Addressing Traffic Congestion

Yooseon Hwang

Gunsan, Republic of Korea

B.A., Economics and Statistics, University of California, Berkeley 2013;
M.A., Economics, Seoul National University 2017;
M.A., Economics, University of Virginia 2018

A Dissertation presented to the Graduate Faculty of the University of Virginia in Candidacy for the Degree of Doctor of Philosophy

Department of Economics

University of Virginia

May 2023

Committee Members:
Kerem Coşar
Jonathan Colmer
James Harrigan
Leora Friedberg
Peter Debaere

Abstract

Traffic congestion is associated with enormous time, fuel, environmental, and health costs. As traffic congestion worsens around the world, it is becoming increasingly crucial to understand how congestion policies affect mobility and the economy.

In the first chapter, I present stylized facts about traffic congestion. I document the extent to which traffic congestion in the US has worsened. I also discuss the implications and limitations of supply side investments such as highway expansions or transit constructions. Lastly, I present case studies of demand side investments, which aim to balance the demands for travel using the existing infrastructure.

In the second chapter, I estimate the effects of highway congestion pricing on traffic using spatial panel data on real-time traffic speed and flow in California. I provide reduced form estimates to determine the degree to which traffic diverts from toll to non-toll lanes using a policy change in the Los Angeles area in which a subset of non-toll lanes on Interstates 10 and 110 were converted to toll lanes with dynamic pricing. Results provide supporting evidence that in the short run, drivers avoid toll costs by mostly switching from toll lanes to non-toll lanes; over time, changes in the spatial distribution of residential and work locations induce further adjustments in driving routes. This implies that individual responses may involve not just changing where they drive, but also where they live or work.

In the third chapter, I estimate the aggregate effects of congestion pricing by taking into account where people live, work and drive. I develop a quantitative urban model with endogenous commuting costs in which residential and commercial locations, driving routes, travel times, and toll costs are simultaneously determined. Based on model estimates, I estimate both the partial and general equilibrium effects of congestion pricing. In the partial equilibrium analysis, which holds the locations of residences and workplaces fixed, congestion pricing induces a spatial leakage of traffic externality as people divert from toll lanes to non-toll lanes; this reduces annual aggregate welfare by \$1.8-\$11.0 million. However, in the general equilibrium analysis, which allows for adjustments in residences, workplaces, and driving routes, congestion in the overall road network decreases because people re-sort to reduce commuting distances. In aggregate, when net toll revenues are redistributed, annual welfare increases by \$2.4-\$11.6 million.

Keywords: traffic, congestion, commuting, economic geography, sorting
JEL codes: R41, R13, L91, H23

Acknowledgement

I would like to thank my advisors, Kerem Coşar, Jonathan Colmer, James Harrigan, and Leora Friedberg. I am incredibly grateful for all their support, guidance, and encouragement. I would not have finished this without Kerem Coşar's guidance and mentorship. I would like to thank Jonathan Colmer for pushing me vigorously and continuosly. Tremendous thanks to Leora Friedberg for patiently guiding me throughout my graduate research. Without her, I wouldn't have started this project. Thanks as well to James Harrigan for your encouragement and guidance.

Thank you to my mother, Tae-Ok, for always supporting me, having faith in me and guiding me in the right direction. Without her, I would not have accomplished this, and I owe everything to her. Thank you to my little sisters, Yoo-Jung and Yoo-Jin, for supporting me and putting up with me. I feel extremely lucky to be part of this family. 정말 많이 사랑합니다 우리 가족.

1 Traffic Congestion: Stylized Facts

1.1 Trends

Traffic congestion is slowing people down in cities. In London, Chicago, Paris, Boston and New York cities, which are the 5 most congested cities in the world as of 2022, the average annual hours lost in congestion were approximately 130 hours per commuter (Pishue, 2023). During rush hour, the average speed in downtown Chicago, Boston, New York City, and Philadelphia was 11 miles per hour in 2022.

Congestion continues to worsen. Figure 1 displays the average annual traffic delay per commuter in US cities - the total annual hours that would have been saved in the absence of congestion. Total delay averaged around 40 hours per year in 1980, but it has nearly doubled by 2019. During the Coronavirus pandemic, however, cities saw an unusual and massive decrease in traffic due to lockdowns and mobility limitations. Figure 2, which shows snapshots of traffic flows in Los Angeles and San Francisco before and after the restrictions, shows a significant reduction in traffic on the majority of the highways. During the Coronavirus restrictions, traffic flows plummeted by 67% and 75% in Los Angeles and San Francisco city, respectively (Marchant, 2020).

Congestion has returned as most cities have relaxed mobility restrictions. Compared to the Coronavirus pandemic, the number of travels to downtown Washington, DC, Seattle, and San Francisco city increased by 13-23% in 2022 (Pishue, 2023). However, post-pandemic traffic hasn't reached the pre-Covid level as working from home has become the new norm. According to the American Community Survey data, before the pandemic, about 5.7% of the workers in the US worked from home while in 2021, 17.9% of the workers worked from home. A growing body of research highlights that remote and flexible working may continue for a few reasons (Choudhury et al., 2021; Emanuel and Harrington, 2021; Barrero et al., 2021). First, people had a better-than-expected experience with remote working; survey data from US workers show that stigma associated with remote working lessened significantly after the epidemic. Second, working from home increases worker productivity in some industries, such as call centers and patent offices. Also, throughout the epidemic, people substantially invested in technology to assist working from home.

1.2 Consequences

Time and fuel costs associated with congestion is enormous. Figure 3 displays a yearly time series plot of time and fuel costs associated with congestion in the US. In 1980, the total time and fuel costs incurred due to congestion was about $1000 per commuter in very large cities; by 2019, it has increased by more than 50 percent. In the rest of this section, I discuss other major externalities.

Environmental costs: Air pollution is one of the major externalities associated with driving. In the US, driving produces 50 percent of carbon monoxide, 34 percent of nitrogen dioxide, and 10 percent of fine particulate matter emissions (Ernst et al., 2002). These pollutants pose serious health risks such as infant mortality and child's health (Currie and Walker, 2011; Dugandzic et al., 2006). And, traffic delays exacerbate air pollution. While cars idle on the road, their fuel consumption and emissions rise. According to Currie and Walker (2011), the installation of electronic toll plazas considerably reduces vehicle pollution from idling engines, which results in a 10% reduction in prematurity and low birth weight within 2 kilometers of the toll plazas.

Mental stress and crime: Congestion induces serious psychological stress. Survey data suggests that an exposure to traffic congestion is positively correlated with numerous health outcomes. A longer commuting distance is positively correlated with a probability of visiting general practitioners and the perceived level of psychological stress (Künn-Nelen, 2016; Roberts et al., 2011). Additionally, Beland and Brent (2018) shows how severe traffic congestion significantly affects domestic violence. Combining fine spatial panel data on real-time traffic and police reports in Los Angeles, it finds that extreme traffic congestion, defined as traffic delay above the 95th percentile, significantly increases the reported domestic violence cases by 9%. The annual costs associated with an increase in domestic violence induced by severe traffic are estimated to be about $5-22 million dollars.

Wages and productivity: Evidence also suggests that commuting and congestion affect compensating differentials and worker productivity. In a quasi-experimental setting, Mulalic et al. (2014) uses the universe of Danish firms and finds that a 1 km increase in commuting distance is associated with about 0.15% increase in wages three years after the relocation. This is consistent with the labor market theory, which hypothesizes that firms possess market power to pay below workers' productivity. Additionally, a recent work argues that commuting distance may deter innovation.

Using a panel data of US inventors, Xiao et al. (2021) finds that a 10 km increase in distance is associated with a 5% decrease in patents, with the effects being stronger for more productive inventors. This implies that policymakers should carefully consider the role of density in urban planning.

1.3 Congestion Mitigation Policies

Governments have used a variety of traffic mitigation investments to reduce the waste caused by traffic congestion and to increase mobility. Supply-side investments, which have historically been more popular, increase the capacity of the existing transportation infrastructure (e.g., highway expansions or transit improvements). Demand-side investments, however, aim to manage the demand for travel without altering the supply of transportation infrastructure.(e.g., congestion pricing). Also, a growing number of cities are adopting micromobility (e.g. electronic scooters or bicycles), which has the potential to transform urban mobility. In this section, I discuss these three types of policies in more detail.

1.3.1 Supply-side investments

Government spending on transportation infrastructure is substantial. According to the annual survey data on state and local finance, federal and state governments spent $204 billion on highways and roads, of which 57 percent went toward the construction of both highways and roads in 2020 (US Census Bureau, 2020b). The total government spending on public transportation was about $79 billion in 2019 (Musick, 2022). In this section, I present examples of highway expansion and rail transit construction and discuss their implications on traffic.

Katy Freeway expansion in Houston: The Katy Freeway expansion is one of the largest road expansions in American history. It was built in 1960 with three lanes in each direction, which could accommodate 80,000 car per day. By 2020, traffic volume had tripled, resulting in chronic traffic delays (US Department of Transportation). From 2003 to 2008, a total of $2.79 billion was spent to add three general lanes and two high occupancy toll lanes in each direction from west of State Highway 6 to the I-10/610 interchange.

Following the expansion, traffic flow on Katy freeway began to increase rapidly. Figure 4 shows yearly time series plots of Average Annual Daily Traffic (AADT) for I-10W, I-10E, and I-69; and Figure 5 shows the locations where AADT was measured. During the pre-expansion period (2000-2009), the AADT on I-10W was around 200,000; once expanded, it nearly doubled by 2018. However, AADT on I-10E and I-69 do not show significant changes during this time period.

Figure 6, and 7 display average traffic speed on major highways connecting suburbs to downtown Houston during 6:00 - 6:30PM.[1] They demonstrate that, prior to its expansion in 2003, I-10W was the busiest highway in Houston; during 6:00 - 6:30PM, its average speed is less than 30 mph, while that of the other highways is greater than 50 mph. A year after the expansion in 2009, traffic speed on I-10W reaches over 50 mph and has the similar level of congestion as the rest of the highways. The improvement in traffic conditions on the Katy Freeway, however, is only temporary, as the average speed drops to less than 30 mph in 2013, 5 years after the expansion.

These findings on changes in traffic flow and speed before and after the expansion are consistent with previous research on the induced demand for travel; while road construction may temporarily alleviate congestion, it increases the demand for travel in the long run (Duranton and Turner, 2011; Cervero, 2002).

Rail transit construction in Los Angeles: As traffic delays in Los Angeles increased significantly in the 1970s, public support for public mass transit grew. In addition, the 1970s saw the start of a movement known as "Freeway Revolt" (Wachs et al., 2015). Residents and drivers agreed that highway construction alone cannot solve traffic problems and disproportionately harms minority neighborhoods. Eventually, a majority of Los Angeles County voters approved Proposition A, which raised the sales tax by a half cent and dedicated 35% of the revenue to rail construction. After construction began in 1986, all of the proposed lines (Gold, Red, Orange, and Green) were completed by 2005.

Empirical evidence suggests that the construction of a rail system significantly increased the number of commuters between residences and workplaces connected by the metro rail system. Severen (2019) finds that for regions served by the rail system, the number of commuters increased by 11–16%. Additionally, by 2000, driving times on routes completely inside 250 meters of lines had decreased by 14%. Other studies have also suggested that a rail transportation system could, in the short term, reduce congestion in the area; Anderson (2014) estimates that during the 47 days that the train system was temporarily shut down due to a labor strike in 2003, highway traffic

[1] Speed map archives are available as video format from the Houston Department of Transportation. Figures are screenshots of the videos in 2003, 2009, and 2013. Speed maps are not available during construction and in 2012.

delays increased by 47%. However, it is unclear whether public transportation systems can persistently reduce congestion in the long run. In fact, Duranton and Turner (2011) does not find significant effects of transit service on congestion in the long-run.

1.3.2 Demand-side investment

Recently, governments have made significant investments on demand side investments in order to better balance the demands for travel using the existing infrastructure. Congestion pricing, which imposes fees on users during specific times of the day or in specific locations, is one of the most popular demand side investments.[2] In this section, I discuss two types of congestion pricing: express toll lanes in Los Angeles and dynamic parking pricing in San Francisco.

Express Toll lanes: The Department of Transportation (DoT) established the Congestion Reduction Demonstration Program in 2006 to reward innovative techniques for reducing urban congestion. Los Angeles and San Francisco were among the cities that received federal funds to experiment with congestion pricing.

One of the major projects in Los Angeles was the implementation of congestion pricing on highway lanes. The existing High-Occupancy vehicles (HOV) lanes on Interstate 10 and 110, which were previously reserved for carpools with multiple passengers, were converted to High-Occupancy Tolled (HOT) lanes. Solo drivers are permitted to use HOT lanes by paying tolls, which vary dynamically based on the level of congestion.[3]

An evaluation of the program during the first 2 years suggests that express lanes reduce travel times and increase travel time reliability (Schroeder et al., 2015). Prior to the implementation of congestion pricing, it took 22 and 23 minutes to drive east and west on I-10, respectively. With the implementation of congestion pricing on express lanes, it dropped to 19 and 17 minutes, respectively.

Drivers' responses to express toll lanes are informative about how they value time. Using transaction-level data on express toll lanes in hedonic estimation, Bento et al. (2020) discovers that preferences for time savings are closely related to the value of urgency. Express lanes save around 3.79 minutes for an average toll cost of approximately \$3.71. Furthermore, the value of urgency accounts for around 87% of

[2] William Vickrey was the first to propose the idea of congestion pricing. In 1952, he suggested to increase subways fares during peak hours.

[3] Carpools are free on HOT lanes.

toll expenses, which implies that commuters are frequently penalized for being a few minutes late.

Dynamic pricing for parking: Parking is a significant contributor to traffic congestion. According to survey statistics based on the central business district of New Haven, cruising for parking accounts for at least 17% of vehicle miles traveled (Huber, 1962). Furthermore, cars are parked roughly 95% of the time on average, reducing available land space (Shoup, 2006).

In an effort to better better balance the demands for parking with the available space, San Francisco experimented with dynamic pricing parking systems, called the *SFpark* program, near downtown area from July 2011 and June 2013. In order to increase the available parking spots and to reduce time cruising for parking, *SFpark* increases parking rates as the available parking spots decrease. It was implemented in 19 garages, and 19 lots based on availability; in total, they accounted for 25% of the city's total parking space. Real-time information about parking space and rates were available on electronic signages and mobile apps.

To evaluate the program, *SFpark* designated 7 areas (Civic Center, Downtown, Fillmore, Fisherman's Wharf, Marina, Mission, and South Embarcadero) as the treated group and two areas (Inner Richmond and Union) as the control group (see Figure 8). Using a difference-in-difference approach with data on transit ridership and parking usage, Krishnamurthy and Ngo (2020) estimates that bus ridership increased by 11% in treated regions, with the effect being greater in the morning and evening peak hours. In addition, the vehicle count - the average number of cars passing over a vehicle detector in a 24-hour period - reduces by 4 vehicles, or 6%, in the treated region. Taken together, these evidence provide that a dynamic parking system encourages some people to use public transportation instead of driving. Hence, dynamic parking pricing could help reduce congestion, increase mobility and improve air quality. In total, the benefits associated with a reduction in emissions and an increase in time savings in two years is about $35 million.

1.3.3 Micromobility

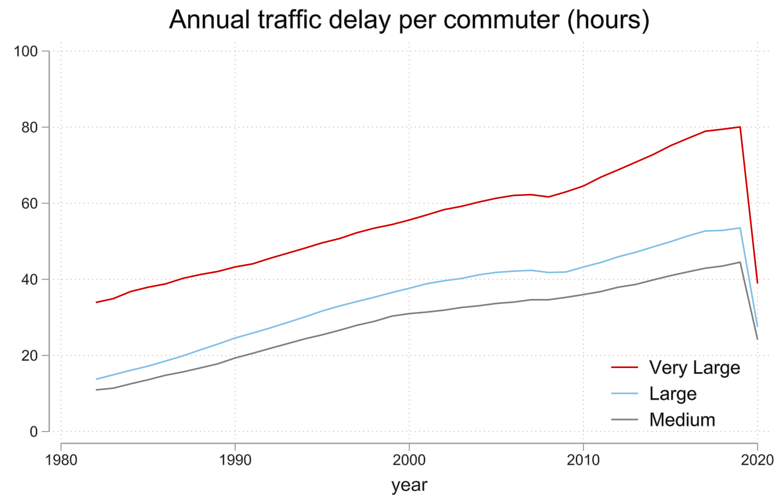
Micromobility, which allows people to rent and share bicycles or scooters, is becoming a popular choice for people living in congested metropolitan areas. Some argue that micromobility could have environmental and economic benefits. Others think it could also pose severe safety concerns. In this part, I address some of the key consequences of implementing dock-based bikesharing.

Dock-based bikesharing: Washington, DC was the first US city to deploy bikesharing in 2010, with 400 bicycles and 49 stations. By the end of 2010, it had more than 100 stations and more than 1000 bicycles (see Figure 9 for locations of stations). Users can pick up and return bicycles at these stations using a kiosk.

A study finds that bikesharing is associated with a significant reduction in traffic congestion. Using spatial panel data on traffic on major roads in Washington DC, Hamilton and Wichman (2018) finds that having a bikesharing station reduces traffic congestion by 4%; the impacts are stronger in more congested regions. These evidence suggests that bikesharing encourage some people to substitue away from driving especially in congested areas. As a result, the estimated economic benefits from reduced travel times and fuel are around $57 per commuter.

1.4 Figures

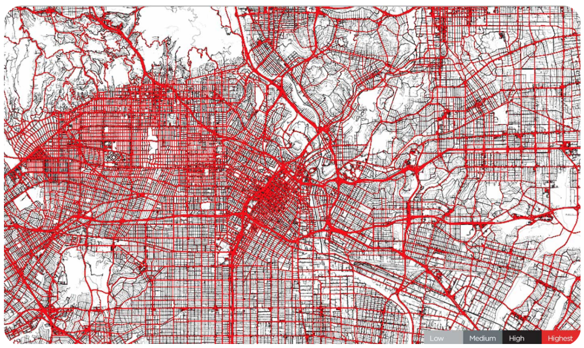
Figure 1: Average Traffic Delay per Commuter



This figure displays the average traffic delay per commuter in the US. Traffic delay is defined as the actual hours spent driving minus the hours spent without congestion. Very large, large and medium urban areas are defined as places with over 3 million population, over 1 million and less than 3 million population, and over 500,000 and less than 1 million population, respectively. Source: Texas A&M Transportation Institute, 2021 Urban Mobility Report, (College Station, TX: 2021), available at http://mobility.tamu.edu

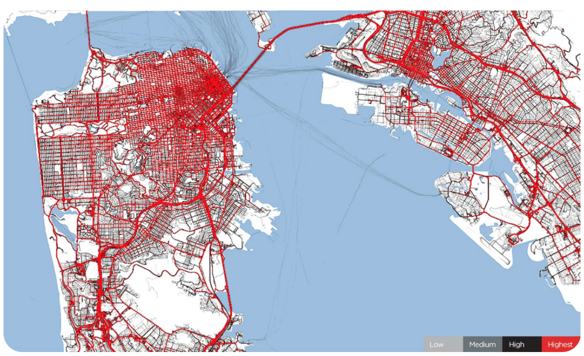
Figure 2: Traffic flows before and after Coronavirus pandemic restrictions

Before Covid: Los Angeles

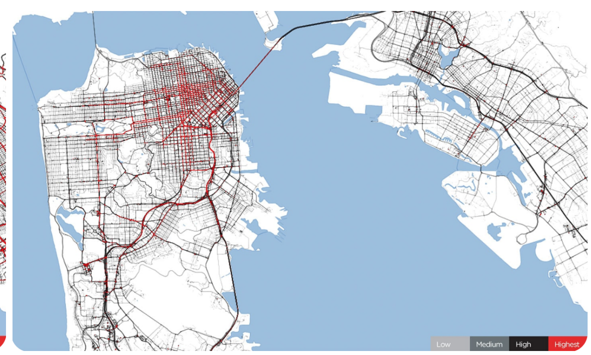


During Covid: Los Angeles

Before Covid: San Francisco

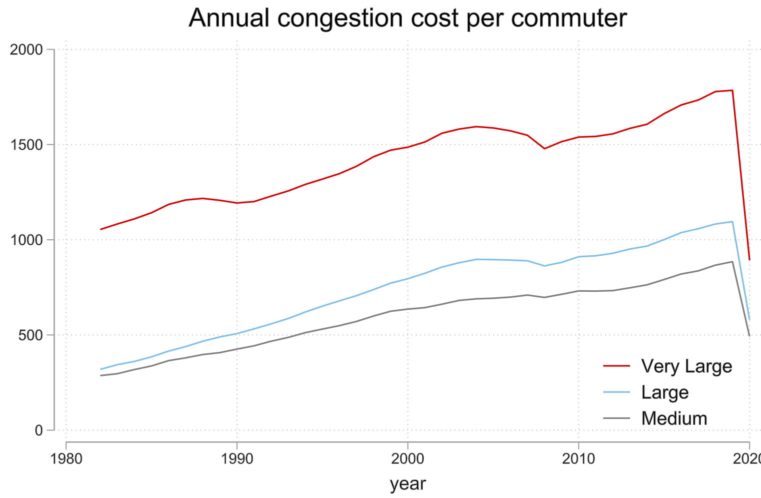


During Covid: San Francisco



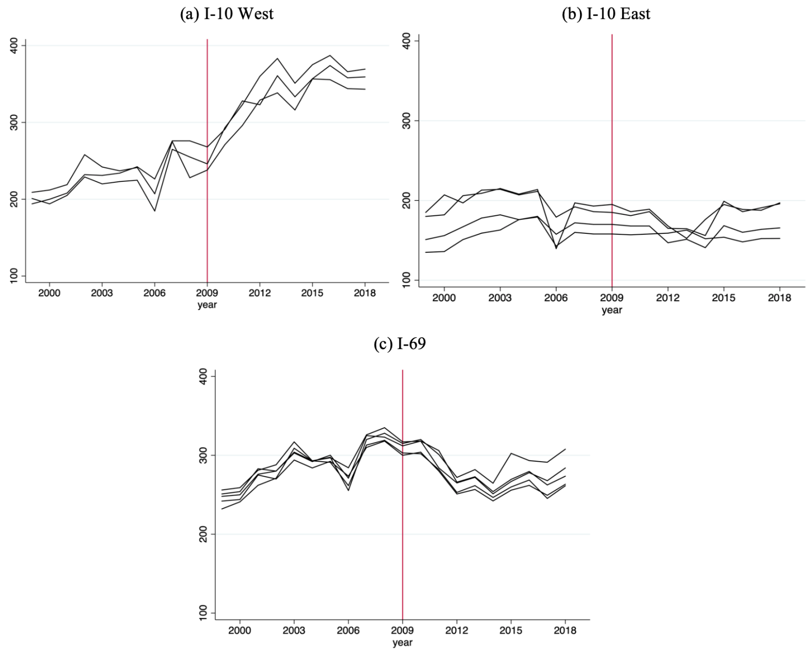
These figures are snapshots of traffic flows in the Los Angeles City and San Francisco on January 24, 2020 (before the restrictions) and April 6, 2020 (after the restrictions). Figures are available from TomTom at https://www.tomtom.com/newsroom/explainers-and-insights/covid-19-traffic/

Figure 3: Costs of Congestion



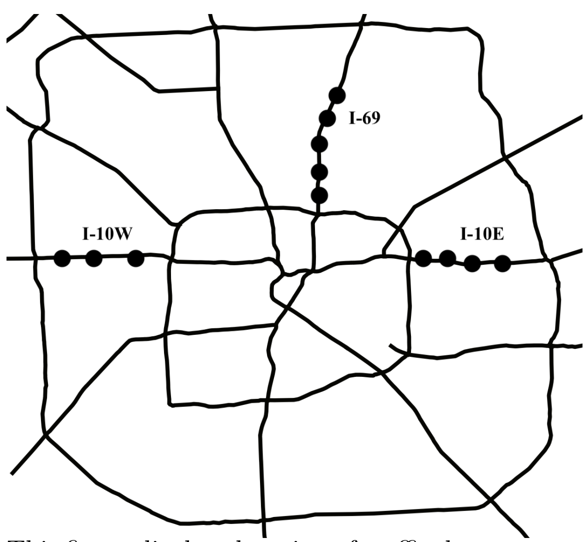
This figure displays a time series plot of fuel and time costs of traffic delay in the US. Very large, large and medium urban areas are defined as places with over 3 million population, over 1 million and less than 3 million population, and over 500,000 and less than 1 million population, respectively. Source: Texas A&M Transportation Institute, 2021 Urban Mobility Report, (College Station, TX: 2021), available at http://mobility.tamu.edu

Figure 4: Average Annual Daily Traffic (in thousands) during 1999-2018



This figure displays Average Annual Daily Traffic (in thousands) during 1999-2018 measured on multiple points on the highways. Unit of observation is point-year. Locations of vehicle detector stations are displayed in Figure 5. I-10 West was expanded in October 2008. I-10 East and I-69 are the non-expanded highways. Source: Texas Department of Transportation.

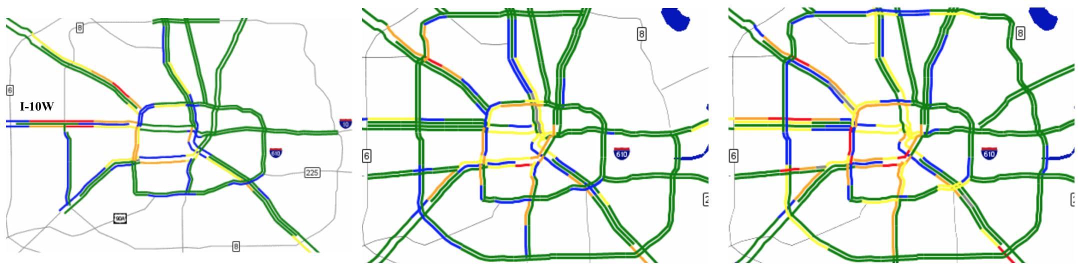
Figure 5: Locations



This figure displays location of traffic detector stations on I-10E, I-10W and I-69.

Figure 6: Speed at 6:00PM-6:15PM

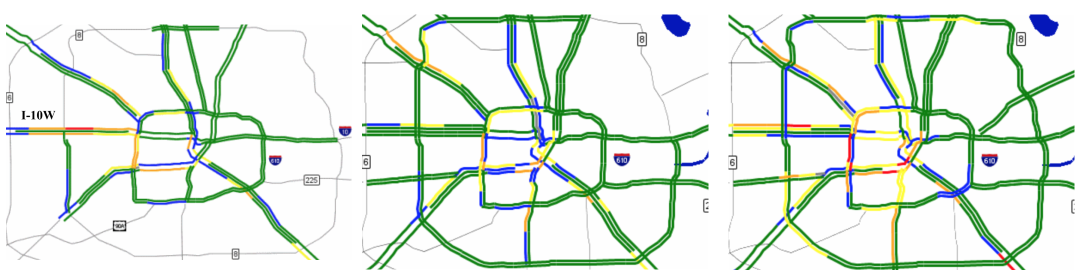
Year 2003 Year 2009 Year 2013



This figure displays vehicle speed maps at 6:00PM-6:15PM in 2003, 2009, and 2013 in Houston. For red, speed ranges from 0 to 19 miles per hour (mph). For orange, it ranges from 20 to 29 mph. For yellow, it ranges from 30 to 39 mph. For blue, it ranges from 40 to 49 mph. For green, it is greater than 50 mph. Grey indicates N/A.

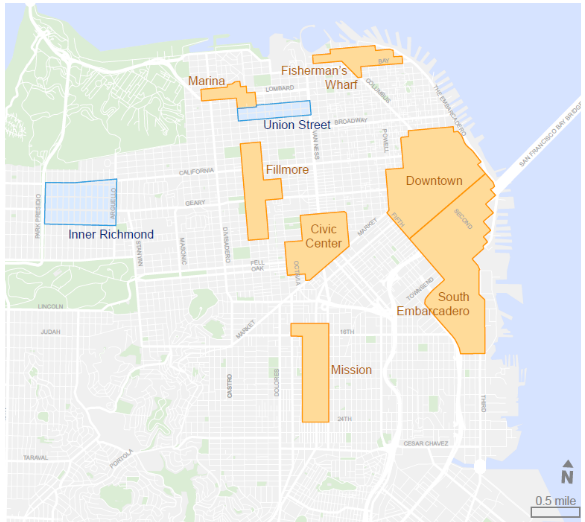
Figure 7: Speed at 6:15PM-6:30PM

Year 2003 Year 2009 Year 2013



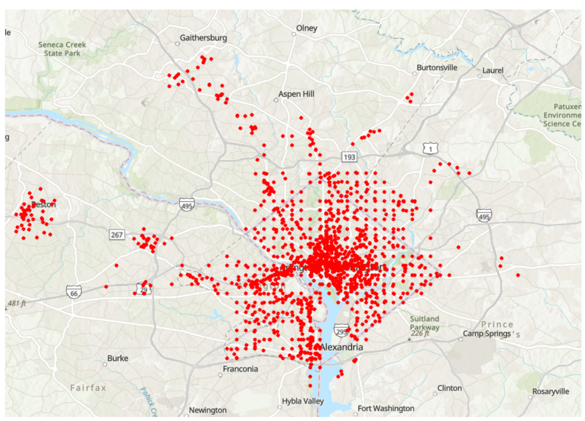
This figure displays vehicle speed maps at 6:15PM-6:30PM in 2003, 2009, and 2013 in Houston. For red, speed ranges from 0 to 19 miles per hour (mph). For orange, it ranges from 20 to 29 mph. For yellow, it ranges from 30 to 39 mph. For blue, it ranges from 40 to 49 mph. For green, it is greater than 50 mph. Grey indicates N/A.

Figure 8: SFpark: Treated and Control Areas



This figure displays locations of 7 treated areas where dynamic parking system was implemented and 2 control areas.

Figure 9: Capital Bikesharing Stations



This figure displays locations of capital bike-sharing stations. Location data is obtained from trip history data available at https://capitalbikeshare.com/system-data.

2 Traffic Effects of Congestion Pricing

2.1 Introduction

An imbalance between the demand for driving and the supply of roads creates traffic congestion. As discussed in chapter 1, a growing number of cities have implemented highway congestion pricing - which varies toll costs based on congestion - to reduce demand for driving.

This chapter provides reduced form estimates to determine the degree to which traffic diverts from toll to non-toll lanes using a policy change in the Los Angeles area in which a subset of non-toll lanes on Interstates 10 and 110 were converted to toll lanes with dynamic pricing. The research design compares changes in traffic in Los Angeles County (the treated region) to changes in traffic in other counties (the control regions) using a difference-in-differences approach. The main empirical challenge is identifying a valid set of control locations with characteristics that resemble Los Angeles, which is more congested than most regions. Therefore, I implement propensity score weighting to create a synthetic sample in which the levels of pre-treatment traffic outcomes are balanced between the two groups.

My reduced-form results show that, for toll lanes, traffic density decreases significantly by 5 cars/mile and speed increases significantly by 2 miles per hour during morning rush hour. These changes are persistent and provide evidence that congestion pricing mitigates congestion on toll lanes in the medium-run. On both non-toll lanes parallel to toll lanes and the remaining non-toll lanes in other locations, traffic density increases significantly: by 7 cars/mile in the first year of implementation. A more than one-to-one increase in traffic in non-toll lanes relative to toll lanes suggests that people re-route and drive longer distances. However, this spillover effect disappears over time in the majority of non-toll lanes. These results provide supporting evidence that in the short run, drivers avoid toll costs by mostly switching from toll lanes to non-toll lanes; over time, changes in the spatial distribution of residential and work locations induce further adjustments in driving routes.

2.2 Background and Data

This section contains information about the ExpressLanes Program, which began implementing congestion pricing on I-10 and I-110 in November 2012. It also explains real-time traffic data from the Caltrans Performance Measurement System (PeMS), which is used for providing reduced-form estimates of traffic effects.

2.2.1 Background

Los Angeles-Long Beach-Anaheim metropolitan statistical area has the second highest population density in the country, with 2,723 people per square mile (US Census Bureau, 2020a). The area also has the highest yearly traffic delay per commuter in the US (Schrank et al., 2019).[4] Despite significant investments in transportation infrastructure throughout its history, the road network in Los Angeles is nearly at capacity (Schroeder et al., 2015).

The empirical setting of this chapter is the Los Angeles Congestion Reduction Demonstration ExpressLanes program, which was led by the Los Angeles County Metropolitan Transportation Authority (MTA) in collaboration with the California Department of Transportation (CalTrans). One of the key CRD projects was the ExpressLanes program, which implemented congestion pricing.[5] High-Occupancy Vehicle (HOV) lanes on Interstates 10 and 110 in the Los Angeles downtown area were converted to High-Occupancy Toll (HOT) lanes in Oct, 2012 and Feb, 2013, respectively. Also, an additional HOT lane was added to I-10 between the I-710 and I-605 interchanges. The total implementation costs, including planning, design, acquisition, and construction, were \$106.76 million, which were funded by the Department of Transportation (Schroeder et al., 2015).[6]

HOT lanes are made up of toll segments with entry and exit sections where vehicles can enter or exit after checking the toll rates displayed on electronic signs. See Figure 10 for locations of entries and exits. To use HOT lanes, all drivers must register their FasTrak transponders. Carpools are permitted to use the lanes toll-free at all times by indicating the number of passengers in the vehicle using the FasTrak transponders.[7] Single-occupancy drivers are automatically charged via a Fastrak transponder.

Tolls ranged from \$0.40 to \$1.40 per mile during the one-year demonstration period; in 2019, they ranged from \$0.25 to \$2.10 per mile. Tolls are adjusted every 5 minutes based on traffic conditions. Once the speed drops below 45 mph, HOT lanes revert to HOV lanes, and solo drivers are not permitted to enter. Although I-10/I-110 congestion pricing began as a demonstration pilot program, in September

[4] The yearly traffic delay per commuter is the extra time spent during congestion during the year divided by the number of people who commute in private vehicles.

[5] Other programs included increasing the frequencies of Metro Rapid service in the I-10 El-Monte Busway and purchasing new buses.

[6] The project was not a public-private partnership.

[7] Depending on the freeway and the time of day, different carpooling conditions apply. For instance, on I-10, vehicles with three or more passengers can always use the toll-free HOT lanes, whereas on I-110, vehicles with two or more passengers can always use the toll-free HOT lanes.

2014, the Los Angeles MTA was granted the authority to operate congestion pricing indefinitely.

Total toll revenues including violation fines were $70 million in fiscal year 2018 (Holliday, 2018). They are first used to pay for maintenance and operation, and the remainder must be reinvested into transportation projects for the I-10 and I-110 corridors (Schroeder et al., 2015).[8]

2.2.2 Data

I obtain real-time hourly traffic data from the Caltrans Performance Measurement System (PeMS) from 2006 to 2019. Traffic speed and flow are measured by more than 30,000 vehicle detector stations (VDS) deployed along interstates, state routes and US highways in major metropolitan areas in California.[9]. Figure 11 depicts the locations of all VDS in California.

Real-time traffic data is available in bulk from the *Dataclearing house* on the PeMS webpage. It contains information about the coordinates of VDS, the freeway number, lane directions, flow, speed, measurement time, and lane types (e.g., non-toll main lanes, HOV lanes, on-ramps or off-ramps). I discard imputed data and observations from the weekends.[10] In addition, I only keep observations from non-toll main lanes, HOV lanes, and HOT lanes. Then, I aggregate hourly speed and flow to monthly average hourly speed and flow. Units of observations are Vehicle Detector Stations.

2.3 Reduced Form Evidence

This section provides evidence on the effects of converting High-Occupancy Vehicle lanes to High-Occupancy Toll (HOT) lanes on traffic in the road network in Los Angeles County using a difference-in-differences. To observe direct and indirect spill-over effects, I estimate traffic effects separately for HOT lanes, non-toll lanes parallel to HOT lanes, and the remainder of the non-toll lanes. To find a valid control group that resembles pre-treatment characteristics of Los Angeles, I apply propensity score weighting to construct a synthetic sample. The remainder of this section defines the treatment, describes the data, and discusses empirical strategy.

[8] For instance, $3.8 million were used for a downtown LA bike share program and $2 million for improving a south LA Metro station.

[9] Anderson and L. W. Davis (2018) and Bento et al. (2020) use traffic data from the Caltrans Performance Measurement System (PeMS) as well

[10] When a detector station fails, it uses traffic data from nearby lanes or lanes with similar historical traffic patterns to impute data.

2.3.1 Definition of Treatment

The treatment is defined as a combination of dynamic toll pricing and better enforcement of carpool regulations. Statistics and anecdotal evidence show that a significant number of single-occupancy vehicles violate carpool regulations.[11] However, since the conversion, it has become more expensive to violate carpool regulations because all vehicles must register FasTrak transponders in order to use HOT lanes. Carpools must also indicate the number of passengers in their vehicles by switching on their FasTrak transponders.

As a result, the treatment is expected to affect traffic on HOT lanes via two offsetting mechanisms. First, improved enforcement is expected to reduce the number of vehicles using HOT lanes by making it more expensive for single-occupancy drivers to violate carpool regulations. Second, dynamic toll pricing, which allows single-occupancy toll-paying drivers to use HOT lanes, is expected to increase the number of vehicles using HOT lanes. Therefore, the observed treatment effects on HOT lanes would be the net effects of these two mechanisms.

Furthermore, the number of vehicles using the remaining non-toll lanes would increase if some single-occupancy drivers switch from toll lanes to non-toll lanes to avoid tolls. The size and duration of the spill-over effects would be determined by potential changes in driving routes, residences and workplaces.

2.3.2 Estimation Strategy

Using observations weighted by inverse propensity scores, I estimate an event study difference-in-differences specification,

$$\mathrm{Y}_{scht} = \sum_{\tau} \beta_\tau \times 1[\mathrm{year}_t = \tau] \times 1[\mathrm{LA}]_c + \gamma_t + \gamma_s + \gamma_h + \epsilon_{scht}, \tag{1}$$

where Y_{scht} is the monthly average speed/hour or cars/mile measured by the Vehicle Detector Station (VDS) s in county c in monthly date t at hour h. $1[\text{year in } t = \tau] = 1$ if the year in monthly date t equals $\tau \in \{2006, 2007, ...,2019\}$, and $1[\mathrm{LA}]_s = 1$ if a station is located in LA County. γ_t are time-fixed effects that capture the effects of macroeconomic shocks common to all areas. VDS fixed-effects γ_s capture any station-level unobservable effects that are time-invariant, such as road conditions or geographic characteristics. γ_h captures hourly trends in traffic. Standard errors are

[11] During the first half of 1989, the California Highway Patrol issued 35,332 tickets for HOV violations (B. McFadden and Innes, 1990). In general, the goal is to keep the violation rate below 10%.

clustered at the county level, which is the level of treatment. I omit October 2012, which is a month prior to the implementation of congestion pricing, as the baseline period.

The identification of β_τ, which captures the average annual effects of converting HOV lanes to HOT lanes, relies on the parallel trend assumption *conditional on observables*. That is, areas with the similar levels of congestion would, on average, experience similar traffic growth over time.

I estimate an additional difference-in-differences specification, in which the post period is divided into two time periods: the first two years and the subsequent years until 2019. An abrupt increase in the number of lane closures related to HOT lane conversion and construction that started in late 2012, as shown in figure 15, motivates the following specification:

$$\begin{aligned} \mathrm{Y}_{scht} = {} & \beta_{\tau 1} \times 1[\mathrm{year}_t \in \{2013, 2014\}] \times 1[\mathrm{LA}]_c + \beta_{\tau 2} \times 1[\mathrm{year}_t \geq 2014] \times 1[\mathrm{LA}]_c \\ & + \gamma_t + \gamma_s + \gamma_h + \epsilon_{scht}, \end{aligned} \tag{2}$$

where $1[\mathrm{year}_t \in \{2013, 2014\}]$ is a dummy that takes a value of 1 for observations in the years 2013 or 2014, and $1[\mathrm{year}_t \geq 2014]$ is a dummy that takes a value of 1 for observations in the subsequent years until 2019. Therefore, $\beta_{\tau 1}$ and $\beta_{\tau 2}$ capture the average annual effects of congestion pricing during the first two years and for the following years, respectively.

2.3.3 Construction of Treatment Group and Control Group

I define the treatment group as major highways in Los Angeles County that could be affected by the conversion of HOV lanes to HOT lanes on I-10 and I-110. An ideal empirical strategy to identify traffic effects is to compare the traffic outcomes for the treatment group to the traffic outcomes in regions with similar levels of congestion. However, finding a valid set of control groups is difficult because Los Angeles is more congested than most regions. To address these issues, I build my sample in two steps.

In the first step, I choose counties with High-Occupancy Vehicle (HOV) lanes as my control counties: Sacramento, San Bernardino, Orange, San Diego, and Riverside County. To mitigate potential spillover effects, I exclude San Bernardino and Orange County, which are directly adjacent to the Los Angeles County.

Then, I estimate propensity scores that are defined as the probability of receiving treatment as a function of pre-treatment covariates (Hirano et al., 2003).[12]

[12] The methodology is similar to Deryugina et al. (2018), which applies propensity score weighting

More specifically, I estimate propensity scores (PS) for each Vehicle Detector Station (VDS) as a function of traffic density and speed during pre-treatment and the census tract-level population in 1990, X: PS $= Prob(Treated = 1|X)$. The weights are $1/PS$ and $1/(1 - PS)$ for the treated units and control units, respectively.

In Panel A of Table 1, I first compare pre-treatment traffic on non-toll lanes between the treated and control regions in the unweighted sample. Panel A1 provides summary statistics throughout the day, and panel A2 only during rush hours. In Panel A1, the average hourly speed is lower by 4 miles per hour, and the average hourly flow is higher by 1,112 cars/mile throughout the day in the treated region than in the control region. In Panel A2, the average hourly speed is lower by 8 miles per hour, and the average hourly flow is higher by 1,145 cars/mile during rush hours in the treated region than in the control region. All of the differences in speed and flow between the two groups are statistically significant as well.

Panel B of Table 1 examines whether propensity score weighting successfully balances pre-treatment traffic on non-toll lanes between the treated and control regions. The number of observations is half that of the unweighted sample, but still above half a million. This is because I impose a common support restriction, which removes observations with propensity scores that fall outside the support of the other group to ensure that the treatment and control groups have a sufficient overlap in propensity scores. The table shows that when the observations are weighted by the inverse propensity scores, the differences in speed and flow are no longer statistically significant throughout the day and during rush hours.

Similarly, Panel A of Table 2 uses an unweighted sample to compare traffic on HOV lanes in the treated and control regions. It suggests that speed on HOV lanes in the treated region is 1 mph slower throughout the day and 4 mph slower during rush hours than in the control region, with the differences being statistically significant. However, Panel B of Table 2 shows that the differences in speed between the treated and control regions are no longer statistically significant in the weighted sample. Therefore, these results show that propensity score weighting effectively balances pre-treatment traffic outcomes.

2.3.4 Main Results

To identify direct and indirect spillover effects separately, I estimate equation 2 by different types of lanes: HOT lanes of I-10 and I-110, non-toll lanes parallel to HOT

to evaluate the effects of Hurricane Katrina on victims compared to the control group.

lanes of I-10 and I-110, and the remainder of non-toll lanes. In the baseline, I use a sample weighted by propensity scores with a common support restriction.

HOT lanes: Figure 12 and *Panel A* of Table 3 present the estimation results for equations 1 and 2 for HOT lanes of I-10 and I-110 at different times of day: morning and evening rush hours. *Panel A* of Table 3 shows that speed on HOT lanes increases significantly by about 2 mph as density decreases by 3 cars/mile during morning rush hour from 2014 to 2019. The changes in traffic appear to last until the end of the sample in Figure 12, providing evidence that converting HOV lanes to HOT lanes reduces the demand for travel and increases speed. Note that speed initially falls very sharply in 2012 and 2013. This is driven by an abrupt increase in the number of lane closures associated with HOT lane conversion and construction beginning in 2012, as illustrated in Figure 15. I do not observe any significant changes in traffic during evening rush hour as the effects are imprecisely estimated.

Non-toll lanes parallel to HOT lanes: Figure 13 and *Panel B* of Table 3 present estimation results for non-toll lanes parallel to HOT lanes on I-10 and I-110 during morning and evening rush hours. Results show that between 2012 and 2013, rush hour density increases significantly by 3-5 cars/mile, reducing speed by 4-6 mph. However, between 2014 and 2019, rush hour density falls by 3-9 cars/mile, increasing speed by 1-2 mph. These findings imply that drivers initially switch from toll lanes to non-toll lanes to avoid tolls. Over time, however, they avoid non-toll lanes parallel to HOT lanes.

The rest of non-toll main lanes: Lastly, I estimate the effects on the remaining non-toll main lanes in Figure 14 and *Panel C* of Table 3. They show that density in the evening rush hour increases significantly by about 7 cars/mile, decreasing speed by 3 mph from 2012 to 2013. A more than one-to-one increase in traffic on non-toll lanes relative HOT lanes in the first two years suggests that people re-route and drive longer distances. However, between 2014 and 2019, I do not find statistically significant effects on traffic, implying that the medium-run changes in driving routes are different from the short-run changes.

Without a common support restriction: Appendix A presents associations between congestion pricing and traffic outcomes using the full sample without imposing a common support restriction.

Figure 16 and *Panel A* of Table 4 suggest that the density on toll lanes decreases significantly by 3 cars/mile and speed increase significantly by 2 mph from 2014 to 2019, implying a persistent decrease in demand for travel on toll lanes.

Figure 17 and *Panel B* of Table 4 show that rush hour density on non-toll lanes parallel to HOT lanes increases significantly by 3-5 cars/mile and speed decreases significantly by 4-6 mph from 2012 to 2013. However, this pattern reverses during 2014 and 2019: a decrease in density by 3-8 cars/mile and an increase in speed by 1-2 mph.

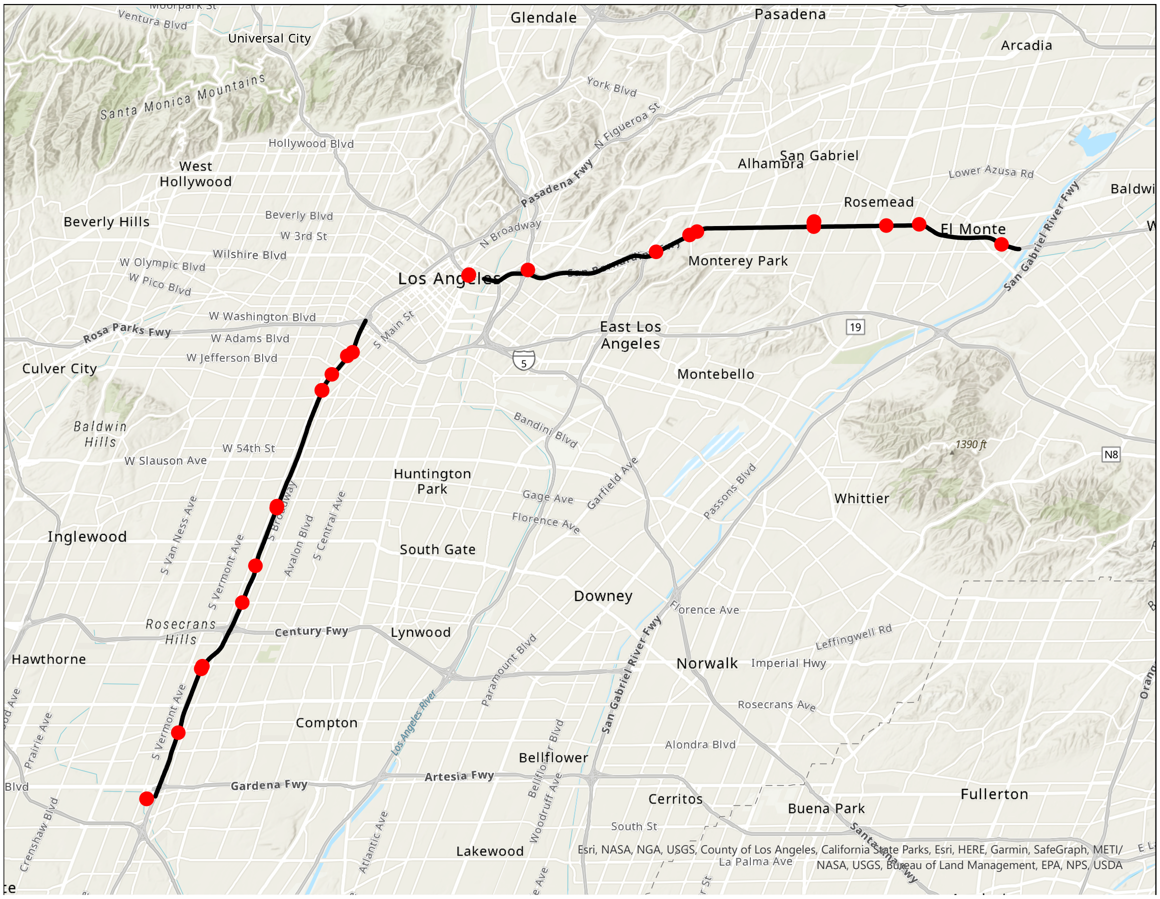
Figure 18 and *Panel C* of Table 4 display the size and duration of spillover effects on the remaining non-toll lanes. In the evening rush hour, density increases by 8 cars/mile and speed decreases by 3 mph in the first two years; but estimates are not statistically significant in the latter years.

These findings are consistent with the results obtained using a common support restriction, supporting the argument that the short-run and medium-run adjustments are different.

Key takeaways: My empirical findings emphasize the two following outcomes. First, converting HOV lanes to HOT lanes significantly and persistently reduces the demand for travel and increases speed. Second, demand for travel on the rest of the non-toll lanes increases significantly during the first two years of implementation, indicating that drivers may shift from HOT lanes to non-toll lanes to avoid tolls. However, these spillover effects disappear from the majority of the non-toll lanes over time. This highlights the possibility of further changes in driving routes caused by changes in residences and workplaces away from toll lanes.

2.4 Table and Figures

Figure 10: Entries and Exits of HOT lanes



High-Occupancy Toll lanes on the I-10 and I-110 in Los Angeles are indicated on the map by black lines. Red dots represent entrances and exits.

Figure 11: Locations of Vehicle Detector Stations in California

The locations of all Vehicle Detector Stations (VDS) in California are shown on the map. Data from these VDS are used by the Caltrans Performance Measurement System (PeMS) to compute traffic performance measures like speed and flow.

Table 1: Summary Statistics of Non-Toll Main Lanes

Panel A: Unweighted sample

Panel A1: all hours	Treated	Control	Difference (s.e)
speed/hour	60.70	64.64	-3.94***(0.50)
flow/hour	4056.70	2944.15	1112.55***(141.50)
N	1,790,272	1,734,645	

Panel A2: rush hours	Treated	Control	Difference (s.e)
speed/hour	53.81	61.45	-7.64*** (0.58)
flow/hour	5552.15	4407.63	1144.52** (269.57)
N	596,622	578,524	

Panel B: Weighted sample

Panel B1: all hours	Treated	Control	Difference (s.e)
speed/hour	63.15	62.98	0.17 (1.22)
flow/hour	3329.68	3515.36	-185.68 (126.88)
N	1,067,356	1,060,708	

Panel B2: rush hours	Treated	Control	Difference (s.e)
speed/hour	57.96	57.27	0.69 (3.18)
flow/hour	4908.47	5072.34	-163.87 (149.01)
N	355,659	353,582	

Panel A shows the unweighted averages of the monthly average hourly speed and the monthly average hourly flow on non-toll main lanes in the treatment and control groups, and the differences with the standard errors in the parenthesis: panel A1 is for all hours of the day and panel A2 is for rush hours only (6AM-10AM or 3PM-7PM). Panel B shows the weighted averages of the monthly average hourly speed and the monthly average hourly flow in the treatment and control groups using the propensity score weighting and a common support restriction. The robust standard errors that are clustered by county. Speed/hour is defined as total miles driven per hour and flow is defined as the total number of cars that pass a Vehicle Detector Station. Observations are at the Vehicle Detector Station-level.

Table 2: Summary Statistics of High-Occupancy Vehicle lanes

Panel A: unweighted sample

Panel A1: all hours	Treated	Control	Difference (s.e)
Speed/hour	60.63	62.05	-1.42* (0.58)
flow/hour	527.50	495.52	31.98 (27.30)
N	880,477	266,176	

Panel A2: rush hours	Treated	Control	Difference (s.e)
Speed/hour	55.62	59.22	-3.60* (1.26)
flow/hour	858.78	704.76	154.02 (68.03)
N	293,436	88,737	

Panel B: weighted sample

Panel B1: all hours	Treated	Control	Difference (s.e)
Speed/hour	60.98	62.91	-1.93 (1.31)
flow/hour	518.84	419.23	99.61 (64.47)
N	1,067,356	1,060,708	

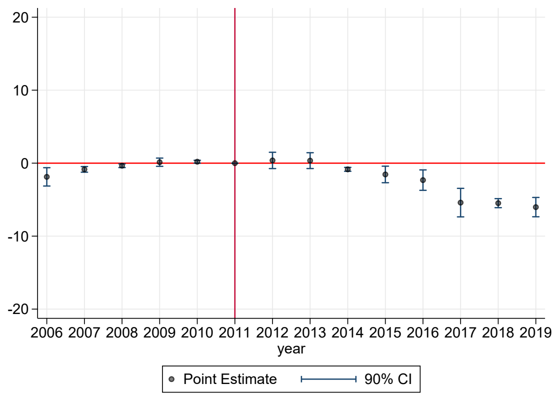
Panel B2: rush hours	Treated	Control	Difference (s.e)
Speed/hour	56.79	56.96	-0.17 (2.32)
flow/hour	783.86	816.87	-33.01 (112.64)
N	355,659	353,582	

Panel A shows the unweighted averages of the monthly average hourly speed and the monthly average hourly flow on High-Occupancy Vehicle lanes in the treatment and control groups, and the differences with the standard errors in the parenthesis: panel A1 is for all hours of the day and panel A2 is for rush hours only (6AM-10AM or 3PM-7PM). Panel B shows the weighted averages of the monthly average hourly speed and the monthly average hourly flow in the treatment and control groups using the propensity score weighting and a common support restriction. The robust standard errors that are clustered by county. Speed/hour is defined as total miles driven per hour and flow is defined as the total number of cars that pass a Vehicle Detector Station. Observations are at the Vehicle Detector Station-level.

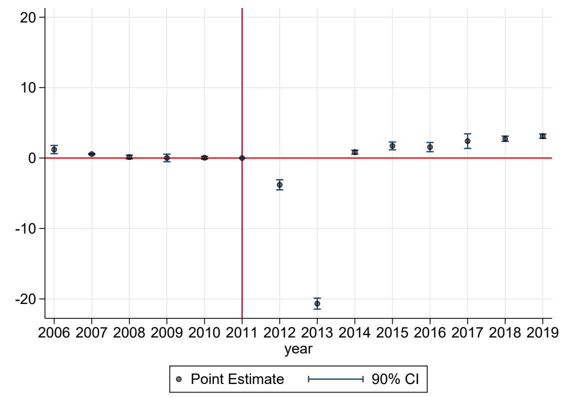
Figure 12: High-Occupancy Toll lanes **with** a common support restriction

Panel A: Morning rush hours

Panel A1: density (cars/mile)

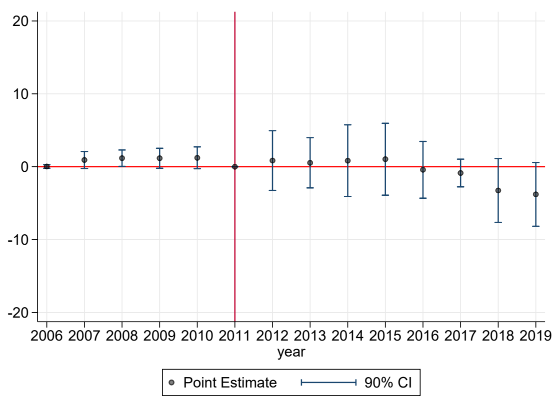


Panel A2: speed (miles/hour)

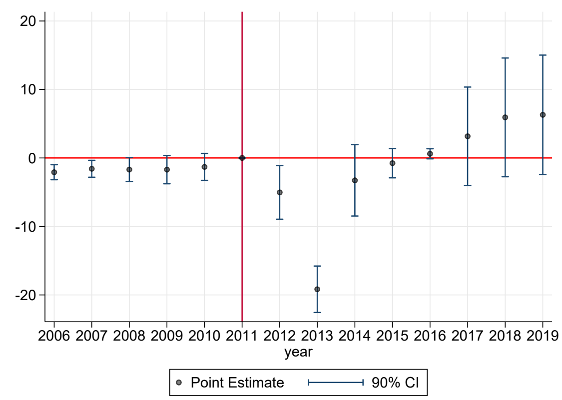


Panel B: Evening rush hours

Panel B1: density (cars/mile)



Panel B2: Speed (miles/hour)

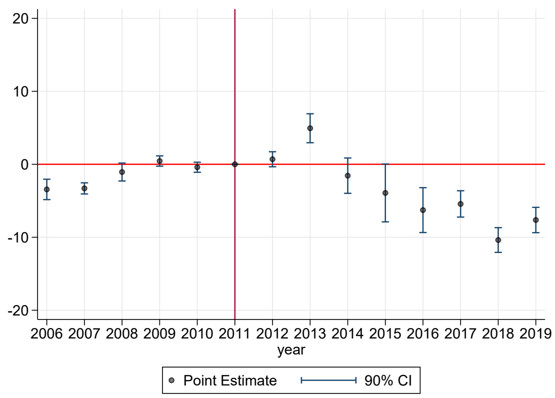


The figures above estimate Equation 1 by types of lanes and times of the day. Panels A shows associations between congestion pricing and traffic outcomes on High-Occupancy Toll lanes during morning rush hours from 6AM to 10AM : Panel A1 for density and Panel A2 for speed. Panels B shows associations between congestion pricing and traffic outcomes on High-Occupancy Toll lanes during evening rush hours from 3PM to 7PM. Observations are weighted by the inverse propensity scores, which are estimated as a function of speed, density, and tract-level population in 2000. I impose a common support restriction, removing observations with propensity scores that fall outside the support of the treatment and control groups. Speed is defined as the total miles driven per hour, and density is defined as the total number of cars occupying a mile on the road. Vertical lines represent 90 percent confidence intervals. Robust standard errors are clustered by county.

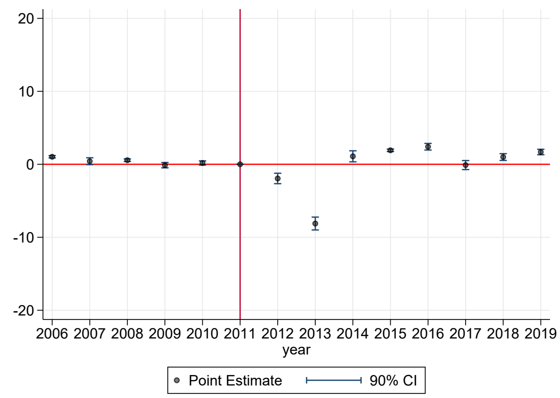
Figure 13: Non-toll main lanes parallel to High-Occupancy Toll lanes **with** a common support restriction

Panel A: Morning rush hours

Panel A1: density (cars/mile)

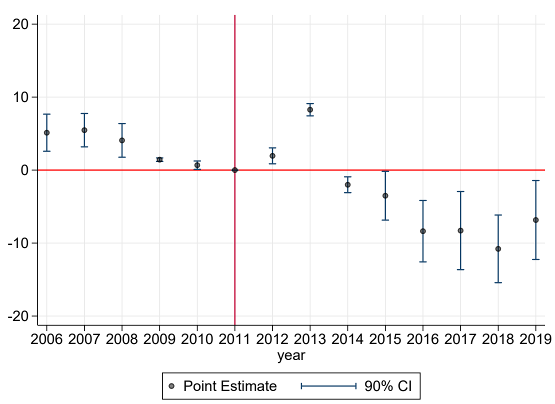


Panel A2: speed (miles/hour)

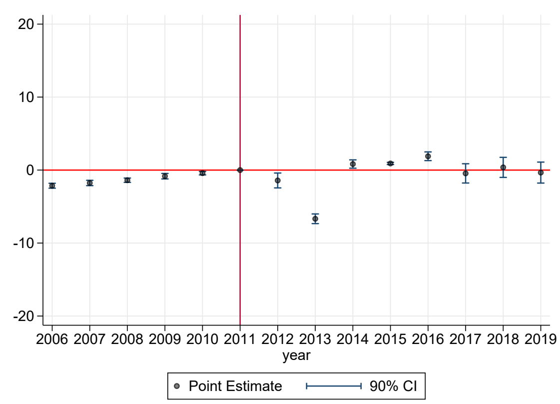


Panel B: Evening rush hours

Panel B1: density (cars/mile)



Panel B2: speed (miles/hour)

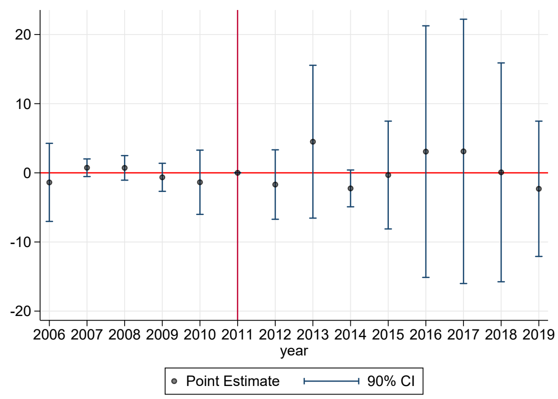


The figures above estimate Equation 1 by types of lanes and times of the day. Panels A shows associations between congestion pricing and traffic outcomes on non-toll main lanes parallel to High-Occupancy Toll lanes during morning rush hours from 6AM to 10AM: Panel A1 for density and Panel A2 for speed. Panels B shows associations between congestion pricing and traffic outcomes on non-toll main lanes parallel to High-Occupancy Toll lanes during evening rush hours from 3PM to 7PM. Observations are weighted by the inverse propensity scores, which are estimated as a function of speed, density, and tract-level population in 2000. I impose a common support restriction, removing observations with propensity scores that fall outside the support of the treatment and control groups. Speed is defined as the total miles driven per hour, and density is defined as the total number of cars occupying a mile on the road. Vertical lines represent 90 percent confidence intervals. Robust standard errors are clustered by county.

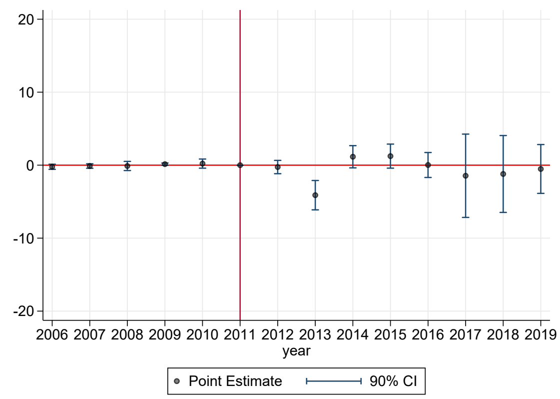
Figure 14: The rest of non-toll lanes **with** a common support restriction

Panel A: Morning rush hours

Panel A1: density (cars/mile)

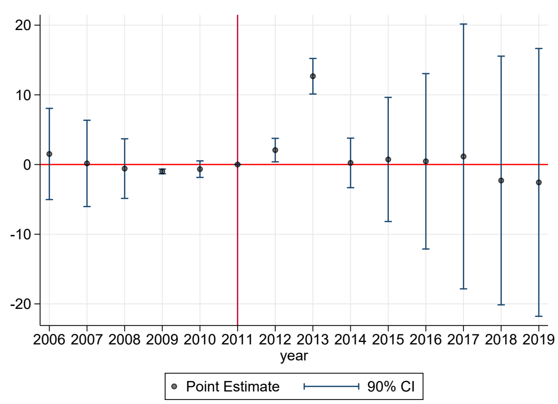


Panel A2: speed (miles/hour)

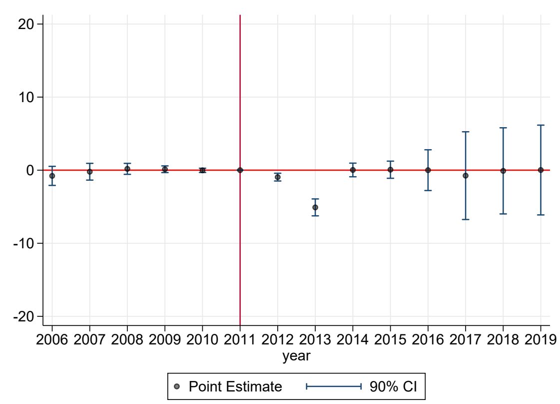


Panel B: Evening rush hours

Panel B1: density (cars/mile)



Panel B2: speed (miles/hour)



The figures above estimate Equation 1 by types of lanes and times of the day. Panels A shows associations between congestion pricing and traffic outcomes on the remaining non-toll lanes during morning rush hours from 6AM to 10AM: Panel A1 for density and Panel A2 for speed. Panels B shows associations between congestion pricing and traffic outcomes on on the remaining non-toll lanes during evening rush hours from 3PM to 7PM. Observations are weighted by the inverse propensity scores, which are estimated as a function of speed, density, and tract-level population in 2000. I impose a common support restriction, removing observations with propensity scores that fall outside the support of the treatment and control groups. Speed is defined as the total miles driven per hour, and density is defined as the total number of cars occupying a mile on the road. Vertical lines represent 90 percent confidence intervals. Robust standard errors are clustered by county.

Table 3: Difference-in-differences estimates **with** the common support restriction

Panel A: HOT lanes

	morning		evening	
	density	speed	density	speed
$1[LA]_c \times 1[Year_t \in \{2012, 2013\}]$	0.74	-15.31***	-0.25	-12.36***
	(0.68)	(0.41)	(1.28)	(1.12)
$1[LA]_c \times 1[Year_t \geq 2014]$	-2.61**	1.54***	-1.29	2.11*
	(0.58)	(0.26)	(0.73)	(0.73)
Observations	25,717	25,717	9,421	9,421
R^2	0.81	0.70	0.91	0.64

Panel B: Non-toll lanes parallel to HOT lanes

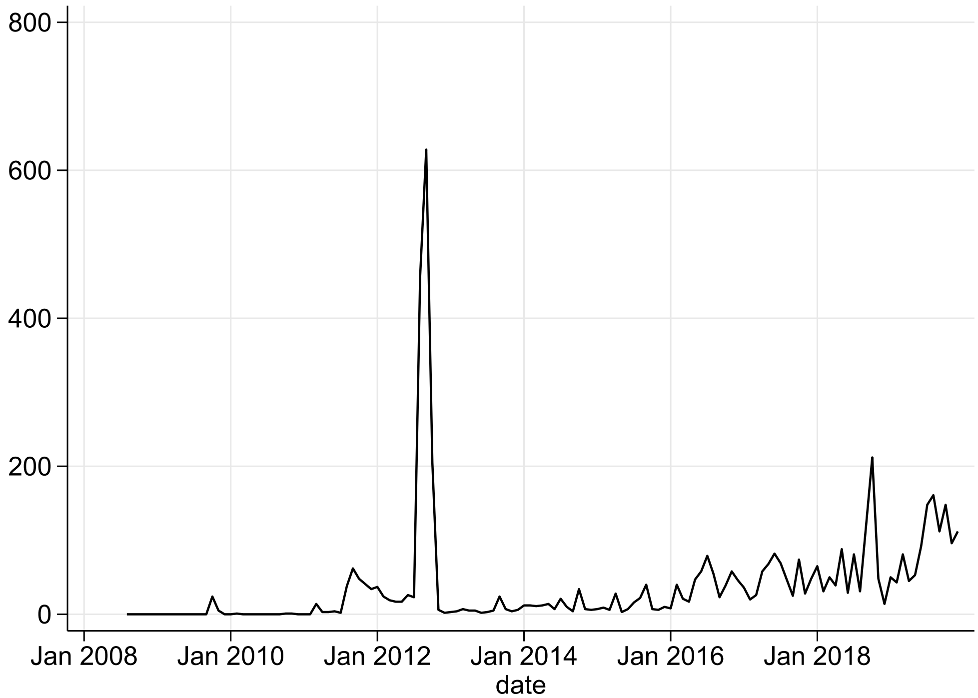
	morning		evening	
	density	speed	density	speed
$1[LA]_c \times 1[Year_t \in \{2012, 2013\}]$	4.61***	-6.10***	2.94**	-3.54***
	(0.57)	(0.32)	(0.64)	(0.41)
$1[LA]_c \times 1[Year_t \geq 2014]$	-3.22**	1.27***	-8.50***	1.96***
	(0.94)	(0.19)	(0.83)	(0.31)
Observations	261,179	261,179	296,212	296,212
R^2	0.87	0.66	0.84	0.70

Panel C: The rest of non-toll lanes

	morning		evening	
	density	speed	density	speed
$1[LA]_c \times 1[Year_t \in \{2012, 2013\}]$	1.67	-2.15**	7.33***	-2.86***
	(2.68)	(0.55)	(0.57)	(0.39)
$1[LA]_c \times 1[Year_t \geq 2014]$	0.42	-0.04	-0.26	0.01
	(5.45)	(0.89)	(4.27)	(1.37)
Observations	568,766	568,766	555,730	555,730
R^2	0.76	0.69	0.79	0.75

Panels A, B, and C estimate associations between congestion pricing and traffic outcomes using Equation 1 for HOT lanes, non-toll main lanes parallel to HOT lanes, and the remaining non-toll main lanes, respectively. Each column represents a separate regression. The dependent variable is either speed (mph) or density (cars/mile). All the regressions include station, hour, and monthly date fixed effects. Morning and evening rush hours are from 6AM to 10AM and from 3PM to 7PM, respectively. Observations are weighted by the inverse propensity scores, which are estimated as a function of speed, density, and tract-level population in 2000. I impose a common support restriction. Vertical lines represent 90 percent confidence intervals. Robust standard errors are clustered by county.

Figure 15: The number of High-Occupancy Vehicle lane closures on I-10 and I-110 in LA by the beginning month



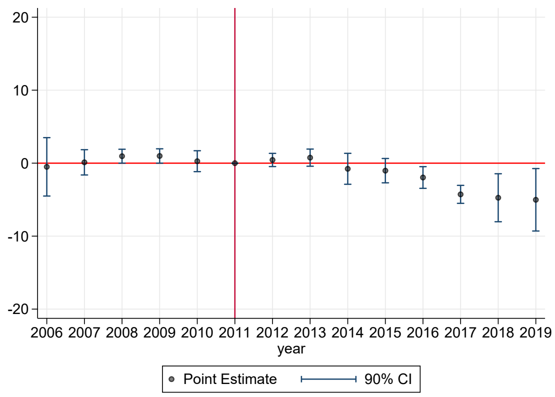
The figure displays the monthly time series plot of the number of High-Occupancy Vehicle lane closures on I-10 and I-110 in Los Angeles (the segments on which HOV lanes are converted to HOT lanes), based on data from the Caltrans Lane Closure System (LCS).

2.5 Appendix: Robustness Checks for Traffic Effects

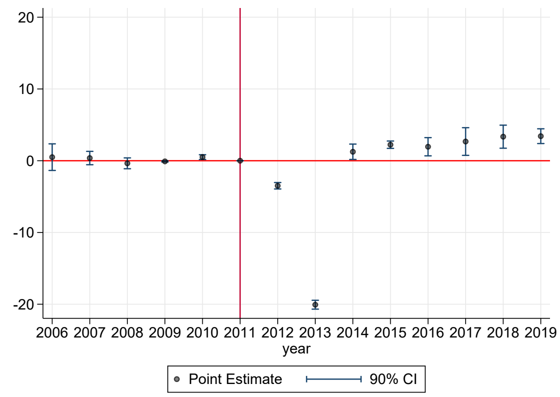
Figure 16: High-Occupancy Toll lanes **without** a common support restriction

Panel A: Morning rush hours

Panel A1: density (cars/mile)

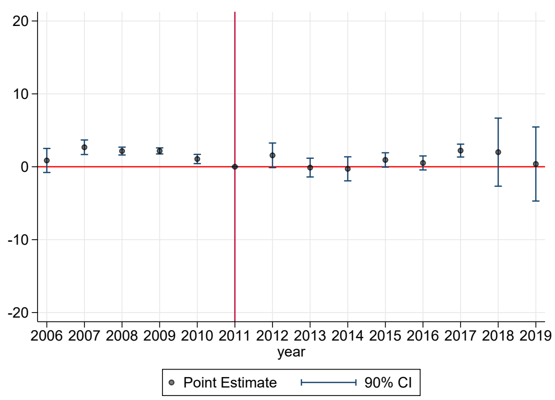


Panel A2: speed (miles/hour)

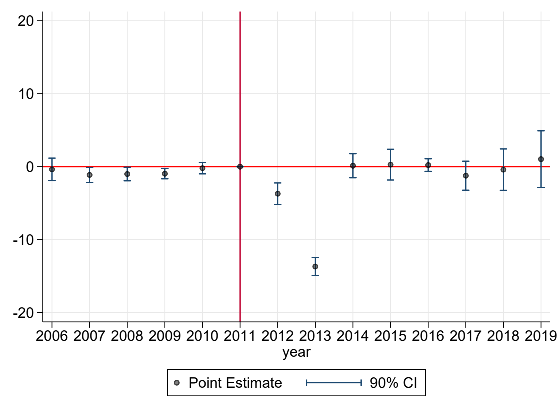


Panel B: Evening rush hours

Panel B1: density (cars/mile)



Panel B2: Speed (miles/hour)

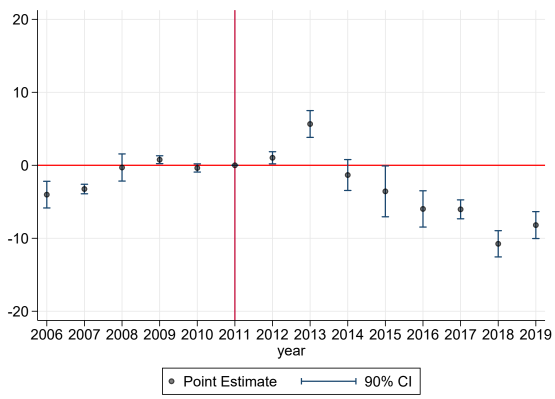


The figures above estimate Equation 1 by types of lanes and times of the day. Panels A shows associations between congestion pricing and traffic outcomes on High-Occupancy Toll lanes during morning rush hours from 6AM to 10AM: Panel A1 for density and Panel A2 for speed. Panels B shows associations between congestion pricing and traffic outcomes on High-Occupancy Toll lanes during evening rush hours from 3PM to 7PM: Panel B1 for density and Panel B2 for speed. Observations are weighted by the inverse propensity scores, which are estimated as a function of speed, density, and tract-level population in 2000. Estimation uses the entire sample as I do not impose the common support restriction. Speed is defined as the total miles driven per hour, and density is defined as the total number of cars occupying a mile on the road. Vertical lines represent 90 percent confidence intervals. Robust standard errors are clustered by county.

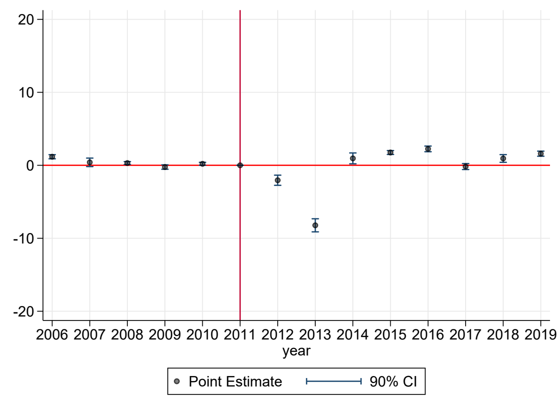
Figure 17: Non-toll lanes parallel to High-Occupancy Toll lanes **without** a common support restriction

Panel A: Morning rush hours

Panel A1: density (cars/mile)

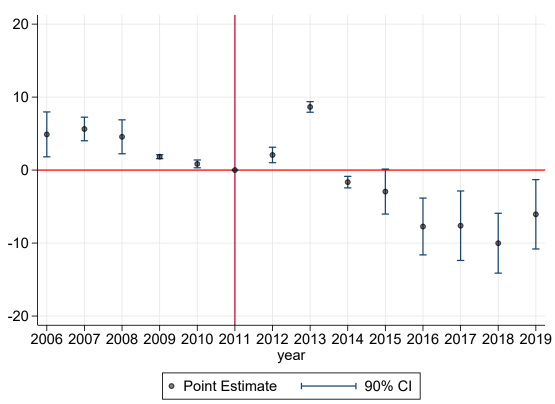


Panel A2: speed (miles/hour)

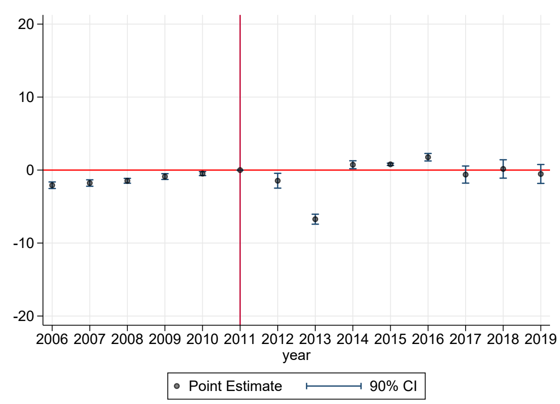


Panel B: Evening rush hours

Panel B1: density (cars/mile)



Panel B2: speed (miles/hour)

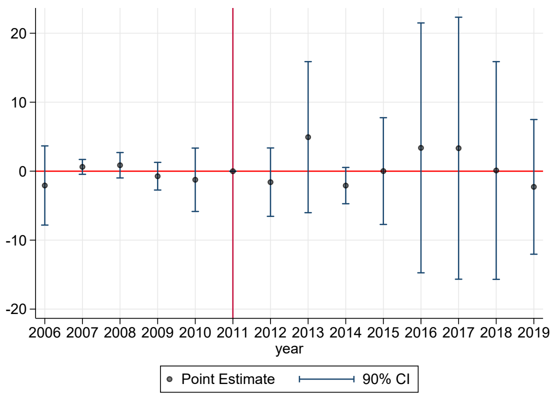


The figures above estimate Equation 1 by types of lanes and times of the day. Panels A shows associations between congestion pricing and traffic outcomes on main lanes parallel to High-Occupancy Toll lanes during morning rush hours from 6AM to 10AM: Panel A1 for density and Panel A2 for speed. Panels B shows associations between congestion pricing and traffic outcomes on main lanes parallel to High-Occupancy Toll lanes during evening rush hours from 3PM to 7PM: Panel B1 for density and Panel B2 for speed. Observations are weighted by the inverse propensity scores, which are estimated as a function of speed, density, and tract-level population in 2000. Estimation uses the entire sample as I do not impose the common support restriction. Speed is defined as the total miles driven per hour, and density is defined as the total number of cars occupying a mile on the road. Vertical lines represent 90 percent confidence intervals. Robust standard errors are clustered by county.

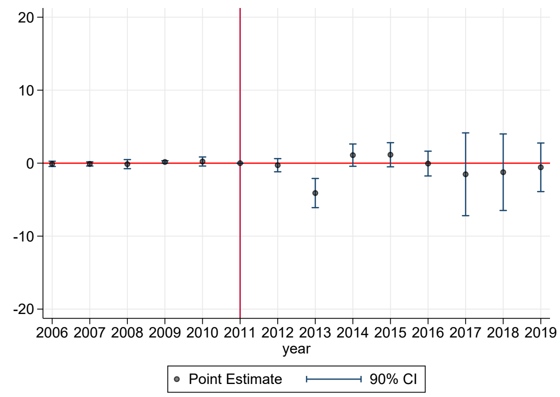
Figure 18: The rest of non-toll lanes **without** a common support restriction

Panel A: Morning rush hours

Panel A1: density (cars/mile)

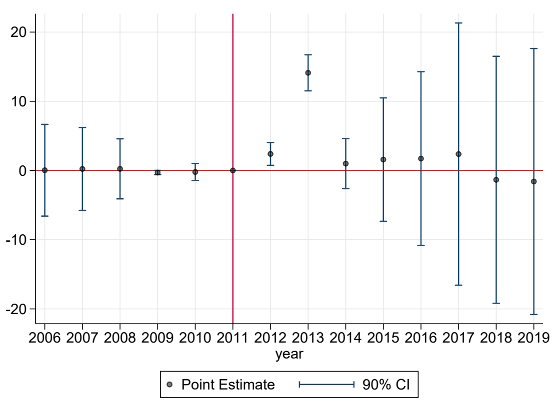


Panel A2: speed (miles/hour)

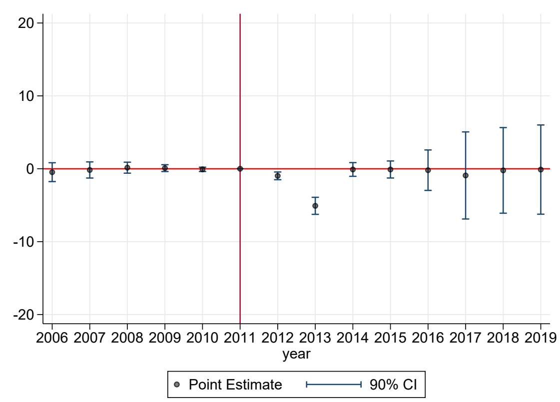


Panel B: Evening rush hours

Panel B1: density (cars/mile)



Panel B2: speed (miles/hour)



The figures above estimate Equation 1 by types of lanes and times of the day. Panels A shows associations between congestion pricing and traffic outcomes on the rest of non-toll main lanes during morning rush hours from 6AM to 10AM: Panel A1 for density and Panel A2 for speed. Panels B shows associations between congestion pricing and traffic outcomes on the rest of non-toll main lanes during evening rush hours from 3PM to 7PM: Panel B1 for density and Panel B2 for speed. Observations are weighted by the inverse propensity scores, which are estimated as a function of speed, density, and tract-level population in 2000. Estimation uses the entire sample as I do not impose the common support restriction. Speed is defined as the total miles driven per hour, and density is defined as the total number of cars occupying a mile on the road. Vertical lines represent 90 percent confidence intervals. Robust standard errors are clustered by county.

Table 4: Difference-in-differences estimates **without** a common support restriction

Panel A: HOT lanes

	morning		evening	
	density	speed	density	speed
$1[LA]_c \times 1[Year_t \in \{2012, 2013\}]$	0.24	-14.51***	-1.05	-8.98***
	(0.26)	(0.28)	(0.50)	(0.78)
$1[LA]_c \times 1[Year_t \geq 2014]$	-2.89***	2.14***	-0.87	0.72
	(0.40)	(0.25)	(0.37)	(0.85)
Observations	49,653	49,653	49,579	49,579
r2	0.75	0.66	0.88	0.66

Panel B: Non-toll lanes parallel to HOT lanes

	morning		evening	
	density	speed	density	speed
$1[LA]_c \times 1[Year_t \in \{2012, 2013\}]$	5.05***	-6.16***	3.05**	-3.57***
	(0.52)	(0.33)	(0.61)	(0.42)
$1[LA]_c \times 1[Year_t \geq 2014]$	-3.23**	1.17***	-8.11***	1.84**
	(0.72)	(0.12)	(0.74)	(0.33)
Observations	323,295	323,295	323,178	323,178
r2	0.87	0.68	0.84	0.70

Panel C: The rest of non-toll lanes

	morning		evening	
	density	speed	density	speed
$1[LA]_c \times 1[Year_t \in \{2012, 2013\}]$	2.02	-2.18**	8.07***	-2.90***
	(2.65)	(0.54)	(0.58)	(0.39)
$1[LA]_c \times 1[Year_t \geq 2014]$	0.70	-0.13	0.60	-0.18
	(5.43)	(0.89)	(4.27)	(1.37)
Observations	587,411	587,411	586,679	586,679
r2	0.77	0.71	0.80	0.77

Panels A, B, and C estimate associations between congestion pricing and traffic outcomes using Equation 1 for HOT lanes, non-toll main lanes parallel to HOT lanes, and the remaining non-toll main lanes, respectively. Each column represents a separate regression. The dependent variable is either speed (mph) or density (cars/mile). All the regressions include station, hour, and monthly date fixed effects. Morning and evening rush hours are from 6AM to 10AM and from 3PM to 7PM, respectively. Observations are weighted by the inverse propensity scores, which are estimated as a function of speed, density, and tract-level population in 2000. Estimation uses the entire sample as I do not impose the common support restriction. Vertical lines represent 90 percent confidence intervals. Robust standard errors are clustered by county.

3 Welfare Effects of Congestion Pricing

3.1 Introduction

Prior research evaluates the economic effects of congestion pricing by focusing on changes in driving behavior. However, as highlighted in chapter 2, individual responses to road pricing may entail not only changing where or when they drive, but also where they live or work. Understanding the welfare effects requires accounting for the full range of adjustments individuals make.

This chapter estimates the aggregate effects of congestion pricing by taking into account where people live, work and drive. To do this, I extend a workhorse quantitative urban model of commuting with endogenous time costs by introducing dynamic congestion pricing and a real estate market. Individuals choose residences, workplaces and driving routes. By choosing a particular driving route, an individual incurs two types of commuting costs: time and toll costs. The key feature of the model is that both time and toll costs are endogenously determined by congestion - that is, the locations of residences, workplaces, and driving routes determine the degree of congestion in a road network and, in turn, characterize travel times and toll costs for all possible driving routes.

I first estimate key objects of the model. I estimate the Fréchet parameter that governs the dispersion of idiosyncratic tastes for residential and commercial locations and driving routes across individuals. For identification, I construct plausibly exogenous Bartik (1991) labor demand shocks as an instrument for wages using tract-level panel data in 1990 and 2000. I exploit variation in differential exposure to national shocks based on the initial share of employment across Census tracts. The instrumental variable estimate of the Fréchet parameter suggests that idiosyncratic tastes are fairly heterogeneous across individuals.

The next object of interest is the elasticity of traffic with respect to the inverse speed that regulates how travel times change with congestion. My estimation strategy exploits temporal and spatial variation in monthly average hourly speed and flow on all major highways since 2006. To account for the simultaneity of speed and traffic, I instrument for traffic using the Covid Stay-Home-Order from March 2020 to June 2021, when traffic abruptly fell by 17.3%. Under the assumption that the order did not coincide with significant changes in lane closures or weather anomalies, the IV estimate is approximately 0.21, which is less than the cross-sectional estimate of 0.4 in the prior literature.[13]

[13]An upward bias in the cross-sectional estimate could be induced by a failure to account for

Lastly, I recover the toll pricing schedule that adjusts toll prices every 5 minutes in response to traffic congestion. Using 5-minute traffic speed data and the posted toll rates in 2019, I observe that toll costs decrease approximately linearly with traffic speed.[14]

The estimated values of the objects determine the magnitude of the equilibrium effects induced by economic shocks. The Fréchet parameter governs how sensitive commuter flows are to changes in wages, rents, and time and toll costs. Changes in commuter flows, in turn, characterize changes in travel times and toll costs in a road network through congestion; how quickly time and toll costs increase with congestion depend on traffic elasticity and the dynamic pricing schedule.

I then use my quantified model to undertake counterfactual simulations of a policy that converts toll lanes back to non-toll lanes in both partial and general equilibrium. A partial equilibrium analysis in which residential and work locations are fixed shows that congestion pricing induces a spatial leakage of traffic externality, which reduces traffic on toll lanes an increases traffic on non-toll lanes. Hence, when net toll revenues are redistributed to the population, annual aggregate welfare decreases by \$1.8-\$11.0 million.[15]

A general equilibrium analysis in which people adjust their residential and work locations and driving routes shows that congestion pricing reduces traffic in the overall road network by reducing commuting distance. This is because people sort based on the value they place on time savings versus toll costs. Some workers who use toll lanes find that toll costs are too costly relative to their wages for the amount of time saved. Hence, they re-optimize their commuting costs by relocating to non-toll areas with faster speed or higher wages and living closer to their workplaces. This has two consequences. First, economic activity increases in non-toll areas with faster speeds or higher wages. Second, traffic on toll lanes decreases more in the general equilibrium than in the partial equilibrium; sorting also reduces traffic on non-toll lanes near toll lanes. When net toll revenues are redistributed to the population, annual aggregate welfare increases by \$2.4-\$11.6 million.

My thesis is related to the urban economics literature, which has focused on explaining patterns of land use and commuting behavior. Classical urban models

---

road or geographic heterogeneity. For example, roads in urban areas have more intersections and crosswalks, which slow traffic.

[14] I run a linear OLS regression of toll costs on traffic speed, and find that tolls per mile range between \$0.50 and \$1.00 when traffic speeds vary between 45 and 70 mph.

[15] Annual aggregate welfare accounts for annual maintenance costs as well as annualized construction costs (Schroeder et al., 2015).

have used stylized settings such as monocentric, symmetric, linear, or circular city structures (Alonso, 1964; Mills, 1967; Muth, 1969; Fujita and Ogawa, 1982; Lucas and Rossi–Hansberg, 2002). Recently, a quantitative urban model that can account for different levels of productivity or amenities across locations has emerged and been applied in a variety of settings (Ahlfeldt et al., 2015; Monte et al., 2018; Heblich et al., 2020; Severen, 2019; Tsivanidis, 2019). I extend this class of models by endogenizing the time cost of commuting.

I extend the recent urban economics literature that incorporates endogenous time costs into a quantitative urban model (Allen and Arkolakis, 2022; Fajgelbaum and Schaal, 2020). In contrast to existing models, in which time costs are the only type of commuting costs, my model incorporates both time costs and monetary toll costs. Furthermore, it confirms the regressive nature of congestion pricing, with toll costs accounting for a greater proportion of wages for workers with lower wages. This is a notable departure from the iceberg commuting costs in existing quantitative urban models, in which everyone's commute costs are an equal fraction of their wages.[16] Therefore, my model is able to deliver plausible predictions about the effects of congestion pricing on city structure and welfare.

Second, my thesis is related to the literature on road pricing. Recent papers have examined the economic implications of toll costs in a partial equilibrium setting (Tarduno, 2021; Kreindler, 2020; Bento et al., 2020). These papers focus on the short-run margins of adjustments, such as departure times or driving routes, to estimate the value of time or characterize optimal road pricing. My contribution is to take into account both residence and employment locations, which allows me to assess welfare in the medium run.[17] Other work estimates the general equilibrium or residential sorting effects of road pricing; for instance, Herzog (2022) examines on flat toll fees in London, and Barwick et al. (2021) studies various transportation policies in Beijing. My contribution is to explicitly model drivers' routing decisions.

Lastly, my thesis is related to the broad literature on the effects of transportation investments on economic activity, such as Baum-Snow (2007); Chandra and Thompson (2000); and Baum-Snow et al. (2020) on the effects of highways on regional development; Donaldson and Hornbeck (2016) on the effects of railways on the agriculture sector; and Donaldson (2018) on the effects of railways on trade. Within this literature, my thesis is also relevant to the set of papers that explore the effect of

[16] The iceberg form makes more sense when time costs are the only type of commuting costs.

[17] My model assumes that the supply of floor space is fixed. To evaluate welfare in the long-run, this assumption needs to be relaxed.

road supply on traffic and find that increasing the supply of road is unlikely to relieve congestion (Jorgensen, 1948; Goodwin, 1996; Cervero, 2002; Duranton and Turner, 2011). Whereas all those papers evaluate supply-side investments - which increase the capacity of the transportation network - my thesis studies the relationship between a demand-side investment and local economic outcomes.

3.2 Setup

To quantify the general equilibrium effects of congestion pricing, I extend a quantitative urban model developed by Allen and Arkolakis (2022), in which time costs are determined simultaneously with individual decisions via congestion. I make the following extensions. First, I incorporate dynamic congestion pricing, which varies toll costs based on traffic speed; this confirms the regressive nature of congestion pricing, as toll costs constitute a higher percentage of wages for lower-wage workers. This is a significant departure from current quantitative urban models, in which time costs are the only type of commuting cost and account for an equal fraction of their wages.[18] Second, I incorporate the market for floor space.

A city is made up of N discrete blocks where economic activities occur. Each block has fixed supply of floor space and a labor market. The city's road network is represented by a $N \times N$ matrix $\mathbf{T} = [t_{mn}]$ where t_{mn} denotes the time cost of traveling from block m to block n. I assume that $t > 0$ for blocks that are directly adjacent to one another; otherwise, $t = \infty$. t is a function of travel time.

An individual who lives in block i and works in block j chooses a particular driving route r of length K. The driving route is represented by a sequence of blocks visited from i to j: $r = \{r_0 = i, r_1, ...r_K = j\}$, where r_k is the kth block that an individual visits. Then, the total time cost incurred from taking the driving route r of length K is:

$$t_{r_0,r_1} \times t_{r_1,r_2} \times ...t_{r_{K-1,K}} = \prod_{k=1}^{K} t_{r_{k-1},r_k}. \tag{3}$$

Also, the total toll cost is:

$$p_{r_0,r_1} + p_{r_1,r_2} + ...p_{r_{K-1,K}} = \sum_{k=1}^{K} p_{r_{k-1},r_k}. \tag{4}$$

[18] The iceberg cost, which assumes everyone's costs are an equal fraction of their wages, is more reasonable if time costs are the only type of commuting costs.

3.2.1 Individuals' Decision

An individual chooses a residence, workplace, and a driving route, and supplies 1 unit of labor inelastically in return for wage w_j in block j. The utility of an individual commuting from residence i to workplace j along a driving route r of length K is determined by the consumption of final goods c_i, residential floor space h_i, and exogenous residential amenity $\bar{a}_i$. It also depends on the total time cost and the total toll cost. As a result, an individual chooses a combination of commuting locations and a driving route (i, j, r), consumption c, and residential floor space h to maximize the Cobb-Douglas utility function, given the prices of residential floor space q_{r_i}:

$$\max_{c_i, h_i, \{ijr\}} \frac{\bar{a}_i}{\prod_{k=1}^{K} t_{r_{k-1}, r_k}} \left(\frac{c_i}{\beta}\right)^{\beta} \left(\frac{h_i}{1-\beta}\right)^{1-\beta} \epsilon_{ijr}, \tag{5}$$

$$s.t \quad q_{r_i} h_i + c_i = \tilde{w}_j,$$

where $\tilde{w}_j = w_j - \sum_{k=1}^{K} p_{r_{k-1}, r_k}$ is the wage net of the total toll cost. The price of the final good is normalized to 1.

Following D. McFadden (1974) and Eaton and Kortum (2002), I assume that individuals draw idiosyncractic preference shocks for residences, workplaces and driving routes ϵ_{ijr} from an independent and identically distributed Fréchet distribution:

$$F(\epsilon_{ijr}) = e^{-E_i G_j \epsilon_{ijr}^{-\theta}}, \tag{6}$$

where the shape parameter θ determines the degree of heterogeneity in idiosyncratic utility shocks.[19] E_i is the mean utility from living in i, and G_j is the mean utility from working in j. After observing idiosyncratic shocks for all the possible combinations of (i, j, r), an individual chooses residence i, workplace j, and a driving route r of length K to maximize the indirect utility V_{ijr}:

$$V_{ijr} = \max_{i,j,r} \frac{\bar{a}_i \times q_{r_i}^{-(1-\beta)} \times \tilde{w}_j \times \epsilon_{ijr}}{\prod_{k=1}^{K} t_{r_{k-1}, r_k}}. \tag{7}$$

It is useful to discuss some of the implications of this problem. First, it assumes

[19] A higher θ is associated with more dispersed idiosyncratic shocks.

that an individual takes the same driving route every day. Second, because the model does not specify the time of departure for work, it assumes that an individual incurs the average travel time throughout the day. Lastly, it abstracts away from trip chains such as grocery shopping or recreational activities.[20]

Utility maximization predicts that the probability of living in i, working in j and taking a particular driving route r of length K is:

$$\pi_{ijr} = \frac{\left(\prod_{k=1}^{K} t_{r_{k-1},r_k}^{-\theta}\right) \times E_i \bar{a}_i^{\theta}\ q_{r_i}^{-\theta(1-\beta)} \times G_j \tilde{w}_j^{\theta}}{\sum_i \sum_j \sum_{r \in R_{ij}^{K \leq K^{max}}} \left(\prod_{k=1}^{K} t_{r_{k-1},r_k}^{-\theta}\right) \times E_i \bar{a}_i^{\theta}\ q_{r_i}^{-\theta(1-\beta)} \times G_j \tilde{w}_j^{\theta}}, \tag{8}$$

where $R_{ij}^{K \leq K^{max}}$ denotes the set of all the possible driving routes with a length less than or equal to K^{max} from i to j.[21] Summing across all the possible routes $r \in R_{ij}^{K \leq K^{max}}$, the bilateral commuting probability is:

$$\pi_{ij} = \frac{\sum_{r \in R_{ij}^{K \leq K^{max}}} \left(\prod_{k=1}^{K} t_{r_{k-1},r_k}^{-\theta}\right) \times E_i \bar{a}_i^{\theta}\ q_{r_i}^{-\theta(1-\beta)} \times G_j \tilde{w}_j^{\theta}}{\sum_i \sum_j \sum_{r \in R_{ij}^{K \leq K^{max}}} \left(\prod_{k=1}^{K} t_{r_{k-1},r_k}^{-\theta}\right) \times E_i \bar{a}_i^{\theta}\ q_{r_i}^{-\theta(1-\beta)} \times G_j \tilde{w}_j^{\theta}}. \tag{9}$$

Using matrix algebra, I show that equation 9 can be simplified as follows:

[20] Miyauchi et al. (2021) finds that a single trip often involves multiple stops using smartphone GPS data in Japan.

[21] Instead of assuming that the length of a driving route could be infinite, as in Allen and Arkolakis (2022), I set K^{max} as the maximum length of all the least cost paths to exclude cases where the proportion of the total toll cost relative to wage is very high.

$$\pi_{ij} \approx \frac{b_{ij,\boldsymbol{j}} \times G_j w_j^{\theta} \times E_i \bar{a}_i^{\theta} q_{h_i}^{-\theta(1-\beta)}}{\sum_i \sum_j b_{ij,\boldsymbol{j}} \times G_j w_j^{\theta} \times E_i \bar{a}_i^{\theta} q_{h_i}^{-\theta(1-\beta)}},$$

$$\text{where} \quad b_{ij,\boldsymbol{j}} = \sum_{K=0}^{K^{max}} A_{ij,\boldsymbol{j}}^{K}, \tag{10}$$

$$\text{and} \quad \mathbf{A}_{\boldsymbol{j}} \equiv \big[t_{mn}^{-\theta} \exp(-\theta \frac{p_{mn}}{w_j})\big].$$

$\mathbf{A}_{\boldsymbol{j}}$ is an adjacency matrix in which each element is a function of time costs and toll costs relative to wages between adjacent blocks. $A_{ij,\boldsymbol{j}}^{K}$ is the i, jth element of matrix $\mathbf{A}_{\boldsymbol{j}}$ raised to the power of K. I refer to $b_{ij,\boldsymbol{j}}$ as the total commuting cost. Let L_{ij} denote the number of commuters between blocks i and j:

$$L_{ij} \approx b_{ij,\boldsymbol{j}} \times E_i \bar{a}_i^{\theta} q_{h_i}^{-\theta(1-\beta)} \times G_j w_j^{\theta} \times \bar{L} \times \Omega^{-1},$$

$$\text{where} \quad \Omega \approx \sum_i \sum_j b_{ij,\boldsymbol{j}} \times E_i \bar{a}_i^{\theta} q_{h_i}^{-\theta(1-\beta)} \times G_j w_j^{\theta}. \tag{11}$$

Equation 11 implies that pairs of residences and workplaces with higher residential amenities, lower floor space prices, higher wages, and higher workplace amenities attract more commuters. The expected utility from living in a closed city with a fixed population $\bar{L}$ is $\bar{U}$:

$$\bar{U} \approx \Gamma\left(\frac{\theta-1}{\theta}\right) \left[\sum_i \sum_j b_{ij,\boldsymbol{j}} \times E_i \bar{a}_i^{\theta} q_{h_i}^{-\theta(1-\beta)} \times G_j w_j^{\theta}\right]^{1/\theta}, \tag{12}$$

where Γ is the Gamma function.

3.2.2 Congestion

Traffic is defined as the total expected number of times people in the city cross the road segment between blocks m and n that are adjacent to one another. To derive the expression for traffic, I first characterize how many times a commuter from i to j traverses the road segment between blocks m and n:

$$\pi_{ij}^{mn} = \sum_{r \in R_{ij}^{K \le K^{max}}} \left[\frac{\pi_{ijr}}{\sum_{r' \in R_{ij}^{K \le K^{max}}} \pi_{ijr'}} \right] n_r^{mn}. \tag{13}$$

It is the product of the probability that a route r is used conditional on commuting from i to j, which is inside the bracket, and the number of times the route r passes the link, which is denoted as n_r^{mn}, summed across all the possible routes.[22] I can rewrite π_{ij}^{mn} as follows by plugging the expression for π_{ijr} obtained in equation 8 into equation 13:

$$\pi_{ij}^{mn} = \frac{1}{b_{ij,\boldsymbol{j}}} \sum_{K=0}^{K^{max}} \sum_{S=0}^{K-1} A_{im,\boldsymbol{j}}^{S} \times t_{mn}^{-\theta} \exp(-\theta \frac{p_{mn}}{w_j}) \times A_{nj,\boldsymbol{j}}^{K-S-1}, \tag{14}$$

where $A_{im,\boldsymbol{j}}^{S}$ is the (i, m)th element of an adjacency matrix $\mathbf{A}_{\boldsymbol{j}}$ to the power of S, and $A_{nj,\boldsymbol{j}}^{K-S-1}$ is the (n, j)th element of an adjacency matrix $\mathbf{A}_{\boldsymbol{j}}$ to the power of $K - S - 1$. Therefore, the final expression for the total traffic on the road segment between blocks m and n is:

$$\Xi_{mn} = \sum_i \sum_j \pi_{ij}^{mn} \times L_{ij}, \tag{15}$$

where L_{ij} is the total commuters from i to j.

3.2.3 Time Costs and Toll Costs

Time Costs: Following Vickrey (1967), I assume that the time cost incurred on the road segment that directly connects blocks m and n is log-linear in travel time:

$$t_{mn} = \left(\frac{distance_{mn}}{speed_{mn}} \right)^{\rho_0}, \tag{16}$$

where ρ_0 is the elasticity of travel time with respect to time costs. The inverse speed

[22] An individual may choose a commuting route with $n_r^{mn} > 1$. This is due to their unique preferences for commuting routes. For instance, on their way to work, they might make a stop at a grocery shop or daycare.

is determined by traffic and road conditions:

$$speed_{mn}^{-1} = m_0 \times \left(\frac{\Xi_{mn}}{lanes_{mn}} \right)^{\rho_1} \times \varepsilon_{mn}, \tag{17}$$

where m_0 and ρ_1 represent the free rate of flow and the elasticity of traffic with respect to the inverse speed, respectively. Combining equations 16 and 17 implies that time costs have an exogenous component $\bar{t}$ and an endogenous component Ξ:

$$t_{mn} = \bar{t}_{mn} \times \Xi_{mn}^{\rho_0 \rho_1}, \quad \text{where} \quad \bar{t} \ \equiv \left(\frac{distance \times m_0 \times \varepsilon}{lanes^{\rho_1}} \right)^{\rho_0}. \tag{18}$$

Toll Costs: Because the precise toll schedule employed by Los Angeles Metro is confidential, I assume that toll costs per mile on the road segment that directly connects blocks m and n vary based on speed, according to the pricing schedule g:

$$p_{mn} = g\left(speed_{mn}\right). \tag{19}$$

3.2.4 Production

A single final good Y_j is produced under perfect competition and constant returns to scale. Production depends on labor L_{f_j}, commercial floorspace H_{f_j}, and exogenous productivity $\bar{A}_j$.[23]

$$Y_j = \bar{A}_j L_{f_j}^{\alpha} H_{f_j}^{1-\alpha}. \tag{20}$$

Firms maximize profit given wage w_j and commercial floor space price q_{f_j}:

$$\pi_j = \bar{A}_j L_{f_j}^{\alpha} H_{f_j}^{1-\alpha} - w_j L_{f_j} - q_{f_j} H_{f_j}. \tag{21}$$

The first order condition produces inverse demands for labor and commercial floor space:

$$w_j = \alpha \bar{A}_j \left(\frac{H_{f_j}}{L_{f_j}} \right)^{1-\alpha}, \tag{22}$$

[23]This is a simplifying assumption. In some models, productivity depends on a production externality, which is determined by the density of workers in the surrounding area. Ahlfeldt et al. (2015), and Heblich et al. (2020) are two examples.

$$q_{f_j} = (1-\alpha)\bar{A}_j \left(\frac{L_{f_j}}{H_{f_j}}\right)^{\alpha}. \tag{23}$$

The zero profit condition determines the equilibrium price for commercial floor space:

$$q_{f_j} = (1-\alpha)\bar{A}_j^{\frac{1}{(1-\alpha)}} \left(\frac{\alpha}{w_j}\right)^{\frac{\alpha}{(1-\alpha)}}. \tag{24}$$

3.2.5 Floorspace Market Clearing

The demand for residential floor space must equal the supply of residential floor space (equation 25), and the demand for commercial floor space must equal the supply of commercial floor space in each location (equation 26):

$$(1-\beta)L_{r_i}\frac{\mathbb{E}\left[w|i\right]}{q_{r_i}} = \bar{H}_{r_i}, \tag{25}$$

$$L_{f_i}\left(\frac{(1-\alpha)\bar{A}_i}{q_{f_i}}\right)^{\frac{1}{\alpha}} = \bar{H}_{f_i}, \tag{26}$$

$$(1-\beta)L_{r_i}\frac{\mathbb{E}\left[w|i\right]}{q_{r_i}} + L_{f_i}\left(\frac{(1-\alpha)\bar{A}_i}{q_{f_i}}\right)^{\frac{1}{\alpha}} = \bar{\delta}_i\bar{H}_i + (1-\bar{\delta}_i)\bar{H}_i = \bar{H}_i. \tag{27}$$

where L_{r_i} and L_{f_i} are the total resident and worker populations in block i. Due to strict zoning regulations, dense construction of residential and commercial units is difficult in the Los Angeles County.[24] Hence, I make two assumption. First, the supply of floor space $\bar{H}_i$ is fixed in each location. Second, the proportions of floor space allocated for residential use, $\bar{\delta}_i$ and for commercial use, $1-\bar{\delta}_i$ are fixed.

3.3 Equilibrium

Given exogenous location fundamentals $\{E_i, G_i, \bar{a}_i, \bar{\delta}_i\}_{i\in L}$, labor endowment $\bar{L}$, floor space endowment $\{\bar{H}_i\}_{i\in N}$, exogenous infrastructure costs $\bar{T} \equiv [\bar{t}_{ij}]_{i,j\in N^2}$, toll pricing schedule g, and the set of parameters $\{\alpha, \beta, \theta, \rho_0, \rho_1\}$, an equilibrium is a set

[24] For example, single-family zoning, which prohibits apartments, is very common in the Los Angeles County (Dedousis, 2020). In the median city, over 80% of residential land is zoned as single-family.

of economic outcomes defined by $\{L_{r_i}, L_{f_i}, w_i, q_{r_i}, q_{f_i}\}_{i\in N}$, commuting probabilities $\{\pi_{ij}\}_{i,j\in N^2}$, toll costs $\{p_{ij}\}_{i,j\in N^2}$, and welfare $\bar{U}$ such that:

1. Given the equilibrium values of the total commuting costs $\{b_{ij,\boldsymbol{j}}\}_{ij\in N^2}$, the equilibrium values of economic outcomes ensure the following statements.
   - In each block, the labor market clears.
   - In each block, the floor space market clears.
   - The total demand for labor equals $\bar{L}$.
2. Given the equilibrium values of exogenous infrastructure costs $\bar{\mathbf{T}} \equiv [\bar{t}_{ij}]_{i,j\in N^2}$ and toll costs $\mathbf{P} \equiv [p_{ij}]_{i,j\in N^2}$, individuals make the optimal routing decision.
3. Given the equilibrium values of economic outcomes, the equilibrium values of times costs $\mathbf{T} \equiv [t_{ij}]_{i,j\in N^2}$ are determined by the equilibrium values of congestion $\boldsymbol{\Xi} \equiv [\Xi_{ij}]_{i,j\in N^2}$
4. Given the equilibrium values of congestion, the equilibrium values of toll costs $\mathbf{P} \equiv [p_{ij}]_{i,j\in N^2}$ are determined by the congestion pricing schedule $\mathbf{g}$.
5. The equilibrium values of the total commuting costs and economic outcomes determine the equilibrium value of welfare $\bar{U}$.

Therefore, an equilibrium satisfies the following system of equations:

$$t_{mn} = \left(\frac{distance_{mn}}{speed_{mn}}\right)^{\rho_0}, \tag{28}$$

$$p_{ij} = g\left(speed_{ij}\right), \tag{29}$$

$$b_{ij,\boldsymbol{j}} \approx \sum_{K=0}^{K^{max}} A^K_{ij,\boldsymbol{j}}, \quad \text{where} \quad \mathbf{A}_{\boldsymbol{j}} \equiv \left[t_{mn}^{-\theta}\exp(-\theta\frac{p_{mn}}{w_j})\right], \tag{30}$$

$$H_{r_i}q_{r_i} = (1-\beta)\sum_j\left(w_j\pi_{ij}\times\frac{\tilde{b}_{ij,\boldsymbol{j}}}{b_{ij,\boldsymbol{j}}}\right)\times\bar{L}, \quad \text{where} \quad \tilde{\mathbf{A}}_{\boldsymbol{j}} = \left[t_{mn}^{-\theta}\exp\left((-1-\theta)\frac{p_{mn}}{w_j})\right)\right], \tag{31}$$

$$\pi_{ij} \approx \frac{b_{ij,\boldsymbol{j}} \times \tilde{E}_i G_j \times w_j^{\theta} \times q_{r_i}^{-\theta(1-\beta)}}{\sum_i \sum_j b_{ij,\boldsymbol{j}} \times \tilde{E}_i G_j \times w_j^{\theta} \times q_{r_i}^{-\theta(1-\beta)}}, \quad \text{where} \quad \tilde{E}_i G_j \equiv E_i G_j \bar{a}_i^{\theta}, \tag{32}$$

$$\Xi_{mn} = \sum_{ij} \left[\frac{\pi_{ij}}{b_{ij,\boldsymbol{j}}} \sum_{K=1}^{K^{max}} \sum_{S=0}^{K-1} A_{im,\boldsymbol{j}}^{S} \times t_{mn}^{-\theta} \exp(-\theta \frac{p_{mn}}{w_j}) \times A_{nj,\boldsymbol{j}}^{K-S-1} \right] \times \bar{L}, \tag{33}$$

$$speed_{mn}^{-1} = \bar{m} \times \Xi_{mn}^{\rho_1}, \quad \text{where} \quad \bar{m} \equiv \frac{m_0 \varepsilon_{mn}}{lanes_{mn}^{\rho_1}}, \tag{34}$$

$$\tilde{\bar{A}}_j = \frac{w_j^{\alpha}}{(1-\alpha)^{1-\alpha} \alpha^{\alpha}} \quad \text{where} \quad \tilde{\bar{A}}_j \equiv \frac{\bar{A}}{q_{f_j}^{1-\alpha}}, \tag{35}$$

$$\bar{U} \approx \Gamma\left(\frac{\theta - 1}{\theta}\right) \left[\sum_i \sum_j b_{ij,\boldsymbol{j}} \times E_i \bar{a}_i^{\theta} q_{h_i}^{-\theta(1-\beta)} \times G_j w_j^{\theta} \right]^{1/\theta}. \tag{36}$$

3.4 Recovering Location and Road Fundamentals

In this section, I show how I recover a unique set of location and road fundamentals given observed data and parameters values.

Given the observed vectors of wages $\{\mathbf{W}\}$ and matrices of speed, distance and commuting probabilities $\{\mathbf{S}, \mathbf{D}, \mathbf{\Pi}\}$, the set of parameters $\{\alpha, \beta, \xi, \theta, \lambda\}$, and the toll pricing schedule g, location and road characteristics are uniquely recovered such that: there exists a vector of location and road fundamentals $\{\tilde{\mathbf{E}}\mathbf{G}, \bar{\mathbf{M}}\}$, a matrix of congestion $\{\mathbf{\Xi}\}$, a vector of adjusted productivity $\{\tilde{\bar{\mathbf{A}}}\}$, and a vector of residential rents $\{\mathbf{H_r q_r}\}$ that are consistent with the set of observed values as an equilibrium.

I provide the solution algorithm. First, given distances and speeds, I recover time and toll costs using equations 28 and equation 29. Second, given wages, time costs and toll costs, I recover the total commuting costs using equation 30. Third, given wages, the total commuting costs, time costs and commuting probabilities, I recover prices for residential floor space using equation 31. Fourth, given wages, commuting probabilities, the total commuting costs, and prices for residential floor space, there exists a unique set of location fundamentals that satisfy equation 32. Fifth, I recover congestion using equation 33. Sixth, given time costs and congestion, there exists a unique set of road fundamentals that satisfy the equation 34. Lastly, given wages, I recover adjusted productivity using equation 35. Then, the welfare is determined using equation 36.

3.5 Model Estimation

There are 6 key objects in the model: a share of household expenditure on residential floor space $1-\beta$, a share of firm costs on commercial floor space $1-\alpha$, time elasticity ρ_0, traffic elasticity ρ_1, toll pricing schedule g, and the Fréchet parameter θ. I set $1-\alpha = 0.2$, and $1-\beta = 0.25$ that are consistent with the estimates in M. A. Davis and Ortalo-Magné (2011), and Valentinyi and Herrendorf (2008). The rest of this section estimates the Fréchet parameter θ, time elasticity ρ_0, and traffic elasticity ρ_1, and recovers the toll pricing schedule g.

3.5.1 Fréchet Parameter & Time Elasticity

The Fréchet Parameter θ and the elasticity of travel time ρ_0 in equation 10 governs the sensitivity of commuting flows with respect to prices and commute times. The existing literature usually presents OLS estimates using cross-sectional data. Instead, I provide panel-estimates by instrumenting for changes in wages with Bartik shocks (Bartik, 1991).

Estimation Strategy: To derive the estimating equation from equation 9, I assume that all the pairs of locations can be directly traveled without passing through other locations, as in Allen and Arkolakis (2022).[25] Then, the estimating equation is:

$$\sum_{r \in R_{ij}^{K \leq K^{max}}} \left(\prod_{k=1}^{K} TravelTime_{r_{k-1},r_k}^{-\lambda} \right) = \left(TravelTime_{ij}\right)^{-\theta\rho_0}, \tag{37}$$

where $TravelTime_{ij}$ is the travel time incurred on the path that directly connects i and j, and $\lambda = \theta\rho_0$. Plugging in equation 37 into equation 9 and taking logarithm

[25] This assumption implies that $K^{max} = 1$.

yields the following estimating equation:

$$\Delta ln(Commuters_{ijt}) = \alpha + \theta \Delta ln(Wage_{jt}) - \theta\rho_0 \Delta ln(TravelTime_{ijt}) + \gamma_{it} + \epsilon_{ijt}, \tag{38}$$

where $\Delta ln(Commuters_{ijt})$ is the change in commute flows from residence i to workplace j at time t, $\Delta Wage_{jt}$ is the change in wage at workplace j at time t, and $\Delta TravelTime_{ijt}$ is the change in travel time from residence i to workplace j at time t. γ_{it} is the residence-year fixed effects. I cluster standard errors by origin and destination county pairs.

Wages may be endogenous due to a potential correlation between unobservables and wages.[26] Hence, I instrument for changes in wages using Bartik shocks, which are widely used as plausibly exogenous local labor demand shocks Δs_{jt}:

$$\Delta s_{jt} = \sum_g \left[\frac{Wage_t^{g,Nat} - Wage_0^{g,Nat}}{Wage_0^{g,Nat}} \times \frac{N_{j,0}^g}{N_{j,0}} \right], \tag{39}$$

where $Wage_t^{g,Nat}$ is the national wage in industry g in the year t, $N_{j,0}^g$ is the total employment in industry g in workplace j in the initial year, and $N_{j,0}$ is the total employment in workplace j in the initial year across all the industries, similar to Saiz (2010), and Diamond (2016). The identification assumption is:

$$E[\Delta s_{jt} \times \Delta \epsilon_{ijt}] = 0. \tag{40}$$

Under the assumption that Bartik shocks Δs_{jt} are correlated with changes in local productivity but uncorrelated with changes in local amenities, using Bartik shocks to instrument for changes in wages identifies θ.[27]

Data: I use the Census Transportation Planning Package from 1990 and 2000 to create tract-level panel data on wages, travel times, and commute flows. I only use data from 1990 and 2000 because more recent data do not include wages. Since CTPP 1990 is defined by Census 1990 geographies, and CTPP 2000 is defined by Census 2000 geographies, I covert wages, commuting flows and travel times in 2000

[26] Estimates of θ would be biased downwards if changes in wages are correlated with changes in characteristics such as parking rates or congestion. Conversely, it would be biased upwards if they are correlated with changes in workplace amenity such as access to dining services or public transportation.

[27] Exclusion restriction can be validated in terms of the initial shares of workers (Goldsmith-Pinkham et al., 2020).

geographies to 1990 geographies. See Appendix B2 for details.

Results: I use the inverse hyperbolic sine to transform the dependent variable because some observations have zero commuting flows. Table 5 displays IV and OLS estimates of the Fréchet parameter θ and time elasticity λ.

The IV estimate of the Fréchet parameter is 0.988, which is significantly greater than the OLS estimates. As previously discussed, potential sources of a downward bias in the cross-sectional estimate include disamenities from urban crowding, such as higher parking fees. My IV estimates are also significantly lower than estimates from other studies, suggesting that idiosyncratic shocks are fairly heterogenous across workers in the Los Angeles metropolitan area.[28] Another interpretation is that workers are relatively immobile because when θ is low, places require a higher wage increase to attract workers. The time elasticity is approximately 0.01, implying that a 10% reduction in commute time corresponds to a 0.1% increase in commuting flows.

3.5.2 Traffic Elasticity

The elasticity of traffic ρ_1 in equation 17 regulates the rate at which travel time increases with an increase in traffic flow. The existing literature, which uses cross-sectional data, may induce the omitted variable bias if traffic is correlated with road or geographic characteristics. Instead, I use a high-resolution spatial panel data on the monthly average hourly traffic in California and employ an instrumental variable approach.

Estimation Strategy: Taking logarithm of equation 17 yields the following estimating equation:

$$log(\text{speed}_{symh}^{-1}) = \beta_0 + \rho_1 log\,(\text{traffic}_{symh}) + \gamma_{sy} + \gamma_m + \gamma_h + log(\epsilon_{symh}), \tag{41}$$

where speed_{symh} and traffic_{symh} are the monthly average hourly speed and flow reported by Vehicle Detector Station s in month m in year y at hour h. Station-year fixed effects γ_{sy} capture the effects of the free rate of flow and the number of lanes, which are allowed to vary by year. I also control for monthly and hourly trends with γ_m and γ_h. I cluster standard errors at the county level.

An OLS estimate may be biased downward due to reverse causality if drivers tend to move from slower roads to faster ones. Therefore, I instrument for endogenous

[28] Allen and Arkolakis (2022) uses a value of 6.83 in counterfactual simulations, which is estimated using cross-sectional data in Ahlfeldt et al. (2015).

variable traffic_{symh} with an instrumental variable covid_{ym} that takes a value 1 for all the observations during the Covid Stay-Home-Order period from March 2020 to June 2021.[29] The exclusion restriction is that the Covid Stay-Home-Order affects speed solely through traffic congestion: $E[\text{covid}_{my} \times \epsilon_{smyh}] = 0$.[30]

Data: I use real-time hourly traffic data from January 2006 to June 2022 that are measured by Vehicle Detector Stations (VDS) on all the major highways in Los Angeles, Riverside, Sacramento, and San Diego counties. I exclude imputed data, and data from the weekends. Also, I keep data from non-toll main lanes, HOV lanes, and HOT lanes only, excluding data from on-ramps or off-ramps. For estimation, I aggregate the hourly speed and flow to the monthly average hourly speed and flow. Observations are at the Vehicle Detector Station-level.

Results: IV estimates of traffic elasticity reported in Panel A of Table 6 are about 0.2, implying that a 10% increase in traffic flow is associated with a 2% increase in the inverse traffic speed. In first-stage regressions, estimates of α_1 are around -0.19, with F-Statistics of 57.32 and 71.35, implying that the initiation of Stay-Home-Order is associated with a 17.3% reduction in traffic flow.[31]

The OLS estimates of traffic elasticity reported in Panel B columns (3) and (4) are significantly lower than the IV estimates and statistically insignificant. This supports the hypothesis that drivers tend to shift from slower roads to faster roads, inducing a significant downward bias.

The IV estimate of 0.2, as far as I am aware of, is the first panel estimate of traffic elasticity. It is lower than the cross-sectional estimate of 0.47 in Allen and Arkolakis (2022). This suggests that a failure to control for unobserved heterogeneity in roads or geographic characteristics may over-estimate the traffic elasticity. Roads in densely populated areas, for example, have more intersections and crosswalks. Because intersections and crosswalks tend to slow traffic, not controlling these factors would result in an upward bias.

[29] During the Covid Stay-Home-Order from March 19, 2020 to June 15, 2021, California ordered residents to stay home unless they are engaged in essential activities.

[30] This could be violated if Stay-Home-Order coincided with a systematic and significant change in lane closures or weather anomalies, which is not true.

[31] In log-linear models, a change in a dummy variable from 0 to 1 is associated with a $100 \times (e^{\alpha_1} - 1)\%$ increase in Y.

3.5.3 Congestion Pricing Schedule

The congestion pricing schedule applied by the Los Angeles Metro that determines toll costs from traffic conditions is a crucial parameter for estimating welfare effects as it governs the rate at which toll costs increases with traffic congestion. Since the precise schedule is confidential, I recover the price schedule with the observed traffic conditions and toll costs.

Data: 5-minute toll costs data obtained through public records request are available at the toll segment level, while 5-minute speed data from the Caltrans Performance Measurement System (PeMS) are available at the Vehicle Detector Stations (VDS) level. To combine these two data, I aggregate them at the toll segment level.

Estimation Strategy: Figure 20 shows a binned scatter plot of toll costs against traffic speed on HOT lanes. It exhibits a strong negative relationship, with toll costs decreasing approximately linearly with speeds. Therefore, I estimate a simple linear OLS regression:

$$toll_{st} = \beta_0 + \beta_1 speed_{st} + \epsilon_{st}, \tag{42}$$

where $toll_{st}$ and $speed_{st}$ are toll costs (in cents) and speeds on toll segment s at time t, respectively.

Results: Table 7 shows that OLS estimates of β_0 and β_1 are 1.91 and -0.02, respectively. As a result, drivers pay 20 cents more for every mile driven on HOT lanes when traffic speed decreases by 10 miles per hour. To put it another way, toll costs increase from \$0.50 to \$1.00 per mile when traffic speeds decrease from 70 to 45 mph.

3.6 Counterfactual Analysis

To goal of this section is to evaluate the effects of congestion pricing on the spatial distribution of economic activity and welfare using "exact hat algebra" developed by Dekle et al. (2007).[32] I first construct data in the observed equilibrium and test the validity of the model. Then, I undertake a counterfactual simulation of a policy that converts toll lanes back to non-toll lanes, and compute changes in endogenous variables. I assume a closed city with a fixed total population, and thus any shocks

[32] "Exact hat algebra" is extensively used in counterfactual exercises in international trade and urban economics. It defines counterfactual equilibrium equations in terms of the observed equilibrium and the relative changes from the observed equilibrium.

to the city cause changes in the level of utility.

3.6.1 Data Construction

I briefly explain how I construct data in the observed equilibrium. Appendix B provides more detail. The units of analysis are 10km-by-10km grids in the Los Angeles-Long Beach-Anaheim, CA Metropolitan Statistical Area, as shown in Figure 22.

Commuting flow: Census block-level bilateral commuting flows in the 2019 LEHD Origin-Destination Employment Statistics (LODES) are normalized to pairs of grids. I develop a crosswalk by intersecting a shapefile of 10km-by-10km grids with a shapefile of the 2019 census blocks. Then, using areal weights, census block-level commuting flows are summed across intersections within pairs of grids.

Wages: I normalize zip-level payroll data from the 2019 County Business Patterns to construct the average annual earnings per employee in each grid. First, I construct a crosswalk by intersecting a shapefile of 10km-by-10km grids with a shapefile of the 2010 ZIP Code Tabulation Area. Then, using areal weights, the zip-level annual earnings per employee are averaged across the intersections within grids.

Commuting days: I adjust the total number of commuting days N so that the total toll revenue predicted by the model equals the actual revenue reported in the Los Angeles County Metropolitan Transportation Authority's 2018 fiscal year budget.[33] Therefore, the following condition holds true in the observed equilibrium:

$$\$70 \text{ million} = \sum_{s \in \text{HOT lanes}} \Xi\left(P(N)\right)_s \times P(N)_s, \tag{43}$$

where $P(N)_s$ and $\Xi\left(P(N)\right)_s$ represent the annual toll costs and total traffic on segment s, respectively.

Adjacency matrix: I obtain driving times between pairs of grids that are adjacent to one another using the HERE routing API calls. Driving times on HOT lanes are not available, which is a limitation of the existing routing API. As a result, I manually create a shapefile of the I-10/I-110 HOT lanes and combine it with the Caltrans Performance Measurement System hourly speed data (PeMS). Using this shapefile, I calculate driving times on HOT lanes between pairs of grids that are

[33] In the 2018 fiscal year, High-Occupancy Toll Lanes generated a total revenue of \$70 million, which included both violation fines and toll revenue.

directly connected by HOT lanes. Then, I combine travel times, wages, and toll costs according to equation 10 to create an adjacency matrix $\mathbf{A}_{\boldsymbol{j}}$.

Traffic: I construct a traffic matrix using wages, annual toll costs, commuting flows, and travel times using equations 14 and 15.

3.6.2 Model Fit

To assess the validity of my theoretical framework, I compare the predicted and the observed values of traffic. I use equations 14 and 15 to construct the predicted values of traffic, as discussed in section 6.1. To obtain the observed values of traffic, I proceed as follows. First, I create a road network in ArcGis Pro using the 2016 Highway Performance Monitoring System (HPMS) data in shapefile format from the Federal Highway Administration, which includes detains on the number of lanes, the Average Annual Daily Traffic (AADT), and speed limits for segments of highways and local roads. Second, using the road network, I solve for least cost paths between pairs of grids that are adjacent to one another. Lastly, I compute the average AADT on these least cost paths as the observed values of traffic. Appendix B provides more detail.

Figure 21 displays a binned scatter plot of the observed traffic against the predicted traffic. Although the model excludes any non-commute related traffic, such as commercial trucking, grocery shopping or leisure activity, they are quite strongly correlated, with a correlation coefficient of 0.64. The plot also shows that in areas with heavy traffic, the model overestimates traffic. A potential explanation is that drivers may avoid roads with heavy traffic to a greater extent because it produces more unfavorable conditions, such as noise or psychological stress, which the model does not account for.

3.6.3 Aggregate Effects

Table 8 reports the effects of implementing congestion pricing on welfare and traffic in a closed city with the fixed population. Panel A shows the percentage change in traffic on toll and non-toll lanes. Panel B shows the net aggregate benefit without or with the redistribution of net toll revenue. To calculate the net aggregate benefit without the redistribution of net revenues, I multiply the percent changes in expected utility by the average wage and the total population in the Los Angeles metro area and subtract the annualized construction costs.[34] Column (1) reports results from a partial equilibrium analysis, in which people are allowed to change driving routes only; column (2) reports results from a general equilibrium analysis, in which people may adjust residences, workplaces or driving routes.

Results from a partial equilibrium analysis show that traffic on toll lanes decreases by about 0.5 percent while traffic on the rest of the remaining non-toll lanes increases by 0.005 percent as some drivers divert from toll lanes to non-toll lanes. As a results of this spatial leakage of traffic externality, congestion pricing lowers aggregate welfare by \$1.8-\$11.0 million.

Results from a general equilibrium analysis show two main findings. First, traffic decreases by about 0.6 and 0.002 percent on toll lanes and non-toll lanes, respectively, indicating an overall reduction in traffic in the road network. Second, compared to the partial equilibrium, the size of the traffic reduction on toll lanes is larger in the general equilibrium. This results from sorting, which reduces the commute distance. Consequently, congestion pricing increases aggregate welfare by \$2.4-\$11.6 million.

3.6.4 Sorting

This section discusses the effects of congestion pricing on the spatial distribution of residents and workers. To understand the mechanism for sorting, it is useful to consider the following equation, which is derived from the model's gravity equation:

$$\begin{aligned} log(\hat{\pi}_{ij}) =& log(\hat{b}_{ij,\boldsymbol{j}}) - \theta(1-\beta)log(\hat{q}_{r_i}) + \theta log(\hat{w}_j) + C_1, \\ \text{where} \quad & \hat{b}_{ij,\boldsymbol{j}} = f\left(\mathbf{T}\hat{\mathbf{T}}, \mathbf{P}\hat{\mathbf{P}}, w_j\hat{w}_j\right) \end{aligned} \tag{44}$$

where $\hat{\pi}_{ij}$, $\hat{b}_{ij,\boldsymbol{j}}$, $\hat{q}_i$, and $\hat{w}_j$ are relative changes in commuting probabilities, commuting costs, prices for residential floor space, and wages, respectively. Also, $\mathbf{T}$, and $\mathbf{P}$, are matrices of relative changes in time and toll costs across the road network or the city. Equation 44 implies that any shocks to time or toll costs affect commute flows directly through $\hat{b}_{ij,\boldsymbol{j}}$, and also indirectly through general equilibrium changes $\hat{w}_j$, $\hat{q}_i$.

After congestion pricing is implemented on toll lanes, some residents and workers find that tolls are too expensive relative to wages for the time saved by using toll

[34]To annualize the initial construction cost, which was funded by the USDOT, I multiply the total costs of \$106.52 with the share of employment in the Los Angeles metro area. Using a 3 percent discount rate and annualizing it over 30 years, this results in an annualized cost of \$0.2 million.

lanes. Hence, they relocate to non-toll areas with faster speed or higher wages or live closer to their workplaces in order to re-optimize commuting costs $b_{ij,j}$. This process of sorting continues until changes in residential floor space $\hat{q}_i$ and wages $\hat{w}_j$ restore the equilibrium, as implied by the equation 44.

As a result, the size of resident and worker inflows among non-toll areas is positively correlated with initial wages but negatively correlated with initial travel times, as shown in figures 23 and 24. This has two consequences. First, economic activity increases in non-toll areas with faster speeds or higher wages. Since some of these non-toll areas have lower wages relative to toll areas (e.g., the North and North East LA), some people earn less after moving, which lowers the aggregate gain. Second, sorting reduces traffic on toll lanes more than in the partial equilibrium; it also reduces traffic on non-toll lanes adjacent to toll lanes.

3.6.5 Sensitivity to the Fréchet Parameter

Table 9 reports the effects of congestion pricing on traffic in *Panel A* and on aggregate welfare in *Panel B* at different values of the Fréchet Parameter θ. There are two key takeaways.

In *Panel A*, for values of $\theta > \theta_0$ where $2 < \theta_0 < 3$, traffic on non toll lanes increases, causing spatial leakage of traffic externality. This is because a higher value of θ is associated with more homogeneous preference shocks across people, causing a greater proportion of residents and workers to relocate from toll areas to non-toll areas. Hence, as non-toll areas receive more inflows of residents and workers, congestion increases.

In *Panel B*, aggregate benefit decreases as θ rises. This is driven by two mechanisms. The first is an increase in spatial leakage of traffic externality, which creates more congestion in non-toll areas, as shown in *Panel A*. Second, as θ rises, a greater proportion of workers move to non-toll areas with relatively lower wages grows.

3.7 Conclusion

Dynamic highway congestion pricing is becoming more common in urban areas. However, little is known about the welfare effects that account for the full range of adjustments individuals make. My thesis fills this gap in the literature by estimating the general equilibrium effects of dynamic highway congestion pricing using a policy change in Los Angeles that converted a subset of non-toll lanes to toll lanes with dynamic pricing.

Both reduced-form estimates from a difference-in-differences approach and model estimates from counterfactual simulations that apply a quantitative urban model provide consistent evidence that changes in traffic patterns differ in the short- and medium-run. In the short-run, when people only adjust their driving routes, congestion pricing induces a spatial leakage of traffic externality because people divert from toll lanes to non-toll lanes, which reduces aggregate benefits by $1.8-$11.0 million. In the medium-run, when people can adjust their residences, workplaces, and driving routes, congestion pricing induces people to sort based on the value they place on time savings versus toll costs. Some workers who live near toll lanes decide that toll costs are too costly relative to their wages and move to non-toll areas or live closer to their workplaces, which reduces the average driving distance. As a result, when net toll revenues are redistributed to the population, annual aggregate welfare increases by $2.4-$11.6 million. Therefore, my thesis emphasizes the importance of accounting for residential and work locations in order to understand the full welfare effects.

References

AHLFELDT, G. M., S. J. REDDING, D. M. STURM AND N. WOLF (2015). "The Economics of Density: Evidence from the Berlin Wall". *Econometrica* 83 (6), 2127–2189.

ALLEN, T. AND C. ARKOLAKIS (2022). "The Welfare Effects of Transportation Infrastructure Improvements". *The Review of Economic Studies*. ISSN: 0034-6527.

ALONSO, W. (1964). "The historic and the structural theories of urban form: their implications for urban renewal". *Land Economics* 40 (2), 227–231.

ANDERSON, M. L. AND L. W. DAVIS (2018). *Two Empirical Tests of Hypercongestion*. Tech. rep. National Bureau of Economic Research.

ANDERSON, M. L. (2014). "Subways, Strikes, and Slowdowns: The Impacts of Public Transit on Traffic Congestion". *American Economic Review* 104 (9), 2763–96.

BARRERO, J., N. BLOOM AND S. J. DAVIS (2021). *Why working from home will stick*. Tech. rep. National Bureau of Economic Research.

BARTIK, T. J. (1991). "Who benefits from state and local economic development policies?"

BARWICK, P. J., S. LI, A. R. WAXMAN, J. WU AND T. XIA (2021). *Efficiency and equity impacts of urban transportation policies with equilibrium sorting*. Tech. rep. National Bureau of Economic Research.

BAUM-SNOW, N. (2007). "Did highways cause suburbanization?" *The quarterly journal of economics* 122 (2), 775–805.

BAUM-SNOW, N., J. V. HENDERSON, M. A. TURNER, Q. ZHANG AND L. BRANDT (2020). "Does investment in national highways help or hurt hinterland city growth?" *Journal of Urban Economics* 115, 103124.

BELAND, L.-P. AND D. A. BRENT (2018). "Traffic and crime". *Journal of Public Economics* 160 (8), 96–116.

BENTO, A., K. ROTH AND A. R. WAXMAN (2020). *Avoiding traffic congestion externalities? the value of urgency*. Tech. rep. National Bureau of Economic Research.

CERVERO, R. (2002). "Induced travel demand: Research design, empirical evidence, and normative policies". *Journal of Planning Literature* 17 (1), 3–20.

CHANDRA, A. AND E. THOMPSON (2000). "Does public infrastructure affect economic activity?: Evidence from the rural interstate highway system". *Regional science and urban economics* 30 (4), 457–490.

CHOUDHURY, P., C. FOROUGHI AND B. Z. LARSON (2021). "Work-from-anywhere: The productivity effects of geographic flexibility". *Strategic Management Journal* 42 (4), 655–683.

CURRIE, J. AND R. WALKER (2011). "Traffic Congestion and Infant Health: Evidence from E-ZPass". *American Economic Journal: Applied Economics* 2 (4), 65–90.

DAVIS, M. A. AND F. ORTALO-MAGNÉ (2011). "Household expenditures, wages, rents". *Review of Economic Dynamics* 14 (2), 248–261.

DEDOUSIS, A. (2020). "Single-Family Zoning and Exclusion in L.A. County: Part 1.City Commentary". URL: https://cityobservatory.org/la_exclusion_pt1/.

DEKLE, R., J. EATON AND S. KORTUM (2007). "Unbalanced trade". *American Economic Review* 97 (2), 351–355.

DERYUGINA, T., L. KAWANO AND S. LEVITT (2018). "The economic impact of Hurricane Katrina on its victims: Evidence from individual tax returns". *American Economic Journal: Applied Economics* 10 (2), 202–33.

DIAMOND, R. (2016). "The determinants and welfare implications of US workers' diverging location choices by skill: 1980-2000". *American Economic Review* 106 (3), 479–524.

DONALDSON, D. (2018). "Railroads of the Raj: Estimating the impact of transportation infrastructure". *American Economic Review* 108 (4-5), 899–934.

DONALDSON, D. AND R. HORNBECK (2016). "Railroads and American economic growth: A "market access" approach". *The Quarterly Journal of Economics* 131 (2), 799–858.

DUGANDZIC, R., D. S. LINDA DODDS AND M. SMITH-DOIRON (2006). "The Association between Low Level Exposures to Ambient Air Pollution and Term Low Birth Weight: A Retrospective Cohort Study". *Environmental Health* 5 (3).

DURANTON, G. AND M. A. TURNER (2011). "The Fundamental Law of Road Congestion: Evidence from US Cities". *American Economic Review* 101 (6), 2616–52.

EATON, J. AND S. KORTUM (2002). "Technology, geography, and trade". *Econometrica* 70 (5), 1741–1779.

EMANUEL, N. AND E. HARRINGTON (2021). "Working remotely? Selection, treatment and the market provision of remote work". *Working Paper*.

ERNST, M., J. CORLESS AND R. GREENE-ROESEL (2002). "Clearing the Air: Public Health Threats from Cars and Heavy Duty Vehicles—Why We Need to Protect Federal Clean Air Laws". URL: http://www.transact.org/report.%20asp?id=227.

FAJGELBAUM, P. D. AND E. SCHAAL (2020). "Optimal transport networks in spatial equilibrium". *Econometrica* 88 (4), 1411–1452.

FUJITA, M. AND H. OGAWA (1982). "Multiple equilibria and structural transition of non-monocentric urban configurations". *Regional science and urban economics* 12 (2), 161–196.

GOLDSMITH-PINKHAM, P., I. SORKIN AND H. SWIFT (2020). "Bartik instruments: What, when, why, and how". *American Economic Review* 110 (8), 2586–2624.

GOODWIN, P. B. (1996). "Empirical evidence on induced traffic". *Transportation* 23 (1), 35–54.

HAMILTON, T. L. AND C. J. WICHMAN (2018). "Bicycle infrastructure and traffic congestion: Evidence from DC's Capital Bikeshare". *Journal of Environmental Economics and Management* 87.

HEBLICH, S., S. J. REDDING AND D. M. STURM (2020). "The making of the modern metropolis: evidence from London". *The Quarterly Journal of Economics* 135 (4), 2059–2133.

HERZOG, I. (2022). "The city-wide effects of tolling downtown drivers: Evidence from london's congestion charge". *Unpublished manuscript.*

HIRANO, K., G. W. IMBENS AND G. RIDDER (2003). "Efficient estimation of average treatment effects using the estimated propensity score". *Econometrica* 71 (4), 1161–1189.

HOLLIDAY, A. (2018). "A Case Study of LA Metro's ExpressLanes as a Congestion Reduction Strategy".

HUBER, M. J. (1962). "Street travel as related to local parking". *Proceedings of the 41st Annual Meeting of the Highway Research Board*, 333–52.

JORGENSEN, R. E. (1948). "Influence of expressways in diverting traffic from alternate routes and in generating new traffic". In: *Highway Research Board Proceedings*. Vol. 27.

KREINDLER, G. (2020). "Peak-hour road congestion pricing: Experimental evidence and equilibrium implications". *Unpublished paper.*

KRISHNAMURTHY, C. K. AND N. S. NGO (2020). "The effects of smart-parking on transit and traffic: Evidence from SFpark". *Journal of Environmental Economics and Management* 99.

KÜNN-NELEN, A. (2016). "Does Commuting Affect Health?" *Health Economics* 25 (8), 984–1004.

LUCAS, R. E. AND E. ROSSI–HANSBERG (2002). "On the internal structure of cities". *Econometrica* 70 (4), 1445–1476.

MARCHANT, A. (2020). "What can traffic data tell us about the impact of the coronavirus?"

MCFADDEN, B. AND J. INNES (1990). *Photographic Enforcement of High Occupancy Vehicle Lanes. Final Report*. Tech. rep.

MCFADDEN, D. (1974). "The measurement of urban travel demand". *Journal of public economics* 3 (4), 303–328.

MILLS, E. S. (1967). "An aggregative model of resource allocation in a metropolitan area". *The American Economic Review* 57 (2), 197–210.

MONTE, F., S. J. REDDING AND E. ROSSI-HANSBERG (2018). “Commuting, migration, and local employment elasticities”. *American Economic Review* 108 (12), 3855–90.

MULALIC, I., J. N. VAN OMMEREN AND N. PILEGAARD (2014). “Wages and Commuting: Quasi-natural Experiments’ Evidence from Firms that Relocate”. *The Economic Journal* 124 (579), 1086–1105.

MUSICK, N. (2022). “Government Spending on Public Transportation and Other Infrastructure”. *Congressional Budget Office*.

MUTH, R. F. (1969). “CITIES AND HOUSING; THE SPATIAL PATTERN OF URBAN RESIDENTIAL LAND USE.”

PISHUE, B. (2023). *Global Traffic Scorecard*. Tech. rep. INRIX.

ROBERTS, J., R. HODGSON AND P. DOLAN (2011). ““It’s driving her mad”: Gender differences in the effects of commuting on psychological health”. *Journal of Health Economics* 30 (5), 1064–1076.

SAIZ, A. (2010). “The geographic determinants of housing supply”. *The Quarterly Journal of Economics* 125 (3), 1253–1296.

SCHRANK, D., B. EISELE AND T. LOMAX (2019). “2019 URBAN MOBILITY REPORT”. *The Texas AM Transportation Institute*.

SCHROEDER, J., R. KLEIN, T. SMITH, K. TURNBULL, K. BALKE, M. BURRIS, P. SONGCHITRUKSA, B. PESSARO, E. S. SANDGREN, E. SCHREFFLER AND B. JOY (2015). “Los Angeles Congestion Reduction Demonstration ExpressLanes Program: National Evaluation Report”. *Department of Transportation*.

SEVEREN, C. (2019). “Commuting, labor, and housing market effects of mass transportation: Welfare and identification”. *The Review of Economics and Statistics*, 1–99.

SHOUP, D. (2006). “Cruising for Parking”. *Transport Policy* 13 (6), 479–486.

TARDUNO, M. (2021). “For Whom the Bridge Tolls: Congestion, Air Pollution, and Second-Best Road Pricing”. *Unpublished manuscript*.

TSIVANIDIS, N. (2019). "Evaluating the impact of urban transit infrastructure: Evidence from bogota's transmilenio". *Unpublished manuscript*.

US CENSUS BUREAU (2020a). "American Community Survey 5-year estimates".

— (2020b). "UC Census Bureau Annual Survey of State and Local Government Finances 1977-2020".

VALENTINYI, A. AND B. HERRENDORF (2008). "Measuring factor income shares at the sectoral level". *Review of Economic Dynamics* 11 (4), 820–835.

VICKREY, W. (1967). "Optimization of traffic and facilities". *Journal of Transport Economics and Policy*, 123–136.

WACHS, M., P. S. CHESNEY AND Y. H. HWANG (2015). "A CENTURY OF FIGHTING TRAFFIC CONGESTION IN LOS ANGELES 1920- 2020". *UCLA Luskin Center for History and Policy*.

XIAO, H., A. WU AND J. KIM (2021). "Commuting and innovation: Are closer inventors more productive?" *Journal of Urban Economics* 121.

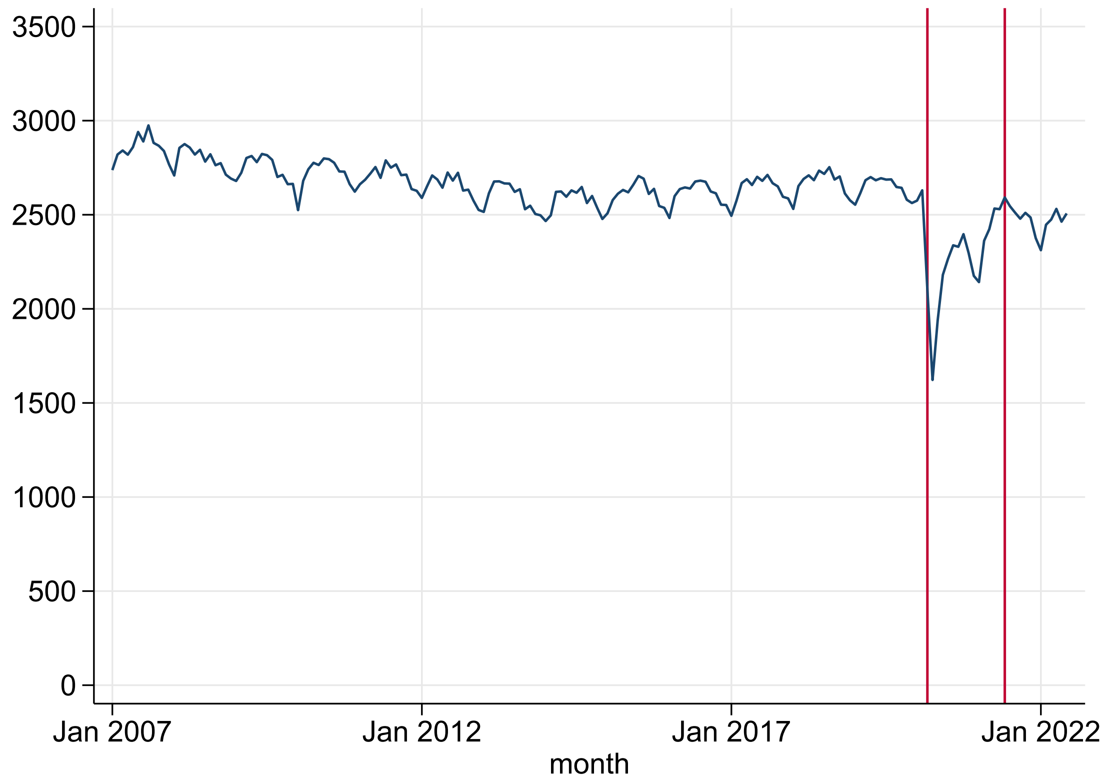
3.8 Tables and Figures

Table 5: Estimates of θ and λ

	(1)	(2)
	$\Delta Ihs(\pi_{ij})$	$\Delta Ihs(\pi_{ij})$
$\Delta log(Wage_j)$	0.988^{**}	0.024
	(0.224)	(0.017)
$\Delta log(TravelTime_{ij})$	-0.007^{***}	-0.007^{***}
	(0.001)	(0.001)
First stage	0.394^{***}	
	(0.087)	
F-statistics	20.39	
Residence×Year	Yes	Yes
N	118,836	118,836

The table estimates the Fréchet parameter θ and time elasticity λ. Each column represents a separate regression. Column (1) reports IV estimates, and column 2 reports OLS estimates. The dependent variables are the changes in the inverse hyperbolic sine of π_{ij}. The independent variables are the changes in the log of wages and travel times. All regressions control for residence-year fixed effects. Standard errors in parentheses are clustered by residence-workplace pairs. The significance levels are indicated as: $*p < 0.10, **p < 0.05, ***p < 0.01$.

Figure 19: Monthly Traffic Flow Per Vehicle Detector Station



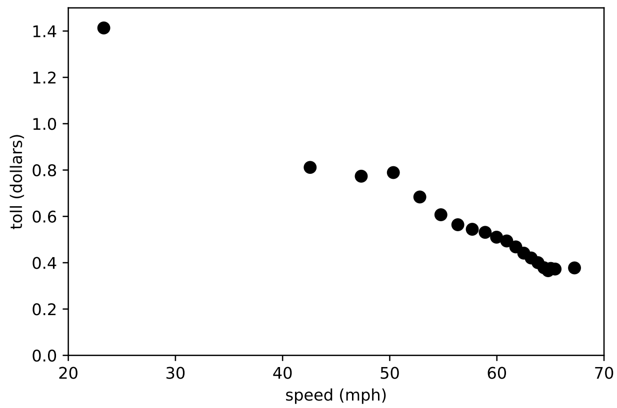
The figure displays a time-series plot of the average monthly traffic flow per Vehicle Detector Station in Los Angeles, San Diego, Sacramento, and Riverside. The stay-at-home order period, which spans from March 2020 to June 2021, is represented by the red vertical lines.

Table 6: The Elasticity of Traffic

	Panel A: 2SLS		Panel B: OLS	
	(1)	(2)	(3)	(4)
	$log(speed^{-1})$	$log(speed^{-1})$	$log(speed^{-1})$	$log(speed^{-1})$
$log(traffic)$	0.214^{***}	0.216^{**}	-0.001	0.004
	(0.036)	(0.039)	(0.008)	(0.006)
First stage	-0.193^{***}	-0.190^{***}		
	(0.025)	(0.022)		
Station-Year FE	Yes	No	Yes	No
Station FE	No	Yes	No	Yes
Year FE	No	Yes	No	Yes
Hour FE	Yes	Yes	Yes	Yes
F-Statistics	57.52	71.35		
N	13,068,312	13,068,318	13,068,312	13,068,318

The table estimates the elasticity of traffic with respect to the inverse speed. Each column represents a separate regression. In all regressions, the dependent variable is the log of the inverse of the monthly average hourly speed. The units of observation are Vehicle Detector Stations. Panel A reports the IV estimates where the log of traffic is instrumented by a dummy variable $covid_{my}$ that takes a value of 1 for observations during the Covid Stay-Home-Order from March 2021 to June 2022. Panel B reports the OLS estimates. All the regressions include hour-fixed effects, and standard errors are clustered by county. The significance levels are indicated as: $*p < 0.10, **p < 0.05, ***p < 0.01$.

Figure 20: Binned scatter plot of tolls against speed



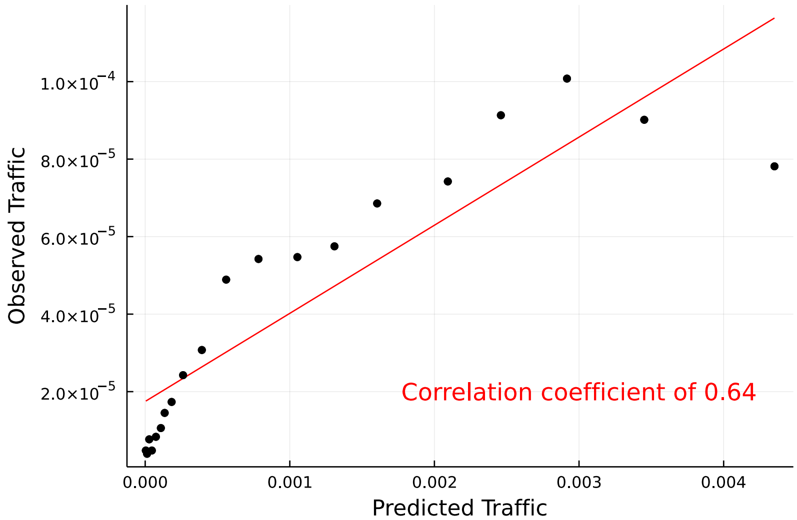
This figure displays a binned scatter plot of 5-minute toll costs (in dollars) on the y-axis against 5-minute traffic speed (in miles per hour) on the x-axis in 2019. The units of observation are toll segments of High-Occupancy Toll lanes of I-10 and I-110.

Table 7: A Regression of Toll Costs on Speed on High-Occupancy Toll lanes

	Tolls
$speed(mph)$	-0.02 ***
	(0.000)
$constant$	1.91 ***
	(0.003)
R^2	0.37
N	361,980

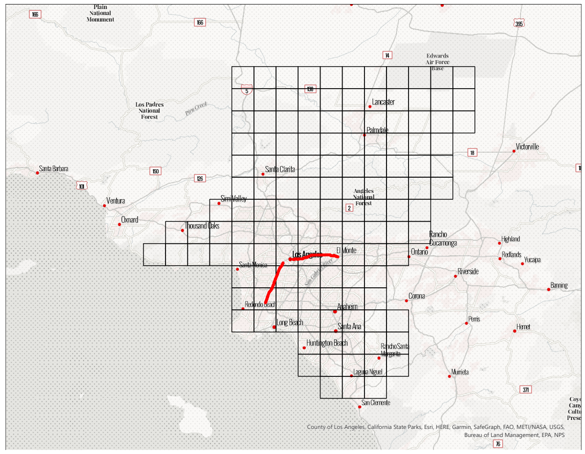
The table estimates the congestion pricing schedule. The dependent is toll costs (in dollars), and the independent variable is traffic speed (in miles per hour). Both are from 2019 and vary every 5 minutes. The units of observation are toll segments of High-Occupancy Toll lanes of I-10 and I-110. No fixed effects are included. The significance levels are indicated as: $*p < 0.10, **p < 0.05, ***p < 0.01$.

Figure 21: Observed Traffic vs. Predicted Traffic



This figure displays a binned scatter plot of observed traffic on the y-axis against predicted traffic on the x-axis. Units of observations are pairs of grids. I obtain the observed values of traffic between an origin and a destination by solving for the least cost path in ArcGis Pro and calculating the Annual Average Daily Traffic along the path. I obtain the predicted values of traffic using equation 15. I normalize both observed and predicted traffic values such that the sum of the total observed traffic and the sum of the total predicted traffic are both equal to 1.

Figure 22: 10km by 10km grids in the Los Angeles-Long Beach-Anaheim, CA, Metro Area



The figure displays the locations of 10km by 10km grids in the Los Angeles-Long Beach-Anaheim, CA Metro Area used as units of analysis in counterfactual exercises. Red lines represent High-Occupancy Toll lanes on I-10 and I-110.

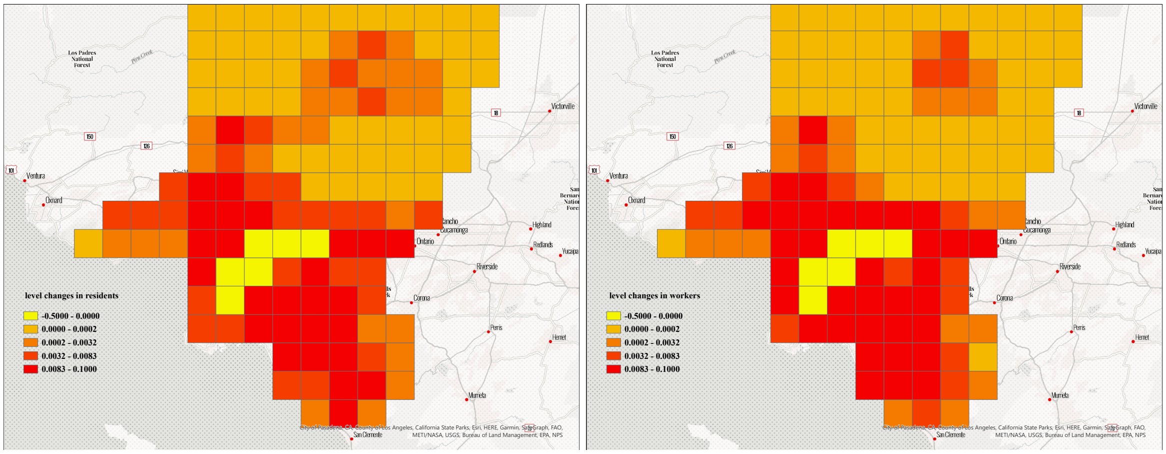
Table 8: The Effects of Implementing Congestion Pricing in a Closed Economy

	partial eq.	general eq.
Panel A: percent changes in traffic		
toll lanes	−0.485	−0.553
non-toll lanes	+0.005	−0.002
Panel B: aggregate benefit in million		
without revenue redistribution	-$37.7	-$24.2
with revenue redistribution	-$11.0 to -$1.8	$2.4 to $11.6

This table estimates the partial and general equilibrium effects of congestion pricing. *Panel A* and *Panel B* report the effects on traffic and on aggregate benefit in million, respectively. To calculate the effects on net aggregate benefit without revenue redistribution, I multiply the percent changes in expected utility by the average wage and the total population, and then subtract the annualized construction cost. To determine the net aggregate benefit with revenue redistribution, I add the net toll revenue and net aggregate benefit without revenue redistribution.

Figure 23: The Effects of Congestion Pricing on Population Distribution

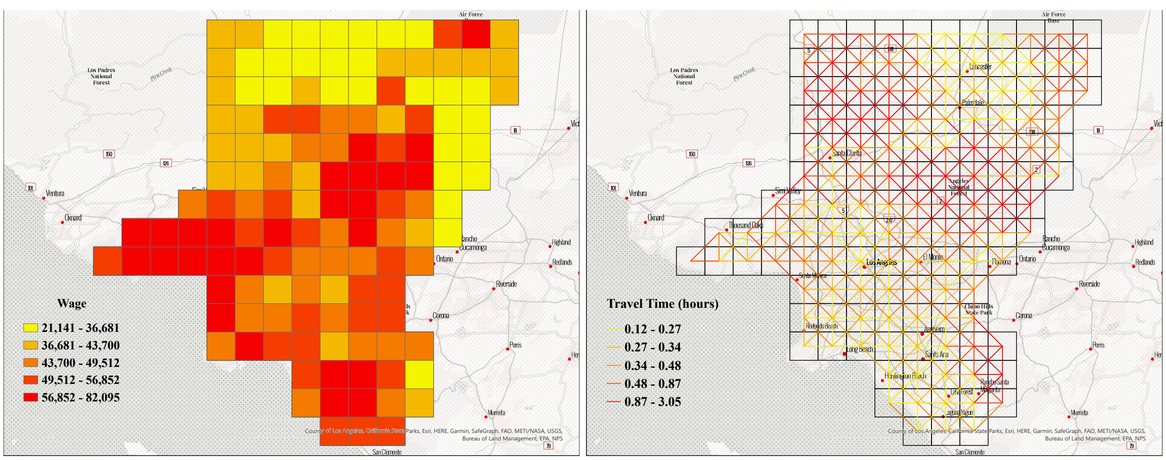
Panel A: Residents *Panel B*: Workers



Panels A and B show level changes in the worker and resident populations as a result of implementing congestion pricing on HOT lanes, which are obtained from a counterfactual simulation. The total population in the city is normalized as 1. Black lines represent High-Occupancy Toll lanes on I-10 and I-110.

Figure 24: Wages and Travel Times Prior to Congestion Pricing

Panel A: Wages *Panel B*: Travel Time (in hours)



Panels A and B show wages and travel times (in hours) between adjacent locations prior to the implementation of congestion pricing. They are calculated using a counterfactual simulation of a policy that converts toll lanes back to non-toll lanes.

Table 9: Sensitivtiy of Welfare to the Fréchet Parameter θ

	values of θ		
	2	3	4
Panel A: percent changes in traffic			
toll lanes	-1.081	-1.592	-2.091
non-toll lanes	-0.001	$+0.001$	$+0.004$
Panel B: aggregate benefit in million			
without revenue redistribution	-\$39.6	-\$50.2	-\$58.1
with revenue redistribution	-\$13.3 to -\$4.1	-\$24.3 to -\$15.1	-\$32.5 to -\$23.3

This table estimates the general equilibrium effects of congestion pricing using different values of the Fréchet Parameter. *Panel A* and *Panel B* report the effects on traffic and on aggregate benefit in million, respectively. To calculate the effects on net aggregate benefit without revenue redistribution, I multiply the percent changes in expected utility by the average wage and the total population, and then subtract the annualized construction cost. To determine the net aggregate benefit with revenue redistribution, I add the net toll revenue and net aggregate benefit without revenue redistribution.

3.9 Appendix A: Derivations

I provide detailed derivations of the model that incorporates monetary toll costs and the floor space market into a quantitative urban model developed by Allen and Arokolakis (2022).

A1: Commuting probability

Let π_{ijr} be the probability that an individual lives in i, works in j, and takes a particular driving route r of length K. Define $R_{ij}^{K\leq K^{max}}$ as the set of all the possible driving routes with a length $K \leq K^{max}$. I set K^{max} as the minimum length of all the least cost paths between pairs of grids. Utility maximization implies that:

$$\pi_{ijr} = \frac{\left(\prod_{k=1}^{K} t_{r_{k-1},r_k}^{-\theta}\right) \times E_i \bar{a}_i^{\theta} q_{r_i}^{-\theta(1-\beta)} \times G_j \left(w_j - \sum_{k=1}^{K} p_{r_{k-1},r_k}\right)^{\theta}}{\sum_i \sum_j \left[\sum_{r \in \in R_{ij}^{K\leq K^{max}}} \left(\prod_{k=1}^{K} t_{r_{k-1},r_k}^{-\theta}\right) \times E_i \bar{a}_i^{\theta} q_{r_i}^{-\theta(1-\beta)} \times G_j \left(w_j - \sum_{k=1}^{K} p_{r_{k-1},r_k}\right)^{\theta}\right]}.$$

I sum π_{ijr} across $r \in R_{ij}^{K\leq K^{max}}$ to obtain the bilateral commuting probability π_{ij}:

$$\pi_{ij} = \frac{\sum_{r \in R_{ij}^{K\leq K^{max}}} \left(\prod_{k=1}^{K} t_{r_{k-1},r_k}\right)^{-\theta} \times E_i \bar{a}_i^{\theta} q_{r_i}^{-\theta(1-\beta)} \times G_j \left(w_j - \sum_{k=1}^{K} p_{r_{k-1},r_k}\right)^{\theta}}{\sum_i \sum_j \sum_{r \in R_{ij}^{K\leq K^{max}}} \left(\prod_{k=1}^{K} t_{r_{k-1},r_k}\right)^{-\theta} \times E_i \bar{a}_i^{\theta} q_{r_i}^{-\theta(1-\beta)} \times G_j \left(w_j - \sum_{k=1}^{K} p_{r_{k-1},r_k}\right)^{\theta}}.$$

Since I set K^{max} as the minimum length of all the least cost paths between pairs of grids, the total toll costs associated with any driving routes are less than 10% of wages. As a result, the following log approximation is correct:

$$
\begin{aligned}
\sum_{r\in R_{ij}^{K\leq K^{max}}}\left(\prod_{k=1}^{K} t_{r_{k-1},r_k}\right)^{-\theta}\left(w_j-\sum_{k=1}^{K} p_{r_{k-1},r_k}\right)^{\theta} =& w_j^{\theta}\sum_{r}\left(\prod_{k=1}^{K} t_{r_{k-1},r_k}\right)^{-\theta}\left(1-\sum_{k=1}^{K}\frac{p_{r_{k-1},r_k}}{w_j}\right)^{\theta} \\
=& w_j^{\theta}\sum_{r}\left(\prod_{k=1}^{K} t_{r_{k-1},r_k}\right)^{-\theta} exp\left[log\left(1-\sum_{k=1}^{K}\frac{p_{r_{k-1},r_k}}{w_j}\right)\right. \\
\approx& w_j^{\theta}\sum_{r}\left(\prod_{k=1}^{K} t_{r_{k-1},r_k}\right)^{-\theta} exp\left(-\theta\sum_{k=1}^{K}\frac{p_{r_{k-1},r_k}}{w_j}\right) \\
\approx& w_j^{\theta}\sum_{r}\left(\prod_{k=1}^{K} t_{r_{k-1},r_k}\right)^{-\theta}\left(\prod_{k=1}^{K} exp\left(-\theta\frac{p_{r_{k-1},r_k}}{w_j}\right)\right) \\
\approx& w_j^{\theta}\sum_{r}\left[\prod_{k=1}^{K} t_{r_{k-1},r_k}^{-\theta}\times exp\left(-\theta\frac{p_{r_{k-1},r_k}}{w_j}\right)\right].
\end{aligned}
$$

I can re-write the commuting probability as:

$$
\pi_{ij}=\frac{\left[\sum_{r\in R_{ij}^{K\leq K^{max}}}\prod_{k=1}^{K} t_{r_{k-1},r_k}^{-\theta} exp\left(-\theta\frac{p_{r_{k-1},r_k}}{w_j}\right)\right]\times E_i\bar{a}_i^{\theta}q_{r_i}^{-\theta(1-\beta)}\times G_j w_j^{\theta}}{\sum_i\sum_j\left[\sum_{r\in R_{ij}^{K\leq K^{max}}}\prod_{k=1}^{K} t_{r_{k-1},r_k}^{-\theta} exp\left(-\theta\frac{p_{r_{k-1},r_k}}{w_j}\right)\right]\times E_i\bar{a}_i^{\theta}q_{r_i}^{-\theta(1-\beta)}\times G_j w_j^{\theta}}.
$$

Define a $N\times N$ matrix $\mathbf{A_j}=[a_{kl,j}]\ \ \forall j=1,2,...N$, where $a_{kl,j}=t_{kl}^{-\theta}\exp(-\theta\frac{p_{kl}}{w_j})$. Each element in the matrix represents the total commuting costs as a function of time costs and toll costs relative to wages. I assume that traveling between grids that are not adjacent to one another is impossible: $t=\infty$. I can rewrite the expression inside the bracket as follows:

$$\sum_{r \in R_{ij}^{K \leq K^{max}}} \left(\prod_{k=1}^{K} a_{r_{k-1} r_k, j} \right) = \sum_{K=0}^{K^{max}} \left(\sum_{k_1=1}^{N} \sum_{k_2=1}^{N} ... \sum_{k_{K-1}=1}^{N} a_{ik_1,j} \times a_{k_1 k_2,j} \times ... a_{k_{K-2} k_{K-1},j} \times a_{k_{K-1} j,j} \right),$$

where k_a is the ath location traversed along the route. Matrix algebra implies that the expression inside the parentheses is equivalent to the $(i,j)th$ element of the matrix $\mathbf{A_{\boldsymbol{j}}^K}$, which is the matrix $\mathbf{A_{\boldsymbol{j}}}$ to the power of K. Therefore, the following is true:

$$\sum_{r \in R_{ij}^{K \leq K^{max}}} \left(\prod_{k=1}^{K} a_{r_{k-1} r_k, j} \right) = \sum_{K=0}^{K^{max}} A_{ij,\boldsymbol{j}}^K,$$

where $A_{ij,\boldsymbol{j}}^K$ is the (i,j) element of matrix $\mathbf{A_{\boldsymbol{j}}^K}$.

Therefore, the probability of commuting from residence i to workplace j is:

$$\pi_{ij} \equiv \frac{b_{ij,\boldsymbol{j}} \times E_i \bar{a}_i^{\theta} q_{r_i}^{-\theta(1-\beta)} \times G_j w_j^{\theta}}{\sum_i \sum_j b_{ij,\boldsymbol{j}} \times E_i \bar{a}_i^{\theta} q_{r_i}^{-\theta(1-\beta)} \times G_j w_j^{\theta}},$$

$$\text{where} \quad b_{ij,\boldsymbol{j}} = \sum_{K=0}^{K^{max}} A_{ij,\boldsymbol{j}}^K.$$

A2: Congestion

This section derives the expression for traffic Ξ, which is defined as the total number of times a road segment is traversed by city residents. Define π_{ij}^{xy} as the expected number of times the road segment that directly connects locations x and y is traversed by a person commuting from i to j. Also, define $R_{ij}^{K<K^{max}}$ as the set of all the possible driving routes from i to j with length $K \leq K^{max}$. Then, π_{ij}^{xy} is expressed as the product of the probability that a particular route r is used conditional on commuting from i to j, which is inside the bracket and the number of times the route r traverses the link, n_r^{xy}.

$$\pi_{ij}^{xy} = \sum_{r \in R_{ij}^{K \leq K^{max}}} \left[\frac{\pi_{ijr}}{\sum\limits_{r' \in R_{ij}^{K \leq K^{max}}} \pi_{ijr'}} \right] n_r^{xy}$$

$$\approx \sum_r \left[\frac{\prod_{k=1}^K t_{r_{k-1,k}}^{-\theta} \exp\left(-\theta \frac{p_{r_{k-1},r_k}}{w_j}\right)}{\sum_r \prod_{k=1}^K t_{r_{k-1,k}}^{-\theta} \exp\left(-\theta \frac{p_{r_{k-1},r_k}}{w_j}\right)} \right] n_r^{xy}$$

$$\approx \frac{1}{b_{ij,\boldsymbol{j}}} \sum_r \left[\prod_{k=1}^K t_{r_{k-1},r_k}^{-\theta} \exp\left(-\theta \frac{p_{r_{k-1},r_k}}{w_j}\right) n_r^{xy} \right].$$

The second line in the equation above is obtained by substituting the expression for the commuting probability π_{ijr}. Define the matrix $\mathbf{A}_{\boldsymbol{j}} \equiv \left[a_{kl,j}\right] \equiv \left[t_{kl}^{-\theta} \exp\left(-\theta \frac{p_{kl}}{w_j}\right)\right]$. Using the fact that for any driving routes of length K, a person can traverse at any length $S \leq K-1$, I rewrite the equation as:

$$\pi_{ij}^{xy} \approx \frac{1}{b_{ij,\boldsymbol{j}}} \sum_{K=0}^{K^{max}} \sum_{S=0}^{K-1} \left[\left(\sum_{r \in R_{ix}^S} \prod_{n=1}^{S} a_{r_{n-1} r_n, j} \right) \times a_{xy,j} \times \left(\sum_{r \in R_{yj}^{K-S-1}} \prod_{n=1}^{K-S-1} a_{r_{n-1} r_n, j} \right) \right],$$

where R_{ix}^S denotes the set of all routes from i to x of length S, and R_{yj}^{K-S-1} denotes the set of all routes from y to j of length $K - S - 1$.

$$\pi_{ij}^{xy} = \frac{1}{b_{ij,j}} \sum_{K=0}^{K^{max}} \sum_{S=0}^{K-1} \Bigg[\left(\sum_{n_1=1}^{N} ... \sum_{n_{S-1}=1}^{N} a_{in_1,j} ... \times a_{n_{S-1}x,j} \right) \times a_{xy,j} \times \left(\sum_{n_1=1}^{N} ... \sum_{n_{K-S-1}=1}^{N} a_{yn_1,j} ... \times a_{n_{K-S-1}j,j} \right) \Bigg].$$

Matrix algebra implies that the expression in the first parenthesis is equivalent to the (i, x)th element of an adjacency $\mathbf{A_j}$ to the power of S, and the expression in the second parenthesis is equivalent to the (y, j)th element of an adjacency matrix $\mathbf{A_j}$ to the power of $K - S - 1$.

$$\pi_{ij}^{xy} = \frac{1}{b_{ij,j}} \sum_{K=0}^{K^{max}} \sum_{S=0}^{K-1} A_{ix,j}^{S} \times t_{xy}^{-\theta} \exp(-\theta \frac{p_{xy}}{w_j}) \times A_{yj,j}^{K-S-1}.$$

The total traffic between locations x and y is the expected number of times an individual traverses the road segment connecting locations x and y multiplied by the total number of commuters, summed across all origin and destination pairs.

$$\Xi_{xy} = \sum_{i,j} \pi_{ij}^{xy} \times \pi_{ij} \times \bar{L}$$

$$= \sum_{i,j} \left[\frac{\pi_{ij}}{b_{ij,j}} \sum_{K=0}^{K^{max}} \sum_{S=0}^{K-1} A_{ix,j}^{S} \times t_{xy}^{-\theta} \exp(-\theta \frac{p_{xy}}{w_j}) \times A_{yj,j}^{K-S-1} \times \bar{L}. \right].$$

A3: Total Residential Rents

The total residential rents collected by absentee landlords in each location must satisfy the following equation, where H_{r_i} is the available residential floor space and q_{r_i} is the price for residential floor space.

$$\begin{aligned}
H_{r_i} q_{r_i} =& (1-\beta) E_{j,r}[w|i] \\
=& (1-\beta) \sum_j \sum_r \left(\pi_{ijr} \left(w_j - \sum_{k=1}^{K} p_{r_{k-1},r_k} \right) \right) \\
\approx& (1-\beta) \sum_j \sum_r \left(\pi_{ijr} w_j \prod_{k=1}^{K} exp(-\frac{p_{r_{k-1},r_k}}{w_j}) \right) \\
\approx& (1-\beta) \sum_j w_j \left(\frac{\tilde{b}_{ij,j} \times E_i \bar{a}_i^{\theta} q_{r_i}^{-\theta(1-\beta)} \times G_j w_j^{\theta}}{\sum_i \sum_j b_{ij,j} \times E_i \bar{a}_i^{\theta} q_{r_i}^{-\theta(1-\beta)} \times G_j w_j^{\theta}} \right) \\
\approx& (1-\beta) \sum_j w_j \pi_{ij} \times \frac{\tilde{b}_{ij,\boldsymbol{j}}}{b_{ij,\boldsymbol{j}}} \quad \text{where} \quad \tilde{b}_{ij,\boldsymbol{j}} = \sum_{K=0}^{K^{max}} \tilde{A}_{ij,\boldsymbol{j}}^{K}
\end{aligned}$$

$\tilde{A}_{ij,\boldsymbol{j}}$ is the $(i,j)th$ element of matrix $\tilde{\mathbf{A}}_{\boldsymbol{j}} = \left[t_{mn}^{-\theta} \exp\left((-1-\theta) \frac{p_{mn}}{w_j}) \right) \right]$. As a result, I calculate total residential rents based on observed wages, toll costs, and commute times.

3.10 Appendix B: Counterfactual Procedure

B1: Counterfactual Algorithm

From the observed equilibrium with congestion pricing, I use exact hat algebra to simulate a conversion of HOT lanes back to non-toll lanes. I assume that residential amenity, the mean utilities at residences and workplaces are exogenous: $\hat{a}_i = \hat{E}_i = \hat{G}_j = 1$. Also, I assume that a city is closed with a fixed population: $\hat{\bar{L}} = 1$. Define $\hat{x} = \frac{x^{'}}{x}$ where $x^{'}$ is the counterfactual value. I simulate the initial shock of eliminating toll costs by setting $\hat{p}_{ij} = 0$. The system of counterfactual equations are as follows:

$$\hat{q}_{r_i} = \frac{(1-\beta)\sum_j w_j \hat{w}_j \pi_{ij} \hat{\pi}_{ij}}{H_i q_{r_i}} \times \hat{\bar{L}} \quad \text{where} \quad H_{r_i} q_{r_i} = (1-\beta) E_{j,r}[w|i] \tag{1}$$

$$\hat{w}_j = \left(\frac{\sum_i \pi_{ij}}{\sum_i \pi_{ij} \hat{\pi}_{ij}} \right)^{1-\alpha} \times \hat{\bar{L}}^{\alpha-1} \tag{2}$$

$$\hat{b}_{ij,\boldsymbol{j}} = \frac{\sum\limits_{K=0}^{K^{max}} \tilde{A}^K_{ij,\boldsymbol{j}}}{\sum\limits_{K=0}^{K^{max}} A^K_{ij,\boldsymbol{j}}} \tag{3}$$

$$\text{where} \;\; \tilde{\mathbf{A}}_{\boldsymbol{j}} \equiv \left[t^{-\theta}_{mn} \hat{t}^{-\theta}_{mn} \exp\left(-\theta \frac{p_{mn} \hat{p}_{mn}}{w_j \hat{w}_j} \right) \;\; \mathbf{A}_j \equiv \left[t^{-\theta}_{mn} \exp\left(-\theta \frac{p_{mn}}{w_j} \right) \right] \right.$$

$$\hat{\pi}_{ij} = \frac{\hat{b}_{ij,\boldsymbol{j}} \times \hat{q}_{r_i}^{-\theta(1-\beta)} \times \hat{w}^{\theta}_j}{\sum_i \sum_j \pi_{ij} \times \hat{b}_{ij,\boldsymbol{j}} \times \hat{q}_{r_i}^{-\theta(1-\beta)} \times \hat{w}^{\theta}_j} \tag{4}$$

$$\hat{\Xi}_{xy} = \frac{\sum_{ij}\left[\sum_{K=0}^{K^{max}}\sum_{S=0}^{K-1} A_{ix,\boldsymbol{j}}^{S}\hat{A}_{ix,\boldsymbol{j}}^{S} \times t_{xy}^{-\theta}\hat{t}_{xy}^{-\theta}\exp\left(-\theta\frac{p_{xy}\hat{p}_{xy}}{w_j\hat{w}_j}\right) \times A_{yj,\boldsymbol{j}}^{K-S-1}\hat{A}_{yj,\boldsymbol{j}}^{K-S-1} \times \frac{\pi_{ij}\hat{\pi}_{ij}\hat{\hat{L}}}{b_{ij,j}\hat{b}_{ij,j}}\right]}{\Xi_{xy}} \tag{5}$$

$$\hat{t}_{ij} = \hat{\Xi}_{ij}^{\rho_0\rho_1} \tag{6}$$

Simulation

1. Guess the initial values of $\{\hat{q}_{r_i}^0\}, \{\hat{w}_j^0\}, \{\hat{t}_{ij}^0\}$. I set them all as 1.

2. Evaluate the initial values of $\{\hat{b}_{ij,j}^0\}, \{\hat{\pi}_{ij}^0\}, \{\hat{\Xi}_{ij}^0\}$ at $\{\hat{t}_{ij}^0\}, \{\hat{w}_j^0\}, \{\hat{q}_{r_i}^0\}$ using equations (4), (5), (6).

3. Main loop for $k \in \{1, 2, 3, ...\}$
   (a) Evaluate equation (1) at $\{\hat{\pi}_{ij}^{k-1}\}, \{\hat{w}_j^{k-1}\}$ to define $\{\hat{q}_{r_i}^{temp}\}$.

   (b) Evaluate equation (2) at $\{\hat{\pi}_{ij}^{k-1}\}$ to define $\{\hat{w}_j^{temp}\}$.

   (c) Evaluate equation (3) at $\{\hat{t}_{ij}^{k-1}\}, \{\hat{w}_j^{temp}\}$ to define $\{\hat{b}_{ij,\boldsymbol{j}}^{temp}\}$.

   (d) Evaluate equation (4) at $\{\hat{b}_{ij,\boldsymbol{j}}^{temp}\}, \{\hat{w}_j^{temp}\}, \{\hat{q}_{r_i}^{temp}\}$ to define $\{\hat{\pi}_{ij}^{temp}\}$

   (e) Evaluate equation (5) at $\{\hat{t}_{ij}^{temp}\}, \{\hat{b}_{ij,\boldsymbol{j}}^{temp}\}$ to define $\{\hat{\Xi}_{ij}^{temp}\}$

   (f) Evaluate equation (6) at $\{\hat{\Xi}_{ij}^{temp}\}$ to define $\{\hat{t}_{ij}^{temp}\}$

   (g) Define $\hat{x}^t = \delta\hat{x}^{t-1} + (1-\delta)\hat{x}^{temp}$ for $x \in \{w_j, q_{r_i}, t_{ij}, \pi_{ij}, \Xi_{ij}, b_{ij,\boldsymbol{j}}\}$

(h) Stop if Δ is sufficiently small.

$$\Delta = \sum_j \left|\hat{w}_j^t - \hat{w}_j^{t-1}\right| + \sum_i \left|\hat{q}_{r_i}^t - \hat{q}_{r_i}^{t-1}\right| + \sum_i \left|\hat{t}_{ij}^t - \hat{t}_{ij}^{t-1}\right|$$
$$+ \frac{1}{N}\sum_{i,j} \left|\hat{\pi}_{ij}^t - \hat{\pi}_{ij}^{t-1}\right| + \frac{1}{N}\sum_{i,j} \left|\hat{\Xi}_{ij}^t - \hat{\Xi}_{ij}^{t-1}\right| + \frac{1}{N^2}\sum_{i,j} \left|\hat{b}_{ij,\boldsymbol{j}}^t - \hat{b}_{ij,\boldsymbol{j}}^{t-1}\right|$$

The change in the expected utility $\hat{\bar{U}}$ in a closed city is:

$$\bar{U} \approx \left[\frac{\sum_i \sum_j b_{ij,\boldsymbol{j}} \hat{b}_{ij,\boldsymbol{j}} \times \tilde{E}_i q_{r_i}^{-\theta(1-\beta)} \hat{q}_{r_i}^{-\theta(1-\beta)} \times G_j w_j^\theta \hat{w}_j^\theta}{\sum_i \sum_j b_{ij,\boldsymbol{j}} \times \tilde{E}_i q_{r_i}^{-\theta(1-\beta)} \times G_j w_j^\theta}\right]^{1/\theta}.$$

B2: Geo-normalization of the Census Transportation Planning Package

The 1990 CTPP is defined by Census 1990 geographies, and the 2000 CTPP is defined by Census 2000 geographies. To construct tract-level panel data, I normalize wages, commuting flows and travel times in the 2000 CTPP to 1990 census tracts, similar to the methodology in Severen (2019).

To normalize wages, I create a crosswalk by intersecting a 2000 census block shapefile with a 1990 census tract shapefile. Using the crosswalk, I normalize 2000 CTPP block-level wages to 1990 census tracts by using areal weights, which are the percentages of 1990 census tracts intersecting with 2000 census blocks.

In the 2000 CTPP, commute flows from 1 to 7 are recorded as 4, while the remaining values are rounded to the nearest 5. However, commute flows in the 1990 CTPP are not rounded. To combine these two datasets, I convert commute flows in the 2000 CTPP from 1 to 4, divide them by 5, and round to the nearest digit. Then, I merge commute flows in the 2000 CTPP with the crosswalk twice, once with the origin and once with the destination. Finally, I sum the block-level 2000 commute flows across the intersections within the 1990 census tract. The weights are the percentages of 2000 census block pairs that intersect with 1990 census tracts. pairs.

In the 2000 CTPP, travel times for many pairs of census blocks are not reported. As a result, the number of 2000 census block pairs intersecting with 1990 census tracts varies significantly across 1990 census tracts, making it difficult to consistently combine travel times across two years. Hence, my preferred method is to use the travel times from the 2000 census block pair that intersects the most with the 1990 census tract pair.

B3: Adjacency Matrix $\mathbf{A}_j$

Each element of an adjacency matrix is a function of time costs and toll costs relative to wages in workplace j.

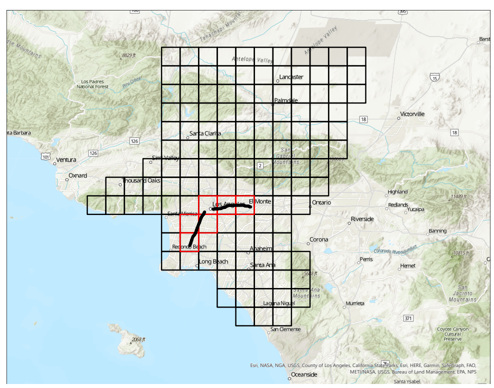
$$\mathbf{A}_j \equiv \left[t_{mn}^{-\theta} \exp(-\theta \frac{p_{mn}}{w_j}) \right]$$

To construct travel times between adjacent grids that are connected by non-toll lanes, I use the HERE routing API. I assume that travel times between non-adjacent grids are infinite: $t = \infty$

Unfortunately, I am not aware of any routing APIs that provide travel times on High-Occupancy Toll (HOT) lanes. As a result, as shown in Figure B1, I first create a shapefile consisting of HOT lanes on I-10 and I-110. Then, I append real-time traffic speed data on HOT lanes in 2019, which is available from the Performance Measurement System (PeMS). I use this shapefile to calculate travel times between adjacent grids that are connected by HOT lanes, as outlined in red Figure B1.

Finally, I construct an adjacency matrix $\mathbf{A}_j$ by combining travel times, wages, and toll costs.

Figure B1: Grids With Direct Access to HOT lanes



The map depicts the locations of 6 grids (in red) with direct access to HOT lanes. Black lines represent HOT lanes of I-10 and I-110.

B4: Traffic Matrix Ξ

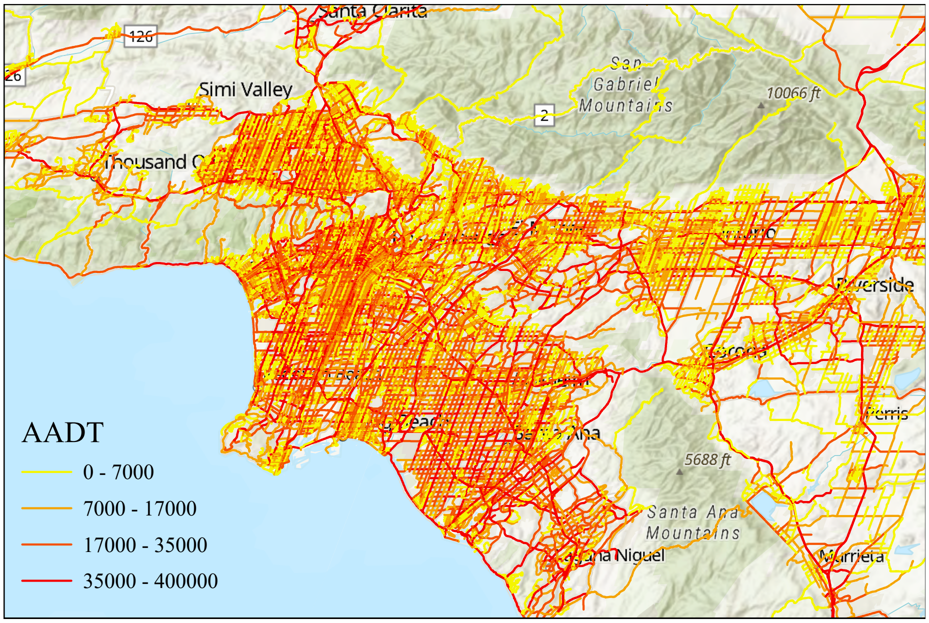
This section discusses details about constructing a traffic matrix Ξ, which follows the method used in Allen and Arkolakis (2022).

The Highway Performance Monitoring System (HPMS) Public Release of Geospatial Data in shapefile format is the primary data set used for constructing a road network in California. It includes information on the number of lanes, Average Annual Daily Traffic (AADT), speed limits, and facility types for road segments along highways and local roads. For some road segments that are parts of dual lanes, AADT, speed limits or the number of the lanes are missing. I impute missing values from their parallel counterparts. In addition, I manually fix all disconnections in ArcGIS Pro.

I define "road sections" as any parts of road not crossed or interrupted by another road. To construct "road sections", I use *"dissolve"* and *"planarize"* tools in ArcGIS pro. Then, for each "road section", I calculate the total Vehicle Miles Traveled (VMT) as the weighted sum of AADT $\times$ Distance. I also compute the number of lanes and speed limit as the weighed averages. Unimpeded driving times are calculated by dividing the length of the road section by the speed limit. Using the cleaned data, I construct a road network in California using the *"create a network dataset"* tool in ArcGIS Pro.

Using the road network in California, I create a traffic flow matrix between adjacent grids; for non-adjacent grids, it takes a value of 0. For adjacent grid pairs, I first solve for the least cost path that minimizes unimpeded travel time between centroids using the *"Origin-Destination Matrix"* tool in ArcGIS Pro. Then, for each least cost path, I sum VMT across all intersecting "road sections" and divide it by the total length of the path to obtain traffic flow values, as shown in Figure B2 below.

Figure B2: The Shapefile of Annual Average Daily Traffic



The figure displays the Shapefile of Annual Average Daily Traffic on "road sections" in Los Angeles Area. I use the Shapefile to construct the road network, and to find the least cost paths between pairs of grids that are directly adjacent to each other.