \Delta s_{bike} + \sum_m (\Delta\text{Avg. speed}_m s_{m0}) \\
&= \sum_{m \neq bike} (\text{Avg. speed}_{m1} \Delta s_m) + \text{Avg. speed}_{bike1} \left(- \sum_{m \neq bike} \Delta s_m \right) + \sum_m (\Delta\text{Avg. speed}_m s_{m0}) \\
&= \sum_{m \neq bike} (\text{Avg. speed}_{m1} - \text{Avg. speed}_{bike1}) \Delta s_m + \sum_m (\Delta\text{Avg. speed}_m s_{m0}) .
\end{aligned}
\tag{B.2}
$$