Evaluating Alternative Strategies for Traffic Reduction in Los Angeles

September 2023

A Research Report from the National Center for Sustainable Transportation

Antonio M. Bento, University of Southern California

Jonathan D. Hall, University of Toronto

Kilian Heilmann, Lyft Inc.

TECHNICAL REPORT DOCUMENTATION PAGE

1. Report No.	**2. Government Accession No.**	**3. Recipient's Catalog No.**
NCST-USC-RR-23-28	N/A	N/A

4. Title and Subtitle
Evaluating Alternative Strategies for Traffic Reduction in Los Angeles

5. Report Date
September 2023

6. Performing Organization Code
N/A

7. Author(s)
Antonio M. Bento, Ph.D., https://orcid.org/0000-0003-1309-4483
Jonathan Hall, Ph.D., https://orcid.org/0000-0003-3835-9097
Kilian Heilmann, Ph.D., https://orcid.org/0000-0002-1364-9576

8. Performing Organization Report No.
N/A

9. Performing Organization Name and Address
University of Southern California
METRANS Transportation Consortium
University Park Campus, VKC 367 MC:0626
Los Angeles, California 90089-0626

10. Work Unit No.
N/A

11. Contract or Grant No.
USDOT Grant 69A3551747114

12. Sponsoring Agency Name and Address
U.S. Department of Transportation
Office of the Assistant Secretary for Research and Technology
1200 New Jersey Avenue, SE, Washington, DC 20590

13. Type of Report and Period Covered
Final Research Report (August 2021 – August 2022)

14. Sponsoring Agency Code
USDOT OST-R

15. Supplementary Notes
DOI: https://doi.org/10.7922/G2JW8C6H
Dataset DOI: https://doi.org/10.7910/DVN/VMJIRX

16. Abstract
Traffic congestion is a major problem in large cities worldwide. This project uses high-frequency data from the Los Angeles metropolitan area combined with an instrument that varies spatially and temporally to estimate the causal impact of an additional vehicle mile traveled on travel times. Specifically, the research team exploits the network structure of the Los Angeles highway system and uses crashes on close alternative routes as exogenous shocks to traffic demand. To do so, the team relies on Google Maps to determine the ideal route and alternatives for over 19,000 real-world commutes. The researchers estimate that at peak times an additional trip reduces speed by, on average, 0.22%. They find the optimal toll at peak times is 33 cents per mile, with the toll being lower, even zero, off-peak. The researchers show how this toll varies over space and time, as well as report on its distributional effects. This toll would more than double highway speeds during peak times and only requires reducing vehicle miles traveled (VMT) at the peak by 10%. The resulting social welfare gains are over two billion dollars per year.

17. Key Words	**18. Distribution Statement**
Big Data, Congestion, Tolls	No restrictions.

19. Security Classif. (of this report)	**20. Security Classif. (of this page)**	**21. No. of Pages**	**22. Price**
Unclassified	Unclassified	43	N/A

About the National Center for Sustainable Transportation

The National Center for Sustainable Transportation is a consortium of leading universities committed to advancing an environmentally sustainable transportation system through cutting-edge research, direct policy engagement, and education of our future leaders. Consortium members include: University of California, Davis; California State University, Long Beach; Georgia Institute of Technology; Texas Southern University; the University of California, Riverside; the University of Southern California; and the University of Vermont. More information can be found at: ncst.ucdavis.edu.

Disclaimer

Acknowledgments

This study was funded, partially or entirely, by a grant from the National Center for Sustainable Transportation (NCST), supported by the U.S. Department of Transportation (USDOT) through the University Transportation Centers program. The authors would like to thank the NCST and the USDOT for their support of university-based research in transportation, and especially for the funding provided in support of this project.

NCST

Evaluating Alternative Strategies for Traffic Reduction in Los Angeles

A National Center for Sustainable Transportation Research Report

September 2023

Antonio M. Bento, Sol Price School of Public Policy, University of Southern California

Jonathan Hall, Department of Economics, University of Toronto

Kilian Heilmann, Lyft Inc.

[page intentionally left blank]

TABLE OF CONTENTS

EXECUTIVE SUMMARY iv

1. Introduction 1

2. A Simple Model of Traffic Congestion 4

3. Data, Methods, and Empirical Strategy 5

  3.1 Data 5

  3.2 Empirical Challenges 7

  3.3 Methods 10

  3.4 Empirical Strategy 10

  3.5 Descriptive Statistics 14

4. Implementation of Empirical Strategy 17

  4.1 Constructing the Instrument 17

  4.2 Functional form 18

5. Results 20

  5.1 First-stage Results 20

  5.2 Second-stage Results 20

  5.3 Heterogeneity 22

6. Policy Counterfactuals 24

  6.1 Temporal and Spatial Substitution of Travel Demand 24

  6.2 The Marginal External Costs of Congestion and Optimal Pricing 26

  6.3 Effects of a Pigouvian Toll 29

7. Conclusions 30

References 31

Data Summary 33

List of Tables

Table 1. Descriptive statistics... 14

Table 2. First-stage regression estimates. 20

Table 3. Second-stage regression results 21

Table 4. Second-stage regression by freeway route.. 23

List of Figures

Figure 1. Standard economic model for congestion. 4
Figure 2. Study area. 5
Figure 3. Average speed standard deviation over time. 6
Figure 4. Frequency of traffic incidents. 6
Figure 5. Fundamental traffic relationships. 8
Figure 6. Example of Google Maps driving directions. 11
Figure 7. Roads used in driving directions. 12
Figure 8. Example of instrument construction. 12
Figure 9. Mean weekday travel speed (mph) by link. 15
Figure 10. System-wide excess capacity by hour of day. 25
Figure 11. Distribution of excess capacity by period. 26
Figure 12. Estimated traffic relationships. 27
Figure 13. Estimated marginal external cost of congestion. 28
Figure 14. Policy counterfactual. 29

Evaluating Alternative Strategies for Traffic Reduction in Los Angeles

EXECUTIVE SUMMARY

Even if post COVID-19 employers provide increased opportunities for telecommuting, Los Angeles traffic will likely continue to be a major problem. Prior to the pandemic, Angelinos spent 104 hours stuck in traffic each year. For a typical worker, this is equivalent to a total loss of 13 working days in a year. In total, the estimates of the social cost of traffic congestion in Los Angeles add up to $9.7 billion dollars per year, or $2,408 per driver. Recently, Los Angeles METRO announced that it is exploring a new approach to tackle congestion, considering congestion pricing coupled with more high-quality transportation options. Several proposals have been discussed: one option consists of introducing a cordon toll around downtown L.A. Another option includes reducing traffic between the L.A. basin and the San Fernando Valley. Yet another would focus on reducing traffic along the I-10 corridor between Santa Monica and downtown L.A.

Using big-data from a rich network of detectors located on all freeways in Los Angeles that measure in real-time speed and flow, this project relies on statistical methods and visualization tools to develop a practical tool for policymakers to infer the effects of alternative strategies for reducing traffic congestion in Los Angeles. Specifically, it provides insights into questions such as: Given typical origin/destination pairs and routes that drivers follow on their journey to work at peak periods, how many vehicles would have to be removed for traffic congestion to be fully eliminated in Los Angeles? What is the resulting level of the tolls needed, and how should these tolls vary spatially and along the hours of the day? The project also sheds light on current proposals being considered by METRO (and other agencies) and contrasts them against this more comprehensive approach to metropolitan-wide congestion. This contrast will help to identify potential unintended effects of some of the more spatially targeted proposals being considered and provide inputs to improve their design. For example, with localized congestion pricing in specific areas, what happens to the traffic that gets displaced elsewhere? Should public transit investments be channeled to different areas to minimize potential unintended effects of traffic displacement? And, if so, how?

Unlike other attempts to measure congestion in cities, this proposed framework offers various advantages. First, because the framework will provide estimates of the speed/flow relationships in real-time at a temporally and spatially disaggregated level, one is able to calculate exactly the number of vehicles needed to be removed at different times from specific routes and locations so that congestion is eliminated. In the future, our estimates should be combined with spatially-explicit data that reflects different key aspects of the existing public transit network, so as to infer the capability of the existing public transit network to absorb such trips, as well as identify locations where investments in public transit would yield the greatest benefits. Given the recent Measure M, knowledge of the returns of public transit investments at different locations is of great importance.

Finally, our proposed framework informs the design of tolls. The estimation of the speed/flow relationships and congestion at a highly disaggregated way leads itself to the calculation of the toll needed to be charged at different times of the day and locations, so that the external costs of congestion are eliminated. Future work, by matching socio-economic data at the census-block level with route and our location-specific estimates of the speed/flow relationships and congestion, can inform how the tax burden vis-à-vis benefit that a congestion toll strategy would generate on different groups of the population.

1. Introduction

Traffic congestion is a perennial issue for cities worldwide. Not only do households spend a large share of their income on transportation but they also spend a considerable share of their time commuting. This time cost of travel is exacerbated by traffic congestion. For example, in Los Angeles, drivers spend on average 104 hours stuck in traffic each year. For a typical worker, this is equivalent to a total loss of 13 working days in a year. Earlier estimates of the social cost of traffic congestion in Los Angeles suggest that congestion costs add up to $9.7 billion dollars per year, or $2,408 per driver.

Since Vickrey (1969), economists routinely think of traffic congestion as one of the largest unpriced negative externalities. This externality is easy to grasp: The more drivers are using the road-system, the lower average speed will be. While drivers entering a road perceive the average driving time as their cost of travel, they do not take into account that their actions slow down other drivers. Economists have therefore recommended congestion pricing to address this issue. It is well established that alternative methods to reduce traffic congestion typically suggested by policymakers, such as road and public transit investments, partial pricing of lanes, and changes in spatial structure are not as effective as congestion pricing. For example, recent studies show that travel demand is largely unresponsive to changes in city structure and public transportation availability (Bento et al. 2005), and that increasing road capacity simply leads to induced demand (Duranton and Turner, 2011).

In this report, we explore the design of optimal congestion pricing using big data, relying on a novel instrumental variables approach. Our goal is to empirically recover the marginal external cost of traffic congestion by estimating the speed-flow relationship for freeways in the Los Angeles metropolitan area. Estimating the speed-flow relationship comes with three empirical challenges: First, since speed and flow are a result of travel demand and supply (highway capacity), an instrumental variables approach that identifies this technical relationship is needed. Second, the potential presence of hypercongestion, which occurs when the relationship between throughput and speed becomes positive (Walters (1961), Hall (2018), complicates the estimation. Finally, economic theory doesn't provide any guidance on the functional form that captures the empirical relationship between speed and flow, and it isn't a priori obvious whether this relationship can be captured by a continuous function, or whether this function differs depending on the level of flow.

We tackle these challenges by combining detailed data that includes traffic speeds and volumes measured by freeway detectors, vehicle accident counts by location, and real-time routing data scraped from Google Maps. In order to distinguish travel supply and demand relationships, we use vehicle crashes on parallel routes as plausible exogenous demand shifts, and therefore recover the causal effect of flow on speed. The idea is that a crash on one route induces some drivers to switch to other routes independently of supply shocks there. As these drivers switch routes, flow levels get altered and speeds adjust.

Our main result points to a significant negative externality of traffic congestion. To demonstrate this, we first provide convincing evidence that our instrumental variables approach is valid. In fact, we show that vehicle crashes shift travel demand to other routes and increase traffic density there by 2-3%.

Importantly, our instrumental variables approach uncovers that the marginal effect of an additional trip is not constant. That is, the speed-flow relationship is well characterized by a structural break: when traffic flow is low, adding an additional trip has little to no effect on travel times. But once the free-flow capacity of a freeway is reached, congestion builds up. At that point, adding a car to the roadway decreases average freeway speeds by roughly 3.11%. As a result, at peak times an additional trip reduces speed by, on average, 0.22%.

Accounting for the heterogeneity of the congestion externality across space and time is also important. Our results underscore that the marginal cost of an additional trip at high levels of traffic increases by more than 60%. Taken together, our instrumental variables approach implies large negative externalities of traffic congestion. And these are significantly more pronounced than the reduced form OLS estimates.

We use the estimates of the speed-flow relationship to perform several policy counterfactuals. We first show that even if we relocated traffic demand over space and time, freeway capacity in Los Angeles would not suffice to provide free flow speeds. This further points towards the need for congestion pricing, especially during peak periods.

We then calculate the optimal Pigouvian congestion charges that recover the social optimum in a static model of freeway travel. Using demand estimates from the literature, we find the optimal toll at peak times is 33 cents per mile. In contrast, tolls would be much lower, even zero, during off-peak periods. We show how this toll varies over space and time, as well as report on its distributional effects. This toll would more than double highway speeds during peak times and only requires reducing vehicle miles traveled (VMT) at the peak by 10%. The resulting social welfare gains are over two billion dollars per year.

Importantly, these results differ substantially from a model that fails to account for the structural break in the speed-density relationship. Correctly estimating the marginal external costs decreases the optimal reduction in VMT by 33%, increases the improvement in speed by 55%, decreases the increase in private costs by 50%, and increases the social welfare gains by 30%. We arrive at these results by contrasting estimates from a standard linear speed-flow relationship with our specification which allows for a structural break in this relationship.

This paper builds on a literature of estimating the negative externality of traffic congestion and optimal road pricing schemes (keeler (1977)). While there is a large literature in the field of transportation engineering that is concerned with the speed-flow relationship, this strand ignores the simultaneity of traffic demand and supply. Akbar et al. (2017), Yang et al (2020), Couture, (2018) have introduced economic rationale for the need to use an instrumental variables approach to recover the causal effect of flow on speed. to

Our approach differs from previous studies in the use of a spatially and temporally disaggregated demand instrument that allows us to quantify the causal relationship between speed and density. While Yang et al. (2020) have provided such an estimate for Beijing, China, we are to our knowledge the first ones to do so for a major metropolitan area in the US. The fine disaggregation of our data and instrumental variable further allows us to estimate speed flow relationships in different parts of the city and link them to socioeconomic characteristics of neighborhoods. We are ultimately interested in the spatial distribution of congestion externalities within cities.

The report is structured as follows. Section 2 describes the methodology by introducing the economic and empirical framework for measuring road congestion. In Section 3 we discuss the data, methods, and empirical strategy. In Section 4 we elaborate on the implementation of the empirical strategy. In Section 5 we present the results. In Section 6 we perform policy calculations, while Section 7 concludes.

While comprehensive pricing of roads is still politically disputed and technologically challenging, several cities have taken the lead and implemented some form of congestion pricing, usually through simple cordon tolling (e.g., London and Singapore). Major cities such as Los Angeles have put forward proposals for comprehensive road pricing. The smart city revolution is expected to reduce administrative costs of implementing comprehensive yet flexible pricing schemes. However, there is still very little known about the spatial and temporal variation in optimal congestion charges.

2. A Simple Model of Traffic Congestion

To help fix concepts, we start with the textbook economic model of traffic congestion. We further extend this model in Section 5. This model considers a single time period and a single road segment of length *l*. Travel speed on the road segment are a decreasing function of the number, or volume, of trips, denoted as $S(V)$. We denote the inverse demand curve as $D(V)$. All travelers have the same value of time, α. The private costs of traveling in the absence of a toll is thus $\alpha. l/S(V)$. The total social cost of V travelers using the road is $V.\alpha. l/S(V)$ and so marginal social cost of an additional traveler is $\alpha. l/S(V) - V.\alpha. lS'(V).S(V)^{-2}$. The first term in the expression for marginal social cost is the private cost born by the additional traveler and the second term is the externality, the additional cost born by all other travelers when the additional traveler increases travel times.

We can use this simple model to illustrate the congestion externality and how tolling can increase social welfare. Since travelers only consider their private costs and benefits when deciding whether to use the road, the no-toll equilibrium occurs where the inverse demand curve intersects the private cost curve. That is, when $D(V_{NT}) = \alpha l/S(V_{NT})$, as shown in Figure 1. This, unfortunately, means there are many travelers for whom the social cost of their trip exceeds their private benefits, and so there is a resulting deadweight loss. In the social optimum, the only people who travel are those whose private benefit exceeds the social marginal cost, and so the social optimum occurs where the inverse demand curve intersects the marginal social cost curve, i.e., $D(V_{SO}) = \alpha l/S(V_{SO}) - V_{SO}\alpha lS'(V_{SO}).S(V_{SO})^{-2}$.

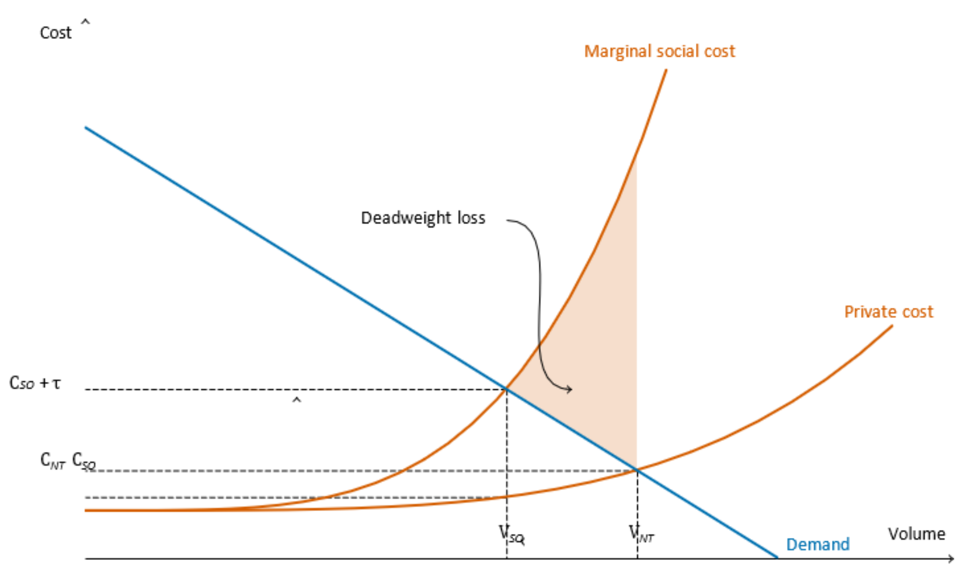


Figure 1. Standard economic model for congestion.

As illustrated in Figure 1, this reduces the volume of trips from V_{NT} to V_{SO}. Since the volume has decreased, the travel time costs fall from C_{NT} to C_{SO}. To reduce the number of travelers to V_{SO}, the social planner must impose a toll, τ , that raises the private cost to $D(V_{SO})$. Imposing this toll increases private costs but also increases social welfare by eliminating the deadweight loss and generating toll revenue.

3. Data, Methods, and Empirical Strategy

Our first goal is to estimate the causal relationship between speed and flow, represented by $S(V)$. This section presents our empirical strategy, and data for doing so. We start by introducing our primary dataset and three empirical challenges in estimating the speed-flow relationship.

3.1 Data

Our primary dataset comes from California Department of Transportation's (Caltrans) Performance Measurement System (PeMS). PeMS contains data on traffic speeds and flows from thousands of inductive loop detectors built into limited-access highways (freeways) across California. This data is collected in real-time at 30-second intervals, and then Caltrans conducts extensive tests of data quality and provides the data in raw and aggregated form. We supplement these data with data on vehicle crashes.

We use the data for 2017 covering Caltrans District 7, which consists of Los Angeles County and Ventura County. These two counties are a subset of the Los Angeles Metropolitan Statistical Area, accounting for over two-thirds of the area's total population. For the sake of brevity, we refer to this area as "Los Angeles". Because some freeways have high-occupancy vehicle lanes, we use the hourly data for mainline stations, where a station is the set of detectors in a given type of lane at a given location and going a specific direction. Loop detectors regularly fail and we only use data from the station-hour observations where all the component detectors are reporting data. We use data from 1,318 stations. Figure 2 plots the locations of the stations we use. Figure 3 plots average lane speeds across the hours of the day, while Figure 4 plots of the frequency of traffic incidents over the course of the day.

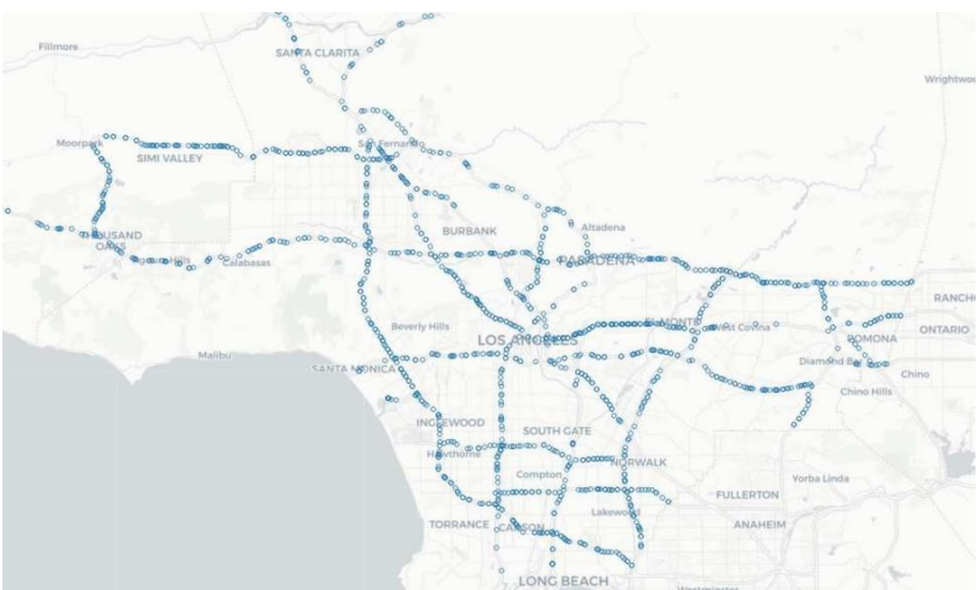


Figure 2. Study area.

Notes: This figure plots locations of in-lane loop detectors that provide data during our study period. To save space, it excludes twelve stations (sets of detectors) north of Santa Clarita.

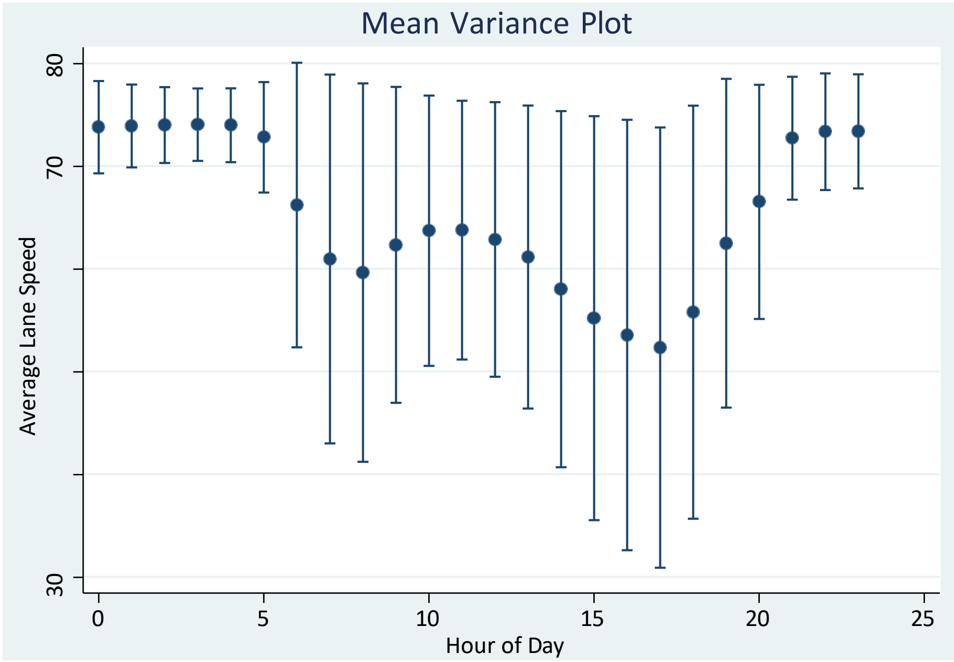


Figure 3. Average speed standard deviation over time.

Notes: This figure plots average lane speed at freeway detectors by hour of day and the associated standard errors.

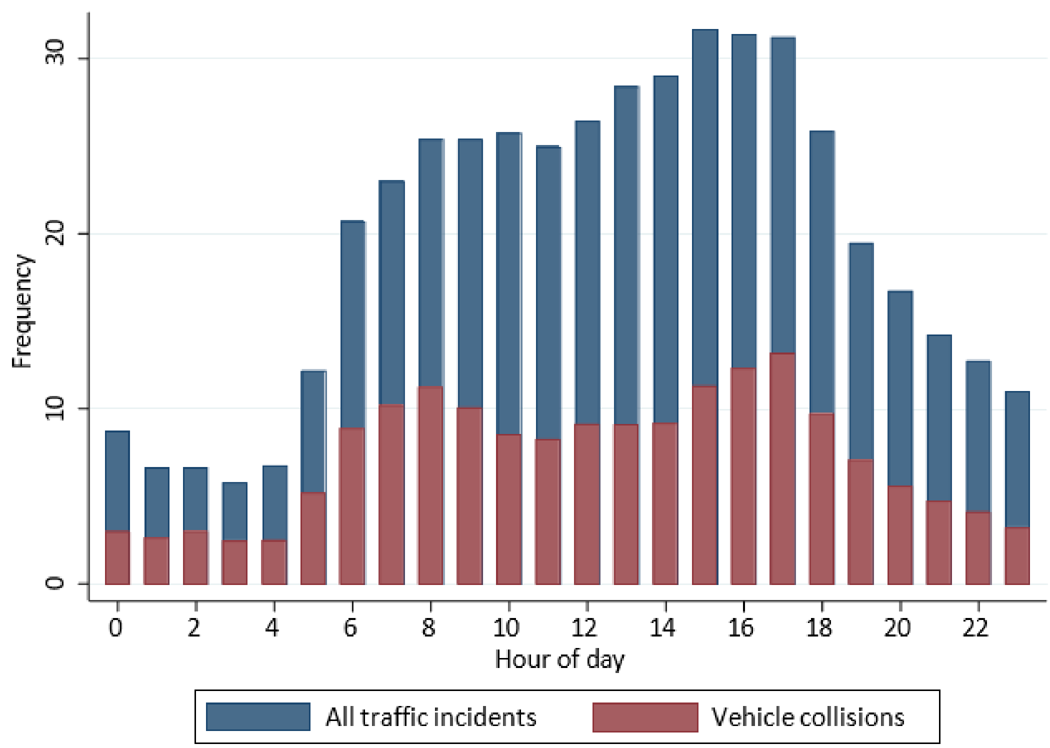


Figure 4. Frequency of traffic incidents.

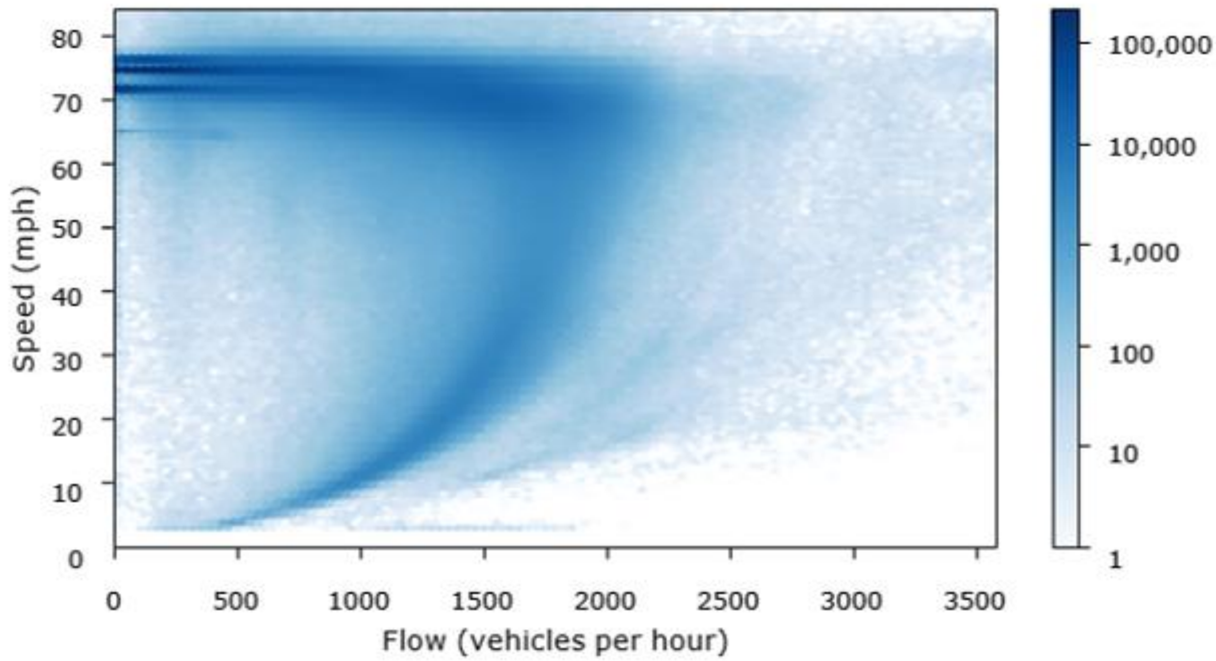
Notes: This figure plots the average frequency of traffic crashes by hour of day in our study area.

3.2 Empirical Challenges

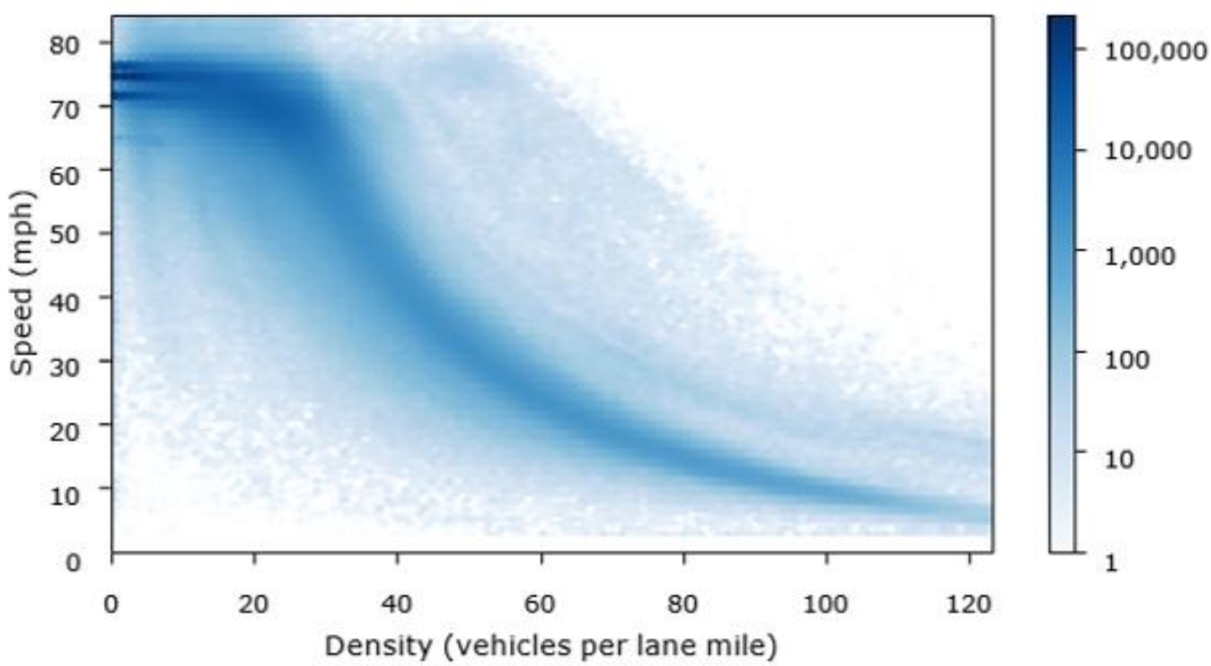
The relationship between speed, flow, and density

Figure 5a depicts the relationship between speed and flow in our data using a heatmap. It shows that speeds of 65–80 miles per hour (mph) are regularly observed with flows from 0–2,000 vehicles per hour per lane (vphpl). However, every level of flow also occurs at lower speeds. This lower portion of the speed-flow relationship is known by traffic engineers as "oversaturated flow", and by economists as "hypercongestion". When traffic flow is oversaturated, higher flows are associated with higher speeds. Oversaturated flow regularly occurs when a downstream road segment is so full of vehicles (i.e., oversaturated) that no more vehicles can fit in, and so vehicles in the upstream segment begin queuing, experiencing low speeds and flow (Verhoef, 2001; Small and Chu, 2003; Lindsey and Verhoef, 2007).

There is an ongoing debate within the literature regarding whether a causal hypercongestion relationship exists, where it is possible to increase both speeds and flow. Hall (2018, 2021) summarize extensive evidence from the traffic engineering literature on the existence of a causal relationship, while Anderson and Davis (2020) find no evidence for one of the causal mechanisms behind hypercongestion. We follow the majority of the literature by being conservative and not treating hypercongestion as a causal relationship (e.g., Lindsey and Verhoef, 2007; Yang et al., 2020).



(a) Relationship between speed and flow



(b) Relationship between speed and density

Figure 5. Fundamental traffic relationships.

Notes: These figures plot the relationship between speed and flow, and speed and density in our data. Specifically, it is a heatmap where the plane is divided into small hexagons, and each hexagon is colored according to the log10 of the number of observations it contains. To reduce figure size, we do not plot the 0.1% largest values on each dimension.

Speed-flow relationship is not single valued

The oversaturated portion of the speed-flow relationship creates our first empirical challenge. That is, the speed-flow relationship is not single valued. To address this, we follow Yang et al. (2020) and estimate the speed-density relationship and use this to calculate the speed-flow relationship. Traffic density is the number of vehicles per mile per lane, and is connected to speed (miles per hour) and flow (vehicles per hour per lane) by an identity:

$$flow = density. speed. \tag{1}$$

This approach is useful, as documented in Figure 5b, since the speed-density relationship is single valued. We use (1) to convert our estimates of the effect of density on log speed to the effects of volume on speed. Therefore 1 will form the basis for our empirical specification.

Speed/Density a result of supply and demand

Our second, and most important, empirical challenge is that observed speed and density depends on both supply and demand. This is the classic challenge in estimating a supply (or demand) curve. While we observe many different equilibrium combinations of speed and density, we do not know whether they vary due to changes is supply, demand, or both. As an example of how this can bias our results, consider a rainstorm. A rainstorm reduces drivers' demand for trips and the capacity of the freeway, which will bias our estimates of the speed-density (and thus speed-flow) relationship towards zero. To overcome this challenge, we need a source of exogenous variation in demand.

Ideal experiment and our solution

To highlight our empirical approach, it is useful to think about the perfect experiment that we try to emulate. In the perfect experiment, we would hire drivers and randomly assign them to drive on various freeway segments. We would then measure the change in speed caused by the additional trips.

Our solution to our second empirical challenge is to approximate this ideal experiment by using traffic crashes on alternate routes as a source of plausibly exogenous demand shocks. Consider, for example, a traveler who has two routes to his destination. When there is a crash on his preferred route, this worsens congestion on this route. The traveler, informed of this congestion either by radio or navigation apps such as Google Maps, instead travels on his alternate route. The crash on the preferred route thus leads to an exogenous increase in traffic flow and density on the alternative route.

This approach has the additional advantage that it solves for the causality issue when dealing with hypercongestion. Hypercongestion can lead to distortions in the speed-flow relationship even when the density transformation is used. The presence of downstream bottlenecks can lead to an increase in density and a simultaneous decrease in speed on upstream routes that is not causal. By using variation in vehicle density on the road that is caused by conditions on alternative routes and unrelated to bottleneck conditions on the current route, we eliminate this simultaneity issue.

3.3 Methods

We implement this approach using an instrumental variables approach, where a crash occurring on an alternate route serves as an instrument for measuring the effect of density on speed. Our exclusion restriction is that a crash on route *i* only effects speeds on route *j* by redirecting additional vehicles to route *j*. This also requires there to be no other factors that cause crashes on route *i* and simultaneously affect traffic speed on route *j*. One example of a violation of this assumption are unobserved weather hazards that affect both traffic speed and crashes. A rainstorm could potentially lead to crashes on the main route and also reduce speeds on the alternative. To address this possible issue, we control for an extensive set of different measurements of weather conditions. Another violation would be time of day or day of week factors that increase traffic flow on all routes, increasing crash risk and reducing travel speeds. We address this by using an extensive set of time of day, day of week, and month fixed effects. We also control for the price of gasoline as speed might be influenced by the cost of driving (Burger and Kaffine, 2009; Bento et al., 2013). To address other possible demand shocks that increase traffic flow on all routes, we also control for hourly vehicle miles traveled (VMT) in District 7. Finally, to account for differences between road segments, we include station fixed effects. Thus, our estimates compare travel speeds at a given station-hour-day of week-month when there is a crash on an alternative route to the same station-hour-day of week-month when there is not a crash.

Thus our baseline instrumental variables (IV) model is described by the following two-equation system:

$$Density_{i,t} = \gamma . Crash_{a(i),t} + X'_{i,t} . \delta + \varepsilon_{i,t} \quad (2)$$

$$Log\left(Speed_{i,t}\right) = \beta . Density_{i,t} + X'_{i,t}\theta + \eta_{i,t} \quad (3)$$

Here, $Density_{i,t}$ is the average density at station i at time t, $Crash_{a(i),t}$ is an indicator variable equal to one if there is a crash on an alternate route for segment i at time t, and $Speed_{i,t}$ is the average speed at station i at time t. In future work, we will generalize the instrument so it can be a continuous variable. That is, instead of an indicator variable, one may consider the number of crashes on alternative routes during a specific time period.

The vector $X'_{i,t}$ contains the control variables. These include a full set of station, hour, day of week, and month indicators, temperature, temperature squared, wind speed, precipitation, gas prices, and District 7 VMT.

3.4 Empirical Strategy

To implement our approach, we require data on crashes, routes and their alternatives, gas prices, and weather. We obtain geo-coded crash reports from PeMS, which reports all crashes investigated by the California Highway Patrol. The data includes the exact location, time (up to a minute), a verbal description of the incident, and the duration until the crash was resolved. We link crashes to the closest station, requiring the distance to be less than a mile, on the correct freeway and direction of travel.

Not all crashes are recorded by law enforcement officers. This under-reporting might vary by traffic density and it is not clear whether this issue is more pronounced at high or low traffic density. In either case, since we are potentially misleadingly counting “treated” stations as “non-treated,” we are introducing a negative bias and therefore our results should be seen as a lower bound of the true marginal effect.

Creating alternative routes

We gather data on routes and their alternatives in three steps. First, we obtain commuting data from the LEHD Origin-Destination Employment Statistics (LODES). This dataset provides the work and home census blocks for all workers covered by state unemployment insurance programs.[4] Second, we query the Google Maps API to find the recommended route and up to two alternative routes for a random sample of 1% of commutes (19,163 origin-destination pairs). Figure 6 shows what Google returns for a single origin-destination pair and Figure 7 shows that this 1% sample covers all of Los Angeles’ major roads. Our third step is to spatially match each route to the PeMS stations they cross. We do so by taking the polyline shapefiles for each route, calculating the non-overlapping portion of each route, and then finding the stations (in the correct direction of travel) crossed by each route. This allows us to define routes as a list of stations crossed.

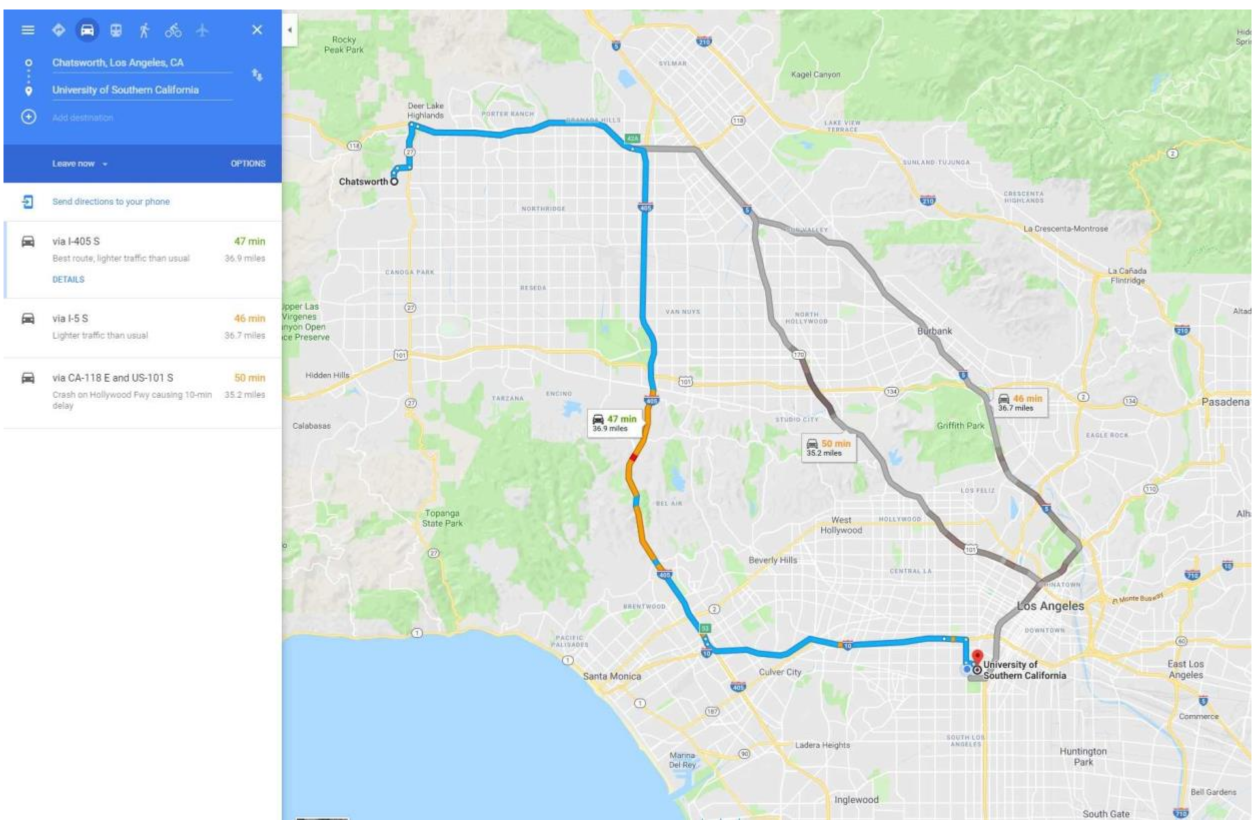


Figure 6. Example of Google Maps driving directions.

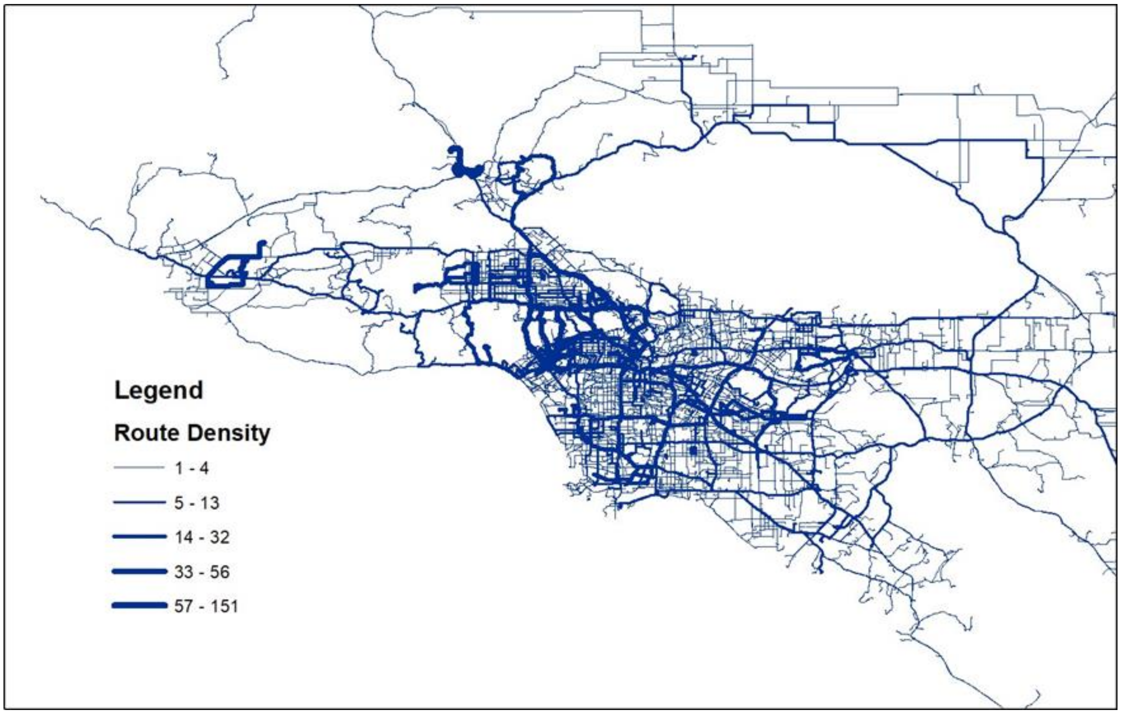


Figure 7. Roads used in driving directions.

Notes: This figure plots the roads used in the driving directions for the 1% sample of all LODES origin-destination pairs.

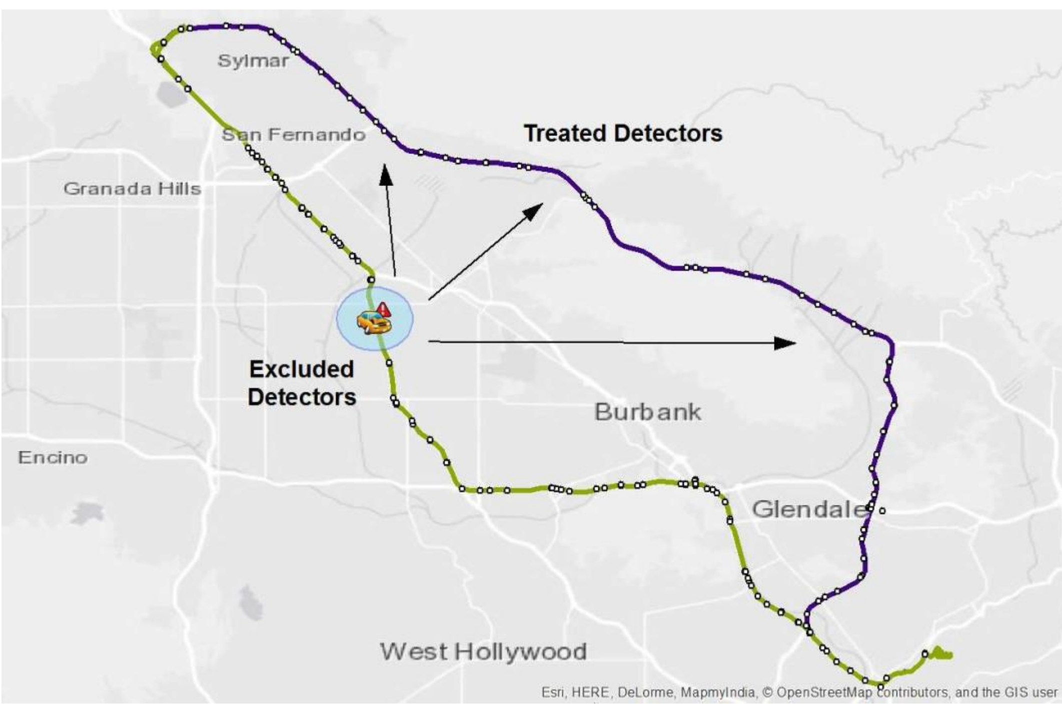


Figure 8. Example of instrument construction.

Notes: This figure shows that when a crash occurs on the green route, detectors along the blue route become treated while the detectors on the same highway link on the green route are excluded.

We define a station as treated in given hour if, during that hour, there is a crash near a station on the non-overlapping portion of an alternate route.[5] Since crashes potentially change the relationship between function of speed and density, we exclude station-hour observations that experience traffic crashes from our regressions. Specifically, we define a freeway "link" as the segment of a freeway between two highway interchanges, and exclude all stations on a link that experiences a crash. Figure 8 shows an example of how this works.

In addition, we also collect weekly retail gas prices for the Los Angeles area from the Energy Information Administration's website and gather data on hourly weather conditions from the National Oceanic and Atmospheric Administration's Integrated Surface Database (ISD) Lite Database. The ISD Lite data contains hourly measurements of temperature, rainfall, windspeed, and visibility from six weather stations in the Los Angeles metropolitan area with varying coverage. We match each station to the closest weather station for which data is provided during that hour.

The speed-density relationship is non-linear

Our third empirical challenge is that, as Figure 5b shows, that the speed-density relationship is highly non-linear. When density is below 20 vehicles per mile per lane (vpmpl), density and speed are uncorrelated, but as density grows beyond 20 vpmpl, then speed begins to fall. This empirical pattern is not unique to our data, the Highway Capacity Manual (Transportation Research Board, 2016) describes the speed-flow relationship as typically having "a range of flow rates over which speed is constant" which ranges from "a flow rate of zero to a breakpoint value" (p. 12-9). This implies a speed-density relationship that has a range of densities over which speed is constant.

Given this, our goal is to estimate the speed-density relationship when density is above the breakpoint. By breakpoint we mean the flow value after which speeds start to decline.

This requires first estimating the location of the breakpoint. We do so by estimating a series of models using ordinary least squares (OLS) where the speed-density relationship has a slope of zero below the breakpoint using the following regression specification:

$$logSpeed_{i,t} = \beta_2 Density_{i,t}.above\_breakpoint_{i,t} + X'_{i,t}\delta + \varepsilon_{i,t} \qquad (4)$$

The vector $X'_{i,t}$ contains the same extensive set of controls as in equations (2) and (3). We then determine the optimal breakpoint by choosing the breakpoint that maximizes model fit as measured by the Akaike Information Criterion (AIC). That is, we run a multitude of OLS regressions varying the level of flow that characterizes the break point and recover this break point. We also contrast our finding against the suggested guidelines for break points in the Highway Capacity Manual (Transportation Research Board, 2016).

Once we know the breakpoint, our main results reports estimates of the speed-density relationship, estimated using (2) and (3) for two subsamples: densities above or below the breakpoint. That is, we estimate two separate sets of models: We estimate the speed-density relationship below the breakpoint to validate the assumption that speed is constant when density is below the breakpoint. And we also report the estimates of the speed-flow relationship above the breakpoint. The latter are our primary empirical estimates. Importantly, a priori there is no reason to believe that the instrument would be needed for the regression below the breakpoint, since at low levels of density, there should be no relationship between density and speed.

This methodology requires that the reduced form OLS regression can inform us of the structural break even in the presence of endogenous regressors. This assumption is based on the notion that the structural break in the congestion function is driven solely by freeway capacity. While the same endogeneity concerns mentioned above are present for the uncongested part of the congestion function (i.e., weather shocks), there is no a-priori reason to believe that the endogenous regressors would distort the location of the structural break.

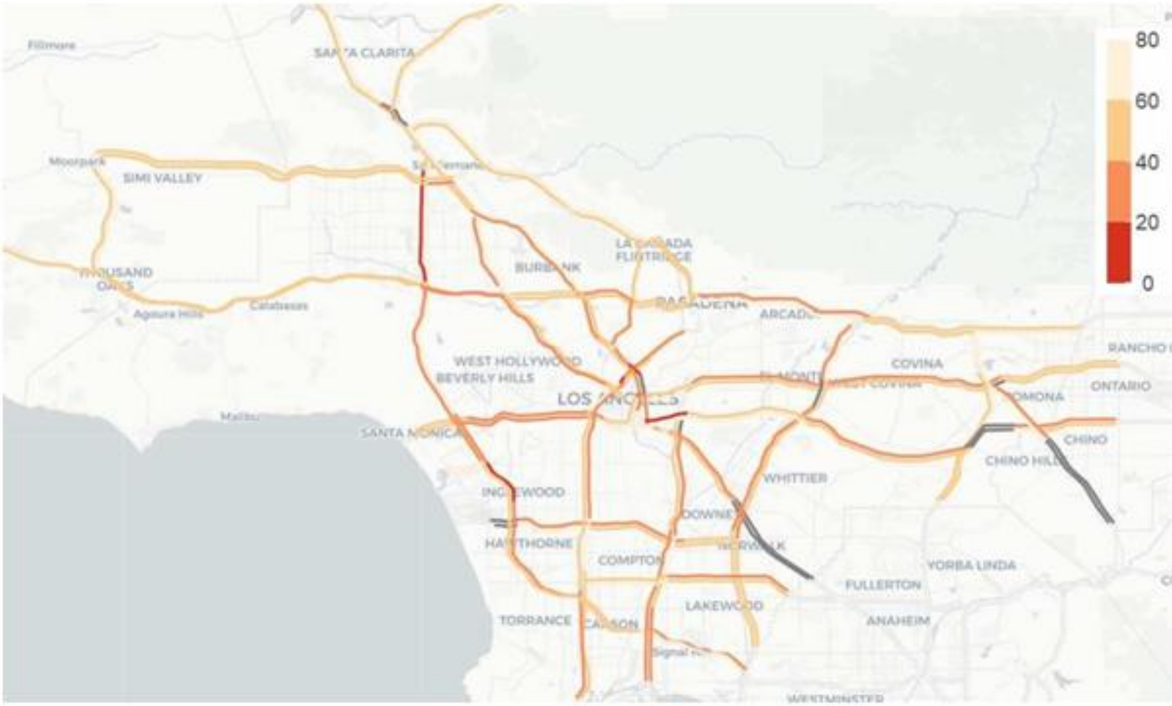
3.5 Descriptive Statistics

Table 1 shows descriptive statistics of the speed-flow data. The first thing which is noticeable in the data is that average speeds are quite high. The unweighted mean of lane speeds on all detectors actually slightly exceeds the maximum speed limit in the study area of 65mph. This average masks high heterogeneity between detectors as certain routes see little traffic for most of the time. The range of speeds varies between 3mph and 86mph. We therefore weight the detector observations by traffic flow to create a typical freeway situation faced by a Los Angeles motorist. This reduces the weight of freeways that see little traffic and therefore provide high speeds. While this weighting scheme reduces the average speed, it is still very high at slightly below 62mph.

Table 1. Descriptive statistics.

	Speeds and Crashes				
Variable	**Mean**	**Std.Dev.**	**Min**	**Max**	**Obs**
Lane Speed (mph)	60.87	17.12	3	86.5	1,362,961
Lane Flow (vehicles per hour)	1,438.30	496.76	1	5173	1,362,961
Lane Density (vehicles per mile)	28.16	18.02	0.01	189	1,362,961
Implied Daily Delay (vehicle hours)	392,989	146,331	59,565	575,902	92
Crashes by Hour	19.6	10.7	0	46	2,208
Crashes by Day	470.9	50	333	562	92

Importantly, Figure 9 makes it clear that traffic congestion is prominently a peak hour phenomenon. The figure depicts average speed and its standard deviation by hour of day. Even in the evening rush hour, the average speed for all detectors does not drop below 50mph. However, we see considerable variance in speeds during peak hour times. The slowest speed occurs during evening peak times at 5pm. The picture is consistent with general traffic patterns. The evening rush hours see lower speeds and last longer than the morning rush hours, consistent with the notion that drivers take more discretionary trips in the hours after work.



(a) 8 am

(b) 5 pm

Figure 9. Mean weekday travel speed (mph) by link.

Notes: This figure plots mean weekday (and non-holiday) travel speed (mph) for each highway link in the Los Angeles area. Travel speed calculated by dividing the length of each link by the travel time from the start to the end of the link, and then averaging across days. A link is defined as the segment between two highway interchanges. The links in grey are those with no valid data.

Crash data

We next describe the characteristic features of the crash data. In our study period, there occurred more than 14,000 incidents within our study area. This corresponds to a daily number of around 450 traffic crashes per day. The temporal distribution of crashes mirrors that of traffic conditions but is more spread out (see Figure 4). Traffic crashes are more likely to occur during daytime hours and start to increase with the onset of the morning peak at 6am. However, they do not die out during noontime and steadily increase throughout the day to the beginning of the evening peak. They decline after 7pm until the minimum is achieved at 3am at night.

Regarding the weather data, our study area experiences very little variation in weather condition during the study period of August 2017. There was no measurable rain during the time period and only 0.25% of all observations experience some trace precipitation. Similarly, wind speeds never exceed 8.2 meters/second which is equivalent to a "moderate breeze" on the Beaufort scale and is unlikely to influence traffic conditions. The only major variation in weather conditions come from changes in temperatures. These vary between 15 and 42.2 degrees Celsius and experience considerable variation over the day.

Additionally, we have very little variation in gas prices. Over our three months study period, the price of regular gasoline varied between $3.07 and $3.27 per gallon.

4. Implementation of Empirical Strategy

In this section, we outline our instrumental variable design and present evidence for the relevance of our instrument. We then proceed to report and discuss the results of our 2SLS regressions and compare them to the OLS estimates.

4.1 Constructing the Instrument

Defining alternatives

We begin by carefully describing the construction of our instrument, namely the occurrence of traffic incident on alternative routes. At first, we need to define what we mean by the term *routes*. We define routes as sets of contiguous freeway detectors that equal a typical commuting path. Since in our data we only observe origin and destination of commutes, but not the actual route taken, we use online routing platforms to determine the likely path of the commute. To do so, we query a random 1% sample (= 19,163 origin-destination pairs) of the universe of all OD commute pairs in the LODES database using Google Maps. The Google Maps API will return up to three different routes and we save these alternative pairs (and triplets respectively) as polyline shapefiles. An example of a commuting origin-destination pair and its geocoded driving route is depicted in Figure 6. Overall, the 1% sample creates driving directions that cover all major roads of the Los Angeles metropolitan area (see Figure 7).

Next, we spatially match these to traffic detectors in the same travel direction using geospatial software. We first calculate the non-overlapping part of the driving routes and then use the Google Maps driving directions to overlay them with the detector shapefile. We then assign each detector that is congruent with the driving directions on the correct direction of the freeway. This allows us to define routes as a list of detectors crossed. Formally, for each origin-destination route $k \in K$ we define up to three $i \in \{1, 2, 3\}$ sets of detectors:

$$R_{ik} = \left\{ d_j | d_j \text{ } is \text{ } along \text{ } route \text{ } K \right\}$$

Where d_j refers to the detector $j = \{1, \ldots, J\}$

This allows us to establish alternatives at the detector level. We denote these exclusive sets of detectors as:

$$E_{ik} = R_{ik} \backslash \left(U_{j \neq i} R_j \right)$$

We define detector pairs that are related if they are on segments that are alternatives to each other. That is, we rely on detectors in an alternative routes if they are located along the alternative routing suggestion from Google.

Instrument construction

We now can link the crashes data to the detectors. We first match every traffic incident report to the closest in-lane detector according to straight-line distance using the coordinates in the PeMS database. This allows us to define detectors that experienced crashes. We define detector d_{jt} as affected by a crash if it is the closest detector to the crash location on the same route and the crash time falls within the hourly time window.

For each detector d_{jt} we define $crash_{-i,t}$ as equal to one if any of its alternatives d_j is affected by a crash in time t. This is a binary indicator equal to one if a crash happened on any alternative route and does not take into account the intensive margin. In future versions of our work we will explore a continuous instrument. Substitution effects could be stronger if there are multiple crashes on alternative routes. The binary instrument will not capture this intensity but measure the average effect of at least one crash occurring. Since crashes potentially change the relationship between function of speed and density, we exclude detector-time observations that experience traffic incidents directly from the regression.

It is well known in the traffic engineering literature that an increase in traffic volume leads to a higher crash rate. We tackle this issue by including an extensive set of time fixed effects. In all specifications, we use seven day of week (DOW) and 24 hour of day (HOD) dummies as well as traffic detector fixed effects. The remaining identifying variation therefore results from changes in traffic density and speeds within the same time of day across different weeks.

4.2 Functional form

To quantify the congestion externality, we need to assure that the functional form is a good proxy of the actual relationship. Empirical evidence suggests that there is no externality at low traffic densities. Traffic engineers refer to the threshold where congestion arises as the capacity of the freeway. This creates a structural break in the relationship between speed and density as suggested by the scatterplot of the raw data.

We operationalize this structural break by splitting the sample into congested and uncongested parts. For the uncongested part, we restrict the slope to zero and then test for the optimal cutoff point. Specifically, we run OLS regressions of the form, each for a different level of vehicles per mile.

$$\log speed_{it} = \alpha_{it} + \beta_2 density_{it} * above_{it} + \beta_3 above_{it} + s_{it} \tag{5}$$

and vary the threshold from 1 to 50 vehicles per mile. That is, fifty different regressions. We then determine the optimal structural break by comparing measures of fit for each of the models using the Akaike Information Criterion. The optimal threshold that maximizes AIC is at density level of *D* = 21. In other words, 21 vehicles per mile.

In the following, we will present estimates of the congestion function for both the lower and upper part of this function. That is, we will split the sample into two subsamples: We assign observations into the congested sample if the detector-day of week-hour of day average of density is above 21 vehicles per mile. Accordingly, we will sort observations that are below that threshold into the uncongested part.

The implicit assumption of this methodology is that the reduced form OLS regression can inform us of the structural break even in the presence of endogenous regressors. This assumption is based on the notion that the structural break in the congestion function is driven solely by freeway capacity. While the same endogeneity concerns mentioned above are present for the uncongested part of the congestion function (i.e., weather shocks), there is no a-priori reason to believe that the endogenous regressors would distort the location of the structural break.

5. Results

5.1 First-stage Results

We begin by showing that our instrument is relevant and that crashes increase traffic density on parallel routes. Table 2 reports the results from regression equation 3. Our first stage regression of traffic density on the occurrence of an incident on parallel routes yields the expected positive signs and shows that there is statistically significant substitution of traffic towards alternatives in case of crashes. These results hold across the board for different controls and sample selections. To control for spatial and temporal auto-correlation of the residuals, we use two-way clustering and cluster our standard errors at both the day and freeway level (Cameron et al., 2011). The point estimate for the model that uses observations below the congestion threshold is 0.033 (column 2 in Table 2), albeit statistically insignificant. Above the congestion threshold of 21 vehicles per mile, we estimate larger and statistically significant coefficients of around 0.638. This result is robust to including weather and gas price controls and the coefficients are highly statistically significant at the 1% level. The F-statistic values for the model above the threshold suggest that we have a very powerful instrument.

Table 2. First-stage regression estimates.

	(1) No kink	(2) Below threshold	(3) Above threshold
Crash	0.620***	0.0330	0.638***
	(0.0762)	(0.0257)	(0.0932)
Detector FE	yes	yes	yes
Day of Week FE	yes	yes	yes
Hour of Day FE	yes	yes	yes
Month FE	yes	yes	yes
Controls	yes	yes	yes
Observations	5538143	3608619	1929517
R squared	0.604	0.851	0.381
F-Statistic	77.12	107.5	68.14
Sample Mean	18.67	10.02	34.85

5.2 Second-stage Results

We now move to the results of our second-stage regression describing the effect of traffic density on average speeds. In Table 3, we summarize the results of estimating regression (4) using both the OLS and the IV estimator. We estimate the average marginal effect of adding one more vehicle per mile of roadway on speed. Given the specifics of the IV approach which excludes detectors where crashes happen, the IV sample size is smaller than the full sample. In the OLS regressions, we force the sample size to be the same as in the IV approach. Again, all standard errors are clustered at both the day and the freeway level. The table presents the results for the OLS and the IV estimates below and above the threshold. We persistently find negative effects of traffic density on speed. For observations below the congestion threshold, we find significant but small effects that are close to zero as expected. In column (1) and (3), the OLS estimates yield nearly identical coefficients of -0.00420 (without controls) and -0.00419 (with controls). This implies that an increase in traffic density by one vehicle per mile of roadway decreases speeds by around half a percent. We next contrast these results with the instrumental variable regression. The IV regression results in columns (2) and (4) are more negative compared to the OLS estimates. Without controls, we estimate a β of -0.0088 and -0.0196 with controls. The standard errors in this regression are notably larger, and the latter coefficient with the controls is not statistically different from zero. Furthermore, we cannot reject the null hypothesis that the IV estimate is the same as the OLS estimate. We conclude that the congestion externality is small at low traffic densities.

Table 3. Second-stage regression results

	(1)	(2)	(3)	(4)
	OLS		OLS	IV
		Panel A: below threshold		
Lane density	-0.0042***	-0.0088**	-0.0044***	-0.0196
	(0.0002)	(0.0024)	(0.0002)	(0.0101)
Detector FE	yes	yes	yes	yes
Day of Week FE	yes	yes	yes	yes
Hour of Day FE	yes	yes	yes	yes
Controls	no	no	yes	yes
Observations	4449577	4449577	3509759	3487822
R squared	0.329	0.274	0.339	0.443
KP F-Statistic	-	37.96	-	5.41
Constant	4.32	4.36	4.30	4.41
Density Sample Mean	9.18	9.18	9.03	9.00
Speed Sample Mean	72.56	72.56	72.54	72.56
		Panel B: above threshold		
Lane density	-0.0255***	-0.0296***	-0.0244***	-0.0327***
	(0.0004)	(0.0011)	(0.0004)	(0.0024)
Detector FE	yes	yes	yes	yes
Day of Week FE	yes	yes	yes	yes
Hour of Day FE	yes	yes	yes	yes
Controls	no	no	yes	yes
Observations	2715770	2715770	2014217	2014217
R squared	0.940	0.925	0.970	0.944
KP F-Statistic	-	55.83	-	15.91
Constant	4.81	4.96	4.74	4.98
Density Sample Mean	36.14	36.14	36.81	36.81
Speed Sample Mean	49.01	49.01	48.16	48.16

Dependent Variable: Log Lane Speed
Standard errors multi-clustered at the freeway and weekly level
$^{*}p < 0.05,^{**}\ p < 0.01,^{***}\ p < 0.001$

We now focus on the congested part of the speed-density relationship. The results in Panel B show distinctively different marginal effects above the density threshold. The OLS estimates in columns (1) and (3) indicate that for each additional vehicle per mile of roadway, we observe a reduction in average speed in that lane by around 2.6%. These coefficients are statistically significant at the 0.1% significance level. In column (2) we present the result from our instrumental variable approach. The estimated coefficient is -0.0296 with a standard error of 0.001. In (4), we check for robustness to additional covariates and again control for weather variables and the price of gas. The point estimate differs only slightly and suggests that every additional vehicle per mile of roadway decreases average speeds by 3.27%.

In each specification pair, the IV estimate is always larger (in absolute values) than the OLS estimate. The differences between the estimates of the two approaches are statistically significant at conventional significance levels. The magnitude of the IV result suggests that our instrumental variable approach reduced the expected upward bias (towards zero) of the OLS estimate. This implies that the OLS reduced form estimate picks up demand shocks and underestimates the true causal impact of traffic density on freeway speeds.

5.3 Heterogeneity

In a next step, we explore the heterogeneous effects and analyze whether the speed-density relationship is stable over time and space. A potential concern is that drivers of different types sort into different routes and hours to a degree that average estimates are distorting policy implications. We use the temporal and geographic detail in our data to look at results for different sub-samples.

We first look at spatial variation of the congestion function. Freeways could experience a differential congestion technology not only due to driver selection but also due to physical differences. Although freeways are built to federal construction standards, they might potentially differ due to other geographic features. We therefore perform separate speed-density IV regressions for the four major freeway routes in our sample (Interstates 5, 10, 110, and 405).

The IV results in Table 4, Panel A do not suggest large heterogeneity of the congestion function over space. The coefficient estimates in Panel A for observations below the density threshold are all close to zero and statistically indistinguishable from another. The largest negative estimate for Interstate 5 at -.015 is imprecisely estimated and comes with a large standard error. Panel B shows the IV results for observations above the threshold. We find remarkable stability between the freeway routes and again, we cannot reject the null that the estimates for all specifications are the same at conventional significance levels. We therefore conclude that the congestion function is stable over space and does not vary between different freeway routes.

Table 4. Second-stage regression by freeway route.

	Panel A: Heterogeneity of IV estimates			
	I-10	I-5	I-110	I-405
	below threshold			
Lane Density	-0.00445*	-0.0153	-0.00506**	-0.00391
	(0.00191)	(0.0138)	(0.00150)	(0.00271)
Density Sample Mean	10.18	9.862	9.400	9.703
	above threshold			
Lane Density	-0.0323***	-0.0339***	-0.0339***	-0.0299***
	(0.00196)	(0.00215)	(0.00228)	(0.00250)
Density Sample Mean	35.17	37.74	33.71	40.19

	Panel B: Heterogeneity by time of day			
	AM Peak	Midday	PM Peak	Night
Lane Density	-0.0346***	-0.119	-0.0306***	-0.0298***
	(0.00513)	(0.700)	(0.00260)	(0.00848)
Density Sample Mean	37.95	34.08	41.85	30.52

Dependent Variable = Log Lane Speed

Standard Errors clustered at the freeway and weekly level

$*p < 0.05, ^{**} p < 0.01, ^{***} p < 0.001$

We next explore heterogeneity over time. We expect potentially heterogeneous results due to driver selection into certain hours of day. We check for divergence of the policy-relevant coefficient estimates by splitting the sample into four large periods of time. We distinguish between morning peak period (5-9am), midday off-peak (9am-4pm), evening peak (4-8pm), and the night off-peak (8pm-5am). Table 4, Panel B presents the coefficient estimates for each time block for observations above the threshold. The results are quantitatively similar to the results from the overall sample with the exception for the midday off-peak period. Here we estimate a marginal effect of -0.119 which is about 3-4 times as large as the earlier estimates. However, this estimate has a very large standard error of 0.7 and we therefore cannot reject that they are statistically the same.

Overall, we find little evidence that the congestion function differs between time and space in a way that would massively distort potential policy schemes based on these coefficient estimates. With the sole exception of the midday hours, we consistently find marginal effects on the order of 3% - 3.4% speed reduction for each additional vehicle per mile of roadway.

6. Policy Counterfactuals

In this section, we use our empirical estimates of the speed-density relationship to provide the potential gains from reshuffling traffic demand over time and space, and from optima congestion pricing schemes.

6.1 Temporal and Spatial Substitution of Travel Demand

We start with evaluating the gains from shifting travel demand over time and space. The large heterogeneity in travel speeds both over space and time displayed in Figure 7 and Figure 8 implies that travel demand is highly localized. In fact, large parts of the freeway network are uncongested during off-peak hours, and there are segments of the system that are rarely at capacity. This suggests that there are potential gains from shifting travel demand from peak hours to off-peak hours and from congested roads to underutilized roads. While the observed travel patterns in the data are a result of differential demand and indicate a higher willingness to pay to travel during peak hours, temporal substitution might be a preferable policy compared to general congestion pricing. Such temporal reshuffling of demand could be induced by e.g., relaxing work schedules. Likewise, spatial substitution of travel demand could be generated by allowing for more flexible zoning ordinances that spread economic activity more evenly over space

We explore the potential gains from substituting travel demand by studying how much substitution is needed to avoid the need for congestion pricing. In this exercise, we keep overall travel demand constant and use our estimated coefficients of the speed- density relationship to evaluate the effect of reshuffling demand on freeway speeds. For each detector and hour of day, we calculate its average "excess capacity" over our study period. This is the amount of traffic density that the detector can accommodate until it becomes congested. If the detector is congested, this value becomes negative and can be interpreted as "excess density", i.e., the amount of traffic that needs to be removed at this detector to reach free-flow speed. This allows us to judge whether all travel demand could be accommodated given the current freeway infrastructure. We start with the most flexible substitution and introduce progressively restrictive constraints on reshuffling. As we ignore a demand reaction, these scenarios will be hypothetical best-case scenarios and thus represent upper bounds on the potential of temporal and spatial substitution.

We start with the most flexible scenario and allow for both substitution over time and space: We shift traffic density between the 24 hours of the day and between individual detectors (weighted by detector length). This represents a scenario where traffic demand is completely flexible and ignores routing dependencies of individual trips. In this most flexible case, the freeway system can provide speeds of about 77mph after reshuffling. This extraordinary high speed is driven by substitution of traffic demand into the nighttime hours. Figure 10 depicts excess capacity by hour of day in relation to a free-flow speed of 65mph. If we restrict the substitution to be within major peak periods of the day, spatial reshuffling can provide speeds of 64.8mph in the morning peak (5-9am), 64.2mph during midday off-peak (9am-4pm), and 53.9mph in the evening peak period (4-8pm).

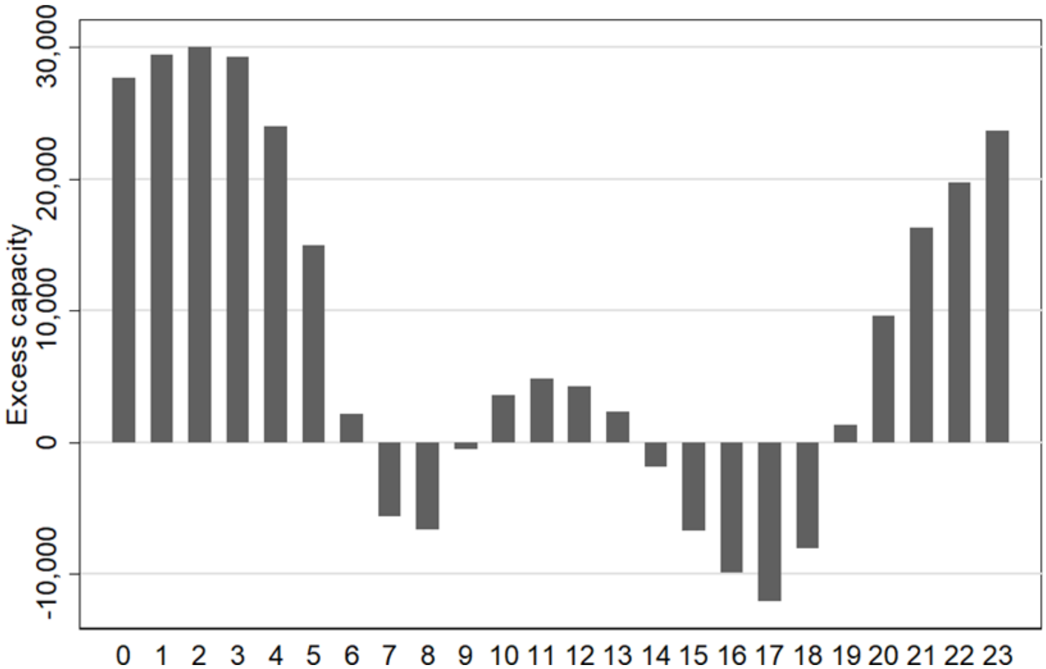


Figure 10. System-wide excess capacity by hour of day.

Notes: Total excess capacity of the freeway system if allowing for complete spatial substitution of traffic demand in relation to free-flow speed of 65mph.

We next look at more realistic scenarios and consider temporal substitution only. For this case, we treat the travel demand for each detector as fixed. This will lead to a distribution of detector excess capacity measures where 26% of detectors remain at speeds below the free-flow level of 65mph even after temporal reshuffling. We finally restrict the substitution to within peak periods. Figure 11 displays the resulting distributions of excess capacity by period. While during the night off-peak period, virtually all detectors can achieve free-flow after temporal substitution, this is not the case during daytime hours. The largest share of detectors that provide speeds below 65mph is in the evening peak as expected. Overall, for the most realistic scenario where travel demand is spatially kept fixed, merely shifting traffic within peak periods does not eliminate congestion or the need for pricing roads.

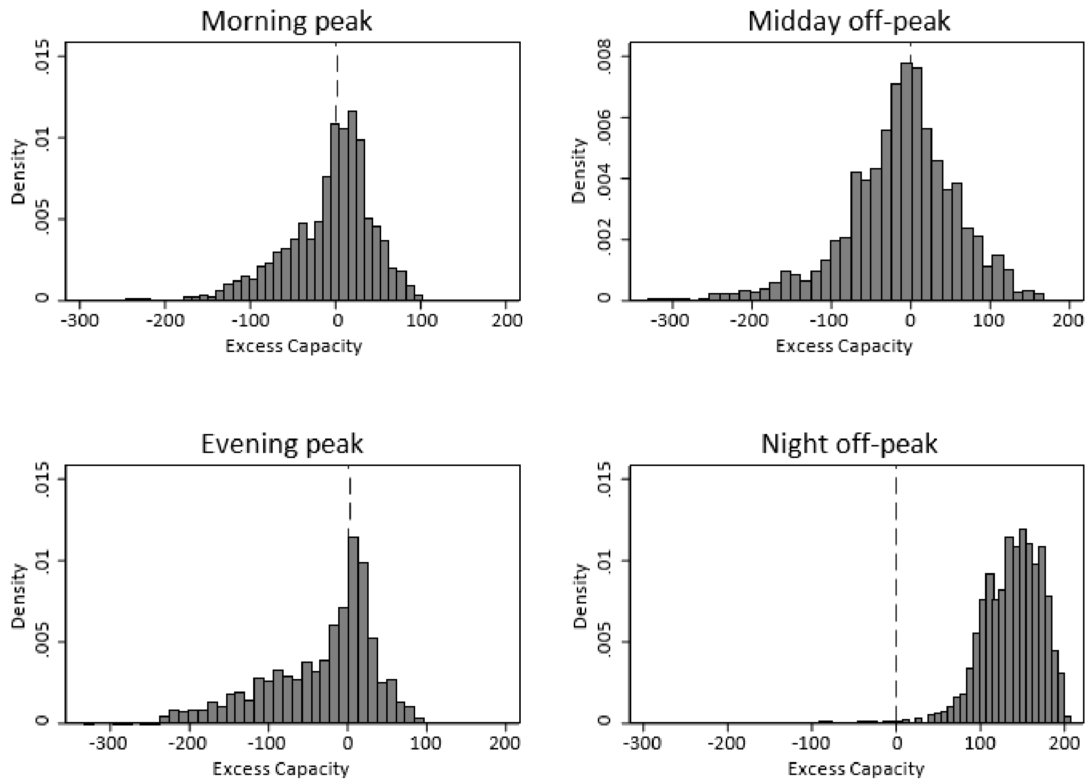


Figure 11. Distribution of excess capacity by period.

Notes: Distribution of detector excess capacity when allowing for complete temporal substitution within each period of day in relation to free-flow speed of 65mph. Morning peak period (5-9am), midday off-peak (9am-4pm), evening peak (4-8pm), night off-peak (8pm-5am).

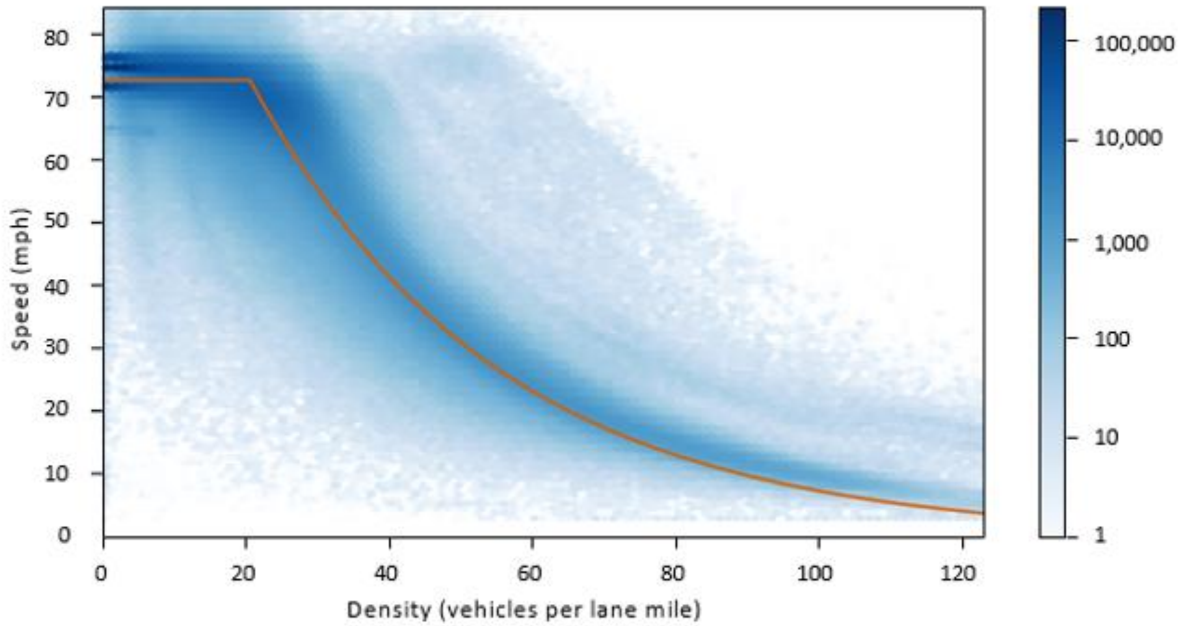
6.2 The Marginal External Costs of Congestion and Optimal Pricing

We now use our estimates to calculate the marginal external cost of congestion and to evaluate the potential welfare gains from pricing roads. To do so, we use the standard static congestion model.

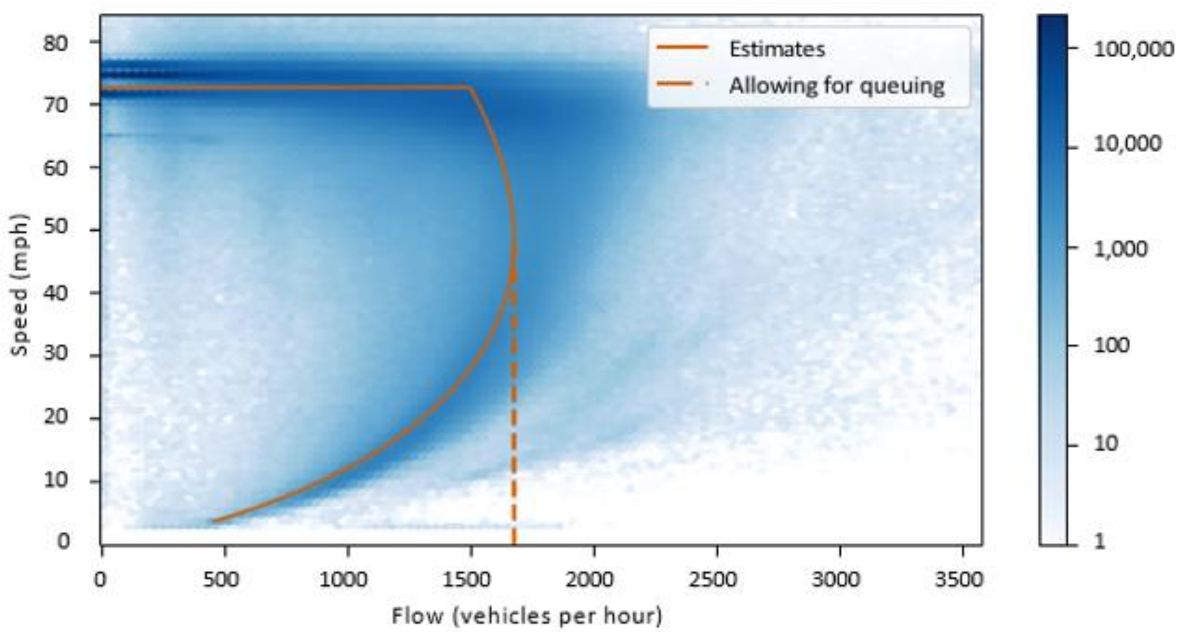
Setup

The first step in doing this is converting our estimates of the marginal effect of an additional unit of density on speeds into a production possibilities frontier fully specifying the relationship between the volume of traffic and speed. Consistent with our estimates, we assume that for densities below the kink point that speeds are constant at the free flow speed. We estimate the free flow speed based on average speed across all Los Angeles highways from 2–4 AM. We assume the marginal effect of an additional vehicle per mile per lane is given by our IV estimate (column (4) of Table 3). Doing this gives us the speed-density relationship plotted in Figure 12a.

Using (2) we can convert these estimates into the relationship between speed and flow. This relationship is plotted in Figure 12b. As the figure shows, when the traffic volume is low, further increases have no effect on speeds. However, once volume crosses 1,441 vehicles per hour per lane (vphpl), each additional vehicle causes a decrease in speed.



(a) Relationship between speed and density



(b) Relationship between speed and flow

Figure 12. Estimated traffic relationships.

Notes: This figure plots our estimates from Table 3.

As discussed earlier, the speed-flow relationship has a portion that is backward bending. This is not a causal relationship where an increase in flow could cause speeds to increase, but rather is caused observing data from within a queue. Additionally, our log-linear functional form for the speed-density relationship mechanically implies a backward bending speed-flow relationship for $S < \exp\,(1 + \alpha)$.[7] To be conservative, we follow Verhoef (2001) in allowing the speed-flow relationship to become vertical once flow has reached its maximum. This represents queuing on the highway, where speeds fall but flow is unchanged. By not allowing for the possibility of a causal hypercongestion relationship, where too many vehicles on the road causes it to jam, and reduces capacity, we will understate the social welfare gains from tolling, and overstate the optimal reduction in flow and optimal toll. Figure 12b plots our estimated speed-flow relationship, shows how it compares to the data, and further shows what it looks like when we allow for queuing.

Having specified the supply side of the model, we now turn to the demand side. In our empirical framework, we do not have exogeneous variation in travel supply to estimate a demand curve, and so instead rely on estimates from the literature. We assume a constant elasticity of demand for travel, so $V = A \cdot P^{S}$, where A is a constant and s_d is the elasticity of demand. We use the estimates of s_d from Anderson (2014), who finds, based on estimates of long-run elasticity of VMT with respect to gas prices or tolls, that $s_d \in [-2.0, -0.67]$. Our results are largely unchanged by which estimate of the elasticity we use, so in the results reported below we use −2.0, which gives more conservative results. We assume the average value of time is $15.00 per vehicle-hour. Figure 13 displays the result.

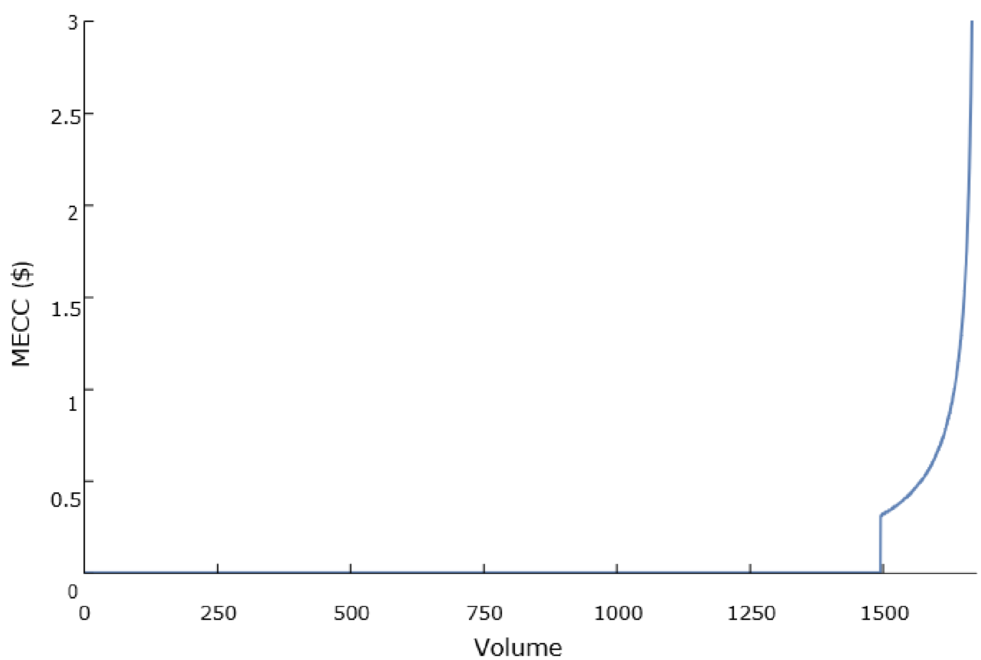


Figure 13. Estimated marginal external cost of congestion.

Notes: Marginal external cost of congestion calculated assuming a value of time of $15.19 and an average vehicle occupancy of 1.6.

6.3 Effects of a Pigouvian Toll

Figure 14 compares the two equilibria. Without tolls, the equilibrium quantity is when the marginal private benefit equals the marginal private cost. Given our estimated relationship between speed and volume and our assumed demand curve, this occurs at a volume of 1560 vphpl and a trip cost of $13, which is all due to the cost of travel time. The social welfare maximizing equilibrium occurs when the marginal private benefit equals the marginal social cost. This occurs at a volume of 1441. By reducing volume by 119 vphpl and eliminating queuing, speeds increase from 30 mph to 69 mph. Achieving this requires charging a toll of $0.81 per mile, and yields improvements in social welfare by $0.68 per vehicle-mile traveled when there was no toll. Since, as is always the case in the standard static model, the toll is greater than the social welfare gain, this toll makes travelers worse off before the revenue is used.

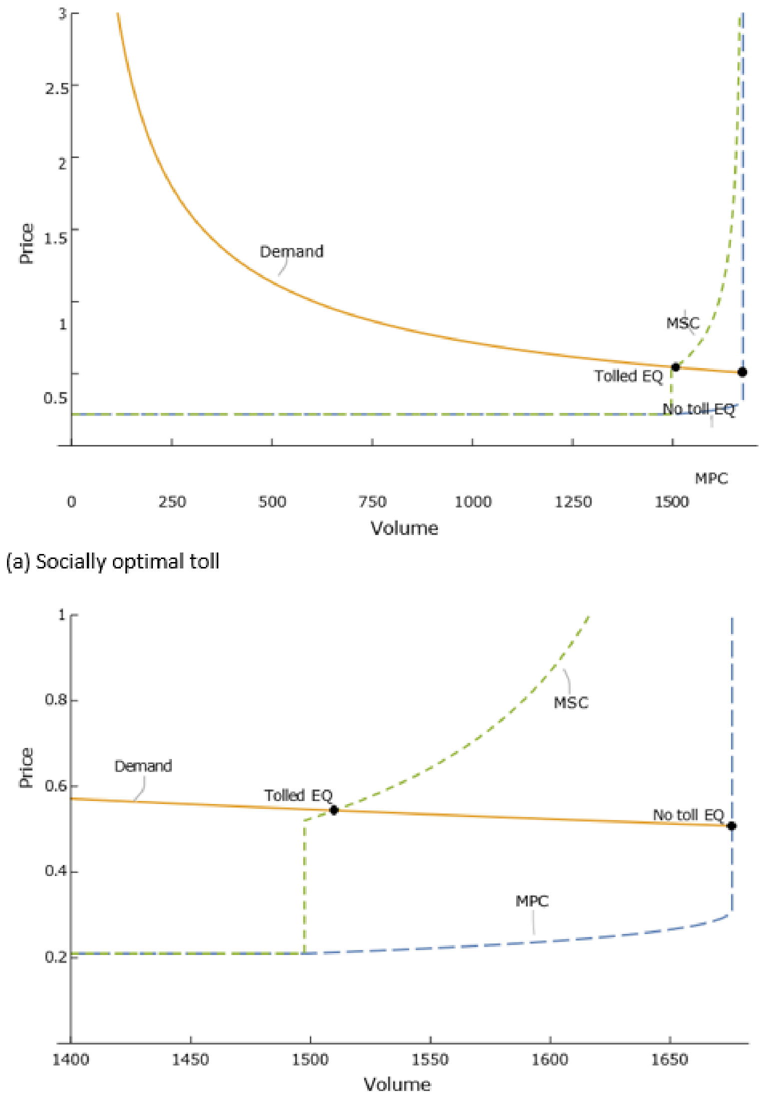


Figure 14. Policy counterfactual.

7. Conclusions

This project employs an instrumental variable design to causally estimate the negative externalities of traffic congestion and calibrate models of optimal congestion pricing. Since the decision to travel is endogenous, ordinary least square will not recover the causal relationship between speed and traffic density. Combining origin-destination commuting data with actual web-scraped driving directions from the Los Angeles metropolitan area, we construct pairs of alternative routes and use traffic crashes on alternatives as demand shocks for the main route. Employing spatially disaggregated freeway detector data collected at high frequency, we show that traffic crashes shift travel demand to other routes. This exogeneous variation in demand allows us to estimate the congestion function relating traffic density to freeway speed and back up the marginal cost of congestion and thus optimal Pigouvian taxes.

Our results indicate that the instrument has power to isolate exogeneous variation in travel demand to estimate a causal speed-density relationship. The instrumental variable estimate suggests that an additional vehicle on the road reduces average freeway speed by about 3.1% once capacity is exceeded. This result implies a larger negative externality from traffic congestion than the reduced form OLS approach. The estimates show considerable heterogeneity that have direct policy relevance: The empirical relationship observes a clear structural break with only small congestion externalities at low levels of traffic.

We use these causal estimates to analyze the potential gains from policy experiments such as congestion pricing and substituting travel demand over time and space. Keeping demand constant, reallocating traffic demand from peak to off-peak hours does not resolve the capacity constraint of the current freeway system in the Los Angeles metropolitan area. This suggests that road pricing is an inevitable policy to alleviate congestion. Employing a static model of freeway travel and common demand parameters, our econometric estimates imply an average optimal congestion charge of 33 cents per mile in the pricing equilibrium. This policy yields large welfare gains and reduces the average social cost of travel by 56%.

We find the optimal toll at peak times is 33 cents per mile, with the toll being lower, even zero, off-peak. We show how this toll varies over space and time, as well as report on its distributional effects. This toll would more than double highway speeds during peak times and only requires reducing vehicle miles traveled (VMT) at the peak by 10%. The resulting social welfare gains are over two billion dollars per year. Ignoring the structural break in the speed-density relationship underestimates the marginal cost of an additional trip at high levels of traffic flow by more than 60%. Correctly estimating the marginal external costs decreases the optimal reduction in VMT by 33%, increases the improvement in speed by 55%, decreases the increase in private costs by 50%, and increases the social welfare gains by 30%. This highlights the importance of correctly specifying the empirical models used to estimate parameters of economic models guiding optimal policy.

References

Akbar, Prottoy and Gilles Duranton (2017) "Measuring the Cost of Congestion in a Highly Congested City: Bogota."

Anderson, Michael L (2014) "Subways, strikes, and slowdowns: The impacts of public transit on traffic congestion," American Economic Review, Vol. 104, No. 9, pp. 2763–96.

Anderson, Michael L. and Lucas W. Davis (2020) "An empirical test of hypercongestion in highway bottlenecks," Journal of Public Economics, Vol. 187, p. 104197

Bento, Antonio M, Jonathan E Hughes, and Daniel Kaffine (2013) "Carpooling and driver responses to fuel price changes: Evidence from traffic flows in Los Angeles," Journal of Urban Economics, Vol. 77, pp. 41–56.

Burger, Nicholas E. and Daniel T. Kaffine (2009) "Gas prices, traffic, and freeway speeds in Los Angeles," The Review of Economics and Statistics, Vol. 91, No. 3, pp. 652–657.

Cameron, A. Colin, Jonah B. Gelbach, and Douglas L. Miller (2011) "Robust inference with multiway clustering," Journal of Business & Economic Statistics, Vol. 29, No. 2, pp. 238–249.

Currie, Janet and Reed Walker (2011) "Traffic Congestion and Infant Health: Evidence from E-ZPass," American Economic Journal: Applied Economics, Vol. 3, No. 1, pp. 65–90

Hall, Jonathan D. (2018) "Pareto improvements from Lexus Lanes: The effects of pricing a portion of the lanes on congested highways," Journal of Public Economics, Vol. 158, pp. 113 – 125

Hall, Jonathan D. (2021) "Can Tolling Help Everyone? Estimating the Aggregate and Distributional Consequences of Congestion Pricing," Journal of the European Economic Association, Vol. 19, No. 1, pp. 441–474

Keeler, Theodore E. and Kenneth A. Small (1977) "Optimal peak-load pricing, investment, and service levels on urban expressways," Journal of Political Economy, Vol. 85, No. 1, pp. 1–25.

Kreindler, Gabriel E. (2020) "The Welfare Effect of Road Congestion Pricing: Experimental Evidence and Equilibrium Implications," Working Paper.

Lindsey, Robin and Erik Verhoef (2007) "Congestion Modelling," in Handbook of Transport Modelling: Emerald Group Publishing Limited, pp. 417–441

Small, Kenneth A and Xuehao Chu (2003) "Hypercongestion," Journal of Transport Economics and Policy (JTEP), Vol. 37, No. 3, pp. 319–352.

Transportation Research Board (2016) Highway Capacity Manual 6th Edition: A Guide for Multimodal Mobility Analysis: National Academies Press, DOI: 10.17226/24798.

TRB, Transportation Research Board (2010) "Highway Capacity Manual," Washington, DC, p. 11.

United Nations, Department of Economic and Social Affairs, Population Division (2019) "World Urbanization Prospects: The 2018 Revision (ST/ESA/SER.A/420)."

Verhoef, Erik T. (2001) “An Integrated Dynamic Model of Road Traffic Congestion Based on Simple Car-Following Theory: Exploring Hypercongestion,” Journal of Urban Economics, Vol. 49, No. 3, pp. 505–542, DOI: 10.1006/juec.2000.2203.

Yang, Jun, Avralt-Od Purevjav, and Shanjun Li (2020) “The Marginal Cost of Traffic Congestion and Road Pricing: Evidence from a Natural Experiment in Beijing,” American Economic Journal: Economic Policy, Vol. 12, No. 1, pp. 418–53. 25

Data Summary

Products of Research

Section 3 of the report describes in detail the dataset. This includes hourly observations on speeds and flows for California's district 7 during 2017. The dataset also includes all vehicle crashes during this time period, as well as additional data used as regressions in the empirical analysis (see section 3).

Data Format and Content

All data is publicly available. The clean dataset used in the project is readily available on stata format.

Data Access and Sharing

The general public can access the data without any restrictions. The dataset is stored at: https://doi.org/10.7910/DVN/VMJIRX

Reuse and Redistribution

Since the data used in the project is publicly available, there are no restrictions on how the data can be reused and redistributed by the general public. The data should be cited as follows:

> Bento, Antonio, 2022, "Replication Data for: Evaluating Alternative strategies for traffic reduction in Los Angeles", https://doi.org/10.7910/DVN/VMJIRX, Harvard Dataverse, V1