Local Pollution Externalities from Driving: Evidence from Roadway Vehicle Sensors

Andrew R. Waxman[a,1], Ruozi Song[b], Rajat Kochhar[c], and Antonio M. Bento[d,e]

[a]University of Texas at Austin
[b]World Bank
[c]University of Chicago
[d]University of Southern California
[e]National Bureau of Economic Research

March 27, 2024

Click here to download latest version

Abstract

This paper estimates the localized effect of automobile congestion on air pollution to better characterize the benefits of anti-congestion policies on human health. We leverage air pollution sensors on Google Street View cars combined with fine grain speed and vehicle density observations on Los Angeles highways, to understand this effect at a level of granularity never previously considered using causally identified methods. We show that higher pollution occurs at very low and very high speeds due to lowered engine efficiency. We show that the success to date of reducing pollution via tailpipe emission standards can be undone on the margin when cleaner vehicles are still operated at inefficient speeds. We present specific guidance on how to address air pollution impacts through the design of anti-congestion policies, speed limits and point to a need to better regulate sources of fine particulate matter pollution from tires and brakes.

Keywords: air quality, traffic congestion, urban transportation, remote sensing, econometrics
[1]Corresponding Author E-mail: awaxman@utexas.edu

Despite increased penetration of electric vehicles, high levels of congestion and transportation system-induced air pollution remain persistent problems in dense urban areas [1]. These pollution impacts continue to have immense impact on human health even as ambient concentrations of particulate matter 2.5 microns or smaller ($PM_{2.5}$) have fallen in most parts of the United States in the past several decades. Concentrated pockets of high pollution areas persist and these are likely to include the most vulnerable to pollution health damages from respiratory infections, lung cancer, stroke and cardiopulmonary disease, dementia and Alzheimer's disease [2, 3, 4, 5]. The vast majority of air pollution damages are linked to premature mortality with the main contributor being $PM_{2.5}$ and its precursors, including nitrogen oxides (NO_x) [6, 7, 8].

As a mechanism to address this concern, transportation policies have had a substantial impact on urban air pollution [9]. This occurs through tailpipe emissions which are regulated directly [10, 11, 12, 13] or indirectly through fuel economy standards [14, 15, 16], gasoline formulation regulation [17], vehicle scrappage policies [18, 19, 20, 21] or similar policies [22]. In addition, policies that address the speed of vehicles can have a substantial impact on urban air quality and are often legislated with this as an objective [8, 23, 24].

Because primary $PM_{2.5}$ emissions come from vehicle brake and tire degradation, which is accelerated in the start-and-stop conditions of urban traffic, even full electrification of the vehicle fleet is limited in its potential to reduce urban air pollution, since even regenerative braking systems on these vehicles still rely on hydraulic disc brake systems that are used during start-and-stop traffic. Some cities have addressed these challenges by making it more difficult for vehicles to drive in urban cores by converting roads into pedestrian spaces [25]. Other cities, such as London, Stockholm and Singapore have opted for congestion cordons which charge drivers a substantial fee to enter the urban core [26]. Another approach has been to limit the entry of large or high emissions vehicles using so-called "low emission zones" [27, 28, 29, 30]. These studies have shown that there is substantial potential to further limit vehicle emissions in urban areas, but transportation policies that have been shown to reduce pollution substantially in some contexts have no effect or even *increase* pollution in others [31].

To correctly predict the effects of these policies and inform their design, we need to better understand the effect of adding vehicles to the road at different speeds on pollution. In this paper, we estimate the effect of speeds on pollution using fine-grained air pollution sensor data, which allows us to visualize the speed-pollution relationship at the road level. Using this data is important to correctly characterize this relationship as panel (A) of Fig. 1 shows, since air pollution levels deviate substantially as one moves a few blocks away. Past work using EPA ground monitors to understand the relationship between vehicle speeds and pollution may then incorrectly measure this relationship [32]. In addition, we use regression methods to causally identify the relationship between speed and pollution. Since variation in speeds and pollution are often determined by a range of observed and unobserved factors (weather, baseline pollution levels, the composition of the vehicle fleet), not accounting for these factors could under- or overestimate the human health benefits of cleaner roads. We exploit plausibly exogenous variation due to accidents so that changes in air pollution concentrations can be attributed to vehicle density and speeds from observational data. While cars today are dramatically cleaner than those of a decade ago from a tailpipe emission perspective [33], we show that cars that are ten years younger would still have their pollution benefits undone if they were driven at substantially slower speeds found in congested areas of Los Angeles.

The approach demonstrated in this paper improves upon the two traditional approaches to characterizing the relationship between vehicle speeds and air pollution. One, laboratory studies in controlled settings allow engineers to characterize the effect of vehicle speeds on engine efficiency and vehicle emissions. These estimates are then incorporated into mobile source

emissions models, such as the U.S. Environmental Protection Agency's (EPA) Motor Vehicle Emission Simulator, which uses a variety of adjustment factors to account for on-road conditions [34]. These can then be used to calculate hypothetical pollution impacts with an eye towards informing ex ante expected pollution reductions for policymakers [35].

Second, *in situ* air pollution monitoring near roadways allows researchers to relate observed variation in pollution to changes in nearby speeds [36, 37, 38, 39, 40, 41, 42]. Since the connection between congestion and pollution is mediated by a range of factors (e.g., climate, vehicle composition, air chemistry, the built environment) that affect how emissions translate into pollution concentrations. As a result, this second approach provides ground truth for the model predictions of the first approach. Indeed, enforcement of the EPA's National Ambient Air Quality Standards (NAAQS) also is based upon attainment of daily and annual maximum pollution levels from the EPA's ground monitor system, where monitor locations are often sited to be close to roadways with a high potential impact on human health through exposure to pollution. Both approaches are used in the design or evaluation of State Implementation Plans to address NAAQS non-attainment, which can involve, among other measures, monitoring of transportation system outcomes like congestion. An advantage of the first approach is that by starting with vehicle emissions, pollution emissions and then concentration changes are directly attributable to cars. While lab studies help to quantify the speed-pollution relationship in a controlled setting and observational studies document the variation of pollution in the field, to accurately predict the policy effects of speed on pollution, researchers and policymakers need to know the *counterfactual* emissions, for which credible estimates require causal inference. In this paper, we bring to bear the most granular estimates of marginal external damages from additional congestion using air pollution observations from vehicle-level sensors. Past work examining the interaction of congestion and pollution has done so at a more aggregate level over weeks not hours and miles not blocks [32, 43, 44, 45, 46]. This aggregate level is useful to characterize pollutants which form over longer time scales, such as ozone, but can mask local-level pollution "hot spots" [47, 48]. As an example of this localized variation, recent work in China has shown that on days when driving restrictions reduce the flow of vehicle traffic, NO_2 levels are 12% lower on average [49]. We describe the general features of these interactions in Figure S1.

We focus on the localized, short-run pollution effect of congestion because these effects are likely to vary depending on the baseline level of speeds, location, weather and the composition of traffic. To estimate the effect of traffic congestion on local air quality, we combine three unique sources of big data: First, real-time spatially and temporally disaggregated data on a variety of air pollutants collected from sensors installed in Google Street View cars in Los Angeles by a startup firm, Aclima. We focus on pollution effects from $PM_{2.5}$ and NO_x given their substantial production by vehicles, their substantial harm to human health and their availability in our data. Nevertheless, Aclima data suffers from two crucial challenges: First, the temporal coverage is limited since vehicles only operate from 7 AM to 7 PM. Second, we only observe pollution observations when the Google Street View cars are in a particular location, which varies over a given day. We match Aclima air quality data to the speed and flow of vehicles from subsurface sensors on California freeways from the California Performance Measurement System (PeMS). Observations are matched based on the closest sensor to reported latitude and longitude of the Google Street View vehicle within a 5-minute interval. PeMS collects real-time speed and flow data from finely spaced loop detectors on California freeways. Then, we leverage machine learning methods to estimate an instrumental variables regression model using California Highway Patrol accident locations and durations as a source of plausible exogeneity for causal estimates of the effect of congestion on air pollution.

Panel (B) of Fig. 1. illustrates this identification strategy for a single accident occurring along the

I-710 freeway, which runs North to South along the right side of the map: congestion reflected by low speeds spreads across the highway and intersecting roads and then abates with time. Panel (C) of Fig. 1 shows how this translates into speeds, while Panel (D) shows how lower speed leads to and increases air pollution. However, as shown in panel (B), pollution effects in response to accidents may be short-lived compared to speed effects. Since observations in our sample that are close to accidents ($< \frac{1}{2}$ mile) may reflect congestion that contributes to congestion (and therefore is endogenous), we use match PeMS and Aclima sensor observations to accidents between $\frac{1}{2}$ and 1 mile away. Observations without an accident within 1 mile are "w/o accident" and act like a control group.

Lastly, we focus on how the regulation of speed could affect localized pollution damages. Increased speeds decrease the marginal external cost of congestion (MECC) as illustrated in (A) of Figure S1. However, since engine efficiency is maximized at speeds around 35-40 MPH, the relationship between speed and pollution is U-shaped, with the biggest pollution levels for low speeds and also substantial pollution at higher speeds. This relationship holds particularly for NO_X (specifically NO and NO_2) which is directly emitted from engines. Also, NO can further convert to NO_2 within minutes. For $PM_{2.5}$, the pollution is also higher at lower speeds but for a different reason. One major source of $PM_{2.5}$ emissions from driving a car is tire friction. During traffic when braking and restarting constantly occur, $PM_{2.5}$ emissions are higher. NO and NO_2 can also convert to ozone but over longer time horizons, depending on background air pollution levels and weather [50].

All three of the pollutants we study are regulated by the EPA under the Clean Air Act and its amendments and have substantial negative health impacts [51]. $PM_{2.5}$ comes not from tailpipe emissions from tire and brake degradation and so can also have a U-shaped relationship with speed since start and stop traffic as well as high speeds both increase degradation rates [52]. This has important implications for the regulation of speeds as we will explore.

Results

We estimate the localized marginal air pollution effect of adding a vehicle to the roadway given different baseline levels of congestion as reflected by speed bins of less than 20 MPH, 20-40 MPH, 40-60 MPH, and greater than 60 MPH. The effect of adding a vehicle, also known as vehicle density, is measured in cars per lane-mile and is a standard measure of congestion in the transportation literature [53]. Density effects on pollution will change based on the baseline level of congestion as we show analytically in the Supplementary Materials and using our data in (B) of Figure S1. In additional results presented in Figure S2, we aggregate our results to consider how these effects to larger time and spatial scales: daily and 2-digit latitude and longitude level. We find larger NO_2 and $PM_{2.5}$ effects, with a similar U-shaped relationship between speed and $PM_{2.5}$ emissions as in our main results. This points to the persistence of the pollution effects measured in our main results, with the caveat that a larger set of weather and air chemistry factors become important at these larger spatio-temporal scales. The magnitude of these effects are slightly larger but of similar order of magnitude to prior studies measuring $PM_{2.5}$ effects of vehicles in Los Angeles [54] and Beijing [55].

Turning to Fig. 2, we show in (A) the effects of adding one vehicle to air pollution effects across four speed bins: In (A), we show using regression estimates that increasing speed generally reduces NO and NO_2 emissions by 89.9% and 86.9% respectively by increasing speed from below 20 MPH to above 60 MPH, while $PM_{2.5}$ emissions have a U-shaped relationship to speed, where they decrease substantially moving from speeds below 20 MPH to those 20-40 MPH, and again

further from 40-60 MPH, but they increase again for speeds above 60 MPH.

To better understand the source of these pollution effects, we decompose the contribution of the marginal pollution effect of adding one vehicle to the road in each speed bin into two components: the effect from changes in engine efficiency and the effect from vehicles slowing down and so taking longer to travel the same lane-mile. Thus, the engine efficiency includes both the direct efficiency of engines across speeds for NO_x, but also the emissions from tires across speeds for $PM_{2.5}$. This decomposition is displayed in (B) and explained in equation (4). For a car traveling below 20 MPH, engine efficiency explains 77% of the marginal effect on PM pollution, while travel time explains 23%. When the car travels above 60 MPH, engine efficiency explains 98% of the marginal effect.

In (C) of Fig. 1, we show two sets of marginal external damages from adding one vehicle to the road at different speeds. On the top, we calculate the marginal external cost of congestion, which ranges between 50 and 90 cents per lane-mile, reflecting that when congestion is high, speeds are low, and adding one more car has a big negative effect since road capacity is constrained. These are comparable to estimates from (*30*) who calculate this externality for all of the US to be 5 cents per lane-mile, where average congestion is much lower and speeds are higher. Accounting for increases in income and inflation, the marginal value of the externality would be closer to 8 cents.

On the bottom half of that panel, we show marginal external damages due to air pollution, specifically NO_2 and $PM_{2.5}$. It is not possible to show this effect for NO since it is short-lived, quickly converting to NO_2. Damages for $PM_{2.5}$ and NO_2 are much lower than for the MECC, ranging from about 3 cents to damages that are statistically indistinguishable from zero in the 20-40MPH speed bin. The higher magnitude of MECC is consistent with this being the most significant source of negative externalities from driving due to the number of people affected and its impact on time allocation [56].

In order to consider the relative effect of emissions intensity, we use estimates from [33] of the effect of vehicle age on tailpipe emissions based on EPA emissions testing on new cars as well as vehicle emissions testing on the active vehicle fleet in the State of Colorado. We do this calculation for NO_2 only, since engine age likely has no effect on $PM_{2.5}$ which comes primarily from tires. In that study, the authors estimate that a vehicle's age increases NO_2 emissions by roughly 8%, on average. Applying these effects to our estimates in each speed bin results in the green dots.

A useful thought experiment, then, is asking what is the relative effect on the marginal pollution externality from improving engine efficiency relative to moving from the first (<20 MPH) to the third (40-60 MPH) speed bin? Comparing the horizontal red to the green lines, indicates that the effect of slower speeds (0.015) is more than ten times the effect of a single year of vehicle age (0.001). In other words, cars that are ten years younger would still have their pollution benefits undone if they were driven at substantially slower speeds found in congested areas of Los Angeles.

Of course, the effects of adding vehicle congestion are likely to vary due to a variety of circumstances, which we explore in Fig. 3, which shows variations to our main results for the effect of vehicle density on $PM_{2.5}$ across speed bins. For one, Los Angeles has substantial variation in local climatic conditions due to topography and coastal weather conditions. In addition, while we include a control of the vehicle fleet in our regression model, these may manifest themselves across regions based on the composition of traffic from nearby industry. To illustrate this, Panel A of Fig. 3 shows the differences in $PM_{2.5}$ effects between Downtown Los Angeles and Long Beach, where the latter has smaller pollution impacts not distinguishable from zero. This could be related to differences in baseline pollution concentrations as the second figure in that panel shows: for lower speeds, these concentrations are higher in Long Beach, which may make the marginal contribution of an additional smaller given sufficient baseline $PM_{2.5}$ saturation. This is consistent with [55] who find smaller marginal pollution contributions in

Beijing on days in which background pollution levels are higher. Nevertheless, this result adds to a growing body of evidence that the spatial distribution of driving-related externalities can be quite heterogeneous even within the same Metropolitan Statistical Area [57].

Panel B of Fig. 3 shows how the estimated effects vary by peak (5-9AM, 3-7PM) versus non-peak hours: lower pollution effects when there is less congestion during the non-peak period. This result highlights the fact that restrictions to more polluting vehicles may be more meaningful during peak hours as is reflected in the time-varying design of the low emission zone in Milan, Italy (now repealed) but not in London, UK [58, 59]. Panel C of Fig. 3 shows the distribution of the marginal effects of our estimates, reflecting that there is substantial heterogeneity even within a particular speed bin in pollution effects. Lastly, Panel D shows how more severe weather (above or below the 75^{th} percentile) affects our estimates, where light grey bars reflect the base weather conditions, and the darker bars show the effect of changes in the weather condition. For example, as in the first panel, for temperatures below 27°Celsius (81°F), the effects across the four speeds are 0.304, 0.299, 0.261, and 0.287. These each increase by 0.034, -0.090, 0.053, and -0.014, respectively, for temperatures above 27°C, resulting in a total effect from higher temperatures of 0.338, 0.209, 0.314, 0.273. We also present the effect of weather on NO_X emissions as well as information about the distribution of temperature in our sample in Figure S3. These results demonstrate that the lowest speeds below 20 MPH generate more pollution, although many weather effects differences across bins are statistically indistinguishable.

Discussion

Our results point to important implications of highway speeds on urban air pollution concentrations. These effects occur at high and low speeds, reflecting a well-documented U-shaped relationship between speeds and air pollution, that is further complicated by greater pollution due to start-and-stop traffic at lower speeds [60]. Air pollution effects are therefore the largest for low speeds, highlighting the need for anti-congestion approaches in dense urban areas such as congestion charges which have been shown to lower air pollution in cities where they have been implemented— Stockholm, London, and Milan [61, 62, 63]. At the same time many cities, particularly in Europe but also in Asia, have implemented low emission zones (LEZs) that only allow cars with very low emission levels to enter. This changes the composition of vehicle emissions making slow speeds less pollution-intensive, although evidence on the effectiveness of these policies is mixed [27, 28, 64]. Both sets of policies have never been implemented in the U.S. beyond a pilot basis but may be in the near future in Manhattan and the Los Angeles area [65, 66].

Our results also point to benefits of anti-congestion policies in the context of climate change. First, these policies can reduce overall driving, resulting in some reduction in greenhouse gas emissions from vehicles. More relevant to our results, the additional pollution created by excess congestion (called the Pollution Interaction Effect and formalized in the Supplemental Materials) will be aggravated by climate change, since higher temperatures can result in higher pollution formation. In our results, we show that the effect on $PM_{2.5}$ pollution of adding one vehicle when speeds are below 20 MPH is 11% higher when temperatures are in the top quartile. The effects on ozone—which is formed over longer time spans and spatial extents and so is beyond the scope of our analysis—is likely to be more significant since longer-term $PM_{2.5}$ or NO_X responses than what we report are likely to be far more dependent on local air chemistry [67].

At the same time, our findings have important implications for the design of urban speed limits, which are traditionally dictated by a balance between reducing accidents and allowing motorists

to drive faster to lower travel times. Past work has shown that, on balance, many speed limits in the U.S. would benefit from being lowered closer to optimal engine efficiency ranges (50-55 MPH) [68]. Beyond lower speed limits, stronger enforcement through speed cameras and variable speed limits are alternative policy approaches. There is mixed evidence that the speed cameras reduce accidents, and speed reductions these effects may be offset by increases further away [69, 70], though pollution reductions have been documented in the Netherlands [71]. Variable speed limits allow municipalities to keep speeds lower in the shoulder of the peak period commuting when congestion does not meaningfully lower speeds, but vehicle volumes are high and pollution from higher speeds may be substantial. They have been implemented in various cities including Seattle, the suburbs of Washington, D.C., and Stockholm [72, 73, 74]. There is also evidence that these policies may have reduced pollution in Barcelona [75]. Despite a vehicle fleet that continues to operate more cleanly due to vehicle emission standards, our work points to a need to leverage a combination of these speed limiting policies alongside anti-congestion approaches to mitigate emissions from high and low speeds.

A final policy-relevant point emerges from the fact that we provide further evidence for substantial *in situ* particulate matter production from vehicles consistent with brake and tire wear emissions. While the effects of these pollution sources is factored into most mobile source emissions models, our study contributes to prior evidence from remote sensing in Southern California that these effects are persistent and large [52]. Regulators have expressed concern about this source of emissions for some time, but there has been limited road-level assessment of their contribution [35, 76]. This study points to the value of studying real-world concentrations related non-exhaust emissions and linking them to congestion on urban freeways to better guide policy [77, 78].

Materials & Methods

Data Sources

Air Pollution

The air pollution data in on Los Angeles freeways in our main results is from Aclima, a technology startup based in San Francisco, California. Their data was collected from pollution sensors installed on Google Street View cars. Street View cars drive along the length of city streets collecting images for Google's mapping products. We have observations from two cars which were simultaneously on the road at different locations. The data allows for unprecedented granularity in measurement, with observations recorded every one to three seconds at a spatial resolution of 4 decimal digit latitude and longitude. This corresponds to a precision of 11 meters at the equator with many repeated observations within a single city block. However, we aggregate the data to 3-digit latitude and longitude quadrants geographically and 5-minute temporally for the regression analysis. This is done for two reasons: (*i)* to ensure a more balanced panel, and; *ii)* because Nitric Oxide (NO) takes tens of seconds to a few minutes to convert to Nitrogen Dioxide (NO_2). As a result, we can capture the impact of density on NO_2 by aggregating to 5 mins. Though we observe data on both freeways and local surface streets, we limit our analysis to only freeways. Throughout the paper, the unit of analysis is a freeway segment, which is approximately 100 meters in length corresponding to the overlap between freeway segments and 3-digit latitude and longitude quadrants. All air pollution values in this paper are reported in units of µg / m^3.

We validate the accuracy of Aclima pollution readings using data from the US Environment Protection Agency's (EPA) ground level monitoring system. The agency provides hourly data on the levels of NO, NO_2 and $PM_{2.5}$ at ten air quality monitors in and around Los Angeles. Each

Aclima reading is matched to pollution levels recorded by an EPA air quality monitor that is closest in terms of time and distance. This allows us to calculate the correlation between pollution captured by a geographically mobile source at a disaggregated level (Aclima) and a fixed monitor capturing pollution at an aggregated level (EPA).

Data on Roadway Conditions

We combine the air pollution data from Aclima with vehicular speed, density, and flow observations from the California Department of Transportation's (Caltrans) Performance Management System (PeMS). PeMS collects traffic information via induction-loop detectors that span the freeway system, with data collected over every 5-minute interval. Therefore, for every detector and 5-minute period, we have information on the sum of flows, average occupancy across all lanes and flow-weighted average speed. Additionally, PeMS also provides spatial data on the geolocation of these detectors, including the freeway and the direction on which it is located. We assign the closest (in terms of geodesic distance) PeMS monitor to every Aclima observation, conditional on the monitor being on the same freeway and direction as the Aclima car. This helps us relate traffic density to pollution exposure for every 5-minute period in our sample, which consists of 1,873 detectors in our study area. Despite the highly detailed nature of our data, we only observe pollution observations at times and locations where the two Google cars are driving. This motivates the empirical approach we detail below, since our ability to compare contemporaneous pollution levels in different locations is limited by the locations of the Google cars. We show the relationship between Google Street View vehicle speeds and PeMS speeds in Figure S4, which are proportionate but not always the same.

As part of our instrumental variable (IV) empirical strategy detailed below, we use the occurrence and duration of accidents more than 0.5 miles away to instrument for endogenous vehicle density. This data on vehicle accidents is reported by the California Highway Patrol (CHP) and made available via PeMS. We observe details on the time, location and type of accident. More details about recovering structural parameters of travel demand and supply as well as additional covariates in our regressions can be found in the Supplementary Materials text.

Regression Models

To recover the relationship between vehicle speeds, roadway densities and local air pollution, we regress Aclima sensor pollution concentrations on freeway segment, i, for Google car, c, and for a 5-imute interval in our data, t, on a flexible function of vehicle density and speed as well as other controls:

$$\begin{aligned}\log\left(p_{ict}\right) = &\sum_{b=1}^{B} \alpha_b \log\left(Cars/Mile_{it}\right) \times 1\left\{Speed_{it} \in S_b\right\} \\ &+ X_{it}\beta + \tau_t + \phi_i + \gamma_c + \epsilon_{ict}.\end{aligned} \tag{1}$$

Here p_{ict} is the log of $PM_{2.5}$, NO, or NO_2 pollution in parts per billion. $Cars/Mile_{it}$ is vehicle density constructed using data from PeMS roadway sensors, which is interacted with an indicator variable for whether road speed, $Speed_{it}$, from PeMS sensors lies within a given speed bin, S_b. In our main specification, we use four speed bins ($B = 4$) corresponding to $S = \{0-20, 20-40, 40-60, >60\}$ in miles per hour. X_{it} is a vector of hourly weather controls including the second-order polynomials of temperature, relative humidity, and wind speed, as well

as wind direction. X_{it} also includes a control for vehicle composition on freeway segment i at time t. Additionally, we also include a rich set of fixed effects for time, τ_t, by hour and day-of-week for freeway, ϕ_i, and Google Street View car, γ_c. As pointed out in the previous section, there were two Google Street View cars deployed by Aclima. Therefore, m of our sample records multiple Google Street View vehicles on the road at different locations in Los Angeles at the same time. Lastly, we account for potential serial correlation and heteroskedasticity in the error term, ϵ_{ict}, by using two-way clustered standard errors by hour and freeway segment following [79].

Finally, contemporaneous pollution levels could be influenced by their lagged counterparts. Specifically, NO_X concentrations are the result of a complex interaction of air chemistry factors. A key determinant of the impact of increased vehicle density on these pollution measures is the existing concentrations of NO and NO_2. Therefore, to account for their influence, we include the ratio of NO to NO_2, which reflects the extent to which atmospheric conditions are saturated by one pollutant or the other. Given the potential influence of background NO_2levels on PM formation via secondary atmospheric production, we also include lagged levels of NO_2 in some of the PM regressions. Figure S6 shows the effect of adjusting the covariates in our regression results, resulting in greater precision and addressing covariates that may bias estimation if they were omitted from the regression.

Econometric Identification

A key challenge to estimating equation (1) is the fact that speed-density changes across the transportation network are not randomly distributed and, therefore, it is possible that temporo-spatial variation in unobservables could simultaneously affect pollution and traffic conditions, biasing our parameter estimates.

To overcome this concern, we take advantage of data on vehicle accidents reported by the California Highway Patrol using accidents as an instrument for changes in congestion levels. This allows us to match the vehicular density in a particular location with all plausibly exogenous accident events, enabling us to account for the influence of accidents elsewhere in the highway network on traffic conditions in our data through the instrumental variable design. The instrument would be valid under the assumption that the accidents are not themselves caused by unobservable determinants of contemporaneous congestion (*i.e.*, travel demand). This is an important concern because accidents are more probable at higher speeds or at locations where speeds are rapidly changing. By using accidents from vehicle collisions that are sufficiently far away, it seems reasonable to assume that our instruments are uncorrelated with the error term in equation (1) and therefore valid. Moreover, if accidents happened *before* our speed-density observations, then it may be reasonable to assume that it is a nearby accident affecting our speed-density observations and not the other way around.

In principle, any accident in the transportation system could result in congestion elsewhere in the road network, though the propagation of slowdowns is likely to decay with distance and time since the accident was cleared from the roadway. In panels (C) and (D) of Fig. 1, we illustrate the propagation of one such accident in our data, documenting the changes in congestion and pollution before and after the accident. It can be seen how speeds drop initially in the vicinity of the accident, but then extend outward along proximate links as time goes by. Similarly, speeds return to their pre-accident levels in the reverse direction as full capacity is returned to the lane. For this reason, we construct an extremely rich instrument set corresponding to variables for the duration and number of incidents in the transportation system, binned by distance to the Google Street View vehicle and time since the accident began. We group distance bins into 0.5-1, 1-2, 3-4, and 4-5 miles. We bin the time since the relevant accident into less than 5, 5-10, 10-15, and

15-20 minutes before the Aclima pollution sensor observation. We plot the distribution of average accident durations in Figure S7.

The instrument set also includes the interaction of these accident measures with weather variables, reflecting the fact that the propagation of accidents might depend upon climatic conditions. In all, the set of candidate instruments from accidents includes 158 variables including weather variable interactions and fixed effects. Nevertheless, many of these instruments are unlikely to be informative about congestion if they are sufficiently far away in time and space from our pollution readings. We use Least Absolute Shrinkage and Selection Operator (LASSO) estimation for the first stage, which selects the minimally informative instrument set. Recent work has used this approach in cases where theory does not provide clear functional form for the instrument set and a large number of regressors is available [80, 81]. Our preferred first-stage instrument set, including interactions, is comprised of 74 regressors for the first and second bins, 86 for the third bin and 88 for the fourth.

References

[1] Parikshit Deshmukh, Sue Kimbrough, Stephen Krabbe, Russell Logan, Vlad Isakov, and Richard Baldauf. Identifying air pollution source impacts in urban communities using mobile monitoring. *Science of The Total Environment*, 715:136979, 2020.

[2] D. W. Dockery, C. A. Pope, X. Xu, J. D. Spengler, J. H. Ware, M. E. Fay, B. G. Ferris, and F. E. Speizer. An association between air pollution and mortality in six u.s. cities. *New England Journal of Medicine*, 329:1753–1759, 1993.

[3] C. A. Pope, M. J. Thun, M. M. Namboodiri, D. W. Dockery, J. S. Evans, F. E. Speizer, and C. W. Heath. Particulate air pollution as a predictor of mortality in a prospective study of u.s. adults. *Am J Respir Crit Care Med*, 151:669–674, 1995.

[4] L. B. Lave and E. P. Seskin. Air pollution and human health. *Science*, 169:723–733, 1970.

[5] K. C. Bishop, J. D. Ketcham, and N. V. Kuminoff. Hazed and confused: The effect of air pollution on dementia. *The Review of Economic Studies*, 2022. rdac078.

[6] N. Z. Muller and R. Mendelsohn. Measuring the damages of air pollution in the united states. *Journal of Environmental Economics and Management*, 54:1–14, 2007.

[7] E. Arceo, R. Hanna, and P. Oliva. Does the effect of pollution on infant mortality differ between developing and developed countries? evidence from mexico city. *The Economic Journal*, 126: 257–280, 2016.

[8] J. Currie and R. Walker. Traffic congestion and infant health: Evidence from e-zpass. *American Economic Journal: Applied Economics*, 3:65–90, 2011.

[9] Soren T Anderson and James M Sallee. Designing policies to make cars greener. *Annual Review of Resource Economics*, 8:157–180, 2016.

[10] Jie Bai, Shanjun Li, Danxia Xie, and Hui Zhou. Environmental protection or environmental protectionism? evidence from tail pipe emission standards in china. In *AEA papers and proceedings*, volume 111, pages 381–385. American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203, 2021.

[11] Matthew E Kahn. The efficiency and equity of vehicle emissions regulation: Evidence from california's random audits. *Eastern Economic Journal*, 22(4):457–465, 1996.

[12] Pierre Mérel, Aaron Smith, Jeffrey Williams, and Emily Wimberger. Cars on crutches: How much abatement do smog check repairs actually provide? *Journal of Environmental Economics and Management*, 67(3):371–395, 2014.

[13] Paulina Oliva. Environmental regulations and corruption: Automobile emissions in mexico city. *Journal of Political Economy*, 123(3):686–724, 2015.

[14] Lawrence H Goulder, Mark R Jacobsen, and Arthur A Van Benthem. Unintended consequences from nested state and federal regulations: The case of the pavley greenhouse-gas-per-mile limits. *Journal of Environmental Economics and Management*, 63(2):187–207, 2012.

[15] Winston Harrington. Fuel economy and motor vehicle emissions. *Journal of environmental Economics and Management*, 33(3):240–252, 1997.

[16] Mark R Jacobsen. Evaluating us fuel economy standards in a model with producer and household heterogeneity. *American Economic Journal: Economic Policy*, 5(2):148–187, 2013.

[17] Maximilian Auffhammer and Ryan Kellogg. Clearing the air? the effects of gasoline content regulation on air quality. *American Economic Review*, 101(6):2687–2722, 2011.

[18] Antonio Bento, Kevin Roth, and Yiou Zuo. Vehicle lifetime and scrappage behavior: Trends in the us used car market. *The Energy Journal*, 39(1):159–184, 2018.

[19] Mark R Jacobsen and Arthur A Van Benthem. Vehicle scrappage and gasoline policy. *American Economic Review*, 105(3):1312–1338, 2015.

[20] Ryan Sandler. Clunkers or junkers? adverse selection in a vehicle retirement program. *American Economic Journal: Economic Policy*, 4(4):253–281, 2012.

[21] Shanjun Li, Joshua Linn, and Elisheba Spiller. Evaluating "cash-for-clunkers": Program effects on auto sales and the environment. *Journal of Environmental Economics and management*, 65(2):175–193, 2013.

[22] Maureen L Cropper, Yi Jiang, Anna Alberini, and Patrick Baur. Getting cars off the road: The cost-effectiveness of an episodic pollution control program. *Environmental and Resource Economics*, 57:117–143, 2014.

[23] Lucas W Davis. The effect of driving restrictions on air quality in mexico city. *Journal of Political Economy*, 116(1):38–81, 2008.

[24] Nicolas Gendron-Carrier, Marco Gonzalez-Navarro, Stefano Polloni, and Matthew A Turner. Subways and urban air pollution. *American economic journal: Applied economics*, 14(1):164–196, 2022.

[25] L. Bertolini. From "streets for traffic" to "streets for people": can street experiments transform urban mobility? *Transport Reviews*, 40:734–753, 2020.

[26] J. Leape. The london congestion charge. *Journal of Economic Perspectives*, 20:157–176, 2006.

[27] L. Ma, D. J. Graham, and M. E. J. Stettler. Has the ultra low emission zone in london improved air quality? *Environ. Res. Lett.*, 16:124001, 2021.

[28] M. Gehrsitz. The effect of low emission zones on air pollution and infant health. *Journal of Environmental Economics and Management*, 83:121–144, 2017.

[29] F. M. Santos, A. Gomez-Losada, and J. C. M. Pires. Impact of the implementation of lisbon low emission zone on air quality. *Journal of Hazardous Materials*, 365:632–641, 2019.

[30] A. Poulhès and L. Proulhac. The paris region low emission zone, a benefit shared with residents outside the zone. *Transportation Research Part D: Transport and Environment*, 98:102977, 2021.

[31] Francisco Gallego, Juan-Pablo Montero, and Christian Salas. The effect of transport policies on car use: Evidence from latin american cities. *Journal of Public Economics*, 107:47–62, 2013.

[32] C. R. Knittel, D. L. Miller, and N. J. Sanders. Caution, drivers! children present: Traffic, pollution, and infant health. *Review of Economics and Statistics*, 98:350–366, 2016.

[33] M. R. Jacobsen, J. M. Sallee, J. S. Shapiro, and A. A. van Benthem. Regulating untaxable externalities: Are vehicle air pollution standards effective and efficient? *National Bureau of Economic Research*, 2022. URL https://doi.org/10.3386/w30702. 30702 [Preprint].

[34] EPA. Overview of epa's motor vehicle emission simulator (moves4). Technical report, U.S Environmental Protection Agency, 2023. URL https://nepis.epa.gov/Exe/ZyPDF.cgi?Dockey=P10186IV.pdf. EPA-420-R-23-019.

[35] CARB. Brake & tire wear emissions vehicle non-exhaust particulate matter sources. Technical report, California Air Resources Board, 2023. URL https://ww2.arb.ca.gov/resources/documents/brake-tire-wear-emissions.

[36] J. S. Apte, K. P. Messier, S. Gani, M. Brauer, T. W. Kirchstetter, M. M. Lunden, J. D. Marshall, C. J. Portier, R. C. H. Vermeulen, and S. P. Hamburg. High-resolution air pollution mapping with google street view cars: Exploiting big data. *Environmental Science & Technology*, 51:6999–7008, 2017.

[37] Sarah E Chambliss, Chelsea V Preble, Julien J Caubel, Troy Cados, Kyle P Messier, Ramon A Alvarez, Brian LaFranchi, Melissa Lunden, Julian D Marshall, Adam A Szpiro, et al. Comparison of mobile and fixed-site black carbon measurements for high-resolution urban pollution mapping. *Environmental Science & Technology*, 54(13):7848–7857, 2020.

[38] Ahmad Bin Thaneya, Joshua S Apte, and Arpad Horvath. A human exposure-based traffic assignment model for minimizing fine particulate matter (pm2. 5) intake from on-road vehicle emissions. *Environmental Research Letters*, 17(7):074034, 2022.

[39] Chirag Manchanda, Robert Harley, Julian Marshall, Alexander Turner, and Joshua Apte. Integrating mobile and fixed-site black carbon measurements to bridge spatiotemporal gaps in urban air quality. *ChemRxiv preprint*, 2023. URL https://doi.org/10.26434/chemrxiv-2023-d4q7n.

[40] Priyanka DeSouza, Amin Anjomshoaa, Fabio Duarte, Ralph Kahn, Prashant Kumar, and Carlo Ratti. Air quality monitoring using mobile low-cost sensors mounted on trash-trucks: Methods development and lessons learned. *Sustainable Cities and Society*, 60:102239, 2020.

[41] Priyanka N deSouza, Sudheer Ballare, and Deb A Niemeier. The environmental and traffic impacts of warehouses in southern california. *Journal of Transport Geography*, 104:103440, 2022.

[42] Priyanka deSouza, An Wang, Yuki Machida, Tiffany Duhl, Simone Mora, Prashant Kumar, Ralph Kahn, Carlo Ratti, John L Durant, and Neelakshi Hudda. Evaluating the performance of low-cost pm2. 5 sensors in mobile settings. *arXiv preprint arXiv:2301.03847*, 2023.

[43] S. Baidar, R. M. Hardesty, S.-W. Kim, A. O. Langford, H. Oetjen, C. J. Senff, M. Trainer, and R. Volkamer. Weakening of the weekend ozone effect over california's south coast air basin. *Geophysical Research Letters*, 42:9457–9464, 2015.

[44] A. Salvo and F. M. Geiger. Reduction in local ozone levels in urban são paulo due to a shift from ethanol to gasoline use. *Nature Geosci*, 7:450–458, 2014.

[45] E. M. Fujita, D. E. Campbell, W. R. Stockwell, and D. R. Lawson. Past and future ozone trends in california's south coast air basin: Reconciliation of ambient measurements with past and projected emission inventories. *Journal of the Air & Waste Management Association*, 63: 54–69, 2013.

[46] J. I. Levy, J. J. Buonocore, and K. von Stackelberg. Evaluation of the public health impacts of traffic congestion: a health risk assessment. *Environ Health*, 9:65, 2010.

[47] A. Bento, M. Freedman, and C. Lang. Who benefits from environmental regulation? evidence from the clean air act amendments. *Review of Economics and Statistics*, 97:610–622, 2015.

[48] A. A. Karner, D. S. Eisinger, and D. A. Niemeier. Near-roadway air quality: Synthesizing the findings from real-world data. *Environmental Science & Technology*, 44:5334–5344, 2010.

[49] N. Zhong, J. Cao, and Y. Wang. Traffic congestion, ambient air pollution, and health: Evidence from driving restrictions in beijing. *Journal of the Association of Environmental and Resource Economists*, 4:821–856, 2017.

[50] D. Jacob. *Introduction to Atmospheric Chemistry*. Princeton University Press, 2000. URL https://press.princeton.edu/books/hardcover/9780691001852/introduction-to-atmospheric-chemistry.

[51] EPA. Naaqs table. Technical report, US EPA Office of Air and Radiation, 2014. URL https://www.epa.gov/criteria-air-pollutants/naaqs-table.

[52] X. Wang, S. Gronstal, B. Lopez, H. Jung, L.-W. A. Chen, G. Wu, S. S. H. Ho, J. C. Chow, J. G. Watson, Q. Yao, and S. Yoon. Evidence of non-tailpipe emission contributions to pm2.5 and pm10 near southern california highways. *Environmental Pollution*, 317:120691, 2023.

[53] Kenneth Small and Erik T Verhoef. *The economics of urban transportation*. Routledge, 2007.

[54] F. Oroumiyeh, M. Jerrett, I. Del Rosario, J. Lipsitt, J. Liu, S. E. Paulson, B. Ritz, J. J. Schauer, M. M. Shafer, J. Shen, S. Weichenthal, S. Banerjee, and Y. Zhu. Elemental composition of fine and coarse particles across the greater los angeles area: Spatial variation and contributing sources. *Environmental Pollution*, 292:118356, 2022.

[55] F. Xia, X. Cheng, Z. Lei, J. Xu, Y. Liu, and Y. Zhang. Heterogeneous impacts of local traffic congestion on local air pollution within a city: Utilizing taxi trajectory data. *Journal of Environmental Economics and Management*, 122:102896, 2023.

[56] I. W. H. Parry, M. Walls, and W. Harrington. Automobile externalities and policies. *Journal of Economic Literature*, 45:373–399, 2007.

[57] C. Nehiba. Correcting heterogeneous externalities: Evidence from local fuel taxes. *Journal of the Association of Environmental and Resource Economists*, 2021. 717418.

[58] C40 Cities. Milan's area c reduces traffic pollution and transforms the city center. Technical report, C40 Cities, 2015. URL https://www.c40.org/case-studies/milan-s-area-c-reduces-traffic-pollution-and-transforms-the-city-center.

[59] TfL. Ulez: Where and when - transport for london. Technical report, Transport for London, 2023. URL https://tfl.gov.uk/modes/driving/ultra-low-emission-zone/ulez-where-and-when.

[60] I. Laña, J. Del Ser, A. Padró, M. Vélez, and C. Casanova-Mateo. The role of local urban traffic and meteorological conditions in air pollution: A data-based case study in madrid, spain. *Atmospheric Environment*, 145:424–438, 2016.

[61] C. P. Green, J. S. Heywood, and M. Navarro Paniagua. Did the london congestion charge reduce pollution? *Regional Science and Urban Economics*, 84:103573, 2020.

[62] M. Gibson and M. Carnovale. The effects of road pricing on driver behavior and air pollution. *Journal of Urban Economics*, 89:62–73, 2015.

[63] E. Simeonova, J. Currie, P. Nilsson, and R. Walker. Congestion pricing, air pollution, and children's health. *Journal of Human Resources*, 56:971–996, 2021.

[64] S. Griffin, S. Walker, and M. Sculpher. Distributional cost effectiveness analysis of west yorkshire low emission zone policies. *Health Economics*, 29:567–579, 2020.

[65] S. Stewart. Congestion pricing plan. https://www.manhattanbp.nyc.gov/initiatives/congestion-pricing-plan/, 2023.

[66] Zero emission delivery zone. https://www.santamonica.gov/zero-emission-delivery-zone, 2023.

[67] D. J. Jacob and D. A. Winner. Effect of climate change on air quality. *Atmospheric Environment*, 43:51–63, 2009.

[68] A. van Benthem. What is the optimal speed limit on freeways? *Journal of Public Economics*, 124:44–62, 2015.

[69] J. Gallagher and P. J. Fisher. Criminal deterrence when there are offsetting risks: Traffic cameras, vehicular accidents, and public safety. *American Economic Journal: Economic Policy*, 12:202–237, 2020.

[70] C. K. Tang. Do speed cameras save lives? http://www.spatialeconomics.ac.uk/, 2017.

[71] M. P. Keuken, S. Jonkers, I. R. Wilmink, and J. Wesseling. Reduced nox and pm10 emissions on urban motorways in the netherlands by 80km/h speed management. *Science of The Total Environment*, 408:2517–2526, 2010.

[72] I-5 variable speed limits and lane control, washington state. https://transportationops.org/case-studies/i-5-variable-speed-limits-and-lane-control-washington-state, 2017.

[73] I-95 northbound variable speed limits - projects. https://www.virginiadot.org/projects/fredericksburg/i-95-northbound-variable-speed-limits.asp, 2023.

[74] Next generation motorway traffic control. https://www.kth.se/ctr/research/current-projects/next-generation-motorway-traffic-control-1.873518, 2023.

[75] G. Bel and J. Rosell. Effects of the 80km/h and variable speed limits on air pollution in the metropolitan area of barcelona. *Transportation Research Part D: Transport and Environment*, 23:90–97, 2013.

[76] Emissions Analytics (EA). Press release: Pollution from tyre wear 1,000 times worse than exhaust emissions. https://www.emissionsanalytics.com/news/pollution-tyre-wear-worse-exhaust-emissions, 2020.

[77] K. Patel. Why tires — not tailpipes — are spewing more pollution from your cars. https://www.washingtonpost.com/climate-environment/2023/07/09/tire-brake-tailpipes-emissions-pollution-cars/, 2023.

[78] J. C. Fussell, M. Franklin, D. C. Green, M. Gustafsson, R. M. Harrison, W. Hicks, F. J. Kelly, F. Kishta, M. R. Miller, I. S. Mudway, F. Oroumiyeh, L. Selley, M. Wang, and Y. Zhu. A review of road traffic-derived non-exhaust particles: Emissions, physicochemical characteristics, health risks, and mitigation measures. *Environ. Sci. Technol.*, 56:6813–6835, 2022.

[79] A. C. Cameron, J. B. Gelbach, and D. L. Miller. Robust inference with multiway clustering. *Journal of Business & Economic Statistics*, 29:238–249, 2011.

[80] Alexandre Belloni, Daniel Chen, Victor Chernozhukov, and Christian Hansen. Sparse models and methods for optimal instruments with an application to eminent domain. *Econometrica*, 80(6):2369–2429, 2012.

[81] E. Derenoncourt. Can you move to opportunity? evidence from the great migration. *American Economic Review*, 112:369–408, 2022.

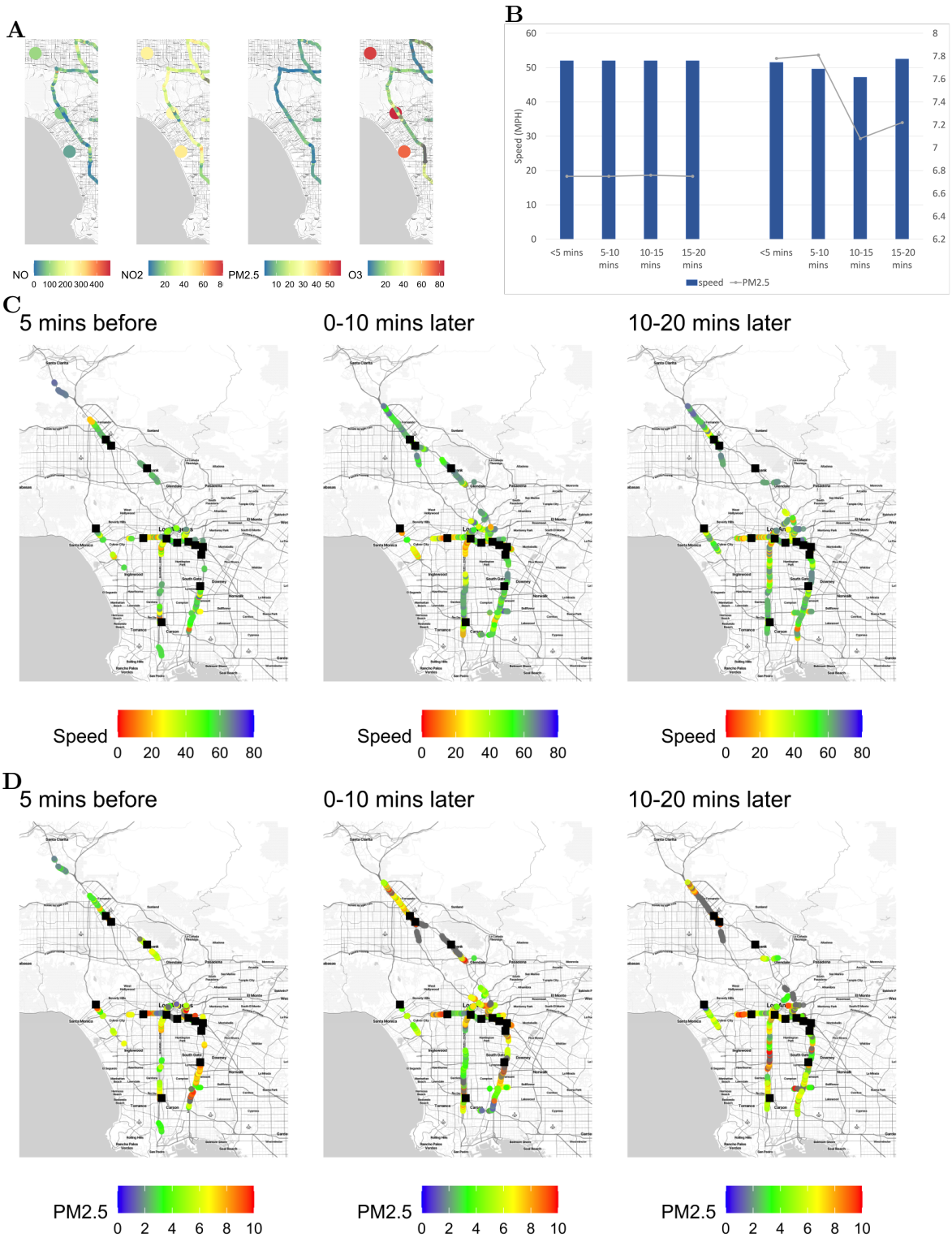


Figure 1: **Value of Road-Level Pollution Monitoring: An Example from Roadway Accidents** (**A**) Hourly maximum of on-road pollution from Aclima sensors along Los Angeles highways and EPA ground monitors (large circles) in October 2016. (**B**) average speed and $PM_{2.5}$ pollution for sensors where with accident means between 0.5 and 1 mile away and without means >1 mile away. (**C**) propagation of congestion from accident (black squares) as shown via freeway speeds, (**D**) propagation of $PM_{2.5}$ pollution from accidents.

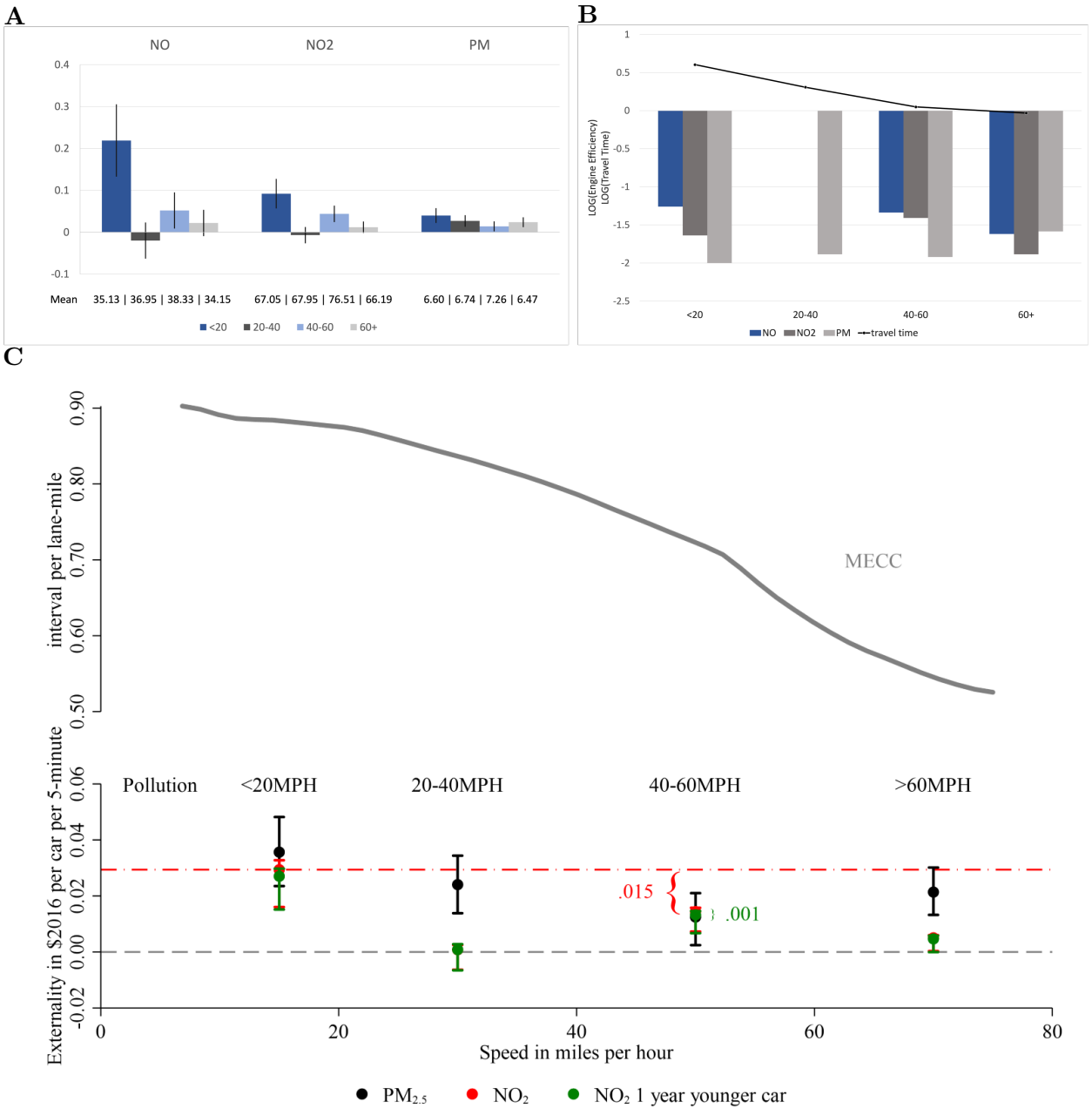


Figure 2: **Estimated effects of Congestion on Pollution** (**A**) coefficient plot of pollution effects of vehicle density by speed bin and pollutant. (**B**) decomposed log of pollution contribution from one vehicle, which is log of engine efficiency (pollution per vehicle per minute) + log of travel time (minutes per mile). (**C**) implied marginal external congestion costs (MECC) and pollution damages from adding one car (same y-axis with broken interval), with effect compared to a 1 year younger car from [33].

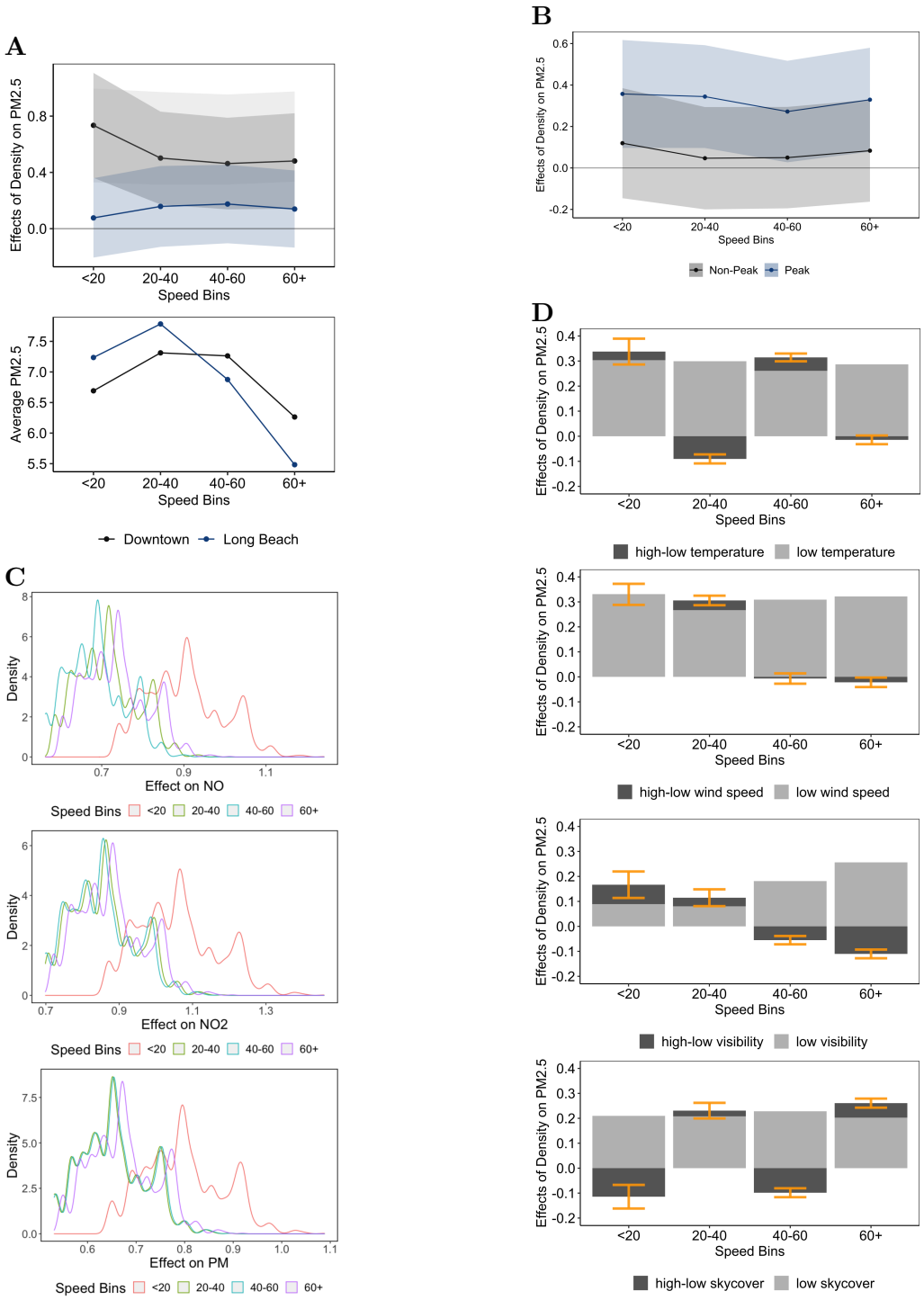


Figure 3: **Heterogeneity of $PM_{2.5}$ Effects of Vehicle Density** (**A**) Coefficient plot of pollution effect of density at two locations in LA (Downtown and Long Beach) and average $PM_{2.5}$ levels. (**B**) Coefficient plot of pollution effect of density during peak and off-peak periods. (**C**) Density of marginal pollution effects. (**D**) Effect of weather severity on $PM_{2.5}$ pollution effect estimates.

Supplemental Material for: "Local Pollution Externalities from Driving: Evidence from Roadway Vehicle Sensors"

Andrew R. Waxman[a,1], Ruozi Song[b], Rajat Kochhar[c], and Antonio M. Bento[d,e]

[a]University of Texas at Austin
[b]World Bank
[c]University of Chicago
[d]University of Southern California
[e]National Bureau of Economic Research

March 27, 2024

Speed-Density Relationship

One challenge to understanding the relationship between congestion is the physical relationship between travel demand and speeds, illustrated in panel (A) of Figure SI1. If we are interested in the pollution effect of adding one additional vehicle to a roadway, this effect depends upon the vehicle's speed. While speed and vehicular density generally have a linear relationship as indicated by the top left panel, vehicle flows (measured in vehicles per hour per lane) have a backward-bending relationship.

Panel (B) of Figure SI1 makes this relationship clear for our data, where the horizontal axis displays the speed of Google Street View vehicles carrying Aclima sensors in our data, while the vertical axis shows the density of vehicles from PeMS in cars per mile. An uncongested freeway begins with relatively high speeds (bottom right corner of figure), which slow down as density increases with congestion. However, this relationship is not linear in practice, and for speeds below 50 MPH, density is increasing by much less and is relatively constant for speeds below 20 MPH. It is this distinction between speed and density as well as the relationship between speed and pollution discussed in the next section that motivates our empirical design's use of speed bins to allow density's effect on pollution to depend on the level of speeds.

Air Chemistry

A concern for our study is the complicated air chemistry that occurs between emissions from the transportation system and other sources. NO is emitted from vehicles, converted to NO_2 and subsequently to ground-level ozone and fine particulate matter ($PM_{2.5}$). Particulate matter 2.5 microns and smaller ($PM_{2.5}$) is produced through two channels: directly from vehicle brakes and tires and other nearby activities (e.g., woodburning or agriculture) or indirectly through atmospheric reactions between NO_x, SO_2 and VOCs. NO_x is a blanket term to describe the concentrations of NO and NO_2, which are primarily the result of combustion of hydrocarbons including internal

combustion engines in motor vehicles. Typically, NO emitted from vehicle tail pipes is rapidly converted into NO_2 in the air within roughly a minute's time. NO_2 is, in turn, converted into ozone in the presence of volatile organic compounds (VOCs), which do not come from the vehicles but from point sources such as industrial facilities. This conversion of NO_x and VOCs to ozone is approximately in fixed proportions and the Los Angeles air basin is VOC-limited. This creates an empirical challenge to our approach since we are unable to control for variation in background level of VOCs, which subsequently result in conversion of NO_x into ozone. It also means that the addition of NO_x to the air from vehicle pollution could either increase or decrease the concentration of ozone depending on whether VOC levels are binding. Lastly, the rate of conversion is determined, in part, by temperature sunlight and the angle of the sun and so we include a rich set of weather controls in our subsequent analysis to address this.

Additional Covariates

We collect hourly data on nine different climatic variables to control for the impact of weather on pollution. These include temperature, precipitation, relative humidity, wind direction, wind speed, sky cover, visibility, dew point and air pressure. Weather data is sourced from National Centers for Environmental Information (NCEI), which provides data for different weather stations at an hourly temporal frequency. We use nine weather stations across Los Angeles: Burbank, Chino, Fullerton, Hawthorne, LA International Airport, Long Beach, Ontario, Santa Monica, and Van Nuys. To avoid issues related to multicollinearity in the regression analysis, we only use four weather variables: temperature, wind speed, wind direction, and relative humidity. For each Aclima observation, we find the closest weather station in terms of the geodesic distance. The temporally closest weather readings from such station are then assigned to the observation.

Though we are interested in estimating the average impact of density on pollution across the entirety of Los Angeles, we also perform heterogeneity analysis by regions. Specifically, we divide Los Angeles into three different geographical regions: Downtown, Long Beach and Santa Monica. The shapefiles for these three regions were constructed using maps provided by the County of Los Angeles (LA County eGIS Program) and the LA Times Neighborhood Boundaries.

The marginal contribution of one vehicle to pollution will depend upon the pollution intensity of that vehicle. As a result, times of the day or segments of freeway with a larger proportion of heavy diesel vehicles will result in higher additional pollution concentrations. To account for this heterogeneity, we control for vehicular composition on freeways at different times of day and locations in our sample using hourly vehicular class census data collected by Caltrans available via PeMS. We include data from 43 census monitors providing hourly information on the flow of different classes of vehicles, ranging from class motorcycles to multi-axle, multi-trailer trucks. Census observations were not taken during our sample period with sufficient coverage, so we use hour-by-location averages from before our sample period in 2010. We plot the distribution of these data in Figure SI5.

More Details on Aclima Sensors & Data

In 2015, Google signed a cooperative agreement with the air quality monitoring startup Aclima to outfit select Google Street View vehicles with high frequency air quality monitoring devices [1]. Street View cars drive urban highways and side streets to take pictures of addresses from the road. Roadway coverage includes almost all the surface roads and adjacent freeways of Downtown Los

Angeles, Santa Monica, Westchester, Playa del Rey, Lakewood, Wilmington and Downtown Long Beach.

Constructing Accident Data

There were 42,449 accidents reported in our CHP data from August to October in 2016. Of the accidents that happened at the same time and location of Aclima observations; 77.2% of them lasted longer than 5 minutes; 59.8% of them last longer than 15 minutes; 51.6% longer than 20 minutes; 44.0% longer than 25 minutes. We merged the accident data by time and location with Aclima observations and removed accidents that lasted less than 5 minutes. We remove the accidents that lasted less than 5 minutes. Two types of IVs are created.

1. $Incidents_{d_1-d_2miles,tmins}$ indicates number of accidents that happened t minutes before the Aclima observation and was d_1 to d_2 miles away.

2. $Dur_{d_1-d_2miles,tmins}$ indicates the average duration of accidents that happened t minutes before the Aclima observation and was d_1 to d_2 miles away.

As reported in the main text, we set d_1 and d_2 to result in bins of 0.5-1, 1-2, 3-4, and 4-5 miles. We bin t into less than 5, 5-10, 10-15, and 15-20 minutes before the Aclima observation.

Congestion-Pollution Relationship

To understand the relationship between vehicle density and pollution, consider the determinants of pollution concentrations measured by Aclima sensors: during hour, t, and location i (e.g., a given mile stretch of the LA freeways), the sensor registers a concentration of a given pollutant of $p_{i,t}$, measured in micrograms per cubic meter ($\mu g/m^3$).

Pollution concentrations in our data are reported as parts per billion (ppb), which can be converted to $\mu g/m^3$ by accounting for the mass of the pollutant molecule.

The pollution concentration is a stock measure of density of pollution in a given volume of air, which is determined by the following relationship:

$$p_{i,t} = p_{i,t-1} + e_{i,t} + \Delta w_{i,t} + \Delta pc_{i,t}. \tag{1}$$

$p_{i,t-1}$ is the pollution concentration in the previous period, $t-1$, $e_{i,t}$ is the total emissions of the pollutant from vehicles in micrograms per hour at location i. $\Delta w_{i,t}$ is the net transfer of pollution molecules by wind and air circulation to location i in micrograms per hour, while $\Delta pc_{i,t}$ reflects net changes in the production or conversion of micrograms of the pollutant via photochemical processes.

The component of interest from (1) is $e_{i,t}$, which will be determined, in part, by traffic congestion. This term is the sum of emissions per vehicle, E_c, from all vehicles c traveling through i (the set $C_{i,t}$) during hour t:

$$\begin{aligned} e_{i,t} &= \sum_{c \in C_{i,t}} E_c \left(speed'(t), ee_c(speed) \right) \\ &\approx V_{i,t} \cdot \bar{\varepsilon}_{i,t} \left(\mathbf{speed}'(t), \mathbf{ee}\left(\mathbf{speed}\right) \right) \\ &= \kappa_i \cdot density_{i,t} \cdot speed_{i,t} \cdot \bar{\varepsilon}_{i,t} \left(\mathbf{speed}'(t), \mathbf{ee}\left(\mathbf{speed}\right) \right) \end{aligned} \tag{2}$$

which itself is determined by the acceleration of that vehicle, $speed'(t)$ and its engine efficiency, $ee_c(speed)$ at its traveling *speed*.

As the second line of (2) these total emissions for i, t can be approximated by the product of the vehicle flow (in vehicles per hour) and an emissions factor, $\bar{\varepsilon}_{i,t}$, measured in micrograms per vehicle, which is the average emissions per vehicle for cars at location i during hour t reflected by the vectors of vehicle acceleration and engine efficiency. As the last line indicates, vehicle flow is the product of vehicle density (cars per lane-mile) and speed (miles per hour), which we can substitute into (2). The parameter κ_i reflects the number of lanes at location i helping to determine the effective capacity. Combining equations (1) and (2) and totally differentiating yields:

$$\begin{aligned}
\frac{\mathrm{d}p_{i,t}}{\mathrm{d}density_{i,t}} &= \kappa_i \cdot speed_{i,t} \cdot \bar{\varepsilon}_{i,t}(\cdot) && \textit{density effect} \\
&+ \kappa_i \cdot density_{i,t} \cdot \bar{\varepsilon}_{i,t}(\cdot) \frac{\mathrm{d}speed_{i,t}}{\mathrm{d}density_{i,t}} && \textit{travel time effect} \\
&+ \kappa_i \cdot density_{i,t} \cdot speed_{i,t} \cdot \sum_{c \in C_{i,t}} \frac{\mathrm{d}speed'_c(t)}{\mathrm{d}density_{i,t}} \cdot \frac{\partial \bar{\varepsilon}_{i,t}(\cdot)}{\partial speed'_c(t)} && \textit{acceleration effect} \\
&+ \kappa_i \cdot density_{i,t} \cdot speed_{i,t} \cdot \sum_{c \in C_{i,t}} \frac{\mathrm{d}ee_c}{\mathrm{d}density_{i,t}} \cdot \frac{\partial \bar{\varepsilon}_{i,t}(\cdot)}{\partial ee_c} && \textit{engine efficiency effect}
\end{aligned} \tag{3}$$

(3) identifies four forces determining the effect of adding a single vehicle on pollution (as measured by $density_{i,t}$ to a stretch of roadway at a particular time): i) a direct density effect, which captures the fact that slower speeds correspond to times when there are more cars and therefore more emissions on the road; ii) an indirect travel time effect, reflecting the fact that slower cars are emitting for a longer duration at location i; iii) an acceleration effect, capturing the fact that at slower speeds there may be more start and stop traffic and therefore more acceleration. The sign of the acceleration effect could be the opposite in instances in which cars have to apply more acceleration to obtain the same speed, such as when driving uphill or into the wind. iv) Lastly, there is likely an engine efficiency effect, which determines the emissions to drive a given car one mile at a given speed: low or high speeds will lower efficiency and heavier vehicles will use more fuel and thus emit more for the same speed change.

Second, Aclima sensors are recording at all locations at all times in our data, but rather are at particular locations at particular times of day. Pollution concentrations are higher closer to downtown and Long Beach although there is considerable variation in space. Another useful comparison is made in panel (A) of Fig. 1, which compares pollution concentrations for four pollutants between our Aclima measures and corresponding measures from EPA Air Quality System ground observations commonly used in the literature on pollution. The EPA measures are hourly maximums, and so can be expected to be higher on average than average Aclima measures.

One thing the data make clear is that pollution levels can vary substantially as one moves away from EPA ground monitor locations, a point which has been made in other related studies.

Channels of Pollution Formation

Using coefficients from panel A of Fig. 2, we decompose the contribution of marginal pollution into two components: an *engine efficiency effect* and *travel time effect*. The engine efficiency effect reflects the fact that at speeds above and below 20-60 MPH, typical passenger vehicle engines are expected to be less efficient and thus emit more pollution per vehicle-mile. For $PM_{2.5}$, engine

efficiency also reflects the fact that more emissions arise from tires at a lower speed in traffic when braking and restarting occur frequently. On the other hand, the *travel time effect* reflects the fact that when roads are congested, speeds will be lower and so the same vehicle will spend a longer period of time along the same mile of road, resulting in a higher concentration of pollution from that vehicle. In equation (SI.4), we write a simple identity to relate the contribution of the engine efficiency and travel time to the pollution from a single vehicle:

$$\text{Pollution contribution of 1 vehicle} = \underbrace{\frac{\text{Pollution contribution of 1 vehicle}}{\text{Travel time}}}_{\text{Engine efficiency effect}} \times \text{Travel time} \tag{4}$$

In the Results section, we report the relative contribution of the engine efficiency effect to travel time in explaining the marginal effect of adding one additional vehicle to the road. From a logarithmic model, this translates to

$$\log(marginal\ effect) = \log(engine\ efficiency) + \log(travel\ time), \tag{5}$$

where *engine efficiency* is the marginal effect per minute and the travel time is in minutes per mile. We define the relative contribution of engine efficiency as reported in panel B of Fig. 2 as:

$$\frac{|\log(engine\ efficiency)|}{|\log(engine\ efficiency)| + |\log(travel\ time)|}. \tag{6}$$

Details on LASSO Estimation

To choose a set of instruments, we estimate the following first-stage regression

$$\min_{\widehat{\theta}} \sum_{i=1}^{n} \left(cars/mile_{ict} - \frac{\widehat{cars}}{mile_{ict}}(\theta) \right)^2 + \lambda \sum_{j=1}^{p} \left| \widehat{\theta}_j \right| \tag{7}$$

where the term in parentheses is a least squares regression, recovering a vector of parameters from equation (1), $\widehat{\theta}$. The regression minimizes the sum of the squared error for actual and predicted vehicle density, the dependent variable. The last term is a penalty function that uses a tuning parameter, λ, to apply a shrinkage penalty to over-specified models. λ is selected as described below.

λ Selection

Figure SI8 shows the effect on our first-stage regression coefficients as the magnitude of for λ increases: the larger the value of λ, the higher proportion of coefficient are set to zero and the more parsimonious the regression equation becomes. Figure S8 also shows how the mean-squared error of the regression increases with $\log(\lambda)$, allowing cross-validation for its selection. As the figure makes clear, the smallest value of λ results in the lowest mean-squared error, but to avoid over-fitting the model, a rule of thumb of choosing a value of λ one standard deviation higher than that at the lowest mean-squared error is often employed, which is indicated by the vertical lines in the figure. We select a value of λ slightly higher that one standard deviation all approximately 0.18. These values occur well before $\log(-4) \approx 0.60$, which is when the mean-squared error begins to grow, so induces greater parsimony without a meaningful impact on prediction error. Figure SI9 shows the first-stage parameter estimates that result from this value of $\log(\lambda)$.

Details on Recovering Pollution Externalities

In this section we describe the procedure for calculating marginal external damages from the air pollution effects estimated from equation (1). These damages capture the negative externality associated with the additional economic cost of air pollution from driving. To gauge the magnitude of these air pollution effects, we benchmark air pollution marginal damages to their equivalent in terms of congestion as reflected by the marginal external cost of congestion (MECC). Details on how the MECC is calculated is presented in the Supplementary Material.

To calculate pollution damages, we use the intake fraction ratio method [2, 3, 4, 5], which approximates the inhalation rate of emissions based on population density and local climate. The intake fraction ratio may be problematic for measuring pollution dispersion over longer distances and time horizons, where a more complicated pattern of interaction with weather and air chemistry affects concentrations. In more localized approaches, as our own, its use may be more justified. Moreover, studies suggest that population exposure is usually, by far, the most important factor [6].

We denote iF_j as the ratio of the inhaled pollution by the exposed population at location j to the total amount of pollution emitted from a specific source and depends on the size of exposed population and meteorological conditions. We use the local intake fraction ratio calculate for Los Angeles in [3] of 43. The steady state approximation is

$$iF_j = Q\frac{LPD_j}{DR_j}, \tag{8}$$

where Q is average breathing rate ($m^3 s^{-1} person^{-1}$), $LPD_j = P_j/\sqrt{A_j}$ is the linear population density ($persons/m$), P_j is population, A_j is urbanized land area, and $DR_j = 1/\left(1/\bar{u_j}H_j\right)$ is the normalized dilution rate (m^2/s) or ventilation coefficient which multiplies wind speed, u_j, by atmospheric mixing height (H_j).

We then calculate marginal external pollution damages as

$$\begin{aligned} MD_{jk} &= \frac{iF_j}{Q} \cdot VSL \cdot \Delta m_k \\ &= \frac{LPD_j}{DR_j} \cdot VSL \cdot \Delta m_k \end{aligned} \tag{9}$$

where MD_{jk} is marginal damage in location j for age group k, VSL is the value of a statistical life, here we assume it is \$8.840 million in 2016 dollars based on [7]. Δm_k is the change in mortality rate for age group k, which is the product of the NO_2 concentration-mortality effect, baseline population and mortality. We use a value of 3% mortality per 10 μg/m^3 from [8], the population of LA, the population of Los Angeles MSA in 2016 (3.958 million) and 6.23 deaths per 1,000 mortality rate [9].

Calculating the Marginal External Cost of Congestion (MECC)

The MECC is the vertical distance between the Marginal Private Benefit MPB, also called average social cost, and the Marginal Social Cost (MSC) as indicated in Figure SI11. The total welfare loss from excess congestion is the area of the triangle shaded in red that lies between socially optimal traffic volumes (ignoring pollution) and the unregulated equilibrium. To calculate the MECC, we need to know the optimal level of vehicle flows, recover the pollution reduction co-benefits of

optimal tolling, and we need to know the co-benefits that occur from moving from unregulated flow levels, V_{it}^0, to the socially optimal level, V_{it}^*.

The co-benefits therefore are

$$\Delta V_{it} \cdot \widehat{\rho}_{it} = \left(V_{it}^0 - V_{it}^*\right) \widehat{\rho}_{it}, \tag{10}$$

where $\widehat{\rho}_{it}$ are flow-adjusted density-pollution estimates from our econometric results. Given the scale at which $\widehat{\rho}_{it}$ are recovered, we define i as being highway segments between PeMS detectors and t as hour by day of week.

We follow past work in the subsequent analysis [10, 11, 12]. First, we solve for optimal traffic volume (cars per hour) by hour, t and road i, V_{it}^*. This consists of 3 steps:

1. estimating empirical speed-density relationship,
2. calibrating a travel demand curve
3. solving for optimal travel demand.

In Figure SI10, step 1 helps recover the slope of the MPC and MSC curves. Step 2 recovers the slope of the demand curve from its intersection with the MPC curve, and optimal volumes come from the intersection of MSC and demand.

Marginal private cost, $MPC\left(V_{it}\right)$, is the cost experienced by commuters and corresponds to average social cost,

$$MPC\left(V_{it}\right) = T\left(V_{it}\right) \cdot o \cdot VOT, \tag{11}$$

where $T\left(V_{it}\right)$ tells the travel time as a function of flow. The marginal social cost is the sum of the marginal private cost and the external cost of congestion, $MECC\left(V_{it}\right)$:

$$\begin{aligned} MSC\left(V_{it}\right) &= MPC\left(V_{it}\right) + MECC\left(V_{it}\right) \\ &= \underbrace{T\left(V_{it}\right) \cdot o \cdot VOT}_{ASC(V_{it})} + \underbrace{T\left(V_{it}\right) \cdot o \cdot VOT \cdot \frac{\epsilon_{it}}{1-\epsilon_{it}}}_{MECC(V_{it})}, \\ &= T\left(V_{it}\right) \cdot o \cdot VOT \cdot \frac{1}{1-\epsilon_{it}}. \end{aligned} \tag{12}$$

The simplest version of a demand curve is

$$V_{it}^D = A_{it} \cdot P_{it}^{-\eta}, \tag{13}$$

where A_{it} is a demand-shifter, P_{it} is the combined time and pecuniary cost of travel, and η is the elasticity of demand with respect to pecuniary travel cost. η needs to be calibrated. We use the following values from [12]: -.44 during peak, -1.32 off-peak, -.88 mid-day.

To recover A_{it}, we can consider the intersection of $MPC\left(V_{it}\right)$ with the inverse demand curve:

$$\begin{aligned} MPC_{it}\left(V_{it}\right) &= P_{it}^D\left(V_{it}\right) \\ T\left(V_{it}\right) \cdot o \cdot VOT &= \left(\tfrac{A_{it}}{V_{it}}\right)^{\frac{1}{\eta}} \\ \frac{2/\widehat{\alpha}}{1+\sqrt{1+4\widehat{\beta}\frac{V_{it}}{\widehat{\alpha}^2}}} \cdot o \cdot VOT &= \left(\tfrac{A_{it}}{V_{it}}\right)^{\frac{1}{\eta}} \\ \Rightarrow A_{it} &= V_{it}\left(\frac{2o \cdot VOT}{\widehat{\alpha}\left(1+\sqrt{1+4\widehat{\beta}\frac{V_{it}}{\widehat{\alpha}^2}}\right)}\right)^{\eta}. \end{aligned}$$

We assume that vehicle occupancy (o) is 1.2 people/car and that the value of time (VOT) is \$10 per hour. To recover optimal flows, V_{it}^*, we want to solve for the locus where travel demand crosses marginal social cost for each i and t :

$$\begin{aligned} P_{it}^D\left(V_{it}^*\right) &= MSC_{it}\left(V_{it}^*\right) \\ \Rightarrow \left(\frac{A_{it}}{V_{it}}\right)^{\frac{1}{\eta}} &= T\left(V_{it}\right) \cdot o \cdot VOT \cdot \frac{1}{1-\epsilon_{it}} \\ &= \frac{2/\widehat{\alpha}}{1+\sqrt{1+4\widehat{\beta}\frac{V_{it}}{\widehat{\alpha}^2}}} \cdot \frac{o \cdot VOT}{1-\epsilon_{it}}, \end{aligned} \tag{14}$$

which can be solved for V_{it}^* using any non-linear solver. From this, optimal tolls can be solved for from $MSC_{it}\left(V_{it}^*\right) - MPC_{it}\left(V_{it}^*\right)$.

We can calculate the marginal external cost of congestion at the social optimum ($MECC\left(V^*\right)$) following [13]:

$$\begin{aligned} MECC\left(V^*\right) &= T\left(V_{it}\right) \cdot o \cdot VOT \cdot \frac{\epsilon_{it}}{1-\epsilon_{it}} \\ &= \frac{2/\widehat{\alpha}}{1+\sqrt{1+4\widehat{\beta}\frac{V_{it}}{\widehat{\alpha}^2}}} \cdot o \cdot VOT \cdot \frac{\epsilon_{it}}{1-\epsilon_{it}}. \end{aligned} \tag{15}$$

Estimation and Calculation of the Speed-Flow-Density Relationship

As Figure SI1 makes clear, estimation of the speed-flow relationship is made challenging by the backward-bending nature of that curve, where the portion of the curve where speeds and flows both decline corresponds to what transportation engineers typically call hypercongestion. We follow much of the recent literature and remove observations for which hypercongestion is likely to occur. To recover the speed-volume elasticity, we need to estimate the following regression:

$$speed_{it} = \alpha + \beta Density_{it} + \mathbf{X_{it}}\gamma + \tau_{\mathbf{it}} + \epsilon_{it}, \tag{16}$$

where $\mathbf{X_{it}}$ are the same whether controls used in equation (1), τ are time and location fixed effects (month, hour, day of week, road/census block group). Here $Density_{it}$ would be measured in vehicles/mile/lane, which is like the density measure we have been using before, but accounting for highway lanes as well (should be in PeMS).

Estimates from (16) of the speed-density elasticity can be recovered in a time and freeway specific format as

$$\epsilon_{it} = -\widehat{\beta}\frac{Density_{it}}{Speed_{it}}. \tag{17}$$

In principle one could recover a single estimate of ϵ from a log-log specification, but there is concern in the literature that this is not sufficiently flexible enough. To recover traffic volumes from speed and density estimates:

$$\widehat{V}_{it} = Density_{it} \cdot \widehat{speed}_{it}, \tag{18}$$

we can also use estimates from (16) to construct travel times as a function of volumes, $T(V_{it})$:

$$T(V_{it}) = \frac{2/\widehat{\alpha}}{1+\sqrt{1+4\widehat{\beta}\frac{V_{it}}{\widehat{\alpha}^2}}}. \tag{19}$$

References

[1] Aclima.io. Aclima comes out of stealth, announces partnerships with google, epa, and lawrence berkeley national labs, 2015. URL https://www.aclima.io/blog/aclima-comes-out-of-stealth-announces-partnerships-with-google-epa-and-lawrence-berkeley-

[2] J. I. Levy, S. Wolff, and J. Evans. A regression-based approach for estimating primary and secondary particulate matter intake fractions. *Risk Analysis*, 22:895–904, 2002.

[3] J. S. Apte, E. Bombrun, J. D. Marshall, and W. W. Nazaroff. Global intraurban intake fractions for primary air pollutants from vehicles and other distributed sources. *Environ. Sci. Technol.*, 46:3415–3423, 2012.

[4] S. Humbert, J. D. Marshall, S. Shaked, J. V. Spadaro, Y. Nishioka, P. Preiss, T. E. McKone, A. Horvath, and O. Jolliet. Intake fraction for particulate matter: Recommendations for life cycle impact assessment. *Environ. Sci. Technol.*, 45:4808–4816, 2011.

[5] I. W. H. Parry, M. D. Heine, E. Lis, and S. Li. *Getting Energy Prices Right: From Principle to Practice*. International Monetary Fund, 2014.

[6] Y. Zhou, J. I. Levy, J. S. Evans, and J. K. Hammitt. The influence of geographic location on population exposure to emissions from power plants throughout china. *Environment International*, 32:365–373, 2006.

[7] EPA. Guidelines for preparing economic analyses. https://www.epa.gov/environmental-economics/guidelines-preparing-economic-analyses, 2000.

[8] R. W. Atkinson, B. K. Butland, H. R. Anderson, and R. L. Maynard. Long-term concentrations of nitrogen dioxide and mortality: A meta-analysis of cohort studies. *Epidemiology*, 29:460, 2018.

[9] CDC. National vital statistics system, mortality: Compressed mortality file 1999-2016 on cdc wonder online database, released june 2017. data are from the compressed mortality file 1999-2016 series 20 no. 2u, 2016, as compiled from data provided by the 57 vital statistics jurisdictions through the vital statistics cooperative program. Technical report, Centers for Disease Control and Prevention, National Center for Health Statistics, 2016. Accessed at http://wonder.cdc.gov/cmf-icd10.html on Jan 26, 2024 2:55:27 PM.

[10] A. Bento, J. D. Hall, and K. Heilmann. Estimating congestion externalities using big data, 2014. 48.

[11] M. L. Anderson. Subways, strikes, and slowdowns: The impacts of public transit on traffic congestion. *American Economic Review*, 104:2763–2796, 2014.

[12] J. Yang, A.-O. Purevjav, and S. Li. The marginal cost of traffic congestion and road pricing: Evidence from a natural experiment in beijing. *American Economic Journal: Economic Policy*, 12:418–53, 2020.

[13] A. Pigou. *The Economics of Welfare*. Routledge, 1920.

Figures & Tables

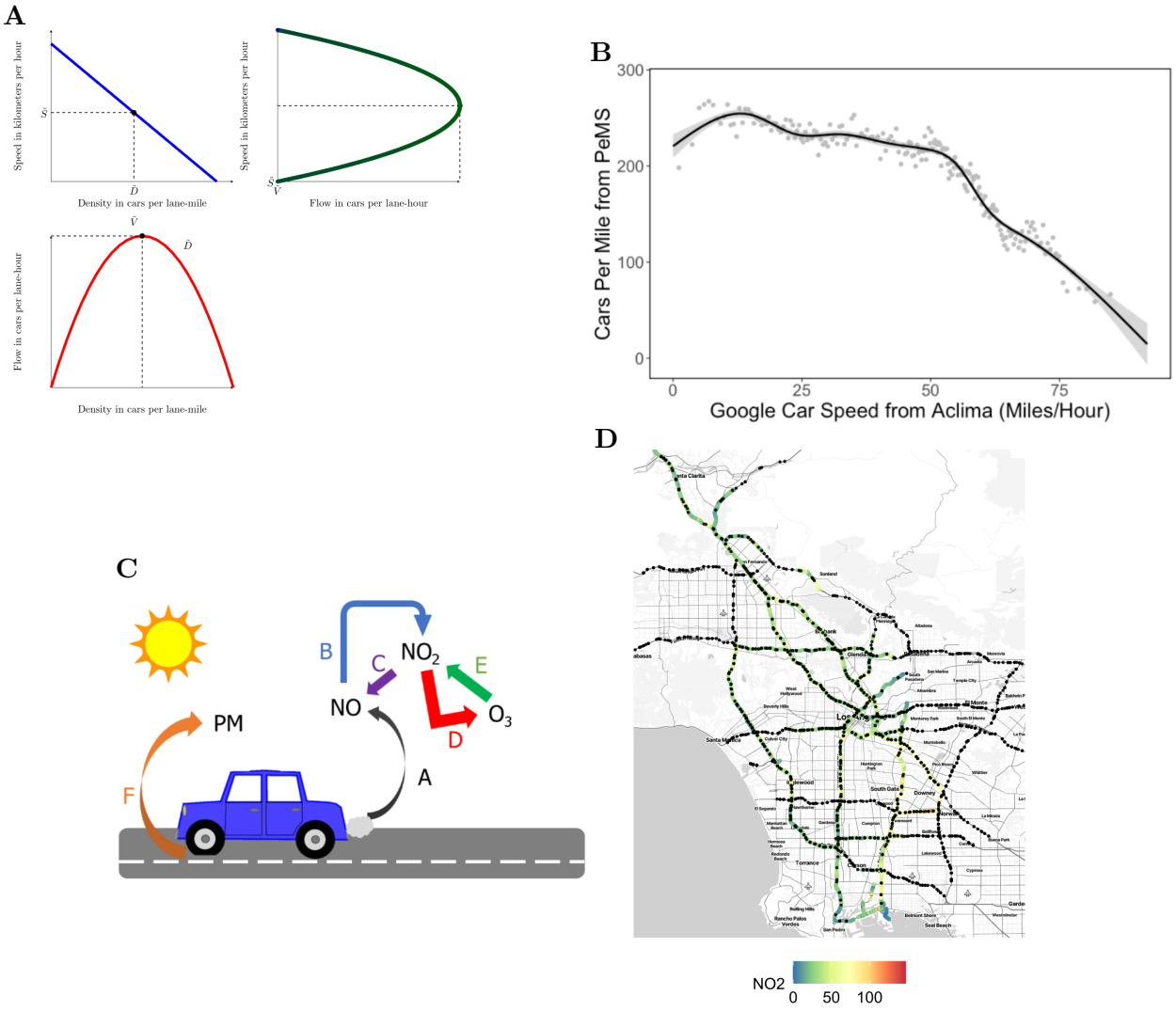


Figure SI1: **Speed-Flow Relationship in Urban Congestion and Pollution Effects** (**A**) illustrates the roadway engineering relationships between speed, density and vehicle flow. Traffic can be characterized by two portions of congestion, standard congestion where, as density increases, the flow increases, and hypercongested flow, where the bottleneck of cars results in flow decrease with density increase. (**B**) illustrates part of this relationship, between Google car speeds from our Aclima data and the density in cars per mile from PeMS observations. (**C**) illustrates the basic air chemistry behind roadway pollution. Arrow $[A]$ shows how NO is a direct emission from vehicle exhaust. This is rapidly converted $[B]$ into NO_2, which, in turn, may convert $[C]$ into O_3 (ozone) given weather conditions and background air chemistry. Both processes may revert (arrows $[C]$ and $[E]$). Lastly, start and stop traffic and acceleration results in particulate matter (PM) formation $[F]$. (**D**) illustrates observations of NO_2 from Aclima sensors and PeMS detectors are indicated with dots.

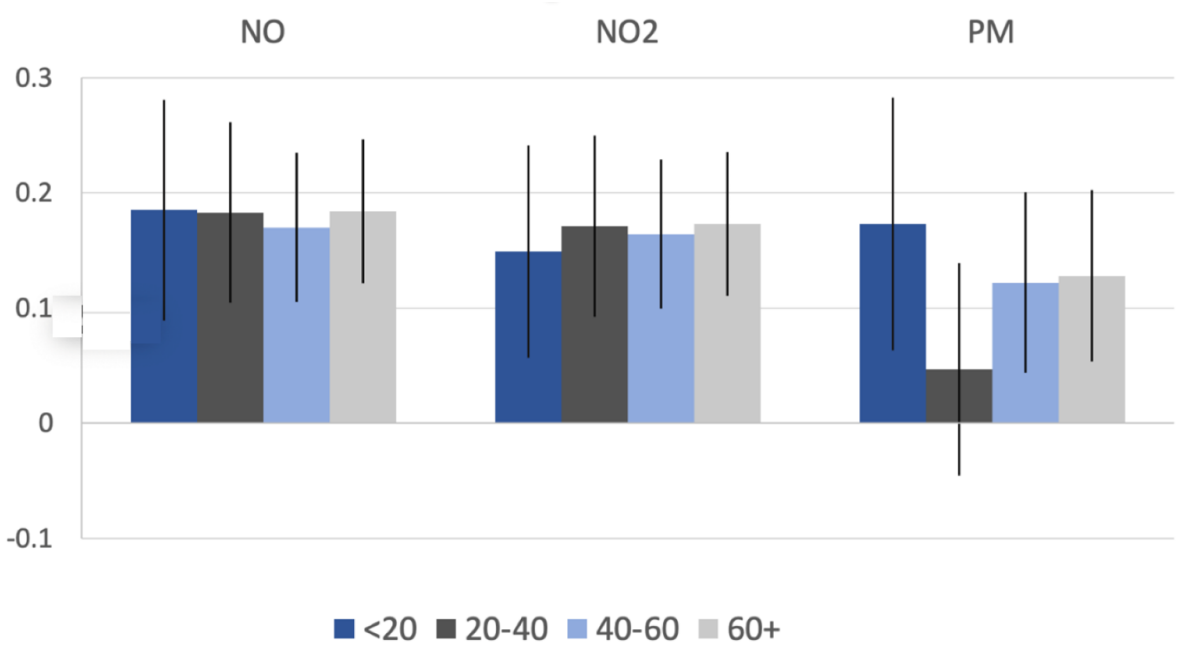


Figure SI2: **Aggregated IV Regressions with LASSO (log-log)** NO, NO_2 and $PM_{2.5}$ Aclima observations are aggregated to daily and 2-digit latitude and longitude level. Accidents that lasted longer than 5 minutes are included. Lagged dependent variables and contemporary ratio between NO and NO_2 are included as controls. Other fixed effects include freeway, month, Google car, fleet composition, and day of the week. Weather controls include second order polynomial of temperature, relative humidity, sky cover, and visibility.

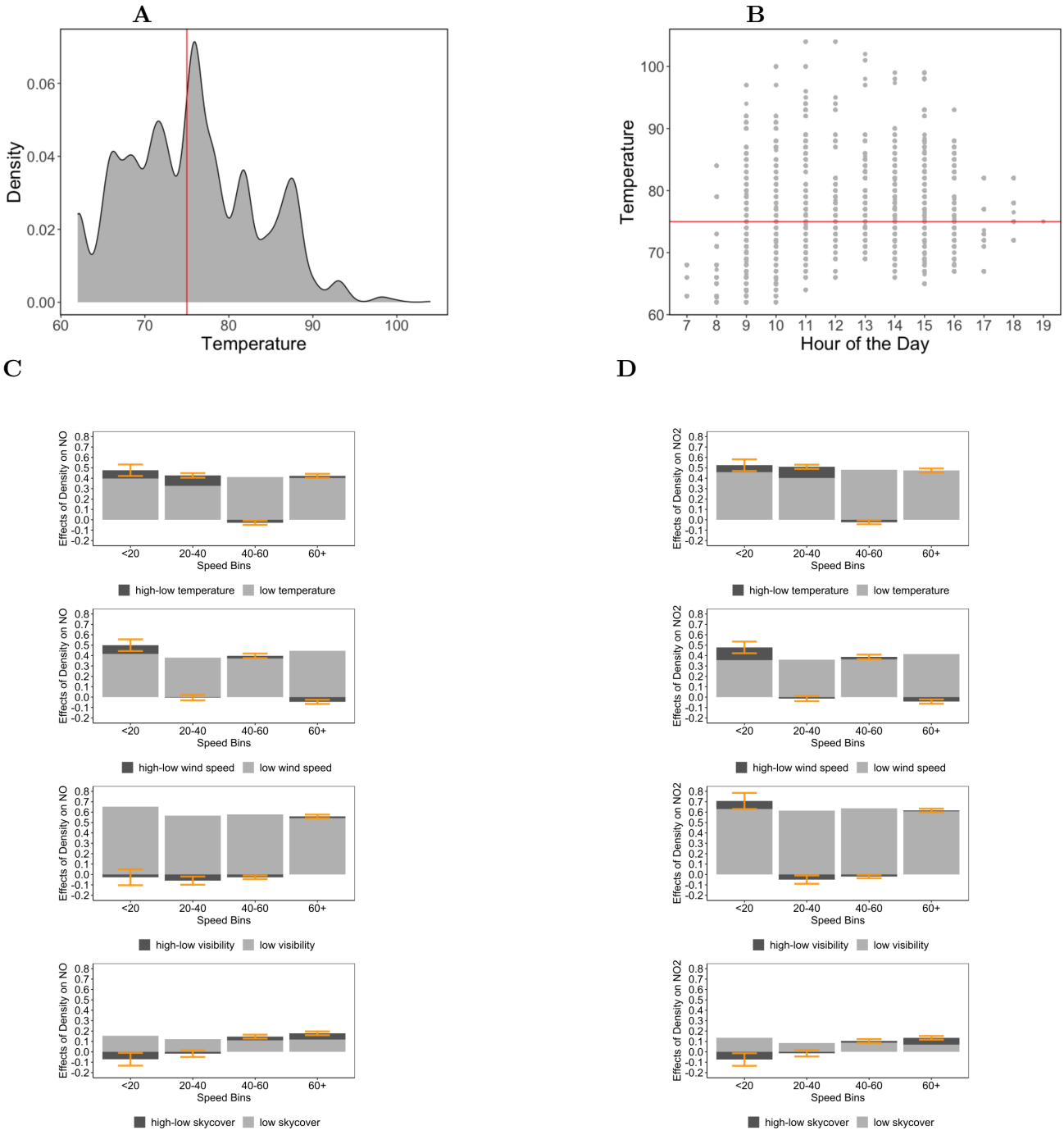
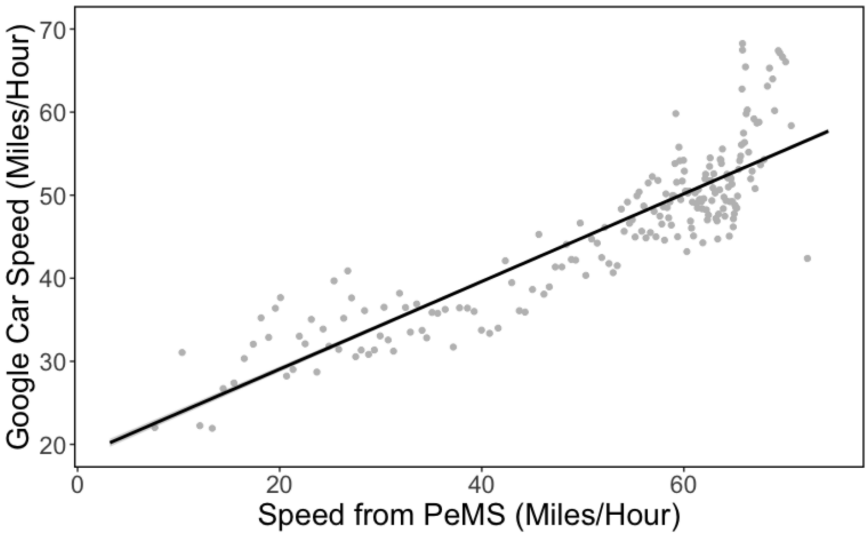


Figure SI3: **Weather variation and its effect on NO_X pollution effect estimates** (**A**) Distribution of temperature (in Fahrenheit) in our sample. (**B**) Distribution of temperature across hours of the day in our sample. (**C**) Effect of weather in top half of distribution on regression results for NO. (**D**) Effect of weather in top half of distribution on regression results for NO_2.

A



B

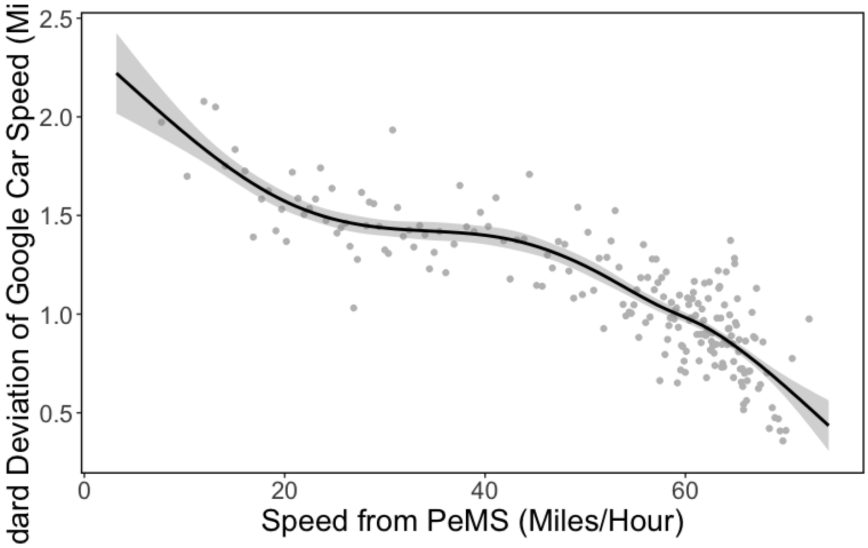


Figure SI4: **Google Street View Vehicle Speeds vs. PeMS Speeds** This figure shows a binned scatter plot and smoothed locally weighted polynomial plot of the relationship between PeMS vehicle speeds (horizontal axis) and the standard deviation of Google Streetview cars (vertical axis) in MPH. 25% of Aclima observations are on freeways. There are 914 PeMS detectors in this study.

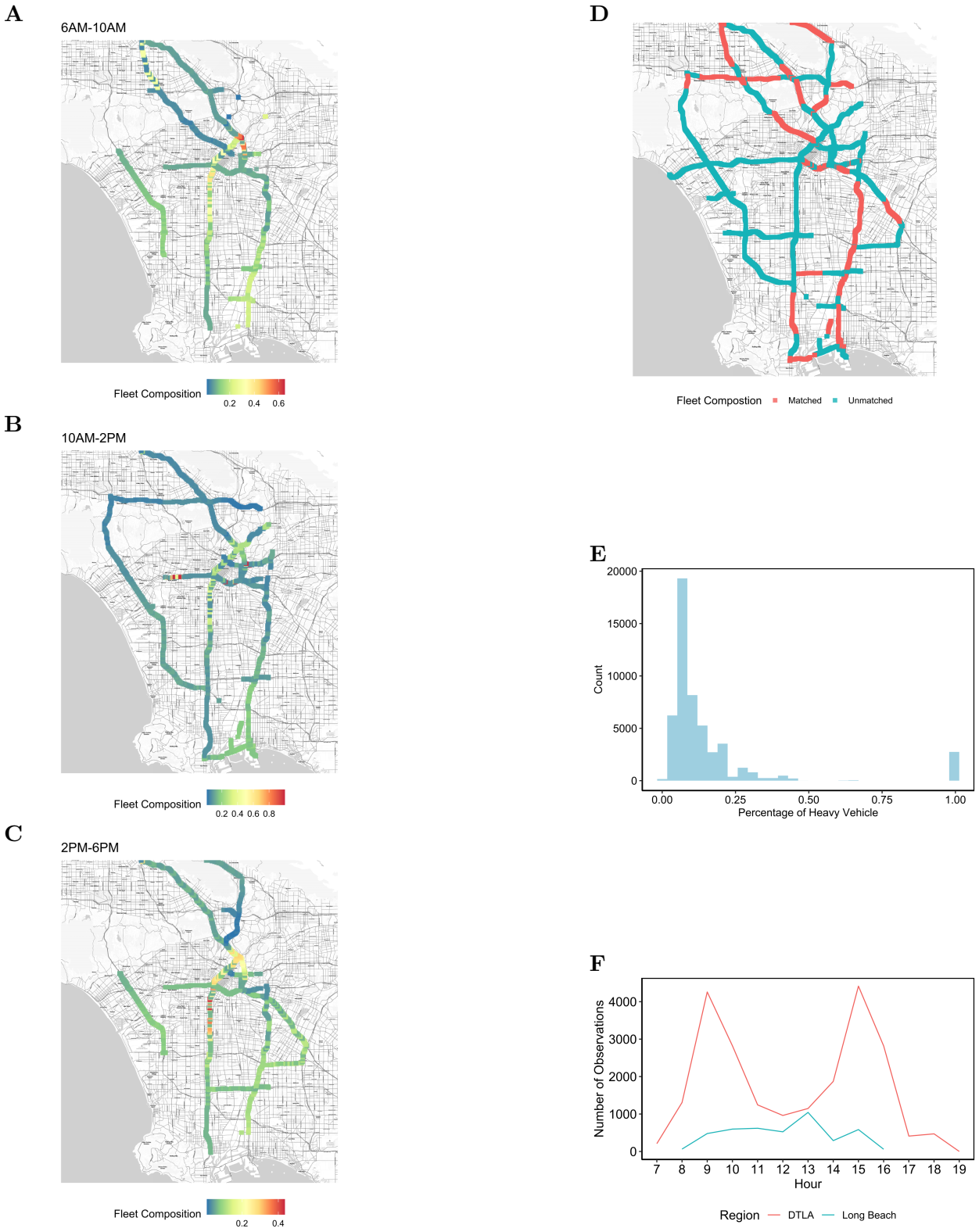
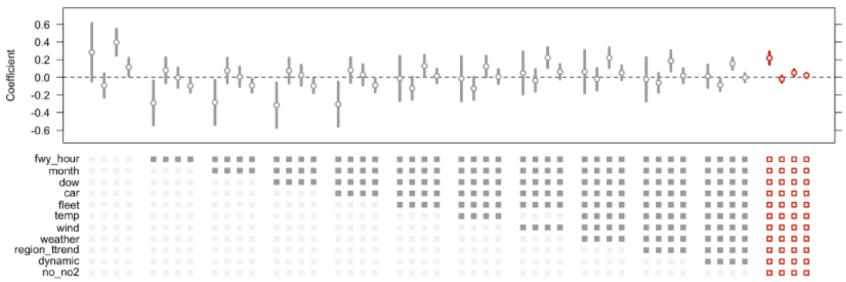
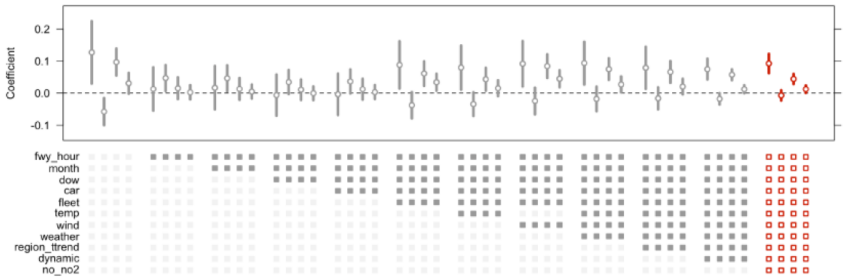


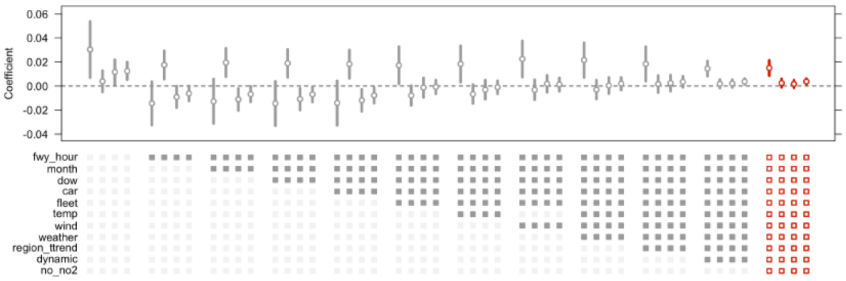
Figure SI5: **Fleet Composition** (**A**) 6AM-10AM, (**B**) 10AM-2PM, (**C**) 2PM-6PM, (**D**) Match Between Aclima and Fleet Composition Data, (**E**) Distribution of Heavy Vehicle Composition Low = 0-8.7%, High = >8.7%. (**F**) Number of Aclima observations by hour and region.



(a) NO



(b) NO_2



(c) $PM_{2.5}$

Figure SI6: **Coefficient Plots of the Effect of Density on Pollution: Robustness** This figure shows coefficient plots for regression models of the impact of vehicle density on air pollution across four speed bins (<20, 20-40, 40-60, >60 MPH) corresponding to each set of four confidence ranges. We repeat the estimation varying the set of regression covariates as indicated by the bullets below each plot. "fleet" are vehicle composition controls, region trend is an interaction between region of LA and a time trend, "dynamic" includes a lagged dependent variable and "no_no2" includes a lagged ratio of NO to NO_2. The rightmost results in red are our preferred specifications shown in the main text.

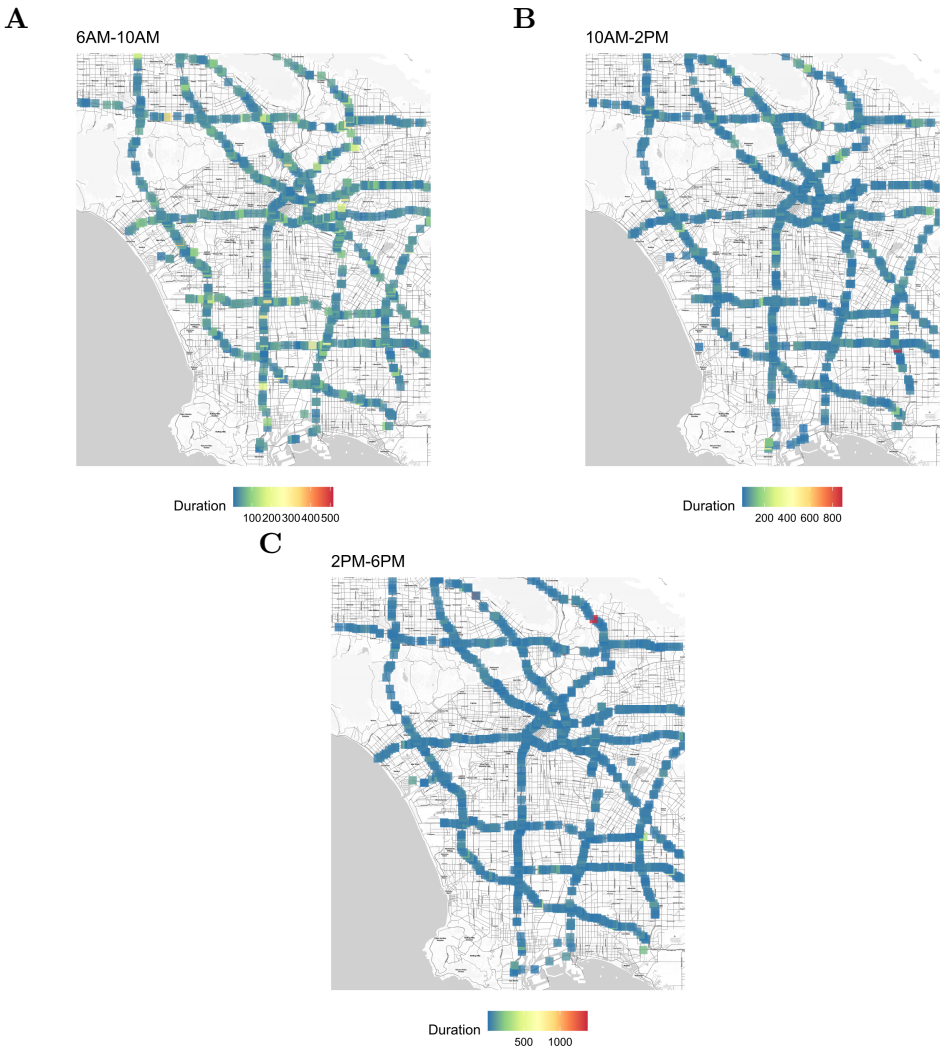


Figure SI7: **Average Accident Durations in Minutes by Time of Day** (**A**) 6AM-10AM, (**B**) 10AM-2PM, (**C**) 2PM-6PM

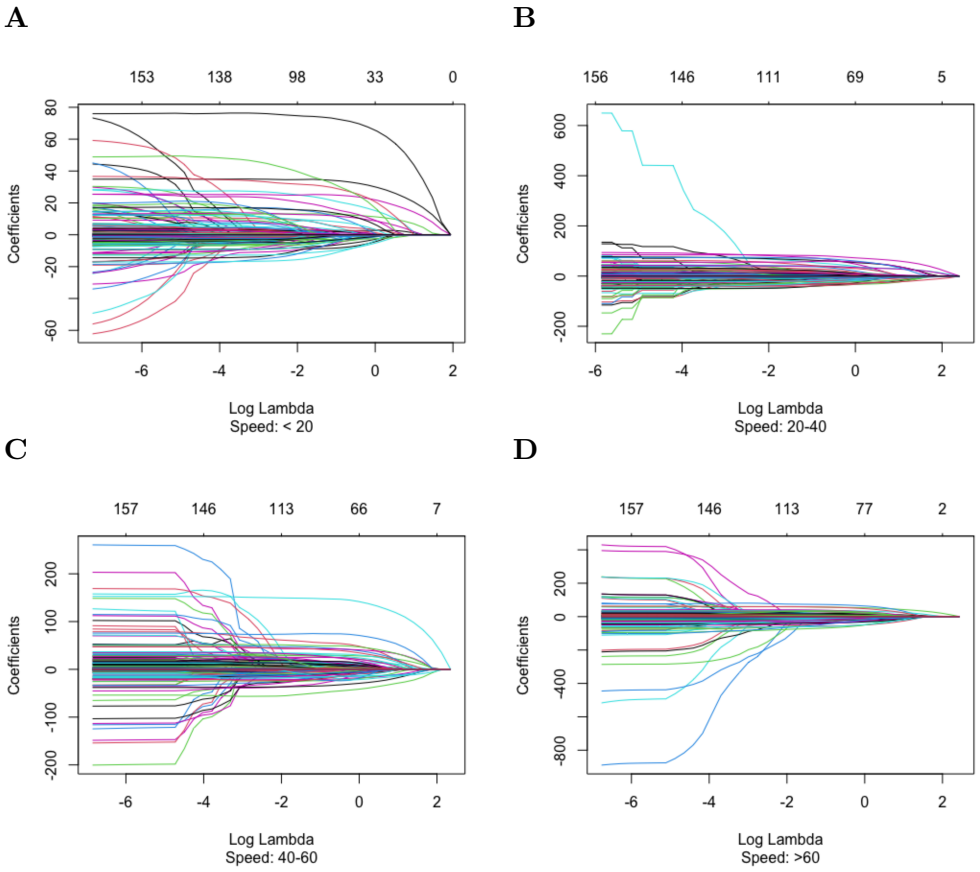


Figure SI8: **Effect of Penalty Parameter on LASSO Coefficients** This figure presents the results of 40 LASSO regressions indicating the values of LASSO parameters for different penalty parameter (lambda) values across the four speed bins. Each curve is a separate parameter.

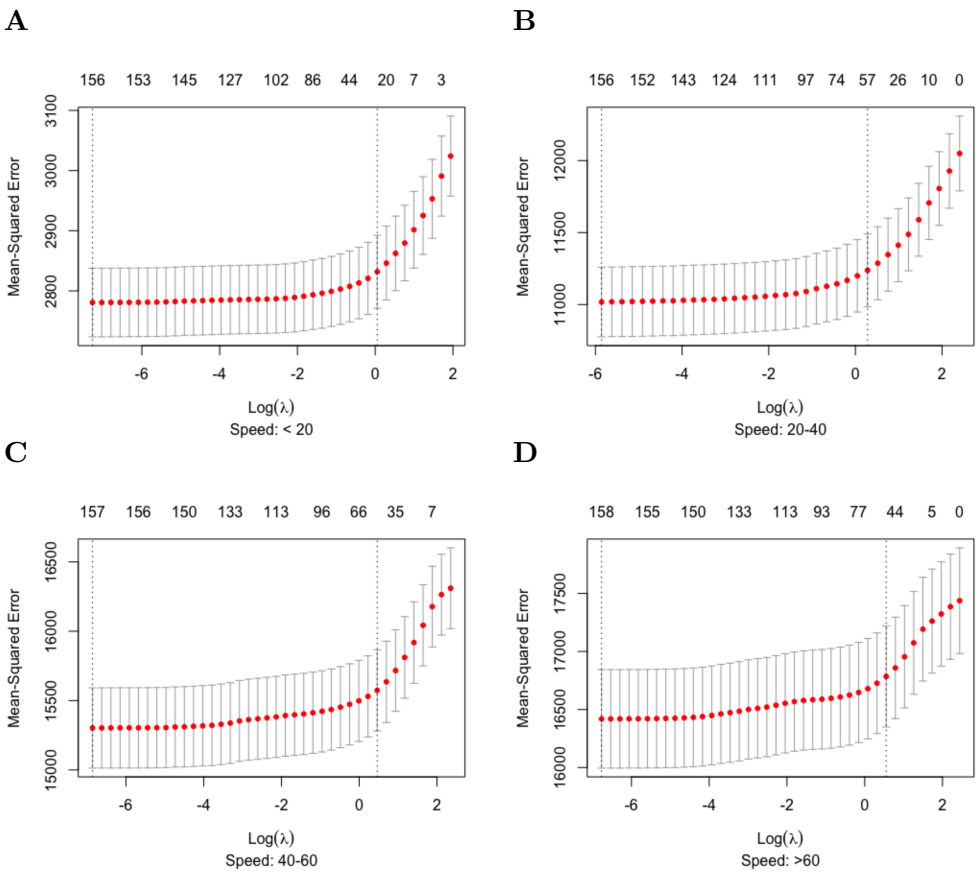


Figure SI9: **Mean-Squared Error from LASSO varying** λ Table SI1 shows estimates of the minimum and 1se values of λ from the 40 LASSO regressions across four speed bins.

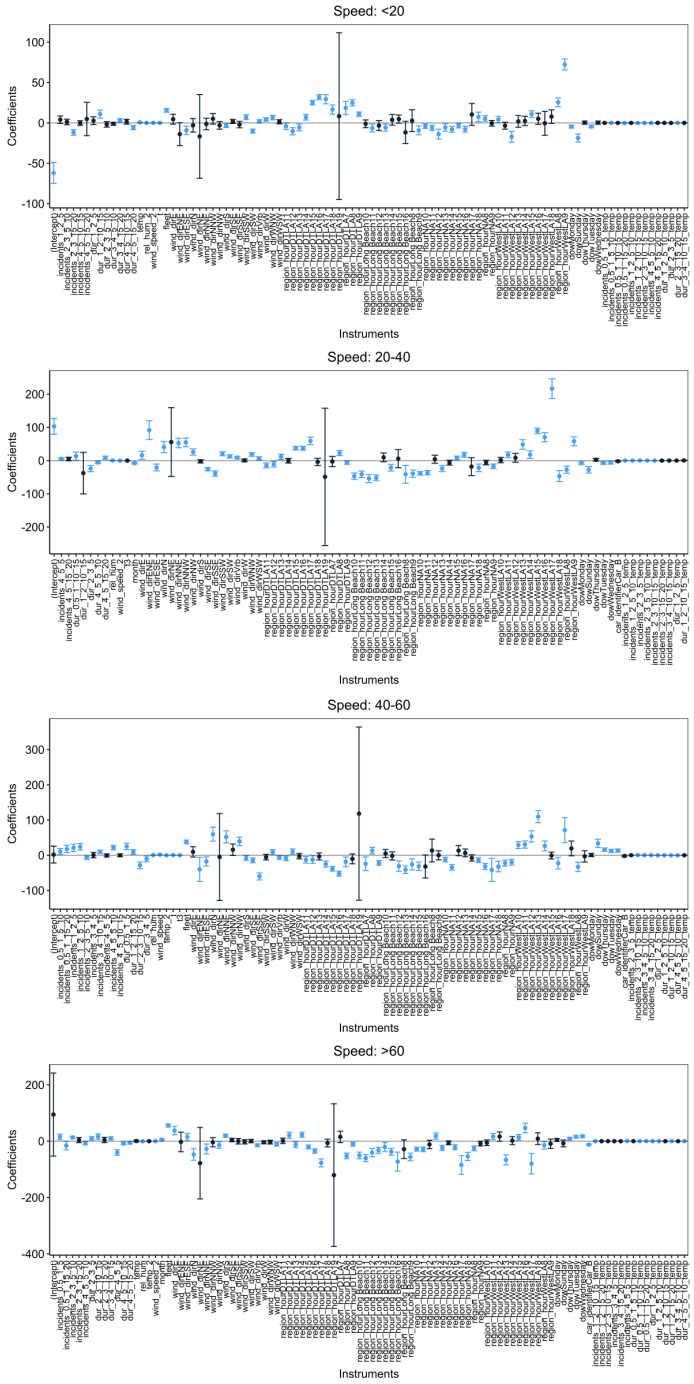


Figure SI10: **First-Stage Estimates from LASSO IV Regression** Each plot displays coefficient estimates from the corresponding instrument set excluding fixed effects for the indicated speed bin interacted with Cars/mile. Instruments include duration and count of incidents within specified distance and time windows as well as their interaction with whether controls. Also included are controls for the relative extent of heavy duty vehicles on that road at that time of day based on vehicle census records. Dark blue estimates are not statistically different from zero at a 95% confidence interval while light blue estimates are.

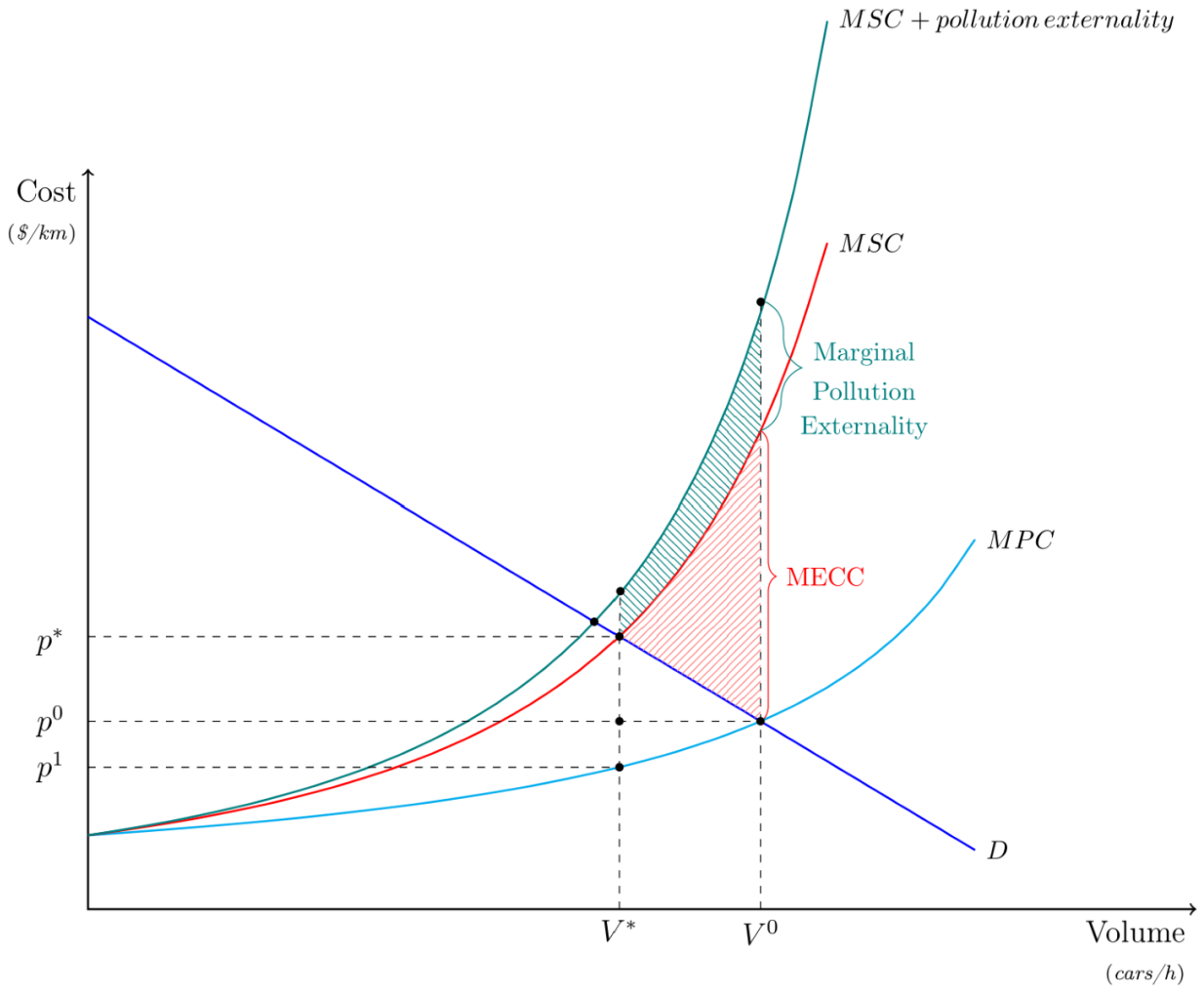


Figure SI11: **Market Diagram of Equilibrium and Optimal Vehicle Flows** The figure shows two equilibrium outcomes for roadway congestion corresponding to congestion levels (in terms of vehicle volume) of V^0, the unregulated equilibrium outcome, and V^*, the socially optimal outcome (without accounting for other externalities such as pollution). Cost refers to the monetized cost of travel including travel time and amortized fixed costs of travel. D is travel demand, which reflects the marginal private benefit of driving. MPC is Marginal Private Cost, which reflects the individual private costs that drivers respond to when making travel demand decisions. It is also equal to the Average Social Cost. MSC is Marginal Social Cost, which includes the contribution of a single driver to congestion levels. Both curves are increasing and convex reflecting the fact that the contribution of an additional driver increases substantially the higher the level of congestion. $MECC$ is the Marginal External Cost of Congestion, which is vertical distance between MSC and MPC. The red dashed area is the total wedge between the social optimum and unregulated equilibrium in terms of unpriced negative externalities from congestion. The Marginal Pollution Externality is the vertical distance between the red and green curves. The green dashed area is the sum of this area between the social optimum and the unregulated equilibrium: a congestion interaction effect. The optimal static congestion charge, ignoring pollution, is the vertical distance between MPC and MSC at the social optimum, corresponding to the vertical distance between p^1 and p^*.

Speed: < 20 MPH				
		MSE	SE	Nonzero
min	0.0007	2,781	57.10	156
1se	1.0526	2,832	60.63	28
	0.2552	2,796	57.94	74

Speed: 20 - 40 MPH				
		MSE	SE	Nonzero
min	0.0029	11,018	242	156
1se	1.3246	11,238	253	57
	0.6522	11,144	248	74

Speed: 40-60 MPH				
		MSE	SE	Nonzero
min	0.0011	15,303	289	157
1se	1.5874	15,574	292	53
	0.4874	15,436	292	86

Speed: > 60 MPH				
		MSE	SE	Nonzero
min	0.0011	16,419	424	158
1se	1.7374	16,785	435	56
	0.4212	16,596	429	88

Table SI1: Selection of LASSO Regression Hyperparameter λ This table is a summary of the results plotted in Figure SI8, where min is the value of lambda at the minimum of mean square error (MSE), while 1se is one standard deviation from the minimum.