EFFECTS OF ON-DEMAND RIDESOURCING ON VEHICLE OWNERSHIP, TRAVEL, ENERGY, AND ENVIRONMENTAL OUTCOMES IN THE UNITED STATES

Jacob Ward
Research Assistant, Department of Engineering and Public Policy and
Department of Mechanical Engineering
Carnegie Mellon University

Jeremy J. Michalek*
Professor, Department of Engineering and Public Policy and
Department of Mechanical Engineering
Carnegie Mellon University
Scaife Hall 324, 5000 Forbes Ave., Pittsburgh, PA 15213
jmichalek@cmu.edu
phone: (412) 268-3765; fax: (412) 268-3757
* Corresponding author

Inês L. Azevedo
Professor, Department of Engineering and Public Policy
Carnegie Mellon University

Constantine Samaras
Associate Professor, Department of Civil and Environmental Engineering and
Department of Engineering and Public Policy
Carnegie Mellon University

Pedro Ferreira
Associate Professor, Heinz College and
Department of Engineering and Public Policy
Carnegie Mellon University

Nicholas Muller
Associate Professor, Department of Engineering and Public Policy and
Tepper School of Business
Carnegie Mellon University

ABSTRACT

We estimate effects of on-demand ride-hailing services Uber and Lyft on vehicle ownership, travel, energy, and environmental outcomes using a set of difference-in-difference propensity score-weighted regression models that exploit staggered market entry across the U.S. from 2010 to 2017. Specifically, we use state-level data to estimate effects of Uber market entry on vehicle registrations, gasoline consumption, travel distances, and emissions, and we use zipcode-level data to estimate effects on vehicle registration patterns, air quality, and transit use in urban areas. We find evidence that TNC entry causes a 3% decline in per-capita vehicle registrations when averaged across states but a 0.7% increase when averaged across urban areas. This difference is due, in part, to heterogeneity in the effects of TNC entry on different cities: TNC entry appears to increase ownership in large dense cities and small family-focused cities with low per-capita vehicle registrations, while the effect on other groups of cities is not statistically significant in our clustering results. Our results regarding transit ridership, travel distances, gasoline consumption, and several air pollutants are not conclusive, but we also find evidence of a negative association between TNC entry and EPA-estimated emissions of highway vehicle volatile organic compounds (VOCs).

Keywords: transportation network company, ride-hailing, vehicle ownership, energy, VMT, emissions, air quality, transit

1. INTRODUCTION

Transportation now contributes more carbon dioxide emissions than any other U.S. economic sector[1], and new personal transportation options are rapidly changing transportation. Transportation network companies (TNCs), like Uber and Lyft, now provide on-demand mobility services that complement and compete with personal vehicle ownership and transit use, changing urban travel patterns and affecting energy and environmental implications of transportation. By 2017, Uber had entered 46% of U.S. urban areas (Figure 1). TNCs made more than 170,000 vehicle trips in San Francisco (15% of all intra-San Francisco vehicle trips) on an average weekday in 2016[2] and more than 90,000 rides in Seattle (more than total average weekday ridership on Seattle's light rail) on an average weekday in 2018[3]. Prior studies have examined effects of this rise in TNC use on outcomes as varied as traffic congestion, drunk driving, local entrepreneurship, ambulance use, and vehicular deaths, but the net effect of these services on vehicle ownership, travel, energy, and the environment is either unexplored or still debated in the literature.

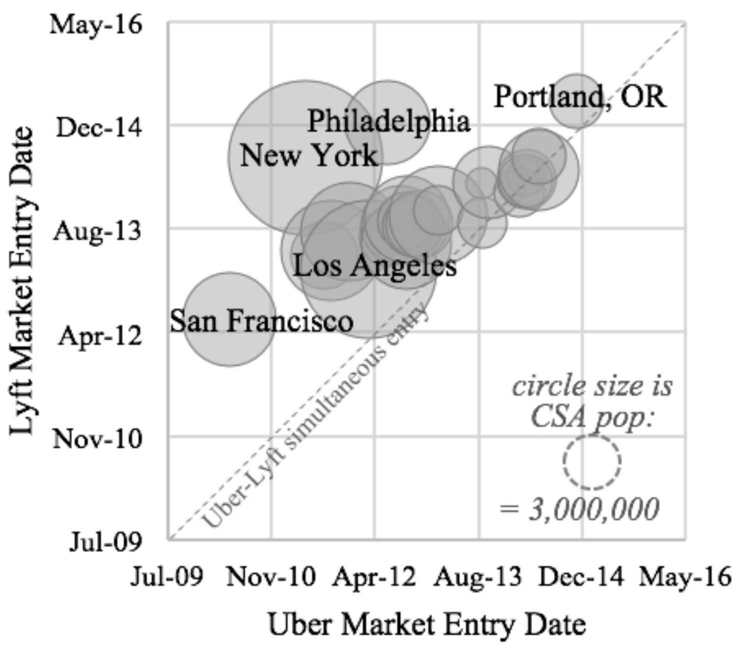

Figure 1 Comparison of Uber and Lyft market launch dates by combined statistical area (CSA). Some CSA labels are omitted for readability; data points, in chronological order, are: San Francisco, New York City, Seattle, Chicago, Washington (DC), Los Angeles, Philadelphia, San Diego, Atlanta, Boston, Dallas-Fort Worth, Denver, Minneapolis-St. Paul, Phoenix, Baltimore, Sacramento, Rhode Island (where Uber entered the entire state at once), Charlotte, Houston, Pittsburgh, Louisville, Cleveland, Tampa Bay, Miami, Orlando, St. Louis, and Portland (OR).

On-demand mobility is part of a larger ongoing transformation of shared mobility—a broader term used to describe a set of transportation modes where passengers travel using vehicles owned by another party on an as-needed basis. Transportation modes such as carpooling, bike-sharing, and shuttle services have long fit into this category. Historically, trends in vehicle travel and transportation-related air pollutant emissions have been relatively predictable: for example, since 2005 vehicle registrations have increased by approximately 1% annually (except for declines during the recession from 2008–2011) and emissions of volatile organic compounds have declined 5% annually (EPA's Tier 2 emissions standards were phased-in from 2004–2009). More recently, car-sharing services have expanded customers' mobility options, introducing such options as renting a fleet-owned vehicle that is regularly available to other customers for either round-trip (e.g., Zipcar) or point-to-point (e.g., car2go) journeys. Furthermore, the growth and capabilities of smartphones enabled TNCs like Uber and Lyft to introduce on-demand mobility. Uber and Lyft launched in March 2010 and June 2012, respectively, in their first market: San Francisco, California. In 2018, Uber announced the completion of 10 billion total trips[4] and Lyft announced one billion total trips[5]. These services opened the door for dynamic ridesharing, where algorithms efficiently route on-demand mobility services to serve several customers with different destinations in the same physical vehicle.

Despite rapid TNC growth in recent years, there is limited knowledge about how they influence vehicle ownership patterns, energy consumption, travel patterns, and environmental outcomes. TNCs may reduce an individual's reliance on a personal vehicle, ultimately resulting in fewer vehicle registrations, or stimulate new vehicle purchases by TNC drivers, increasing registrations. TNCs may increase VMT by requiring vehicles to travel between passenger trips ("deadheading") and by increasing travel demand or shifting demand from mass transit to light-duty vehicles. But they may also reduce vehicle miles traveled (VMT) through ride pooling, by providing a "first/last-mile" solution that encourages partial use of public transportation, or by providing travelers with the option to pay per trip as an alternative to making a "lumpy" investment in a personal vehicle and observing low marginal costs of additional travel. TNCs might increase or decrease energy consumption and emissions by changing VMT, by shifting VMT to vehicles with different efficiency and emissions rates, and by changing the portion of VMT traveled at hot operating temperature, when vehicles are more efficient and have lower emission rates.

1.1. Prior Literature

Peer-reviewed studies of the effects of TNCs on vehicle ownership, travel, energy, and environmental outcomes are limited: Rayle et al.[6] found that while find 33% of surveyed TNC users in San Francisco would have traveled via bus or rail if the TNC service were not available, "ridesourcing probably did not influence car ownership behavior". Hall et al.[7] use a difference-in-difference econometric model in 147 U.S. metropolitan areas and conclude that, while transit ridership does not change immediately after Uber entry, transit ridership increases by five percent two years after Uber entry, on average, and that this heterogeneous effect is larger in big cities with small transit agencies. They also find that Uber entry decreases commute times for transit users while increasing vehicular congestion. There are no peer-reviewed journal publications of TNC effects on energy or emissions, to our knowledge.

Some working studies and internal reports have suggested that TNCs have affected vehicle ownership, use, and emissions, but the estimated effects vary. Both Hampshire et al (2017)[8] and Clewlow and Mishra (2018)[9] use survey methods to infer a *reduction* in overall vehicle ownership attributable to Uber and Lyft: Hampshire et al. surveyed former users of Uber after Uber left Austin, TX in 2016 and found a 9% increase in reported vehicle ownership among those former Uber users, and Clewlow and Mishra report that 9% of survey respondents who use ride-hailing across a group of 7 U.S. metropolitan areas disposed of one or more household vehicles. In contrast, Schaller (2018)[10] and Gong et al (2017)[11] find that Uber is associated with an *increase* in vehicle ownership: Schaller observes that while TNCs were operating in the nine largest U.S. metropolitan areas from 2012–2016, growth in vehicle ownership outpaced that of population, and Gong et al. apply a difference-in-difference regression model in China and estimate an 8% increase in new vehicle registrations associated with Uber entry.

Vehicular travel effect estimates from working studies and internal reports have also varied (the two peer-reviewed studies mentioned earlier found different and even heterogeneous effects). Li et al (2016)[12] find that TNCs are associated with *reductions* in some travel metrics: they use a difference-in-difference regression to estimate a 1.2% decline in overall congestion and associated travel times and fuel consumption. But other studies suggest an *increase*: Clewlow and Mishra (2018) suggest, based on survey responses from ride-hailing users across a group of 7 U.S. metropolitan areas, that 49% to 61% of ride-hailing trips are associated with an increase in VMT; Hampshire et al. (2017) find a 23% reduction in the likelihood to take a trip among former Uber users surveyed in Austin, TX that transitioned to a personal vehicle after Uber and Lyft left; and Schaller (2018) finds, based on a comparison of eight surveys from other working studies, that 60% of ride-hailing trips would have otherwise happened via transit, walking, or biking (or not have happened at all) in a group of nine U.S. metropolitan areas.

TNC services can have effects not only on the number of vehicles registered, but also on how those vehicles are used. Recent analysis suggests that less than 60% of miles traveled by a TNC vehicle are productive miles spent moving a passenger from an origin to a destination—the remaining 40% of TNC vehicle empty-mile travel is spent cruising in search of the next fare, driving to passenger pick-up, or driving after passenger drop-off[14]. Additionally, the travel demand that is shifted to vehicles from other modes (i.e., from walking, biking, and transit) due to the convenience of on-demand ridesharing services was estimated to be as high as 85% in Denver, CO[14], though Hall (2018) concludes that Uber is more of a complement to transit. Despite potential increases in the number of trips and the total number of miles travelled to complete each trip, chaining trips in the same set of vehicles may reduce criteria air pollutant emissions by reducing the number of cold starts[15].

In summary, literature of the effects of TNCs on vehicle ownership, travel, energy, and environmental outcomes is inconclusive, and there are few peer-reviewed studies. We contribute to this literature by exploiting the staggered entry timing of Uber and Lyft across U.S. cities seeking to identify causal relationships between TNC entry and our outcomes of interest.

2. METHODS

We use difference-in-difference (DiD) models to estimate effects of the intervention (TNC entry) by comparing the trends of treated and untreated groups before and after the intervention occurs. DiD methods have been used previously to evaluate the effect of TNCs on other outcomes, including traffic congestion[12], vehicle-related homicides[17], entrepreneurial activity[18], and new vehicle ownership in China[11].

2.1 Difference-in-Difference Model

Our regression model is informed by models used in prior literature for our outcomes of interest. Regression analysis is conducted using inverse probability of treatment weighting (described below) and the following baseline specification:

$$y_{gt} = \boldsymbol{\beta}^\top \mathbf{x}_{gt} + \boldsymbol{\alpha}^\top \mathbf{z}_{gt} + \gamma_g + \delta_t + \varepsilon_{gt} \tag{1}$$

where y_{gt} is the dependent variable of interest for group g and year t. At the state level g indexes U.S. states, and we examine four types of dependent variables: 1) vehicle registrations per capita; 2) VMT per capita; 3) gasoline use per capita, or 4) per capita passenger vehicle emissions estimates for each of the following: CO, NH_3, NO_x, PM_{10}, $PM_{2.5}$, SO_2, and VOCs. At the urban area level g indexes urban areas, as defined by the U.S. Census, and y_{gt} is 1) vehicle registrations per capita, 2) the percentage of registered vehicles that are electric, 3) concentrations of each of several vehicle-related air pollutants (carbon monoxide, oxides of nitrogen, benzene, toluene, and xylene), or 4) transit ridership. $\mathbf{x}_{gt}$ is the vector of treatment effects (in our base model the vector has length 1 and represents the presence or absence of Uber in group g in year t) with coefficient vector $\boldsymbol{\beta}$. $\mathbf{z}_{st}$ is a vector of controls[i], with corresponding coefficients $\boldsymbol{\alpha}$. γ_g and δ_g are fixed-effects dummies for group g and year t, respectively, and ε_{gt} is unobserved error.

The estimates of a difference-in-difference model provide unbiased causal effect estimates when its assumptions are satisfied, including that the intervention is exogeneous, trends are parallel, and there are no spillover effects. We discuss each in turn.

- **Exogeneous Intervention:** A potential concern arises if treatment (TNC entry) is conflated with other attributes of the treated and untreated groups (e.g.: if densely populated cities are treated more frequently than less densely populated cities). To control for systematic differences between treated and untreated groups, we apply both control variables and inverse probability of treatment weights in a weighted least-squares model. This model compares post-treatment trends in treated units with weighted trends in non-treated units, probabilistically weighted to resemble the treated states along attribute dimensions that are correlated with treatment[iii]. After estimating the probability of treatment, we compare measures of balance to confirm that the propensity score weights succeed in matching the control states' weighted pretreatment characteristics to those of the unweighted treatment states (that is, that the weighted control and unweighted treatment group are balanced).
  An additional concern arises if the decision to treat a location is influenced by changes in the dependent variable (e.g.: if changes in vehicle registrations in a region encourage Uber to enter that region). To address this possibility, we perform event studies to identify whether in any case the change in dependent variable preceded treatment. Through informal discussions with Uber we also learned that early decisions to enter U.S. cities used information including Google searches for "Uber" and "Lyft" to help determine where to enter first. It is plausible that changes in some of our dependent variables (e.g.: registrations) may be correlated with Google searches for "Uber" and "Lyft", which could bias our estimates. Publicly available Google Trends data are too imprecise during this time period to be useful in our analysis, and we are still seeking usable Google search history trends to control for this possibility.
- **Parallel Trends:** To examine the parallel trends assumption, we plot outcomes for individual states and groups of states to compare trends prior to intervention. We also examine a model variant that includes different linear time trends for each group. Finally,

[iii] For our state-level analysis, these are state population, income, gasoline price, emissions standards, and largest city population, density, and GDP. For our urban area and zipcode level analysis, these are population, portion of population over age 16 and over 65, population density, unemployment rate, income, and percent of population commuting by transit.

we use an event study to check whether or not we find evidence of an effect following treatment without requiring the parallel trends assumption.

- **Spillover:** The model assumes that treating one location will not affect other locations. It is plausible that experience with ridehailing services during travel to other cities could affect vehicle ownership or travel behavior in a home city that does not have access to ridehailing services, but we assume such effects are negligible.

2.2 Propensity Score

We estimate propensity scores using gradient boosting[20], which previous studies have shown as superior to simple logistic regression models for propensity score estimation[21], to approximate the logistic model:

$$\log\left(\frac{p_{gt}(\mathbf{z}_{gt})}{1-p_{gt}(\mathbf{z}_{gt})}\right) = \sum_m f_m(\mathbf{z}_{gt}) + \epsilon_{gt} \tag{2}$$

where p_{gt} is the probability of treatment for group g and year t; $\mathbf{z}_{st}$ is a vector of covariates for group g and year t,[i] and ϵ_{gt} is unobserved error. We estimate the additive function f_m using gradient boosting, given the treatment and covariate data, and compute estimated probability of treatment $\hat{p}_{gt}$ for each state and year. The resulting estimates for probability of treatment are then used in a weighted regression for Eq(1).

For the particular case where y_{gt} is a measure of air pollution concentrations at nearby air quality monitoring sites, we are hesitant to apply IPTW because air quality monitoring sensors are not distributed randomly – rather, locations are "treated" with sensors for specific reasons, such as to comply with regulation or monitor an industrial facility. Because of this, we abandon the attempt to estimate causal effects for this case and examine only associations identified in an unweighted OLS model. We discuss implications in the results section. We are continuing to investigate methods to account for non-random sensor placement in future work.

2.3 Robustness

We apply several statistical tests to check model assumptions and test for robustness (see SI Sections 4 and 5)[v]. Model assumptions are informed by generalized additive models (GAMs) for independent variable functional form, and final model fit is checked using visual inspection of residual errors to confirm no structural error. Additionally, for each model, we subject our results to four robustness checks:

(1) We introduce linear time-varying fixed effects into the regression model (i.e., an additional term in equation (1) above) to allow for different trends in different groups;
(2) We conduct randomized treatment tests to ensure that the effects we estimate are unique to the particular observed pattern of treatments, rather than a result of the structure of the model. Estimated effects are considered robust if they fall in the tails (>95%) of the distribution of randomized treatment-estimated effects;
(3) We conduct leave-one-out tests to ensure that our estimates do not hinge on the data of any one state. Estimated effects are considered robust if they remain significant when systematically leaving each state out;

v The supplemental information document is available from the authors upon request

(4) We conduct leave-multiple-out tests to ensure that our estimates do not hinge on outliers. Estimated effects are considered robust if they do not change in magnitude (i.e., 95% confidence intervals still overlap) or significance level;

Additionally, we perform several sensitivity analyses appropriate for each case, including:
(1) alternative dependent variable normalization (i.e., per licensed driver or per urban population),
(2) alternative period of analysis (2009–2015),
(3) alternative treatment encoding (annualizing between June and July instead of December and January),
(4) additional control variables (indicators for Uber leasing/incentive programs, Lyft market entry, and transit), and
(5) alternative specifications with lagged treatment (by one and two years).

3. DATA

We describe and identify data sources for dependent variables, treatment, and control variables in turn:

3.1 Dependent Variables:

State-Level Analysis

- *Vehicle registrations (measured):* We use vehicle registration data for each state and for each year for light-duty passenger vehicles from Ward's Automotive[22]. Ward's data are based on data published in U.S. DOT's State Statistical Abstracts and Highway Statistics Series,[23,24] which is the set of official vehicle registration data published by state DOTs.
- *Gasoline consumption (measured):* DOT's State Statistical Abstracts and Highway Statistics Series also report Federal Highway Administration estimates of annual private and commercial vehicle state level on-highway motor fuel based on reports of aggregate motor fuel sales from state motor fuel tax agencies.
- *VMT (estimated):* VMT data comes from DOT's State Statistical Abstracts, which are tracked and reported annually as a function of figures reported by state agencies. State agencies estimate aggregate VMT based on vehicle count data measured on representative roadways and distributions of roadway type within the state (while DOT issues a Traffic Monitoring Guide, individual state methods may differ). VMT (table VM-2) has been published in DOT's State Statistical Abstract series since 2008; earlier data are available in DOT's Highway Statistics Series. Interpretation of statistical inference based on these VMT data is constrained by the representativeness of the underlying VMT estimation (rather than direct measurement) methods.
- *Emissions (estimated):* State-level emissions data are published annually in the EPA's State Average Emissions Trend report, which is informed by EPA's National Emission Inventory, which, in turn, relies on EPA's Motor Vehicle Emission Simulator (MOVES) model. The MOVES model estimates vehicular emissions based on vehicle population and fleet characteristics, vehicle speed distributions, and relative hour- and day-type VMT distributions at the county level and aggregated. Emissions attributable to highway vehicles are estimated by the EPA annually[25]: 2008, 2011, and 2014 estimates were

developed in conjunction with the National Emissions Inventory for those years; 2005, 2007, 2009 and 2010 estimates were updated using additional MOVES modeling; and 2006, 2012, and 2013 were interpolated. EIA estimates an annual series of State Carbon Dioxide Emissions based on energy consumption data contained in the State Energy Data System (SEDS). Transportation sector estimates are published without highway or light-duty vehicle detail after an approximately 2-year lag[26]. Interpretation of statistical inference based on these emissions data is limited to factors considered as part of emissions estimation modeling (rather than direct measurement).

We divide each of the four quantities above by state population each year to compute per-capita values. Annual state-level population estimates are from DOT's State Statistical Abstract and Highway Statistics series and, as such, they align with VMT data and are related to Ward's Automotive vehicle registration data (the ultimate source for which is also these DOT publications). DOT population reports match U.S. Census statistics in census years and are no more than 0.6% different than Census Bureau's annual estimates of the resident population in intercensal years[27], which the Census calculates assuming geometric interpolation with some exceptions[28].

Urban-Area Analysis

- *Vehicle registrations (measured):* IHS Markit (formerly Polk) collects and sells vehicle registration information from U.S. State agencies responsible for registration data[29]. We rely on a version of the dataset that reports, by ZIP Code, vehicle make, model, and engine size for the approximately 240 million light-duty vehicles registered in the U.S.
- *Air pollutant concentration (measured):* U.S. EPA generates data tables for the measurements from the monitors at 20,000 sites around the U.S. that comprise its Air Quality System (AQS)[30]. We extract annual summary measures of several vehicle-relevant pollutants: carbon monoxide, oxides of nitrogen, several species of volatile organic compounds (benzene, toluene, and xylene), and particulate matter.
- *Transit ridership (measured):* U.S. DOT's Federal Transit Administration (FTA) reports annual summary statistics on more than 660 transit providers receiving federal funding in the National Transit Database (NTD)[31]. We focus on transit providers that consistently report data for all years of this analysis and aggregate individual transit agencies by urban area, per classification in the database.

3.2 Treatment Variables:

- *Uber and Lyft entry dates (state, urban area, and ZIP Code level analyses):* We adopt data from previous sources that aggregated and published a time-series of Uber market entry dates. A 2014 Forbes article first aggregated Uber launch dates from 2010–2014[32] by service area, as originally announced on Uber's official blog (on a post no longer available) and/or in local media from each new service area. Forbes continued to update that dataset to reflect additional Uber markets launched through December 2015. Those dates are cross-referenced against Uber market launch date data that were independently gathered and published in two later studies[16,32,33]. Burtch et al. include a table of market launch dates for UberX—Uber's lower-cost, on-demand service provided in the driver's personal vehicle, which the authors compiled directly from the Uber Blog[18]. Lyft market launch dates were requested from and provided by Lyft[34]. A comparison of Uber and Lyft market launch date time-series is depicted by combined statistical area in Figure 1.

Because Lyft market entry years are the same or later than Uber market entry years in all cases, we use Uber entry dates in our analysis to represent on-demand mobility availability in the state.

3.3 Control Variables

State-Level Analysis:

- *State-level control variables:* Our control variables include: (i) population, reported annually in DOT's State Statistical Abstract and Highway Statistics series, (ii) percentage of a state's population that is urbanized[35], (iii) state average real personal income, reported annually by the Bureau of Economic Analysis[36]; (iv) state average gasoline price data, reported annually by the U.S. Energy Information Administration (EIA), and (v) an indicator for whether each state has adopted California's more stringent vehicle emissions control requirements, pursuant to Section 177 of the Clean Air Act[37]. Additionally, recognizing that TNC market entry and use is primarily a city phenomenon, additional control variables are included for the largest city within each state, including: (vi) population[38], (vii) population density[38], and (viii) GDP[39].

Urban-Area Analysis:

- *Urban area- and ZIP Code-level control variables:* Control variables at the urban area and ZIP Code level are 5-year American Community Survey (ACS) estimates reported by the U.S. Census and include: (i) population, (ii) portion of population over age 16 and over 65, (iii) population density, (iv) unemployment rate, (v) income, (vi) and percent of population commuting by transit.

While these control variables are intended to help reduce bias, the possibility of omitted variable bias cannot be overlooked. Sensitivity analyses were conducted using several additional potentially relevant independent variables (number of licensed drivers, Lyft market entry, transit ridership, and Uber/Lyft leasing incentive programs), as well as two variations on Uber treatment encoding; none greatly affected the magnitude or the significance of effects reported as significant and robust.

Variable encoding and summary statistics for each data source above are shown in Table 1. On average, population steadily increases, criteria pollutant emissions steadily decrease, and vehicle registrations and income generally increase, except for a dip in 2009–2010 corresponding to the Great Recession. Gasoline price is volatile and non-monotonic.

Table 1 Variable encoding descriptions and associated summary statistics (U.S. totals, except where averages are shown, as noted) for 2005, 2010, and 2015. Monetary values are reported in current dollars (as indicated).

Parameter	Unit	Description	2005	2010	2015
Population	million persons	Coded as log state population	296	309	321
Light-Duty Vehicles	million vehicles	Coded as log light-duty vehicles per capita	234	232	241
Gasoline Use	billion gallons	Gasoline taxed by states as used by non-public, non-exempt vehicles	133	131	130
VMT	trillion miles	Coded as log vehicle miles traveled per capita	2.99	2.97	3.10
CO	million tons	Coded as per-capita highway carbon monoxide emissions	42.4	28.3	19.7
NH3	million tons	Coded as per-capita highway ammonia emissions	0.14	0.12	0.10
NOx	million tons	Coded as per-capita highway nitrous oxides emissions	8.30	5.70	4.12
PM10	million tons	Coded as per-capita highway particulate matter emissions	0.38	0.28	0.30
PM2.5	million tons	Coded as per-capita highway particulate matter emissions	0.31	0.20	0.15
SO2	million tons	Coded as per-capita highway sulfur dioxide emissions	0.17	0.04	0.02
VOC	million tons	Coded as per-capita highway carbon monoxide emissions	3.41	2.77	1.97
Income	trillion $ (current $)	Coded in regression as real personal income per capita	10.6	12.5	15.5
s177	binary	A state's Section 177 status (whether it has adopted California's more stringer mobile-source emissions regulations)	5	11	13
Katrina	binary	Indicator for potential vehicle hurricane damage (2005 only)	1	0	0
Sandy	binary	Indicator for potential vehicle storm damage (2012 only)	0	0	0
Clunkers	Number of vehicles scrapped	Number of participants in "Cash for Clunkers" vehicle scrappage program (2009 only)	0	0	0
Treat	% states	Uber indicator, binary	0%	2%	90%
(For parameters below, values shown are averages across states)					
Gas Price	$/gal (current $)	Average gasoline price	2.08	2.63	2.34
Pop_u	% pop, state avg.	% of state population that is considered Ubran by the Census (coded relativel to the average % urbanization for 2005-2015, which is 74%)	73%	74%	75%
Citypop	thousand persons	Population of center city in a state's largest metropolitan statistical area	652	703	714
Citydensity	persons per square mile	Population density of center city in a state's largest metropolitan statistical area	4120	4483	4539
CityGDP	billion $ (current $)	GDP of state's largest metropolitan statistical area	127	140	177
Treatpop	% pop, state avg.	Uber indicator, weighed by % of state population with Uber access	0%	0%	21%

4. RESULTS

Table 2 summarizes results for the effect of TNC entry on several of our dependent variables at the state and urban area levels. The state model suggests that, on average, Uber market entry (in any portion of a state) decreases per-capita vehicle registrations by 3.1% (95% confidence interval: 0.7% to 5.5%) over the period examined (relative to per-capita registration had the TNC not been introduced). Conversely, the urban-area model suggests that, on average, Uber market entry increases per-capita vehicle registrations by 0.7% (95% confidence interval: 0.1% to 1.3%) over the period examined (relative to per-capita registrations absent TNC entry). We interpret these results in the context of heterogeneous effects across urban areas later. The state model indicates a decline of 4.2% (95% confidence interval: 1.0% to 7.4%) in EPA-estimated vehicular VOC emissions after Uber enters any portion of a state. All of the statistically significant findings here are robust when subjected to our robustness tests (details reported in SI Sections 4 and 5).

Table 2 Weighted least-squares regression model treatment effect estimates for per-capita vehicle registrations, EV registration percentage, per-capita gasoline use, per-capita VMT, and per-capita transit trips. Coefficients estimated for control variables (state population, urban population percentage, income, gasoline price, emissions standards, and largest city population, density, and GDP, as well as indicators for Hurricane Katrina, Cash for Clunkers, and Superstorm Sandy and fixed effects for state and time at the state level and population, portion of population over age 16 and over 65, population density, unemployment rate, income, percent of population commuting by transit at the urban area level) are excluded from the table for brevity; weights are calculated as described in equation (2).

	Dependent variable:					
	Veh. Reg.	VOCs	EV Reg.	Gas. Use	VMT	Transit Trips
State-Level Model						
Treatment Effect	−0.031**	−0.042**		0.001	−0.003	
	(0.012)	(0.016)		(0.004)	(0.003)	
Observations	550	550		550	550	
Deg. Freedom	474	474		474	474	
Adjusted R-Sq.	0.844	0.962		0.840	0.834	
Urban Area-Level Model						
Treatment Effect	0.007**		−0.0001			−0.001
	(0.003)		(0.0002)			(0.012)
Observations	3,402		3,402			1,848
Deg. Freedom	2,903		2,903			1,570
Adjusted R-Sq.	0.913		0.705			0.998

Note: *p<0.1; **p<0.05; ***p<0.01

Table 2 also shows the estimated effects of TNC entry on EV market penetration, gasoline consumption, VMT, and transit use, none of which are statistically significant. Not shown are estimated effects on EPA-estimated per-capita emissions of carbon monoxide, oxides of nitrogen, and particulate matter, as well as GHGs at the state level, as none were found to be significant. We also examined the effect on concentrations of CO, NO_X, PM_{10}, $PM_{2.5}$, and several VOCs at nearby air quality monitors using an unweighted regression and found mixed results.

Additional research is needed to refine the air quality results to address the non-randomness of air quality monitor locations, so we do not present any preliminary results for air quality here.

4.1 Robustness

We subject our results to a variety of checks including a set of robustness checks, sensitivity analysis, event studies, and unweighted regression. We discuss each in turn.

Robustness Checks and Sensitivity Analysis:

The battery of robustness checks and sensitivity analyses that we apply support our findings. Both the estimated vehicle registration and VOC emission effects at the state level are robust (or "near-robust", as slightly crossing the threshold for the level of significance of the vehicle registration or VOC emissions effect estimates is sensitive to whether Ohio or Indiana, respectively, are included in the sample) to randomized treatment, leave-one-out, alternative treatment encodings, and leave-multiple-out checks (all described previously in the Methods section), as is summarized in the SI. Furthermore, similar state-level effects are estimated even when regressions are specified to test potential sensitivity to alternative dependent variable normalization (i.e., per licensed driver or per urban population), timeframe (2009–2015), treatment encoding (annualizing between June and July instead of December and January) and additional control variables (indicators for Uber leasing/incentive programs, Lyft market entry, and transit); and, finally, a set of results examining the effect of lagged treatment (by one and two years) (details are included in the SI). Comparable robustness checks and sensitivity analyses at the urban-area level are still in process.

Table 3 Summary of robustness checks results at the state level.

	Coefficient	RT	LOO	Enc	LMO
Vehicle Registrations	-3.1% **	●	◒	●	●
VOC	-4.2% **	●	◒	●	●

*Notes: **RT** - Randomized Treatment; **LOO** - Leave-one-out; **Enc** - Uber treatment alternative encodings; **LMO** - Leave-multiple-out; ● robust, ◒ "near-robust", ○ not robust*

Event Study:

We conduct event studies to test whether or not effects estimated in our difference-in-difference model can be observed without making the assumptions underlying the difference-in-difference model. Figure 2 shows event studies at the state and urban area level, where time for each state or urban area is normalized relative to the year that Uber entered (time zero). At the state level (left) there is no statistically significant change in registrations in years prior to Uber entry, but we find a statistically significant decrease in registrations after Uber entry. This result provides additional evidence of the effect identified in the difference-in-difference model without assuming parallel trends. At the urban area level (right) there is no statistically significant change in registrations in years prior to Uber entry, but we find a statistically significant increase in registrations several years after Uber entry. This result provides additional evidence of the effect identified in the difference-in-difference model without assuming parallel trends; however, the continuous shape of the estimates offer weaker support than if a step change had been found. The event studies do not control for other time-varying factors and, as such, serve only as an additional look at the data without the parallel trends assumption. These results are consistent with our difference-in-difference estimates at the state and urban area level, respectively.

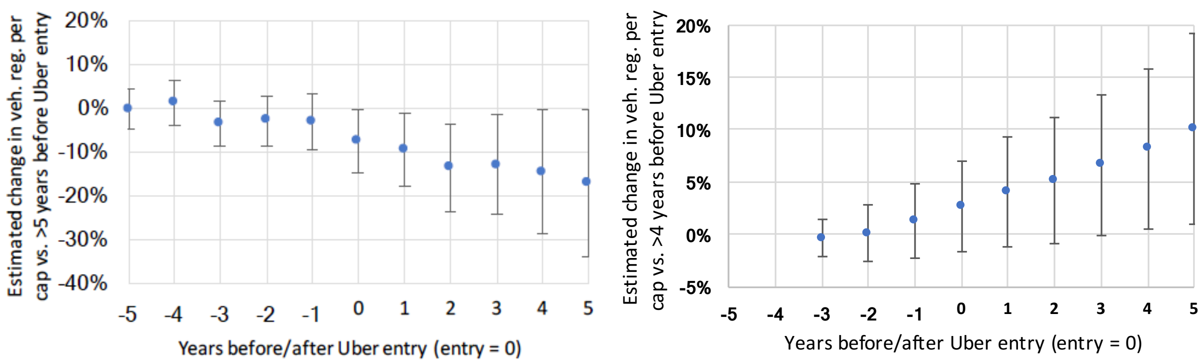

Figure 2 Event study results at the state level (left) and urban-area level (right) showing both no evidence of significant pre-treatment changes in per-capita vehicle registrations and significant evidence of changes at some point in time after treatment.

Unweighted OLS Results:

In Table 4, we compare the IPTW results from Table 2 with the treatment effect estimated using ordinary least squares (OLS) with same model specification, i.e., equation (1), as well as the effect estimated after adding time-varying group fixed effects to equation (1).

$$y_{gt} = \boldsymbol{\beta}^\top \mathbf{x}_{gt} + \boldsymbol{\alpha}^\top \mathbf{z}_{gt} + \gamma_g + \delta_t + \zeta_g t + \varepsilon_{gt} \qquad (3)$$

The first comparison against an OLS model is meant to demonstrate whether finding a significant effect is dependent on the weights used in the IPTW model, and the second comparison against an IPTW model with time trends is meant to indicate whether the results are robust after controlling for potentially different time trends in different locations. At the state level, the OLS and time-trends models result in estimates with the same sign, somewhat smaller magnitude, larger standard errors, and a resulting loss of statistical significance. At the urban area-level, OLS and time-trends models produce similar statistically significant estimates (p-values increase from p=0.045 to p=0.062 and p=0.055).

Table 4 Comparison of regression models specified using weighted least-squares (using inverse probability of treatment weights), ordinary least-squares, and weighted least-squares with time trends (i.e., time-varying group fixed effects).

	Dependent variable: log(Veh. Reg. per cap)		
	IPTW	OLS	IPTW w/ time trends
State-Level Model			
Treatment Effect	−0.031**	−0.025	−0.021
	(0.012)	(0.015)	(0.017)
Observations	550	550	550
Deg. Freedom	474	474	425
Adjusted R-Sq.	0.844	0.782	0.894
Urban Area-Level Model			
Treatment Effect	0.007**	0.007*	0.006*
	(0.003)	(0.004)	(0.003)
Observations	3,402	3,402	3,402
Deg. Freedom	2,903	2,903	2,412
Adjusted R-Sq.	0.913	0.913	0.954
Note:			*p<0.1; **p<0.05; ***p<0.01

Figure 2 compares the treatment and control groups for both the state-level and urban area-level analyses before and after weighting along a set of parameters used to calculate propensity scores[i]. At the state level, weighting is shown to reduce mean differences between the treatment and unweighted control group parameters by 70% to 100%. The differences between treated and untreated states are statistically significant when unweighted, but, as desired, become not statistically significant in the weighted sample (even at the p=0.10 level). Weighting is nearly as effective in the urban area case, and, while the algorithm fails to achieve a statistically indistinguishable unemployment rate in the weighted control group compared to the treatment group, the means for each group (8.1% for the weighted control group versus 7.9% for the treatment group) have comparable practical significance.

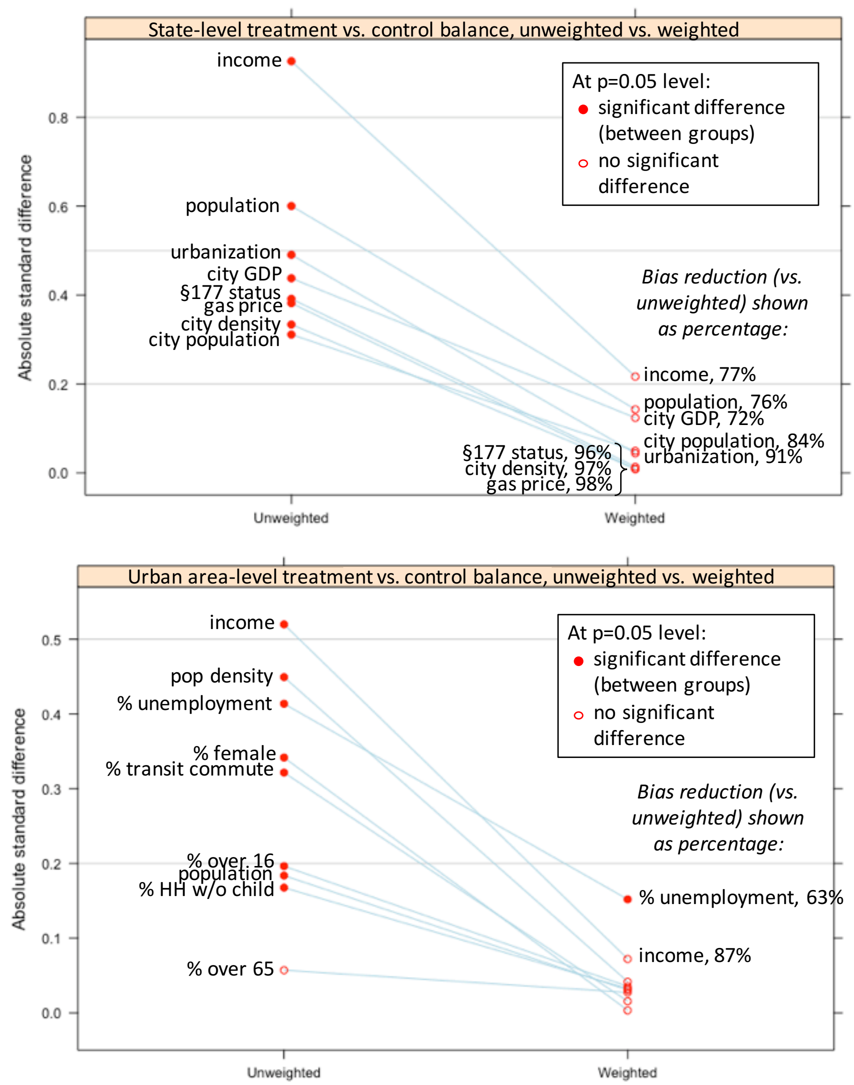

Figure 2 Effect size plot comparing the treatment states and control states (top) and urban areas (bottom) before and after weighting. Closed red circles indicate a statistically significant difference before weighting; open circles reflect no significant difference after weighting.

4.2 State-vs.-Urban Area Comparison

The effect of TNC entry on vehicle registrations estimated at the state level, a reduction of 3.1%, would correspond to a reduction in vehicle ownership of 4.1%, on average, across all urbanized areas if we assume no effect in rural areas (recognizing that TNC market entry and ridership is generally an urban phenomenon). A reduction in ownership is consistent with survey results from Hampshire et al (2017)[40] and Clewlow and Mishra (2018)[41], who find, respectively, a 9% increase in vehicle ownership among former Uber users after Uber left Austin, TX and a reduction in household vehicle ownership among 9% of households that use ride-hailing services in seven U.S. metro areas.

The effect of TNC entry on vehicle registrations estimated at the urban area level is an increase in 0.7%. An increase in ownership is consistent with the findings of Gong et al (2017) in China.

It is not necessarily inconsistent that our results find a *negative* TNC market entry effect on vehicle registrations at the state level and a *positive* effect at the urban-area level. To verify that the different result is not an artifact of using a different data source, we replicate the state level analysis using urban area-level data by aggregating (or population-weighting) urban area data by state and re-specifying the state-level regression model. Table 5 compares the effect estimates from the state- and urban area-level analyses with the effect estimated using urban area-level data aggregated to the state level. We find that the urban area data produces a significant negative estimate when aggregated to the state level, consistent with the state-level analysis. This suggests that the different data source is not the cause of finding different results at the state versus urban area level. Rather, the different result when averaged across different units of observation suggests heterogeneity: If TNC entry has different effects in different cities, averaging effects across urban areas can yield different results than averaging effects across states.

Table 5 Comparison of state-level analysis results and urban area-level analysis results and reproduction of state-level results using urban area data aggregated to the state level (i.e., arithmetic or population-weighted sums).

	log(Veh. Reg. per cap)		
	State data at state level	UA data at UA level	UA data at state level
Treatment Effect	−0.031**	0.007**	−0.010**
	(0.012)	(0.003)	(0.005)
Observations	550	3,402	287
Deg. Freedom	474	2,903	229
Adjusted R-Sq.	0.844	0.913	0.963
Note:		*p<0.1; **p<0.05; ***p<0.01	

4.3 Heterogeneous Effects

We investigate heterogeneity of the TNC entry effect across urban areas using 1) regression models that interact treatment with selected urban area characteristics to determine whether these characteristics explain differences in TNC entry effects across urban areas, and 2) cluster analysis, which identifies clusters of similar cities and estimates of TNC entry effects for each. In future work we plan to also apply latent class / profile analysis as an alternative approach to characterizing heterogeneity.

First, we specified a series of regression models identical to equation (1) but added interactions between treatment and one of three urban area covariates: a continuous measure of population, population density, or unemployment rate. In no case did we find a statistically significant interaction effect. In future work we plan to investigate categorical representations of these attributes and to use results from clustering and latent class analysis to inform our selection of urban area covariates.

Next, we use hierarchical clustering to identify groups of urban areas that are similar in terms of the covariates in our dataset[i]. We employ a divisive (rather than agglomerative) algorithm, in hopes of finding larger groups of similar urban areas, and [dis]similarity across urban areas is computed using Euclidean distances and Ward's minimum variance method[42]. For a given number of clusters, we re-specify our regression with an interaction between the treatment indicator and an urban-area cluster indicator. Doing so allows for the estimation of a baseline treatment effect for the first cluster and a series of interaction effects quantifying the difference between the effect in that baseline cluster and each other cluster. We sweep from two to nine clusters and estimate cluster-specific TNC entry effects as described. As Figure 3 shows, we confirm the presence of heterogeneous effects across urban areas. These effects range from –0.7% to 3.4%; though, only the clusters with positive effects that are large in magnitude (2.2% to 3.4%) are statistically significant.

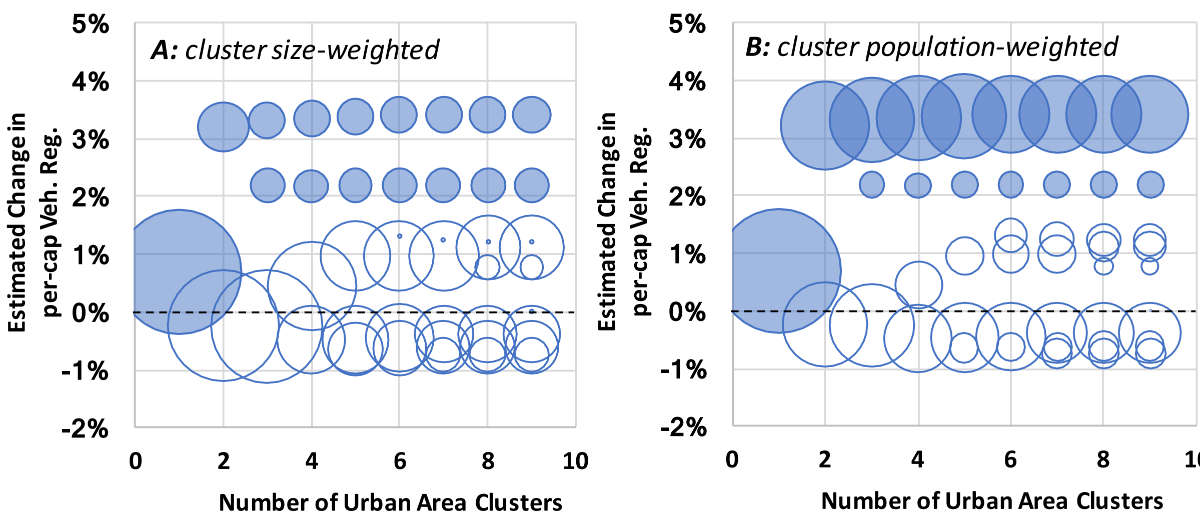

Figure 3 TNC treatment effect on the change in per-capita vehicle registrations varies by urban area typology, from as low as –0.7% to as high as 3.4%. Statistically significant effects are shown as shaded, and estimates that are not significant are open. In A (at left), the size of each circle reflects the number of urban areas in each cluster; whereas, in B (at right), the size of each circle reflects the total population in each cluster. Note that in the urban area-number plot, the average of effects is consistent at 0.7% across the number of clusters, which also aligns with the average estimate in Table 2.

While Figure 3 shows a change in the TNC effect estimated as each of the first five urban area clusters are added, the pattern appears to stabilize beyond five clusters, and including more than five clusters results in a cluster that contains just one urban area. Accordingly, we explore the case of five clusters for illustrative detail in Table 6. For each of these five clusters, the estimated TNC market entry effect is presented alongside the mean value of the (scaled) characteristics of the urban areas that comprise each cluster; each cluster is identified by the name of the largest city in that cluster. Cluster 1, New York, NY-like urban areas, and cluster 3, Riverside, CA-like urban areas, are the two clusters for which TNC effects are estimated as significant and positive. Table 6 makes clear that one thing urban areas in both of these clusters share is a relatively low number of per-capita vehicle registrations. The first cluster has higher average population, population density, commuters by transit, income, electric vehicle ownership, and trips by bus and rail as well as lower per-capita vehicle registrations than the other clusters. We refer to this cluster as "large dense cities". The third cluster has higher average unemployment and percentage of households with children and lower vehicle registrations per capita than the other clusters and is primarily composed of small to medium sized cities in California and Texas. We refer to this cluster as "small family-focused cities".

In summary, it appears that TNC entry tends to increase vehicle ownership in large dense cities and small family-focused cities with low per-capita vehicle registrations, but the effect on other types of cities is not statistically significant in this clustering. In future work we aim to examine robustness of our heterogeneity characterization to alternative clustering approaches and latent class / profile analysis and to investigate whether the urban area attributes identified by clustering produce statistically significant interaction effects with treatment in the base model.

Table 6 Mean values of regression covariates (scaled) and estimated TNC market entry effects for a five-urban-area-cluster case, sorted from largest-population cluster (New York, NY-like urban areas) to smallest-population (Tulsa, OK-like urban areas)

cluster	center city	TNC effect estimate	number of urban areas in cluster	vehicle registrations per capita	population	population density	% of commutes by transit	% unemployment	income per capita	% households w/o children	electric vehicle percentage	rail trips per capita	bus trips per capita	paratransit trips per capita
1	**New York, NY**	**3.4% *****	**45**	**-0.76**	**1.53**	**1.48**	**1.70**	**-0.21**	**1.46**	**-0.07**	**1.98**	**0.30**	**0.91**	**0.07**
2	Tampa, FL	-0.5%	150	0.10	0.01	0.14	0.02	-0.16	0.25	-0.10	-0.19	-0.39	-0.26	-0.08
3	**Riverside, CA**	**2.2% ****	**39**	**-0.94**	**-0.14**	**0.14**	**-0.20**	**1.30**	**-0.06**	**-2.05**	**0.22**	**-0.47**	**-0.37**	**-0.25**
4	San Antonio, TX	-0.6%	94	-0.20	-0.26	-0.30	-0.12	-0.66	-0.40	0.23	-0.32	N/A	0.70	0.17
5	Tulsa, OK	1.0%	157	0.48	-0.26	-0.41	-0.39	0.29	-0.40	0.49	-0.25	-0.50	-0.45	0.09

*Note: *p<0.1; **p<0.05; ***p<0.01*

5. DISCUSSION

Our results suggest that access to TNC services is associated with a significant effect on per-capita vehicle registrations: a decrease when averaged across states and an increase when averaged across urban areas. The effect flips direction when averaged over different units of observation, in part, because of underlying heterogeneity of the effects of TNCs in different types of cities. Our cluster analysis suggests that TNC entry tends to increase per-capita vehicle registrations in large dense cities and in small family-focused cities with low per-capita vehicle registrations. Effects on our other clusters is not statistically significant in our cluster analysis, though additional research is needed to assess robustness of the characterization of heterogeneity to alternative approaches. We also find a negative effect of TNC entry on EPA-estimated emissions of volatile organic compounds from passenger vehicles in U.S. states.

Interpreting these effects as causal relies on three key assumptions: 1) exogeneous intervention, 2) parallel trends, and 3) no spillover. Our event studies provide evidence supporting the exogeneous intervention and parallel trends assumptions both at the state level and at the urban area level because they find no statistically significant effect before treatment and a statistically significant effect after treatment having the same sign as our difference-in-difference results without assuming parallel trends. We also examine a model variant that includes different linear time trends for each group, relaxing the parallel trends assumption, and we find similar effect estimates (with a drop in statistical significance). Additionally, our application of IPTW successfully produces balanced or near-balanced treatment and control groups, mitigating conflation of treatment with group attributes. While this evidence is encouraging, trends are not strictly parallel across all states, even after applying our controls, so we do not eliminate the possibility of spurious results. Further, while our event studies do not indicate that changes in the dependent variable (registrations) preceded treatment, we cannot rule out the possibility that the decision to treat was influenced by changes in omitted variables. In future work we are seeking data on Google Trends that would allow us to control for one factor that we understand influenced Uber's decision to enter urban areas: local web searches for Uber and Lyft. For the coefficients reported as significant findings, the application of several diagnostic methods—visual inspection of regression residual errors as well as randomized treatment, leave-one-out, TNC market launch encoding, or excluding-outlier robustness checks—yields no evidence of systematic error or potential misspecifications. We assume that spillover effects are negligible (e.g.: that residents in one city do not change vehicle ownership patterns in response to experiences with TNCs in other cities).

Our results do not identify robust, statistically significant effects of TNC entry on gasoline consumption, vehicle miles traveled, or emissions other than VOCs, but this does not imply that TNCs have no effect on these outcomes. It is possible, for example, that TNCs have had substantial impact on these outcomes in particular U.S. cities (especially in light of the heterogeneous effects detected among the urban area clusters) without producing robust, statistically significant patterns across U.S. states or urban areas that are identified with our analysis. Further, our analysis does not capture the mix of trends that may lead to these net results, such as competing factors that act both to increase and to decrease VMT or changes in the fleet mix that result in fewer vehicle registrations overall but not necessarily fewer new vehicle purchases.

Our analysis focuses on net effects to overall outcomes after TNCs enter urban areas. We cannot identify changes to vehicle fleet mix with the available data, and there are potentially multiple alternative—and sometimes competing—narratives that might explain these trends. For example, it is possible that TNCs reduce VOC emissions primarily by shifting VMT away from older, less efficient personal vehicles toward newer, more efficient TNC vehicles that operate under hot steady-state conditions for a large portion of VMT, but it is also possible that the VOC emissions decline detected here results from the fewer vehicles (also detected here) used as an input to the models that EPA uses to produce published highway emissions data. Newer vehicles are associated with lower pollutant emissions: CO, NO_x, VOC, and PM emissions in light-duty transportation have declined 30-50% over the past ten years[43]. The EPA emissions estimates we use do not account for potential changes in cold start vs. hot operation ratios induced by TNCs, so any signal captured by our linear models and data is potentially attributable to a vehicle fleet transition but not likely to drive-cycle changes. As another example, it is possible that TNCs increase VMT on a per-trip basis due to "deadheading", or empty miles traveled between passenger trips, and trips induced from other travel modes and that TNCs simultaneously decrease the total number of trips traveled, since the perceived cost per trip is higher in a TNC than in a personal vehicle (where vehicle capital costs are "sunk"). Depending on their relative magnitudes, these dynamics could yield a near-zero net effect. Additional study on the effect of TNC market entry on vehicle fleet composition and distribution of VMT across the fleet is needed for deeper insight about the mechanisms that produce these outcomes.

ACKNOWLEDGEMENTS

This work was supported in part by Alfred P. Sloan Foundation and the U.S. Department of Energy through the National Bureau of Economic Research. This study was also supported in part by a grant from Carnegie Mellon University's Scott Institute for Energy Innovation and by the Center for Climate and Energy Decision Making (SES-0949710 and SES-1463492) through a cooperative agreement between the National Science Foundation and Carnegie Mellon University and developed in part under Assistance Agreement No. R835873 awarded by the U.S. Environmental Protection Agency. Jacob Ward is a Technology Manager at the U.S. Department of Energy`s Vehicle Technologies Office, who supported him in that role while conducting this work; he was also supported by a Dwight David Eisenhower Transportation Fellowship. This study has not been formally reviewed by these sponsors; the views expressed in this document are solely those of authors and do not necessarily reflect those of the sponsors; and the sponsors do not endorse any products or commercial services mentioned in this publication.

Additionally, the authors thank Kenneth Gillingham and Kate Whitefoot for their thoughts and guidance on regression specification and robustness checks and Albert Presto for his expertise on air quality data.

DECLARATION OF INTEREST

The authors declare no competing interests.

REFERENCES

1. Energy Information Administration (EIA). (2017). "Power sector carbon dioxide emissions fall below transportation sector emissions". Today in Energy. https://www.eia.gov/todayinenergy/detail.php?id=29612
2. San Francisco County Transit Authority. (2017). "TNCs Today". http://www.sfcta.org/tncstoday
3. Gutman, D. (2018). "How popular are Uber and Lyft in Seattle? Ridership numbers kept secret until recently give us a clue". Seattle Times. Nov 5. https://www.seattletimes.com/seattle-news/transportation/how-popular-are-uber-and-lyft-in-seattle-ridership-numbers-kept-secret-until-recently-give-us-a-clue/
4. Uber. (2018). "5 Billion Trips." Uber Newsroom. Jun 29. https://www.uber.com/newsroom/10-billion/
5. Lyft. (2018). "One Billion Rides. One Billion Connections." Lyft Blog. Sep 18. https://blog.lyft.com/posts/one-billion-rides
6. Rayle, L., Dai, D, Chan, N., Cervero, R., and Shaheen, S. (2016). Just a better taxi? A survey-based comparison of taxis, transit, and ridesourcing services in San Francisco. Transport Policy 45 (2016) 168–178.

7. Hall, J.D., Palsson, C., and Price, J. (2018). Is Uber a substitute or complement for public transit. Journal of Urban Economics 108 (2018) 36–50. https://doi.org/10.1016/j.jue.2018.09.003
8. Hampshire, R. C., Simek, C., Fabusuyi, T., Di, X., and Chen, X. (2017). Measuring the Impact of an Unanticipated Suspension of Ride-Sourcing in Austin, Texas. 31 May. http://dx.doi.org/10.2139/ssrn.2977969
9. Clewlow, R., and Mishra, G. S. (2017). Disruptive Transportation: The Adoption, Utilization, and Impacts of Ride-Hailing in the United States. University of California–Davis Research Report – UCD-ITS-RR-17-07.
10. Schaller Consulting. (2018). The New Automobility: Lyft, Uber and the Future of American Cities. 25 Jul. http://www.schallerconsult.com/rideservices/automobility.htm
11. Gong, J., Greenwood, B. N., and Song, Y. (2017). Uber Might Buy Me a Mercedes Benz: An Empirical Investigation of the Sharing Economy and Durable Goods Purchase. 19 May. http://dx.doi.org/10.2139/ssrn.2971072
12. Li, Z. and Hong, Y. and Zhang, Z. (2016). Do On-demand Ride-sharing Services Affect Traffic Congestion? Evidence from Uber Entry. 30 Aug. http://dx.doi.org/10.2139/ssrn.2838043
14. Henao, A. (2017). Alejandro Henao, "Impacts of Ridesourcing - Lyft and Uber - on Transportation Including VMT, Mode Replacement, Parking, and Travel Behavior" PhD diss., University of Colorado Denver, 2017, ProQuest 10265243.
15. Drozd, G. T., Zhao, Y., Saliba, G., Frodin, B., Maddox, C., Weber, R., M.-C. O. Chang, Maldonado, H., Sardar, S., Robinson, A. L., and Goldstein, A. H., (2018). Time Resolved Measurements of Speciated Tailpipe Emissions from Motor Vehicles: Trends with Emission Control Technology, Cold Start Effects, and Speciation. Environ. Sci. Technol. 2016, 50, 13592−13599
17. Greenwood, B. N., & Wattal, S. (2017). Show Me the Way to Go Home: an Empirical Investigation of Ride-sharing and Alcohol Related Motor Vehicle Fatalities. MIS Quarterly 41(1), 163–187.
18. Burtch, G., Carnahan, S., and Greenwood, B. N., (2018). "Can You Gig It? An Empirical Examination of the Gig Economy and Entrepreneurial Activity." Management Science. Ahead of Print. 27 Feb. https://doi.org/10.1287/inte.2017.090610.1287/mnsc.2017.2916
20. Friedman, J., Hastie, T., Tibshirani, R. (2000). Additive Logistic Regression: A Statistical View of Boosting. The Annals of Statistics 28(2): 337–407
21. McCaffrey, D. F., Griffin, B. A., Almirall, D., Slaughter, M. E., Ramchand, R., and Burgette, L. F. (2013). A Tutorial on Propensity Score Estimation for Multiple Treatments Using Generalized Boosted Models. Stat Med. 2013 Aug 30; 32(19): 3388–3414.
22. WardsAuto. (2017). Data and Insights: U.S. Total Vehicle Registrations by State. http://wardsauto.com/data-center
23 Bureau of Economic Analysis. (2015). Annual State Personal Income and Employment. https://www.bea.gov/itable/
24 Environmental Protection Agency. (2011). Sales of California Vehicles for 2011 Model Year and Beyond. https://iaspub.epa.gov/otaqpub/display_file.jsp?docid=24724&flag=1
25 U.S. Bureau of Economic Analysis. (N.D.) Table 1. Current-Dollar Gross Domestic Product (GDP) by Metropolitan Area. Interactive Data Application. https://www.bea.gov/itable/index.cfm

26 U.S. Census Bureau. (N.D.) Annual Estimates of the Resident Population for Incorporated Places of 50,000 or More. American FactFinder. https://factfinder.census.gov/faces/nav/jsf/pages/index.xhtml
27. U.S. Census Bureau. (2017). American FactFinder. https://factfinder.census.gov/faces/nav/jsf/pages/index.xhtml
28. U.S. Census Bureau. (2012). Methodology for the Intercensal Population and Housing Unit Estimates: 2000 to 2010. https://www2.census.gov/programs-surveys/popest/technical-documentation/methodology/intercensal/2000-2010-intercensal-estimates-methodology.pdf
29. IHS Markit. (2019). Vehicle Market Analysis: Registrations and Vehicles-in-Operation. https://ihsmarkit.com/products/automotive-market-data-analysis.html
30. U.S. EPA. (2019). Air Quality System (AQS). https://www.epa.gov/aqs
31. Federal Transit Administration (FTA). (2019). The National Transit Database. https://www.transit.dot.gov/ntd
32. Bi, F. (2014). Uber Launch Cities and Dates. https://github.com/voxmedia/data-projects/tree/master/verge-uber-launch-dates
33. Brazil, N. and Kirk, D. S. (2016). Uber and Metropolitan Traffic Fatalities in the United States. American Journal of Epidemiology, Volume 184, Issue 3, 1 August 2016, Pages 192–198, https://doi.org/10.1093/aje/kww062
34. Gigante, P. (2016). Personal communication.
35. U.S. Census Bureau. (N.D.) 2010 Census Urban and Rural Classification and Urban Area Criteria. https://www.census.gov/geo/reference/ua/urban-rural-2010.html
36. Bureau of Economic Analysis. (2015). Annual State Personal Income and Employment. https://www.bea.gov/itable/
37. Environmental Protection Agency. (2011). Sales of California Vehicles for 2011 Model Year and Beyond. https://iaspub.epa.gov/otaqpub/display_file.jsp?docid=24724&flag=1
38. U.S. Census Bureau. (N.D.) Annual Estimates of the Resident Population for Incorporated Places of 50,000 or More. American FactFinder. https://factfinder.census.gov/faces/nav/jsf/pages/index.xhtml
39. U.S. Bureau of Economic Analysis. (N.D.) Table 1. Current-Dollar Gross Domestic Product (GDP) by Metropolitan Area. Interactive Data Application. https://www.bea.gov/itable/index.cfm
40. Hampshire, R. C., Simek, C., Fabusuyi, T., Di, X., and Chen, X. (2017). Measuring the Impact of an Unanticipated Suspension of Ride-Sourcing in Austin, Texas. 31 May. http://dx.doi.org/10.2139/ssrn.2977969
41. Clewlow, R., and Mishra, G. S. (2017). Disruptive Transportation: The Adoption, Utilization, and Impacts of Ride-Hailing in the United States. University of California–Davis Research Report – UCD-ITS-RR-17-07.
42. Murtagh, F. and Legendre, P. (2011). Ward's Hierarchical Clustering Method: Clustering Criterion and Agglomerative Algorithm. arXiv. https://arxiv.org/pdf/1111.6285.pdf
43. Davis, S. C., Williams, S. E., and Boundy, R. G. (2016). Transportation Energy Data Book. Oak Ridge National Laboratory Report ORNL-6992. http://cta.ornl.gov/data/index.shtml