Mobility and Congestion in US cities: Evidence from Google Maps

Principal Investigators:
Prottoy A. Akbar, PhD Student, University of Pittsburgh
Victor Couture, Assistant Professor, University of California, Berkeley
Gilles Duranton, Professor, University of Pennsylvania
Adam Storeygard, Associate Professor, Tufts University

Traffic congestion is an increasing concern in many American cities. New York City recently decided to become the first American city to implement congestion pricing. Many other cities have imposed limits on ride-sharing vehicles, and policy-makers regularly defend proposals for new transit and road infrastructure based on their expected effect on road traffic. Data on urban mobility have proliferated dramatically, thanks both to expanded administrative and survey data collection by governments, as well as GPS data from now-ubiquitous mobile phones. These data provide an excellent opportunity for characterizing urban traffic patterns. However, characterizing such patterns requires careful attention to the appropriate aggregation of data to reflect the nature of choices travelers make.

In this project, we harness novel methods and data from Google Maps to tackle these issues. The core questions that we propose to answer are: to what extent does variation in motor vehicle speed across US cities reflect congestion versus non-congestion factors, and how does that vary across cities and city characteristics such as infrastructure stocks, road layout and the availability of public transit?

In addition, we will provide evidence on several other questions of interest:

- How do results based on our high-volume, scalable and cheaply repeatable Google Maps data compare to estimates using the richer and representative but much more costly National Household Travel Survey (NHTS) data?

- How does travel speed vary *within* a city, for example by distance to the city center, type of destination, trip distance, or the class of road being used? Are these patterns stable across all hours of the day?

- How do mobility indices based on private vehicle speeds compare to analogous measures for public transit trips?

- Aside from congestion delays, how important is unreliability: how much does the duration of a trip vary from one day to the next (or one week to the next) when leaving at the same time of day?

- How does the US transportation system compare to those of other developed countries in these features?

These questions have important policy implications. To the extent that peak congestion drives speed variation across cities, congestion pricing and similar policies could have a large impact on mobility. To the extent that other factors are more important, other policies, including infrastructure development, may be more important. Note also that our data collection method has a tiny cost relative to the NHTS. If its results are consistent with the NHTS, and especially if it can detect differences across years, it could prove useful as a continuous interim study between the full NHTS rounds, which have been carried out every 5-8 years since 1969. Higher temporal resolution could allow for a better understanding of transportation trends, including responses to changes in infrastructure and land use patterns, as well as natural disasters and other short-term events. Our method is also uniquely well-suited to measuring unreliability in travel time, because we can query the same trip at the same time on different weekdays. Traditional methods (loop detectors, gps devices, or recall diaries) are not well-suited for measuring unreliability, but the cost of unexpected late arrival at work is potentially large and distinct from that of a predictably long commute. Finally, other countries with comparably developed road transportation systems provide potential lessons for US policymakers. Obtaining truly comparable data across countries is an important first step in this analysis.

Data and core methods

Our core objects of study are carefully defined city-specific indices of the speed of travel in each city, with and without traffic, and the differential between the two, congestion. Our key insight here, as described in our work on India (Akbar et al. 2018), is that such a speed index is akin to a price index, as speed is the price of travel in terms of time.[1] This has two important implications. First, the relevant "good", *chosen* by a traveler, is a trip with an origin and a destination, not travel along a specific block or highway segment that is used in some other studies. Second, the "market basket" of trips must be made comparable across cities.

Our primary data source is a set of 60 million trip instances in 139 American cities we have queried on Google Maps between June and October 2019. Each trip instance is one of 2 million randomly selected pairs of origins and destinations (based on several sampling schemes) queried at one of roughly 30 randomly selected times of day and days of the week (based on the distribution of departure times in the NHTS). In each case, we recorded Google's estimate of the duration and distance on the most recommended and fastest routes. Crucially, we record both an estimate reflecting real-time traffic conditions at the time of the query, and one in the hypothetical state of no traffic. Since we continue to query trips through the middle of the night, as an alternative measure of the duration without traffic, we can select the fastest speed for each origin-destination pair, across all times. For each pair, we also make sure to query at the same (randomly selected) time of day for 14 consecutive days. This allows us to study the variability, or reliability, of travel durations. Google Maps has recently introduced estimated real-time transit speeds, even for cities where GPS locations of buses and trains are unavailable, so we can generate analogous transit indices as well.

[1] Travel has other costs, most notably in fuel, and externalities in the form of pollution and accidents. We do not study these explicitly, but argue that they are roughly proportional to time costs.

In addition to the randomly sampled trips, we simultaneously queried an additional 50 million instances of driving trips that were reported in the 2017 NHTS. While we do not know exact origins and destinations in the NHTS, we calibrate our trips to match the reported distances while connecting census block groups where the trips start and end. We query each trip at their reported departure time of day on roughly 50 different days (on average), making our measure of average travel duration robust to local shocks like rain and accidents. We also query each trip at roughly 30 randomly selected times of day and days of the week, making them comparable to our trips with random origin-destination pairs.

Using these data, we will regress trip speed on trip distance and distance to the city center, plus indicators for time of day, day of the week, city, and in some regressions the class of the road being taken. These key controls ensure that we are comparing like trips across cities. The estimated city effects form our baseline index of city speed. Predicted values from these regressions on city-specific samples allow us to construct Laspeyres, Paasche and CES-like speed indices more directly analogous to price indices as well. Running the analogous regression using speeds without traffic generates a city-level speed index without traffic, and the difference between this and our baseline speed index is an index of congestion. These two sub-indices provide an exact decomposition of the baseline speed index. Each index can further be related to city characteristics including the state of its infrastructure and public transit, drawn primarily from NHTS and OpenStreetMap.

We believe that such estimated relationships are useful even though establishing causality will be challenging. First, descriptive analysis is instructive here on its own, given the novelty of the data. Second we can follow standard approaches used by past literature that rely on the importance of historical factors that are not relevant to today's travel decisions within cities but nonetheless determined the transportation infrastructure of American cities or the fact that a given area was served because it is located between some key points of interest. Third, we will look for policy experiments that might provide credible event study identification, such as the forthcoming congestion pricing scheme in New York City or the opening of a new transit line.

Related work, hypotheses, and extensions

In Akbar et al. (2018), currently in revise and resubmit status at the *American Economic Review*, we carry out a related exercise for 154 cities in India. We show that 1) Google Maps provides reliable estimates of the speed of travel, and 2) the resulting indices of mobility and congestion are remarkably stable when subjected to a wide variety of radical resampling, suggesting strongly that our results are not due to our particular sampling scheme. In India, we show that most variation across cities is driven by non-congestion factors, most likely related to infrastructure.

We expect that congestion will play a larger but not exclusive role in the variation across US cities. For example, even in the absence of congestion, travel along the centuries-old curved, narrow, short-block streets of downtown Boston will be slower than along their counterparts in downtown Houston regardless of traffic conditions.

In the United States, unlike in India, we have access to the 2017 NHTS, which contains data on real trips travelled by US household members. The NHTS has the advantage of being representative, and containing important information about the nature of trips and the households taking them, as well as the location of their origins and destinations. However, it costs tens of millions of dollars to carry out; hundreds of dollars per household. We sample Google trips at less than 0.01% of the per-trip cost, a total of less than a thousand dollars for tens of millions of sampled trips. This means that it is feasible to collect the same data every year, or even on a continuous basis. In the second part of our analysis, we propose to consider precisely this tradeoff. Our second key question asks how our replication of NHTS trips with Google data compare to actual NHTS trip data. If they provide consistent results, they might prove useful as an interim analog to the full survey. Using Google data, we are also interested in assessing the validity of our trip sampling strategies by comparing them to NHTS trips also queried using Google data.

We note that our Google data for the trips we sample have (at least) two advantages of their own. First, trip departure is at an exact time and trip duration is measured in seconds, with none of the rounding present in survey data. Second, our trips are exogenously imposed. This is important, for example for comparability across cities. Even if a certain type (time, length, location) of trip is rarely taken in New York, for example because of extreme congestion from longer trips, we must measure its speed for comparability with other cities, because it is a potential trip not taken due to its high (time) cost.

Contingent on funding, we will repeat our data collection in the late spring/summer of 2020 in order to consider changes over time. Such short run changes are interesting in their own right, and will also help us better understand the link between the 2019 Google Maps data and the 2017 NHTS.

Because our data acquisition technology is highly scalable, we have collected similar data for several other countries. We will thus also run analogous analyses for Canada and the European Union, to see how congestion in their cities compares to the United States. If data for the 2018-19 French survey analogous to the NHTS (Enquête nationale transports et déplacements, ENTD) is made available in an appropriately analogous format in time, we will use it in a similar comparison.

We are also interested in mass transit information as real-time travel time data for transit is becoming more broadly available. Each of our counterfactual trips and every actual NHTS trip is being queried for transit in addition to private vehicle travel. This should allow for detailed comparisons between modes, including how congestion varies across modes over time. For NHTS trips for which the mode is known, we can estimate standard travel demand models with much more precise data.

Research team

The proposed team is well-suited to carry out this research. We have been working together using a related methodology to study Indian cities with Google Maps data for the past three years. Couture and Duranton have used the 2009 NHTS extensively to study US cities, most notably in Couture, Duranton and Turner (2018), which uses related methods but without the Google Maps data. In ongoing work, Akbar and Duranton (2017) have combined Google Maps and a travel survey for Bogotá, Colombia that is similar to the NHTS. Our methodologies are thus highly refined and can be easily applied to the new data, which are of the same structure as those we have already used.

References

Akbar, Prottoy A., Victor Couture, Gilles Duranton and Adam Storeygard, 2018. "Mobility and Congestion in Urban India," NBER Working Papers 25218, National Bureau of Economic Research, Inc., November.

Akbar, Prottoy A. and Gilles Duranton, 2017. Measuring the cost of congestion in a highly congested city: Bogotá. Development Bank of Latin America (CAF) Working paper 2017/04.

Couture, Victor, Gilles Duranton, and Matthew A. Turner, 2018. Speed. *Review of Economics and Statistics* 100(4): 725-739, October.

Conflicts of interest

We, Prottoy A. Akbar, Victor Couture, Gilles Duranton and Adam Storeygard, each declare that we have no relevant or material financial interests that relate to the research described in this proposal.