Wp

WORKING PAPERS

MEASURING THE COST OF CONGESTION IN HIGHLY CONGESTED CITY: BOGOTÁ
04/2017
N° 2017/04

MEASURING THE COST OF CONGESTION IN HIGHLY CONGESTED CITY: BOGOTÁ

Akbar, Prottoy A.
Duranton, Gilles

CAF DEVELOPMENT BANK OF LATIN AMERICA

WORKING PAPERS

MEASURING THE COST OF CONGESTION IN HIGHLY CONGESTED CITY: BOGOTÁ

Akbar, Prottoy A.
Duranton, Gilles

CAF – Working paper N° 2017/04
04/2017

ABSTRACT

We provide a novel approach to estimate the deadweight loss of congestion. We implement it for road travel in the city of Bogotá using information from a travel survey and counterfactual travel data generated from Google Maps. For the supply of travel, we find that the elasticity of the time cost of travel per unit of distance with respect to the number of travelers is on average about 0.06. It is close to zero at low levels of traffic, then reaches a maximum magnitude of about 0.20 as traffic builds up and becomes small again at high levels of traffic.
This finding is in sharp contrast with extant results for specific road segments. We explain it by the existence of local streets which remain relatively uncongested and put a floor on the time cost of travel. On the demand side, we estimate an elasticity of the number of travelers with respect to the time cost of travel of 0.40.
Although road travel is costly in Bogotá, these findings imply a small daily deadweight loss from congestion, equal to less than 1% of a day's wage.

WORKING PAPERS

MIDIENDO LOS COSTOS DE CONGESTIÓN EN UNA CIUDAD ALTAMENTE CONGESTIONADA: BOGOTÁ

Akbar, Prottoy A.
Duranton, Gilles

CAF - Documento de trabajo N° 2017/04
04/2017

RESUMEN

Este trabajo provee un enfoque novedoso para estimar la pérdida irrecuperable de eficiencia a causa de la congestión. La metodología se implementa para el caso de los viajes que ocurren en las vías de la ciudad de Bogotá, mediante el uso de una encuesta de viajes y el cálculo de viajes contrafactuales generados a partir de Google Maps. Para el caso de la oferta de viajes encontramos que la elasticidad del costo del tiempo de viajes por unidad de distancia con respecto al número de viajeros es, en promedio, 0.06. La elasticidad es cercana a cero para niveles bajos de tráfico, alcanza un máximo de 0.20 a medida en que el tráfico aumenta, y vuelve a ser chica para niveles altos de tráfico. Este resultado contradice la evidencia existente para segmentos de vía específicos. Esta discrepancia se explica por la existencia de vías locales, que permanecen relativamente sin congestión poniendo una cota mínima al costo del tiempo de viaje. Para el caso de la demanda, estimamos una elasticidad del número de viajeros con respecto al costo del tiempo de viaje de -0.40. Aunque el uso de las vías en Bogotá es costoso, estos resultados implican un pérdida irrecuperable de eficiencia diaria baja, menor al 1% del salario diario.

Measuring the cost of congestion in a highly congested city: Bogotá*

Prottoy A. Akbar†
University of Pittsburgh

Gilles Duranton‡
University of Pennsylvania

First draft: April 2017

Abstract: We provide a novel approach to estimate the deadweight loss of congestion. We implement it for road travel in the city of Bogotá using information from a travel survey and counterfactual travel data generated from Google Maps. For the supply of travel, we find that the elasticity of the time cost of travel per unit of distance with respect to the number of travellers is on average about 0.06. It is close to zero at low levels of traffic, then reaches a maximum magnitude of about 0.20 as traffic builds up and becomes small again at high levels of traffic. This finding is in sharp contrast with extant results for specific road segments. We explain it by the existence of local streets which remain relatively uncongested and put a floor on the time cost of travel. On the demand side, we estimate an elasticity of the number of travellers with respect to the time cost of travel of -0.40. Although road travel is costly in Bogotá, these findings imply a small daily deadweight loss from congestion, equal to less than 1% of a day's wage.

Key words: congestion, deadweight loss of externality, travel demand, travel supply

JEL classification: R41

*We thank Martin Adler, Isabela Manelici, Antonio Russo, Chris Taber, Jos van Ommeren, Juan Vargas, Tony Venables and conference and seminar participants at Torcuato Di Tella , Harvard, NARSC, Oxford, and Wisconsin for their comments and suggestions. We are grateful to Hugh MacMullan and Wharton Research Computing for their outstanding technical help, Magda Biesiada, to Antonio Pinzon for providing us with the Bogotá Travel Survey, and to Jungsoo Yoo for his incredibly valuable research assistance. We are also grateful to Wharton Global Initiative and CAF for their funding.

†Department of Economics, University of Pittsburgh, 230 S. Bouquet St., Pittsburgh, PA 15260, USA (e-mail: `prottoyamanakbar@gmail.com`).

‡Wharton School, University of Pennsylvania, 3620 Locust Walk, Philadelphia, PA 19104, USA (e-mail: `duranton@wharton.upenn.edu`; website: `https://real-estate.wharton.upenn.edu/profile/21470/`). Also affiliated with the Center for Economic Policy Research.

1. Introduction

We provide a novel approach to estimate the deadweight loss of congestion. We implement it for road travel in the city of Bogotá using information from a travel survey and counterfactual travel data generated from Google Maps. For the supply of travel, we find that the elasticity of the time cost of travel per unit of distance with respect to the number of travellers is on average about 0.06. It is close to zero at low levels of traffic, then reaches a maximum magnitude of about 0.20 as traffic builds up and becomes small again at high levels of traffic. This finding is in sharp contrast with extant results for specific road segments. We explain it by the existence of local streets which remain relatively uncongested and put a floor on the time cost of travel. On the demand side, we estimate an elasticity of the number of travellers with respect to the time cost of travel of -0.40. Although road travel is costly in Bogotá, these findings imply a small daily deadweight loss from congestion, equal to less than 1% of a day's wage.

Estimating the cost of congestion for an entire city (or part of a city) is important. Cities in developing countries face considerable challenges. The management of their transportation infrastructure is certainly one of them. These cities often have an infrastructure tailored to a population that is half or less of what it is now. Economic growth also leads to fast-rising rates of motorisation in many places. Traffic congestion is already infamous in many developing cities from Bangkok to Buenos Aires, Bombay, or Bogotá, and many worry it is only going to get worse as population grows and rates of motorisation increase. In turn, congestion may impede the growth of those cities and have negative aggregate implications for development.

Policy makers have used a variety of policies to attempt to address this problem. These policies include increased road provision, the development of new subway lines or bus rapid transit systems, and quantitative restrictions that prevent cars from operating at certain hours depending on their plate number. Some of these policies are extremely expensive and their efficiency has been repeatedly questionned.[1] Economists often advocate the use of congestion taxes but making drivers pay for using urban roads is politically unpopular and perhaps technically hard to implement in many emerging cities.

In this study, we want to take a step back and assess first how big a problem congestion is.

[1] Elevated highways are expensive and are now decommissioned in many places. Subways and light rail often offer a poor cost benefit ratio (e.g., Gomez-Ibanez, 1996). Quantity restrictions do not seem very effective at reducing congestion either (Davis, 2008).

The first challenge is to develop a methodology to estimate the deadweight loss of travel congestion in a city. Simple economics indicates that we first need to measure the wedge between what travellers actually pay and what they should optimally pay. Because the roads in Bogotá are free and congestible just like in most cities, those that travel using the regular roadway with their car, a taxi, or a small bus, pay the average of the aggregate time cost of travel for each unit of distance travelled. What they cost to society is instead the (higher) marginal time cost of travel. While the issue is conceptually simple, the real difficulty lies instead in the empirical estimation of the supply and demand for travel.

The supply of travel in a city is essentially a cost function that describes how the time cost of travel is affected by the aggregate distance travelled at that time. Because aggregate distance travelled per unit of time is equal to the number of travellers divided by the time cost of travel (or multiplied by speed, keeping in mind that speed is the inverse of the time cost of travel), we can recover the supply of travel by estimating how number of travellers affects the time cost of travel. Transportation economists refer to this relationship as the speed-density curve.

Regressing a measure of the time cost of travel on the number of travellers is subject to a simultaneity issue. This is an equilibrium relationship and we expect the number of travellers to adapt to travelling conditions. Shocks like traffic accidents, road works, or thunderstorms will affect both the number of travellers and travelling conditions. There may also be a selection issue as intrinsically slow trips may be taken at times when the number of travellers is particularly high (or low). To tackle these identification challenges, we first gather data about the traffic shocks that can be observed such as weather shocks. Second, we consider a broad area of study instead of a few roads, as we expect drivers to re-route when facing a traffic shock that slows down traffic locally such as a road repair. Third, we replace the actual time cost of trips by a counterfactual time cost for the same trip measured on different days but for the same departure time to break any correlation between the error term of the regression and the dependent variable. Finally, we consider counterfactual time costs at alternative times of the day to be able to compare a trip with itself under different traffic conditions.

The demand for travel in a city aggregates individual demands. In turn, individual demand describes how far a traveller is willing to travel on a trip given its time cost of travel. Using again the fact that distance travelled is equal to the duration of travel divided by the time cost of travel, we can estimate demand by assessing whether a traveller travels on a trip or not at a given point in time depending on the time cost of travel at this time.

In our estimation of demand, we obviously worry that individual demand is correlated with aggregate demand. Traffic is slow at peak hours when everyone wants to travel. To tackle this problem, we allow for different demand intercepts by time of the day and, in some extensions, by areas of Bogotá. The estimation of individual travel demand is also subject to the same simultaneity problem as travel supply. A traffic shock may affect both the propensity to travel on a trip and the time cost of travel. Like with our estimation of supply, we use direct measures of traffic shocks as controls. We also rely on counterfactual time costs of travel instead of actual time costs. The counterfactual time costs of travel should not be affected by the exact travel conditions at the time at which a traveller chooses to travel unlike actual time costs. Finally, we also worry about possible unobserved trip or traveller characteristics correlated with the time cost of travel. We can condition them out using trip fixed effects. This said, when computing the deadweight loss of congestion, we contemplate the possibility that the demand for travel may be more elastic than we estimate it to be.

The second challenge is a data challenge. The methodology just described to estimate both the demand and supply of distance travelled is extremely data intensive. It requires knowing about the actual trips that travellers take including their origin, destination, time of departure, time of arrival, mode of travel, and, ideally, further trip and traveller characteristics. Beyond this, our approach also requires knowing about counterfactual time costs of travel at the same time a trip took place (but on different days) as well as at alternative times. For actual trips, the Bogotá Travel Survey (BTS) contains the information we need for each trip taken by a sample of Bogotá residents. Like its American counterpart (the National Personal Transportation Survey) or its French counterpart (Enquête Transport), the data of the BTS are built from travel diaries. For counterfactual trips, we developed an infrastructure to query Google Maps for the same trips at the same time on different days as well as at alternative times. This website reports precise trip durations at the time of the query. An issue is that the data for the BTS is from 2011 while we queried counterfactual trip durations in the summer of 2015. Fortunately, there was no major change in the transportation infrastructure in the intervening period (except for some minor extensions of the bus rapid transit and the start of the formalisation of informal buses which ewe discuss below).

Once we know about demand, about supply with the congestion externality (i.e., the average cost of travel), and about supply without the externality (i.e., the marginal cost of travel), it is easy to compute the deadweight loss of congestion as already argued. This quantity is interesting since it provides an estimate of how big a problem urban congestion is in welfare terms. Ideally, we would expect a city like Bogotá to implement a congestion charge. The estimated wedge between the average cost of travel and the marginal cost of travel is directly informative of the optimal congestion charge. Even if this politically controversial policy is not implemented, the output of this study is informative of the gains from other policies such as more staggered departure times coming, for instance, from more flexible working hours in the public sector or from a time-differentiated pricing of taxis or parking.

One would have expected economic estimates of the welfare loss of congestion to have been provided a long time ago since our conceptual framework essentially dates back to Pigou (1920). Nonetheless, as far as we know, our approach is novel. We actually know very little about the economic costs of traffic congestion. There exist a number of estimates for the US such as those provided by the Texas Transportation Institute's annual report on the costs of congestion in US cities (Schrank, Eisele, Lomax, and Bak, 2015). These numbers are unfortunately unhelpful for our purpose. First, the Texas Transportation Institute computes the hypothetical saving of resources that would be generated by moving from the current traffic situation to a situation where the overall amount of travel would remain the same but traffic would flow freely. This type of number makes little sense economically because we expect demand to respond to any improvement in traffic conditions, because forcing free flow traffic at peak hours as opposed to allowing an optimal amount of congestion would be grossly suboptimal, and because in many central parts of cities, a 'free-flow' speed of 60 kilometres per hour is impossible even in absence of traffic. Second, travel delays are seemingly worse in developing cities than in American cities and it is unclear how much worse they really are. Extrapolating to Bogotá or Bangkok some figures for Chicago (a city of roughly the same size) may not make much sense.

Related to our supply estimation, there is a literature measuring the wedge between the private and social marginal cost of travel. Keeler and Small (1977) provide an unusually careful study of this wedge for a collection of roads in the San Francisco Bay area and use it to assess whether the provision of highways is efficient. Dewees (1979) proposes a first attempt to measure congestion for an entire city in Toronto but his work relies on a traffic simulation model rather than actual data. In a similar spirit, much of recent literature on the social cost of congestion relies heavily on theoretical models for which parameter values are taken from previous literature (e.g., Parry and Small, 2009). Another strand of literature assesses the short-run effect of transit strikes on congestion (Anderson, 2014, Adler, Liberini, Russo, and van Ommeren, 2016).

There is also a large literature estimating speed-flow or speed-density curves for particular roads. See Small and Verhoef (2007) for a selective review. While it is important to know how travel is produced at the level of specific road segments and be able to measure the intensity of use of an infrastructure, these estimates are unlikely to be useful to estimate the city-wide cost of congestion. The main reason is that, as traffic worsens on a particular road, travellers using their own vehicles have the option of taking alternative routes. In order to know about the deadweight loss of congestion, we need to collect data about travel and travellers in a given areas rather than on specific segments.[2]

In a precursor to our data work, Ardekani and Herman (1987) use aerial photography of the central business districts of Austin and Dallas in Texas. More specifically, they exploit snapshots taken at short time intervals at various times of the day to measure both the density of vehicles and their speed on the main arterial roads. They cross-checked their measures of speed against ground-level speed measurements. Their findings suggest that the elasticity of speed with respect to the number of vehicles is slightly below -0.30.[3] In more recent work, Geroliminis and Daganzo (2008) estimate a speed-density curve for the center of Yokohama in Japan using fixed censors that count cars and moving censors installed in taxis. Their findings are suggestive of an elasticity of speed with respect to vehicle density of about -0.65.[4] Despite much interest for the "macroscopic fundamental diagram of traffic congestion" in transportation engineering, the work of Geroliminis and Daganzo (2008) appears to be nearly unique in its reliance on observational data instead of data simulated through a computer. As far as we know none of that literature attempts to identify supply separately from demand. In a different vein, Couture, Duranton, and Turner (2016) estimate an aggregate production function of travel from a cross-section of US metropolitan areas.

There is an extremely large transportation literature concerned with urban travel demand. Most of it, however, is not concerned with the estimation of the demand for distance travelled as a

[2]Using a simulation model, Safirova, Gillingham, and Houde (2007) show that congestion measured on a single link is a poor predictor of total congestion costs imposed by travel on that link.

[3]For unexplained reasons, they prefer to fit their observations of speed and number of vehicles against a speed-density curve with an exogenously imposed elasticity of -0.50.

[4]This elasticity is not explicitly calculated by Geroliminis and Daganzo (2008). We measured it as an arc elasticity using their plots. We view this elasticity of -0.65 as somewhat problematic. In our data, the number of travellers increases by a factor of more than a hundred between the most and least busy times of day. This elasticity would imply a nearly twentyfold increase in the time cost of traffic. In our data, the ratio of the time cost of traffic between the fastest and slowest hour is only about 1.7.

function of the time cost of travel. Instead, a large part of that literature focuses on the choice of travel mode (Small and Verhoef, 2007). The reason is that mode choice is easily observable and can be readily estimated as a function of the characteristics of trips that were taken. Our demand estimation requires knowing about the time cost of travel when the traveller is not travelling, an innovation of our data collection. There is also a recurring concern with the effect of gasoline price on travel demand (Graham and Glaister, 2002). Another large strand of the literature focuses on the effect of population density and other measures of urban form on distance travelled (Ewing and Cervero, 2010, Duranton and Turner, 2016). Much attention has also been devoted to induced demand (e.g., Duranton and Turner, 2011).

Finally, there are also several strands of literature preoccupied with various other urban externalities. First, there is an important body of work that attempts to estimate the agglomeration benefits of cities. See Combes and Gobillon (2015) for a recent review. Much attention has also been devoted to the estimation of neighbourhood effects and various forms of externalities associated with the physical layout cities. See for instance Ioannides and Topa (2010) for a review on neighbourhood effects; Rossi-Hansberg, Sarte, and Owens III (2010) for an example of a housing externality; or Campbell, Giglio, and Pathak (2011) for foreclosure externalities. There is also some research on environmental externalities, some of which is associated with transportation. See for instance Chay and Greenstone (2005) for an assessment of the costs of urban air pollution and Davis (2008) for the effect of driving restrictions on air pollution. This type of work on urban externalities is usually cross-sectional in nature or sometimes uses low-frequency time variations. Because congestion varies at high frequency, we rely instead on within-day variations in our estimation.

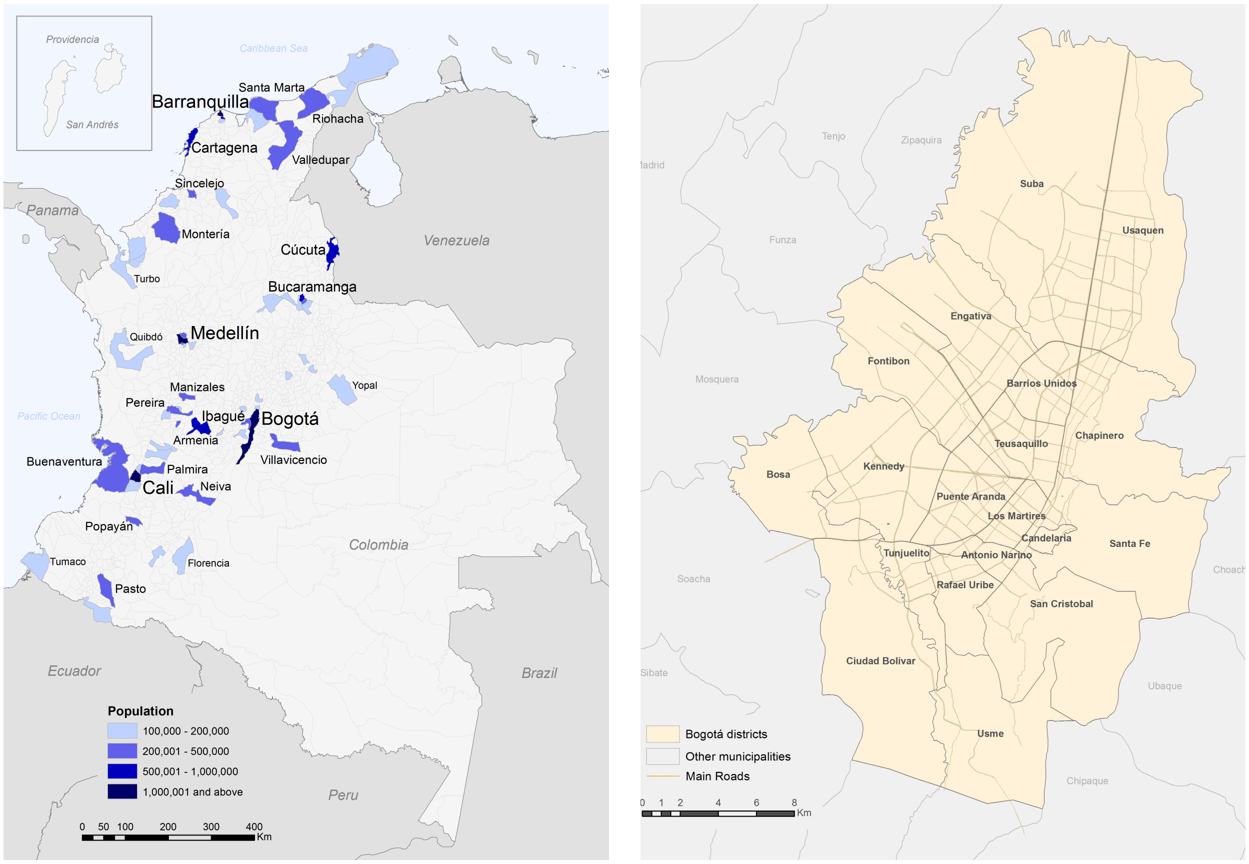
2. Data

2.1 Bogotá

Bogotá is the economic and political capital of Colombia. It is located inland on a high plateau between two branches of the Andes. Panel A of figure 1 provides a map of metropolitan areas in Colombia.[5] Panel B of the same figure represents municipal Bogotá and surrounding municipalities. The 2010 population of municipal Bogotá is 7.36 million. Using commuting patterns across

[5]The map represents the entire capital district which contains the city of Bogotá to the North and an inhabited mountainous region to the South (most of it is a national park).

Figure 1: Colombia and Bogotá



Panel A: Metropolitan areas in Colombia

Panel B: Municipal Bogotá

Note: Metropolitan areas are as defined in Duranton (2015).

municipalities, Duranton (2015) estimates that the metropolitan area of Bogotá is composed of 23 satellite municipalities grouped around their core, municipal Bogotá. Total population for the metropolitan area is 8.67 million in 2010. The annual GDP per capita in Bogotá is approximatively 12,000 USD, which is about 50% higher than the national average. Despite the city being highly congested, the ownership of motorised vehicles remains low with 165 vehicles per 1,000 inhabitants (Secretaría Distrital de Movilidad, 2013). On the other hand, population density in Bogotá is more than twice that of large European cities (Samad, Lozano-Gracia, and Panman, 2012).

To understand urban travel in Bogotá, it is useful to keep in mind that the city 'hits a wall' (literally) to its east, is further surrounded by mountains to the south and west, and opens up to a vast plateau to the north. The southern half of the city is poor while the more recently developed northern neighbourhoods along the mountains are the richest residential areas of the city. The historical center retains key administrative and government functions but much of formal private sector employment has moved to a new central business district in the northern half of the city.

2.2 *The Bogotá Travel Survey*

The 2011 Bogota Travel Survey (BTS) is a survey of household travel inside the Bogotá 'region'.[6] The initial sample contains 16,157 households (with on average 3.44 persons per household) who were asked to fill a travel diary for an entire day. For each traveler, we know his or her age, education, gender, seven household income indicators, and industry of occupation. For each trip, the traveller should normally report the address of the origin, the address of the destination, departure time, arrival time, the purpose of the trip, and its mode. We have travel information for 15,417 households, 11,994 of whom report journeys within Bogotá. We focus only on municipal Bogotá in our analysis.[7] The BTS also queried households in 15 other 'urban nodes' in surrounding municipalities. To take an example, one of these nodes, Gachancipá has an urban population of only about 8,000 and is located more than 40 kilometres away from Bogotá. To avoid trying to measure congestion where there might be none, we focus only on travel taking place within the municipality of Bogotá.

The BTS reports information for 104,441 trip instances. They correspond to 65,121 unique origin-destination pairs that we refer to as (unique) trips. In the data, going to work, going from work to the supermarket, and then back home should constitute three separate trips. However, for 14,138 of the unique trips, the reported origin is the same as the reported destination as some survey respondents appear to have confused arrival time at their destination with return time at the origin. We also eliminated trips with an origin or a destination that is missing or outside Bogotá and ended up with 36,309 trips, which we refer to as well-defined trips.

Table 1 reports a number of descriptive statistics for the BTS. These statistics are for well-defined trips as well as for all trips in the survey. Note also that table 1 reports statistics for both 'all trips' and only 'motorised' trips. By motorised trips, we mean trips taken by privately-owned vehicles, taxis, and small informal buses using roads. There are two other types of trips: walking trips and Transmilenio trips. Transmilenio is a bus rapid transit system that uses large buses on dedicated lanes. We do not consider walking trips nor Transmilenio trips in our main analysis of congestion

[6] The survey is in two parts. It contains the travel diary information that we exploit here and around 136,000 travel interceptions where travellers are surveyed as they travel in their car or using transit. We do not exploit information from travel interceptions as intercepted trips are not representative of the trips that Bogotá residents take.

[7] As implied by the figures reported above, 85% of the residents of the Bogotá metropolitan area live in the municipality of Bogotá. Using cell-phone data, Coscia, Neffke, and Lora (2016) find that commuters from outside municipal Bogotá represent only about 5% of employment in the municipality.

Table 1: Descriptive statistics from the Bogotá Travel Survey

	Mean	Median	Variance	25th pctile	75th pctile
Panel A: All trips (65,121 trips)					
Trip duration (all modes)	37.63	25	1,659	10	60
Trip duration (motorised trips)	54.13	45	1,914	30	70
No. of trips per person	2.69	2	1.71	2	3
No. of motorised trips per person	1.15	1	1.83	0	2
Panel B: Trips within Bogotá (36,309 well-defined trips)					
Trip duration (all modes)	37.93	30	1,462	10	60
Trip duration (motorised trips)	53.21	45	1,579	30	70
No. of trips per person	2.04	2	2.61	1	2
No. of motorised trips per person	0.91	0	1.67	0	2
Mode	Walk	Private vehicle	Taxi	Transit	Transmilenio
Panel C: All trips					
Share of trips	0.456	0.155	0.038	0.234	0.075
Share of trips longer than 15 min	0.289	0.188	0.048	0.326	0.106
Panel D: Trips within Bogotá					
Share of trips	0.448	0.154	0.044	0.245	0.082
Share of trips longer than 15 min	0.286	0.185	0.055	0.334	0.113

Notes: Statistics above are computed on all trip instances but are not weighted by sampling expansion factors. Trip durations are in minutes.

below since these trips do not use the regular roadway and, for Transmilenio trips, are subjects to a different type of congestion that we cannot measure.

As can be verified from the table, all trips and well-defined trips are very similar. The mean traveller takes about 2.7 trips per day, each lasting on average 38 minutes. Motorised trips are much longer. They last 54 minutes on average and 70 minutes at the top quartile. While many short trips are walking trips, trips longer than 15 minutes are for 29% of them walking trips, 19% by private vehicle, 5% by taxi, 33% by regular (informal) transit, 11% by the official bus rapid transit, Transmilenio, and the remainder by all other modes of transportation, mostly bikes. These statistics are close to those reported by Secretaría Distrital de Movilidad (2013). We also keep in mind that all selected trips are weekday trips.

2.3 *Mapping the Bogotá Travel Survey*

After the initial cleaning of the data, our next treatment is to map the addresses reported in the Bogotá Travel Survey. In principle, this task should be relatively straightforward because addresses in Bogotá follow a grid system. For instance, C21#42-25 refers to 21st Street (Calle in Spanish)

Table 2: Geocoding trips from the Bogotá Travel Survey

	All	Motorised	$\geq$15 min
Panel A: Share of all trips within Bogotá			
Trips with well-defined address-pairs (origin$\neq$destination)	0.71	0.89	0.76
Panel B: Share of well-defined trips within Bogotá			
Geocoded origin	0.53	0.47	0.50
Geocoded destination	0.58	0.47	0.53
Geocoded (origin and destination)	0.38	0.32	0.35
Geocoded with trip distance of at least one kilometre	0.37	0.31	0.34
Panel C: Share of geocoded trips within Bogotá with trip distance of at least one kilometre			
With weekday road-traffic data	0.84	0.93	0.91
With weekday road-traffic data at survey departure time	0.83	0.92	0.90

Notes: Statistics above are not weighted by sampling expansion factors. By survey departure time, we refer to the 15-minute interval which includes the reported departure time. All distance are Euclidean distances.

at 42nd avenue, 25 metres from the intersection. In practice, the address system in Bogotá is more complicated because of renumberings taking place in some areas as they (re-)develop, the existence of many exceptions to the grid such as 'diagonals', and named streets and avenues. The reporting of directions in the survey is also fairly poor as travellers often report directions like 'my daughter's school' or 'supermarket X in neighbourhood Y', etc. For instance, we counted more than 15 different ways to refer to the airport.

We have 39,888 addresses that correspond to 36,309 trips within Bogotá. Google Maps will sometimes return several locations for an address (including, in some cases, locations far from Bogotá). We only retained the 26,023 addresses that we could successfully geocode with a high degree of confidence.[8] These addresses correspond to 16,735 trips. For geoscraping, we only consider the 14,410 geocoded trips longer than one kilometre.[9] See table 2 for descriptive statistics regarding this process.

We worry about possible selection issues for the trips we could geocode since they represent only about 40% of all well-defined trips. Table 3 reports some descriptive statistics for geocoded trips. These descriptive statistics can be directly compared to those for all trips in table 1. Again, although we queried all geocoded trips, our analysis of road congestion below relies solely on

[8] That is, we get a unique answer for either a full address or an intersection or, if we get several returns, the first two are within 100 metres of each other.

[9] Short trips are hard to query and geoscraping often leads to problematic results. These trips are also disproportionately walking trips.

Table 3: Survey Statistics for geocoded trips

	Mean	Median	Variance	25th pctile	75th pctile
Panel A					
Trip duration (all modes)	38.20	30	1,235	15	60
Trip duration (motorised trips)	47.78	40	1,243	30	60
No. of trips per person	1.37	1	0.59	1	2
No. of motorised trips per person	0.71	1	0.65	0	1
Mode	Walk	Private	Taxi	Transit	Transmilenio
Panel B					
Share of all trips	0.357	0.204	0.061	0.251	0.102
Share of trips longer than 15 min	0.215	0.233	0.072	0.323	0.133

Notes: Statistics above are not weighted by sampling expansion factors. We only include geocoded trips that are at least one kilometre in Euclidean distance. Trip durations are in minutes.

motorised trips. Motorised trips that could be geocoded have a shorter duration by about 10% relative to all motorised trips.

2.4 *Google Maps data*

We then queried geocoded trips using Google Maps (GM) over March-August 2015. For Bogotá, GM reports real-time trip duration as well as trip distance.[10] To understand whether it is appropriate to use this information, two separate questions must be examined. The first regards the reliability of Google Maps. While Google gives out very little detail about how its travel information is produced, it is generally accepted to be reliable and accurate. It is the most popular app in the world, used by more than 50% of smartphone users.[11] While real-time travel times from Google Maps are only predictions, these predictions are trusted by more than a billion users. The second issue is whether travellers conform to the routes indicated by Google Maps. Anecdotal evidence suggests that that Google Maps and its sister community-driven application Waze are widely used by all professional drivers in Bogotá. Small buses have less flexibility in adapting their route to traffic. While actual travel durations reported in the BTS may on average overstate true trip durations, our counterfactual queries from Google Maps might slightly understate these true durations. It is hard to believe that this bias is large enough to modify our count of travellers in a

[10]While Google sells various business applications that allow subscribers to retrieve trip durations easily, these are projected durations rather than real-time durations. To obtain real-time trip durations one needs to use one of GM's websites which are much harder to scrape.

[11]`https://en.wikipedia.org/wiki/Google_Maps` (20 February 2017).

major way or that it varies enough throughout the day in a way that is correlated with the number of travellers to affect our estimations.[12]

We also keep in mind that two changes took place in the transportation infrastructure between the initial survey and our web data collection. The first is the introduction of the third phase of Transmilenio in 2012-2013. It consisted mainly in the opening or extensions of the 6th street, 26th street, and 7th avenue corridors as well as some new stations in Soacha, just outside of Bogotá. The second change is the gradual roll-out of the Bogotá integrated transit system (SITP, its Spanish acronym) which aims to formalise informal transit and consolidate the number of operators. In practice, this is about the gradual replacement of small informal buses by larger buses operating less flexibly along fixed routes with formal stops. Until after our data collection was completed, these changes had been limited in scope and slow to occur following some disfunctions with the city leadership.[13]

Using GM, we collected 18,092,738 counterfactual real-time trip durations by road including 13,403,699 for weekday travel for the 14,410 trips that we could geocode. For all but 15 of these trips we have at least one record of GM trip duration by road at the time of departure as stated in the BTS. Some of these observations correspond to multiple answers received from the same query.[14] In total, we observe at least one trip duration for 6,251,884 queries each corresponding to an instance (day and time) of a unique trip. Overall, we queried each trip on average 445 times including 54 times for the same departure time as reported in the survey for the actual trip.

Some descriptive statistics regarding these counterfactual trip durations are reported in table 4. The main difference between counterfactual and reported trip durations is that counterfactual durations are shorter. For the same motorised trips at the same time of the day, the BTS reports an average of 48 minutes whereas GM reports only 28 minutes. This large difference has two main sources. The BTS asks about journey duration and not the time spent in traffic. Recall also

[12]Recall also that a large fraction of motorised trips are done with small buses. A worry is that they may be slower than cars and taxis. Sizeable speed differences between cars and small buses seem hard to believe. Actually, the 'excessive' speed of small buses in Bogotá is a perennial complaint of other road users and a source of frequent accidents. A query of 'accidentes de busetas en Bogotá 2016' on Google News returned 3,150 press articles (15 February 2017).

[13]The mayoral tenure of Samuel Moreno (2008-2011) was plagued by a major corruption scandal involving the theft of much of the infrastructure budget earmarked for transit expansion. Moreno was eventually sentenced to 18 years in jail. His successor, Gustavo Petro was impeached following a conflict over a garbage collection reform before being later reinstated as mayor.

[14]These multiple returns correspond to different routes. Because trip durations using these routes are very similar (generally within a minute or two), we averaged the returns we obtained to generate a single observation for each query. We also average across queries when the same trip is queried twice or more within the same 15-minute time interval on the same day.

Table 4: Google Maps Statistics

	Mean	Median	Variance	25th pctile	75th pctile
Trip duration (all modes) at stated time	28.76	27	239.84	16	39
Trip duration (all modes) at any time	27.26	24	239	15	37
Trip distance (all modes) at stated time	10.93	9.9	46.53	5.3	15.2
Trip distance (all modes) at any time	11.26	10.4	46.78	5.6	15.6
Trip speed (all modes) at stated time	22.55	22.12	27.5	18.85	25.71
Trip speed (all modes) at any time	25.16	23.84	54.32	19.81	29.54
Trip duration (motorised trips) at stated time	28.49	26	208.75	17	37
Trip duration (motorised trips) at any time	27.18	24	217.13	16	36
Trip distance (motorised trips) at stated time	10.93	9.9	46.53	5.3	15.2
Trip distance (motorised trips) at any time	11.26	10.4	46.78	5.6	15.6
Trip speed (motorised trips) at stated time	23.13	22.65	28.89	19.35	26.41
Trip speed (motorised trips) at any time	25.53	24.23	54.46	20.19	30

Notes: Only Geoscraped trip-times for trips that are at least one kilometre in Euclidean distance. Durations are in minutes; distances are in kilometres; speeds are in kilometre per hour.

that a large fraction of motorised trips are by taxi or small buses and involve some waiting times (including for connections).

The trip durations obtained from GM are more useful for our purpose than overall journey durations since we are primarily interested in the time spent on the road, not the overall duration of a journey that also involves reaching one's vehicle, getting into traffic, and reaching one's final destination after parking a vehicle.[15] The counterfactual trip durations we collected are also arguably more precise as respondents to travel surveys tend to round both their time of arrival and time of departure up or down, sometimes to the next half hour.

Finally, we also collected hourly weather data for the city of Bogotá during our study period from Weather Underground which produces weather data from both official meteorological sources and personal weather stations.

3. Congestion in a city

Consider an area at time t. There is a number of vehicles N_t which we consider to be exogenous for now and equal to the number of travellers. Because we consider a fixed area, this number of vehicles can also be interpreted as a density. Since the roads are congestible, the speed of travel S_t

[15] One might be concerned with the particular issue of cruising for parking. Cruising is arguably a much less important issue in Bogotá relative to most American cities because keeping a car safe requires leaving it in a dedicated parking lot. Parking lots are broadly available in most parts of Bogotá. This is a second-order issue that is unlikely to affect our relative number of travellers throughout the day.

decreases with N_t.[16] This relationship is known as the speed-density curve and it is represented in panel B of figure 2. It is part of the 'fundamental diagram of traffic congestion'.[17]

The flow of vehicles at time t is then given by $V_t = S_t N_t$. Because speed is a function of the number of vehicles and this relationship can be inverted, we can express speed as a function of the flow. This relationship is known as the speed-flow curve and it is represented in panel A of figure 2. We note that speed is potentially a multi-valued function of the flow. This is because an increase in the number of vehicles has a positive direct effect on the flow of vehicles and a negative indirect effect through reduced speed. The region where the direct effect dominates is conventionally referred to as the congested region. The region where the indirect effect dominates is the hypercongested region. In this region, the elasticity of speed with respect to the number of vehicles is below minus one and the speed-flow curve bends backwards as the flow of vehicles declines with their number.[18]

In a simplified situation where the main cost of travel is time spent travelling, the private cost of travel faced by travellers at time t is given by $C_t = 1/S_t$ after normalising the value of time to unity. The function $C(N_t)$ represents the cost of travel at time t as a function of the number of travellers. This cost is taken as given by the travellers. Using the fact that $V_t = N_t/C_t$, we can use this cost function to derive an (inverse) supply curve that represents the time cost of travel at time t (the 'price' of travel) as a function of the flow of travel at the same time (the 'quantity' quantity). This supply curve is represented in figure 3.

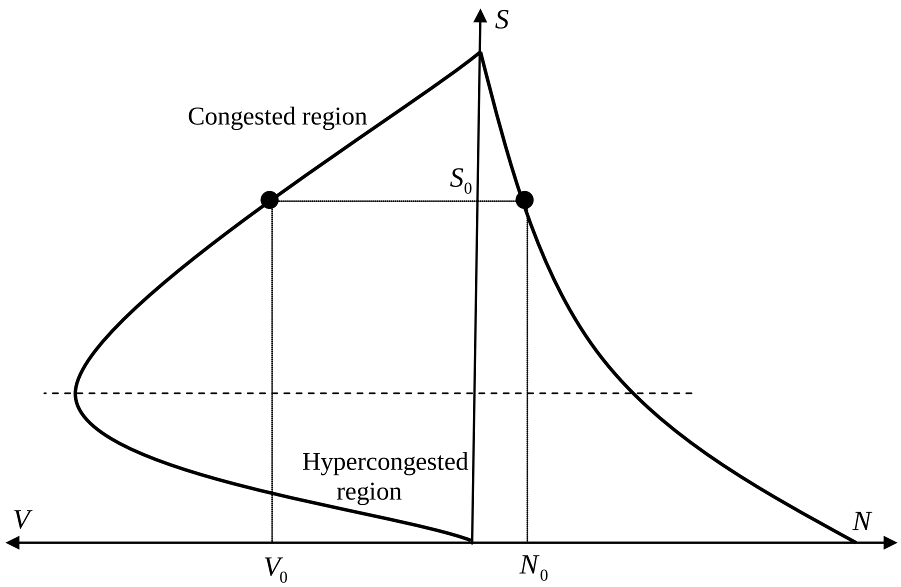
In an equilibrium where access to the roads is not priced, travellers impose a congestion cost upon other travellers but do not pay for this cost. They nonetheless pay for the congestion cost that is imposed to them by other travellers. That is, travellers pay the average cost of travel, not the marginal cost of travel. Put differently, the private cost paid by travellers is the average cost of travel, which these travellers take as given, while the social cost is the marginal cost $MC(N_t)$

[16]Given the data at hand, we only propose a simple 'static' model of congestion for each time t. The main alternative would be to use a dynamic approach building on the bottleneck model (Vickrey, 1969, Arnott, De Palma, and Lindsey, 1993). A representation of the simplest version of the bottleneck model in the supply and demand diagram of figure 3 below implies a supply curve that is horizontal below the capacity threshold and then vertical at this threshold. Our results suggest instead an S-shaped supply curve, which we explain by the re-routing of travellers around heavy traffic. Given this, a more advanced version of the bottleneck model would thus be needed. This is left for future work.

[17]This diagram was first derived by Walters (1961) and panel A was first dubbed the fundamental diagram of traffic flow by Haight (1963). See Small and Verhoef (2007) for further discussions.

[18]The point of maximum flow defines the capacity of the road network. There is a theoretical debate about whether hypercongestion constitutes a stable equilibrium. See, among others, Small and Chu (2003) , Arnott and Inci (2010), and Fosgerau and Small (2013) for contributions. See also Small and Verhoef (2007) for a summary discussion.

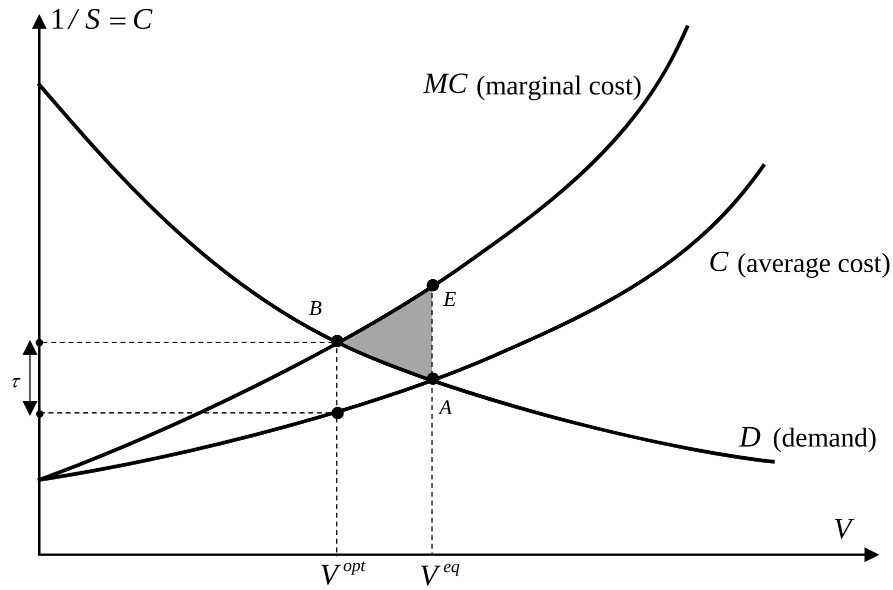
Figure 2: The fundamental diagram of traffic congestion



Panel A: the speed-flow curve

Panel B: the speed-density curve

Figure 3: The demand and supply for urban travel



associated with the total cost function $TC(N_t)$.[19] For instance, when $TC(N_t)/N_t \equiv C(N_t) = N_t^{\sigma}$, the congestion externality is a fraction σ of the average cost of travel: $MC(N_t) = (1+\sigma)N_t^{\sigma} = (1+\sigma)C(N_t)$.

As argued long ago by Pigou (1920) using a closely related example, having travellers pay the average cost of travel leads to an inefficient equilibrium for a given demand for travel. This inefficient equilibrium is represented by point A in figure 3.[20] To reach an efficient outcome, represented by point B on figure 3, travellers ought to be additionally charged the cost of congestion that they impose upon other travellers. This quantity is equal to the difference between the average cost of travel paid by travellers and the marginal cost of travel. It is represented by τ in figure 3. We also note that the deadweight loss associated with congestion is given by the area of the triangular shape ABE. Hence, estimating the deadweight loss of congestion and the optimal congestion charge requires estimating both the supply and demand for travel.

Further comments are in order. First, we note that we can obtain an upper bound for the deadweight loss of congestion by imposing a perfectly elastic demand for travel. Second, it is important to understand what this deadweight loss includes and what it does not include. Given our data, we only measure lost time, not lost gas, increased depreciation of cars, or increased pollution. Any calculation of the deadweight loss of congestion is also sensitive to the time interval over which we measure demand. If we measure the demand for travel over an entire day, rescheduling costs will not be accounted for since a trip that would have ideally taken place at 8 am may have eventually taken place at 11 am. Demand is likely to be more elastic over a short time period. To account for rescheduling costs, we need to measure demand over relatively short time intervals.

4. The demand and supply of urban travel

4.1 Supply

The supply of travel in a city is a relationship between a quantity of travel that is produced and its price. More specifically, (inverse) supply in our context describes how the aggregate amount of travel affects the time cost of travel per unit of distance. Because aggregate distance travelled

[19] To remain consistent with the fundamental diagram, with common sense intuition and with the estimation we perform below, we posit that the externality is in the number of travellers. The welfare calculation implicit in figure 3 will later require using travel flows.

[20] The point is also made by Beckmann, McGuire, and Winsten (1956) and by Vickrey (1963).

per unit of time is equal to the number of travellers divided by the time cost of travel, it is more convenient to estimate the effect of the number of travellers on the time cost of travel. For this purpose, it is natural to use the variation in the number of travellers over time. We start with the assumption that the time cost of travel C_{ijt} per unit of distance traveller i faces for trip j at time t is given by:

$$\log C_{ijt} = A_0 + A_1 X_i^{'} + A_2 Y_j^{'} + A_3 Z_t^{'} + \sigma \log N_t + \epsilon_{ijt}\,, \tag{1}$$

where X_i is a vector of characteristics for traveller i, including gender, age, education, income, and industry of occupation, Y_j is a vector of characteristics for trip j, including its length, mode of travel, purpose, and number of travelling companions, Z_t is a vector of characteristics at time t, including weather variables, day indicators, and time of day indicators in some specifications, and N_t is the number of travellers at time t.[21] Because we can only count the number of travellers instead of the number of vehicles, our estimations regard the number of travellers and account for possible adjustments in the intensity of vehicle occupancy. While it would be interesting to study this margin, this is beyond the scope of our work here.

In equation (1), the coefficient of interest, σ measures the elasticity of the time cost of travel with respect to the number of travellers. The main identification challenge is that the number of travellers and the time cost of travel may be simultaneously determined. Any supply shock ϵ_{ijt} is likely to affect both the number of travellers and the time cost of travel. What might these supply shocks be? First travel conditions may vary throughout the day for reasons unrelated with the number of travellers but correlated with it. For instance, drivers may slow down at night when there are fewer travellers. We can tackle this issue by focusing on variations taking place only within certain parts of the day.

Weather shocks like a rainstorm are a second source of shocks. Rainstorms occur regularly in Bogotá. They make driving hazardous and can flood some areas. These shocks arguably affect driving in the entire city. As a result the time cost of travel during rainstorms will be particularly high and the number of travellers particularly low. This will lead to estimates of the effect of the number of travellers on the time cost of travel that are biased downwards. To deal with supply shocks, our strategy is to control for weather conditions at the time a trip is taken through the vectors Z_t in equation (1).

[21] Relative to our empirical implementation, we slightly abuse notations here. A trip starts at time t_o and finishes when reaching its destination at time t_d. Both the time cost of travel and the number of travellers are computed for this time interval.

Accidents and road works are another type of supply shocks. These shocks are harder to measure and more local in nature. Accidents increase the number of travellers by reducing their flow. Accidents also often induce travellers to take alternative routes. Unexpected road works will be like accidents. Expected road works may have the opposite effect and reduce the number of travellers as they plan around them. To avoid our estimations being contaminated by the re-routing of trips, we consider all trips taking place in the city (or part of the city) without distinguishing them across specific roads or small neighbourhoods.[22]

Aggregating the analysis to the level of an entire area will attenuate the problem of the simultaneous determination of the number of travellers and time cost of travel but not fully eliminate it. First, a shock to the time cost of travel mechanically affects our count of number of travellers at a given point in time as these travellers reach their destinations earlier or later. To tackle this problem, we can use a counterfactual trip duration measured at the same time of the day on different days instead of actual trip duration. This allows us to construct a measure of the time cost of travel for a trip under 'regular' travel conditions. In turn, we can use the counterfactual durations of trips under regular travel conditions to compute a counterfactual number of travellers at each point in time. We can use this counterfactual number of travellers either as an instrument for the actual number of travellers computed using reported departure and arrival times or as an alternative explanatory variable of interest.

Beyond the duration of the trip at the extensive margin, shocks to the time cost of travel may also affect the decision of whether to take a trip at the intensive margin. A shock to the time cost of travel can affect the willingness of a traveller to travel when this shock is known to the traveller at the time of departure. For instance, if travellers are notified of major road works that slow down travel, they may elect to stay home and avoid travelling. To break down this correlation between traffic shocks and the time cost of travel, we do not have an uncontaminated measure of the number of travellers at hand. However, we can use the counterfactual time cost of travel instead of the actual time cost of travel as explanatory variable. In this case, the error term of the regression is no longer correlated with the explanatory variable.[23]

[22]Beyond biases caused by re-routing, cross-sectional results would also be driven by unobserved supply differences across those areas. See below for more on this.

[23]However, the solution is not perfect as it still leaves some measurement error with respect to the explanatory variable of interest. As our demand estimations below show, the measurement error on the number of travellers associated with this specific category of shocks is unlikely to be a first-order concern since the propensity to travel is fairly inelastic with respect to the time cost of travel.

We also note that using counterfactual time costs of travel and counts of travellers will also remove any bias associated with unobserved traveller characteristics. For instance, slow travellers may travel when the time cost of travel is low. Counterfactual trip durations and counterfactual numbers of travellers are independent of traveller characteristics.

To sum-up this discussion, rather than estimating equation (1), for each actual trip instance, we can estimate

$$\log \overline{C}_{ijt} = A_0 + A_1 X_i^{'} + A_2 Y_j^{'} + A_3 Z_t^{'} + \sigma \log \overline{N}_t + \epsilon_{ijt}\,, \tag{2}$$

where $\overline{C}_{ijt}$ is the counterfactual time cost of travel for trip j taken at time t by traveler i and $\overline{N}_t$ is the counterfactual number of travellers at time t where both counterfactuals are computed under average travel conditions.[24]

Our last worry is that particularly slow trips may be taken at specific times of the day and thus be correlated with the time cost of travel. While we control for trip mode, distance, and purpose, unobserved trip characteristics may be correlated with their time cost of travel and with the number of travellers. To tackle this problem, we rely on counterfactual trip durations measured at all times of the day instead of only at the time the actual trips take place. More specially, we can first re-estimate equation (2) using all counterfactual observations of a given trip instead of only those taking place at the same time as the actual trip instance. If the times at which counterfactual trips are sampled were random, there should be no correlation between unobserved trip characteristics and number of travellers. This is not exactly the case in our situation where we over-sampled times close to the reported time of a trip. To avoid this possible sampling issue, we can include trip fixed effects ξ_j and estimate:

$$\log \overline{C}_{jt} = A_0 + \xi_j + A_3 Z_t^{'} + \sigma \log \overline{N}_t + \epsilon_{jt}\,. \tag{3}$$

Finally, we note that the relationship between the time cost of travel and the number of travellers may not be well approximated by a log log form as in equations (1)-(3). It is of course easy to consider richer alternative functional forms such as Box-Cox transformations, polynomial forms, or even semi-nonparametric forms.

[24]Although counterfactual trip durations are uncorrelated with household characteristics, we still include traveller characteristics in the regression as these may proxy for trip characteristics.

4.2 *Demand*

To compute the deadweight loss from congestion, we also need to know how individual travel responds to changes in the time cost of travel. We can then aggregate individual demand across travellers to obtain aggregate demand. As highlighted in Couture *et al.* (2016), a full modelling of travel demand is well beyond what can be currently achieved since it involves a myriad of decisions regarding the choice of a destination, the choice of a departure time, the choice of a mode of transportation, the choice of a route, all this accounting for possible complementarities and substitutabilities across trips. Prior to these decisions, the choice of purchasing a vehicle or not and the choice of a residential location would also need to be modelled.

Our approach estimates whether an individual travels on a given trip at each time of the day, depending on cost of travel that he or she is facing for that trip. More specifically, to model the demand for actual trip j taken by person i and for time t — with for instance $t = 1, 2, ..., 24$ — let us start with the following expression:

$$\mathbb{T}_{ijt} = B_t + B_1 Y_j^{\prime} + B_2 X_i^{\prime} + B_3 Z_t^{\prime} - \eta \, \log \overline{C}_{jt} + \mu_{ijt} \,, \tag{4}$$

where $\mathbb{T}_{ijt}$ is an indicator variable taking value one if traveller i is travelling on trip j at time t and zero otherwise, B_t is a time-interval intercept (say, hourly), X_i is a vector of traveller characteristics, Y_j is a vector of trip characteristics, Z_t is a vector of characteristics at time t, $\overline{C}_{jt}$ is the time cost of travel for trip j at time t, and μ_{ijt} is an error term. While we know from the survey the time cost of travel at the time a trip is taken, the time cost of travel for a given trip at any other time of the day requires using counterfactual time costs.[25]

Equation (4) estimates for all times of the day (or a subset of them) the propensity of traveller i to travel on a given trip j that he or she actually takes as a function of the hour of the day, the characteristics of traveller i, the characteristics of trip j, some characteristics associated with time t, and the time cost of travel for that trip.[26] We thus estimate 'many' demand curves corresponding to each time of the day.

[25] This estimation considers each trip independently to avoid overwhelming complexity. It thus ignores the constraint that a traveller can only travel on one trip at a time.

[26] The demand estimated here depends on policy-imposed restrictions. The city of Bogotá imposes driving restrictions during some hours of the day based on plate numbers ('pico y placa'). For instance, a driver might be 'forced' to leave work after the restriction is lifted rather than during the restriction period when traffic might be lighter. Thus, the 'policy-constrained' demand that we estimate may appear as less elastic than unrestricted demand would be. Demand under institutional constraints is nonetheless the one we need to estimate for our purpose. This nonetheless prevents us from recovering the deadweight loss of driving restrictions since they are treated as if they were part of preferences.

Before going further with equation (4), note that we can sum travel across trips for a traveller and across all travellers to obtain an aggregate relationship which informs us of the number of travellers, N_t, depending on the time cost of travel at this time. We can then use the fact that $V_t = N_t / C_t$, to derive aggregate distance travelled per unit of time as a function of the time cost of travel. This is an aggregate demand curve, as per figure 3.

Next, note that after introducing trip fixed effects we can modify equation (4) and write:

$$\mathbb{T}_{jt} = B_t + \delta_j + B_3 Z_t' - \eta \, \log \overline{C}_{jt} + \mu_{jt} \, . \tag{5}$$

This equation no longer contains vectors of trip and traveller characteristic. This is because a trip is essentially specific to a traveller, traveller fixed effects are unnecessary and would not be identified.[27]

There is an important difference between equations (4) and (5). Equation (4) identifies η primarily from the cross-section of trips and relies on the notion that different trips face a different time cost of travel as they take place in different locations at time t. Hence in equation (4), η measures the propensity to take a given trip depending on its cost relative to other trips. Equation (5) relies instead on the within-day variation in the time cost of travel for a given trip. In this equation, η measures the propensity to take a given trip at a given time depending on its time cost at this time relative to its time cost at alternative times. As made clear below, both sources of variation are important to estimate our key parameter of interest, the slope of the demand for travel, η. Then, the time effects will determine the intercepts of the demand curve for each hour of the day.

Equations (4) and (5) call for a number of comments. First, the dependent variables is a binary outcome for which an ordinary least-squares estimation is inappropriate. We instead estimate equations (4) and (5) with a Poisson regression. Our choice of Poisson regressions as our preferred estimation method is driven by computational tractability in situations where we include a large number of fixed effects and cluster our standard errors by time intervals. In any case, we check the robustness of our estimates using alternative estimation techniques for discrete outcomes such as logit and negative binomials.

Second and most importantly, we are worried about a possible correlation between unobserved individual demand factors and the time cost of travel. Actually, individual and aggregate demands

[27]Strictly speaking, trips are not truly specific to travellers as the same trip is sometimes taken by multiple travellers and some travellers take multiple trips. However, we do not have enough connected trips and travellers for a separate identification of traveller and trip fixed effects here. A key reason is that in their immense majority, trips taken by multiple travellers are taken by travellers that belong to the same household.

are positively correlated by construction. People tend to travel at times when the time of cost of travel is higher – peak hours – because everybody else travels during those hours. To limit this problem, we include hourly fixed effects in equations (4) and (5). In robustness checks, we experiment with shorter time periods and allow for demand intercepts to vary across different parts of the city.

In equation (4), the elasticity of demand is appropriately estimated when differences across trips in their time cost of travel only reflect different (local) supply conditions, not demand factors specific to one trip relative to another. Concretely, this condition will not be satisfied if, say, the demand (and thus the time cost) for going from A to B is much higher than the demand for going from B to A. Then, more travellers may take the slower trip despite a downward-sloping demand curve. This condition will also fail if travellers sort across locations. For instance, travellers who have a high propensity to travel may elect to live in locations where travel costs are particularly low. To limit these possible correlations, we include many traveller and trip characteristics, including measures of the weather in equation (4).

The identification threats associated with the estimation of equation (4) result from the comparison of different trips with different unobserved demand factors possibly correlated with different time-costs of travel. These problems can be conditioned out by the inclusion of trip fixed effects in equation (5). The inclusion of trip fixed effects nonetheless raises different identification threats. Appropriate identification of the elasticity of demand in equation (5) requires that unobserved demand for a given trip is not correlated with its time cost after conditioning out the time of the day. To be concrete, this condition will not be satisfied if, for instance, a traveller takes a slow trip to the stadium at exactly the same time as his neighbours to attend a game while traffic in the rest of the city remains fine and this trip is fast during the rest of the day. To limit this problem and eliminate any bias associate with a temporary demand shock, recall that we rely on counterfactual time costs of travel. These counterfactual time costs of travel should be uncorrelated with shocks that happened to this trip at the time the actual trip we taken.[28]

To draw our conclusions, we also rely on the fact that equations (4) and (5) are subject to different simultaneity threats. This said, we keep in mind that a correlation between unobserved

[28]Ideally we would like to measure how regular travel behaviour is affected by regular travel conditions. Instead, we measure how the actual travel status, which may have been affected by shocks on the day the trip was taken, is determined by the counterfactual time cost of travel, which we interpret as the usual cost of travel. Because this measurement problem affects the dependent variable we do not expect it to affect our results.

individual demand factors and the time cost of travel is still possible and would lead us to underestimate the true elasticity of the demand for travel. In our computations of the deadweight loss of congestion below, we also provide results for a perfectly elastic demand.

Next, we may be concerned that the coefficient on the time cost of travel η may vary over time during the day. For instance, we may expect travel to be less responsive to travel condition at peak hours because the trips taken at those hours are less discretionary.[29] There is also an issue of functional form. We estimate a log linear form in our baseline but consider alternative forms in robustness checks.

Note finally, that we measure demand using trips that were actually taken using information about traffic conditions when they were taken relative to when they were not taken. There are arguably other trips that could have been taken under better traffic conditions. We do not know about these trips but note that they were trips for which the willingness to pay was less than what we observe at equilibrium. Simply put, these trips are lower down on the demand curve. While it would be important to know more precisely about those trips to assess, for instance, the effects of traffic improvements, we only need to know about the part of the demand curve above the equilibrium to compute the deadweight loss of congestion (as per figure 3).

5. Results for the supply and demand

5.1 Supply

We first report results for actual motorised trips in table 5. Our variable of interest is the number of travellers. To compute this number for each trip, we first sum the number of travellers for each each five-minute interval throughout the day across all motorised trips. For instance, if a traveller reports leaving home at 7.21 and arriving at work at 7.52, this traveller will be counted among those that travel during the six five-minute intervals from 7.20 to 7.50. For each trip, we then compute the number of travellers by taking the mean of the number of travellers over all the five-minutes intervals that compose this trip.

[29]On the other hand, travel at peak hours may respond to the time cost of travel through other margins such as household location. Consistent with this argument, Duranton and Turner (2016) find that the elasticity of commute distance with respect to population and job density (arguably a correlate of the time cost of travel) is more than twice as large than the corresponding elasticity for total vehicle-kilometre travelled.

Table 5: Regressions on actual survey trips

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
log travellers	0.058[a]	0.92[a]	-22.9[a]	-23.4[a]	-21.9[a]	-21.8[a]	-23.3[a]	-20.3[a]
	(0.023)	(0.34)	(3.86)	(3.21)	(3.11)	(3.14)	(3.24)	(3.75)
log travellers squared		-0.036[b]	2.08[a]	2.14[a]	2.01[a]	2.00[a]	2.13[a]	1.85[a]
		(0.014)	(0.34)	(0.28)	(0.27)	(0.28)	(0.29)	(0.33)
log travellers cubed			-0.062[a]	-0.064[a]	-0.060[a]	-0.060[a]	-0.064[a]	-0.055[a]
			(0.010)	(0.0083)	(0.0080)	(0.0082)	(0.0085)	(0.0098)
log trip length				-0.68[a]	-0.71[a]	-0.73[a]	-0.73[a]	-0.72[a]
				(0.011)	(0.011)	(0.010)	(0.011)	(0.015)
Transit					0.26[a]	0.24[a]	0.21[a]	0.17[a]
					(0.015)	(0.016)	(0.018)	(0.022)
Taxi					-0.27[a]	-0.28[a]	-0.29[a]	-0.28[a]
					(0.023)	(0.023)	(0.024)	(0.031)
Travellers in trip						-0.031[a]	-0.056[a]	-0.033
						(0.012)	(0.014)	(0.021)
Trip purpose	N	N	N	N	N	Y	Y	Y
Traveller demographics	N	N	N	N	N	N	Y	Y
Industry	N	N	N	N	N	N	N	Y
Observations	8,996	8,996	8,996	8,996	8,996	8,996	8,121	4,758
R^2	0.01	0.01	0.01	0.31	0.36	0.37	0.38	0.37
Mean marginal effect	0.058	0.012	-0.061	-0.086	-0.13	-0.13	-0.12	-0.054
Max marginal effect	0.058	0.34	0.38	0.43	0.35	0.35	0.38	0.35

Notes: OLS regressions with a constant in all columns. The log time cost of travel per kilometre (i.e., minus log speed) is the dependent variable in all columns. Standard errors in parentheses. *a*, *b*, *c*: significant at 1%, 5%, 10%. We have three trip purpose indicators: work, study, health (default is everything else). Traveller demographics include age, education, gender, and seven household income category indicators. We also have 17 industry indicators.

Column 1 of table 5 regresses the log of the time cost of travel (or minus log speed) on the log number of travellers. The estimated elasticity of the time cost of travel with respect to the number of travellers is 0.058. Columns 2 and 3 add a quadratic and a cubic term for log density. Because, it is not easy to read elasticities from the estimated coefficients of a non-linear specification, we report the mean and maximum elasticities at the bottom of the table. While the mean elasticity of travel with respect to the number of traveller is slightly negative, the maximum elasticity in column 3 is 0.38.[30]

Column 4 also adds log trip distance as explanatory variable. Longer trips are faster because there is a fixed time cost associated with every trip and because longer trips tend to use faster

[30] To a large extent, the negative mean elasticity is driven by the S-shaped form of the speed-density curve where the first and last region (where most of the traffic takes place) are slightly downward sloping with the cubic form used here. We return to these issues below.

roads. As Couture *et al.* (2016), we note the explanatory power of this variable. Column 5 includes indicators for the (motorised) mode of transportation. As expected, taxis, which do not require to be parked, are faster than regular vehicles whereas transit is unsurprisingly slower. Column 6 further considers the purpose of trips and the number of travellers. Work and study trips are modestly slower just like trips taken with more travellers. Columns 7 and 8 additionally include socio-demographic characteristics of travellers and their industry of employment. While we find that richer travellers travel faster, none of these additional controls affect the coefficients on the number of travellers nor the explanatory power of the regression. This suggests that in our context we can treat trips separately from the travellers that undertake them as a first approximation.

The results of table 5 use the within-day variation in the number of travellers. In results not reported here, we duplicated the specifications of this table and added a fixed effect for each hour of the day to estimate the elasticity of the time cost of travel with respect to the number of travellers from variations within each hour. The specification analogous to column 1 estimates a similar elasticity of 0.053 instead of 0.058.

We now turn to results obtained using counterfactual trips collected from GM. Recall that this change affects both the dependent variable and the computation of the number of travellers since we now use GM trip durations instead of reported trip durations from the BTS. We start in table 6 with trips observed at the same time as the actual trip. This table reproduces the structure of table 5 but also includes in some regressions an indicator variable for each fortnight during which the data collection took place and eight variables measuring the weather at the time of travel. In the simplest specification of column 1, we estimate a (constant) elasticity of time cost of travel with respect to the number of travellers of 0.098 instead of 0.058 in the same column of table 5. We attribute the higher elasticity estimated here to the greater precision of traveller counts computed from counterfactual GM trip durations (instead of reported trip durations from the BTS) and to a reduction of the simultaneity bias.[31] Table 6 also estimates slightly lower maximum elasticities and mean elasticities around zero.

Interestingly, in table 6 the elasticity of the time cost of travel with respect to trip distance is now only about -0.2 instead of about -0.7 in table 5. The latter elasticity uses reported trip durations and

[31] When we repeat the estimation of column 1 of table 5 using again the actual time cost of travel as dependent variable and the number of travellers computed using counterfactual trip durations, we estimate a highly significant coefficient on the log number of travellers of 0.077 instead of 0.058. If instead we use the actual number of travellers and instrument it with the number of travellers computed from counterfactual travel durations, the estimated coefficient is 0.088.

Table 6: Regressions on GM time cost of travel at survey times

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
log travellers	0.098^a	1.11^b	-18.8^a	-18.5^a	-14.8^a	-13.9^a	-13.2^a	-13.1^a
	(0.041)	(0.53)	(2.96)	(2.55)	(2.23)	(2.12)	(2.06)	(2.14)
log travellers squared		-0.042^c	1.72^a	1.70^a	1.36^a	1.27^a	1.21^a	1.21^a
		(0.023)	(0.27)	(0.23)	(0.20)	(0.19)	(0.19)	(0.20)
log travellers cubed			-0.051^a	-0.051^a	-0.041^a	-0.038^a	-0.036^a	-0.036^a
			(0.0081)	(0.0069)	(0.0062)	(0.0059)	(0.0057)	(0.0061)
log trip distance				-0.21^a	-0.21^a	-0.21^a	-0.21^a	-0.21^a
				(0.0047)	(0.0043)	(0.0040)	(0.0041)	(0.0053)
Fortnight	N	N	N	Y	Y	Y	Y	Y
Mode	N	N	N	N	Y	Y	Y	Y
Weather	N	N	N	N	Y	Y	Y	Y
Purpose / # trip travellers	N	N	N	N	N	Y	Y	Y
Traveller demographics	N	N	N	N	N	N	Y	Y
Industry	N	N	N	N	N	N	N	Y
Observations	299,481	299,481	299,481	299,481	299,481	299,481	270,486	158,741
R^2	0.02	0.03	0.05	0.31	0.32	0.33	0.33	0.34
Mean marginal effect	0.098	0.044	-0.022	-0.043	-0.0086	0.0067	0.010	0.020
Max marginal effect	0.098	0.43	0.36	0.36	0.31	0.30	0.29	0.30

Notes: OLS regressions with a constant in all columns. The log time cost of travel per kilometre (i.e., minus log speed) is the dependent variable in all columns. Standard errors clustered by time intervals in parentheses. *a*, *b*, *c*: significant at 1%, 5%, 10%. There are 18 indicators for each fortnight of our data collection period. We have three trip purpose indicators: work, study, health (default is everything else). Traveller demographics include age, education, gender, and seven household income category indicators. We also have 17 industry indicators and eight weather variables (temperature, humidity, atmospheric pressure, visibility, wind speed, fog, rain, and thunderstorm.

reflects that longer trips are on average faster because the fixed cost of each trip is spread over a greater distance and because longer trips use faster roads. Instead, the elasticity of -0.2 in table 6 only reflects differences in roads taken since GM travel durations only measure time in traffic. The elasticity of -0.7 in table 5 can be compared to elasticities of -0.3 to -0.4 estimated for car travel in the US by Couture *et al.* (2016). The difference is mainly due to the importance of transit trips in Bogotá, which can involve significant waiting times.

Some columns of table 6 also include a variety of weather variables and indicator variables for the fortnight during which the data was collected. Although we do not report the details of all these coefficients, three results are noteworthy. First, the weather variables are always significant when included in the regression but their coefficients are generally small. The largest effects are unsurprisingly observed for thunderstorms during which travel is 5 to 7% slower. We also find some expected effects for the fortnights during which the observations were collected. For the two

Table 7: Regressions on GM time costs of travel at all times

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
log travellers	0.11^a	-0.23^a	-6.96^a	-6.61^a	-5.04^a	-5.01^a	-4.96^a	-5.01^a
	(0.0029)	(0.078)	(0.89)	(0.86)	(0.70)	(0.70)	(0.69)	(0.72)
log travellers squared		0.016^a	0.65^a	0.62^a	0.47^a	0.47^a	0.46^a	0.47^a
		(0.0037)	(0.085)	(0.082)	(0.067)	(0.067)	(0.066)	(0.068)
log travellers cubed			-0.020^a	-0.019^a	-0.014^a	-0.014^a	-0.014^a	-0.014^a
			(0.0027)	(0.0026)	(0.0021)	(0.0021)	(0.0021)	(0.0021)
log trip distance				-0.21^a	-0.21^a	-0.21^a	-0.21^a	-0.21^a
				(0.00013)	(0.00013)	(0.00013)	(0.00013)	(0.00018)
Fortnight	N	N	N	Y	Y	Y	Y	Y
Mode	N	N	N	N	Y	Y	Y	Y
Weather	N	N	N	N	Y	Y	Y	Y
Purpose / # trip travellers	N	N	N	N	N	Y	Y	Y
Traveller demographics	N	N	N	N	N	N	Y	Y
Industry	N	N	N	N	N	N	N	Y
R^2	0.28	0.29	0.30	0.48	0.49	0.49	0.50	0.50
Mean marginal effect	0.11	0.15	0.050	0.051	0.060	0.061	0.062	0.066
Max marginal effect	0.11	0.19	0.22	0.21	0.18	0.17	0.17	0.18

Notes: OLS regressions with a constant in all columns. 8,690,424 observations in columns 1 to 6, 7,842,426 in column 7, and 4,673,952 in column 8. The log time cost of travel per kilometre (i.e., minus log speed) is the dependent variable in all columns. Standard errors clustered by time intervals in parentheses. *a*, *b*, *c*: significant at 1%, 5%, 10%. There are 18 indicators for each fortnight of our data collection period. We have three trip purpose indicators: work, study, health (default is everything else). Traveller demographics include age, education, gender, and seven household income category indicators. We also have 17 industry indicators and eight weather variables (temperature, humidity, atmospheric pressure, visibility, wind speed, fog, rain, and thunderstorm.

fortnights from mid-June to mid-July, we estimate that the time cost of travel is about 10% less during these vacation periods.

In results not reported here, we also duplicated the specifications of table 6 and added fixed effects for each hour. These specifications estimate similar elasticities for the time cost of traffic. For instance, the specification analogous to column 1 estimates an elasticity of the time cost of travel of 0.11 instead of 0.098 while for column 2 the mean elasticity is 0.073 instead of 0.043.

Table 7 duplicates table 6 but uses all counterfactual GM trips regardless of the time of the day instead of only considering GM trips observed at the same time as the actual trip. Column 1 of table 7 estimates an elasticity of the time cost of travel with respect the number of travellers of 0.11 instead of 0.098 for the corresponding column of table 6. This difference is small. Columns 2 to 8 consider terms of higher order for the log number of travellers. The results are now more stable

Table 8: Regressions on GM time costs of travel at all times, fixed effects estimations

	(1)	(2)	(3)	(4)	(5)
Panel (A): Trip fixed effects					
log travellers	0.11^{a}	-0.21^{a}	-6.33^{a}	-6.34^{a}	-4.93^{a}
	(0.000043)	(0.00079)	(0.010)	(0.010)	(0.010)
log travellers squared		0.015^{a}	0.59^{a}	0.59^{a}	0.46^{a}
		(0.000037)	(0.00096)	(0.00096)	(0.00096)
log travellers cubed			-0.018^{a}	-0.018^{a}	-0.014^{a}
			(0.000030)	(0.000030)	(0.000030)
Fortnight	N	N	N	Y	Y
Weather	N	N	N	N	Y
R^2	0.45	0.46	0.48	0.49	0.51
Mean marginal effect	0.11	0.15	0.056	0.056	0.064
Max marginal effect	0.11	0.19	0.21	0.21	0.18
Panel (B): Trip fixed effects and hour fixed effects					
log travellers	0.034^{a}	-0.10	-1.59^{a}	-1.58^{a}	-1.55^{a}
	(0.0081)	(0.063)	(0.52)	(0.51)	(0.50)
log travellers squared		0.0065^{b}	0.15^{a}	0.15^{a}	0.15^{a}
		(0.0033)	(0.052)	(0.051)	(0.051)
log travellers cubed			-0.0045^{a}	-0.0045^{a}	-0.0044^{a}
			(0.0017)	(0.0017)	(0.0017)
Fortnight	N	N	N	Y	Y
Weather	N	N	N	N	Y
R^2	0.58	0.58	0.58	0.59	0.59
Mean marginal effect	0.034	0.053	0.032	0.032	0.033
Max marginal effect	0.034	0.071	0.070	0.070	0.070

Notes: OLS regressions with 8,690,356 observations in all columns. The log time cost of travel per kilometre (i.e., minus log speed) is the dependent variable in all columns. Standard errors clustered by time intervals in parentheses. *a*, *b*, *c*: significant at 1%, 5%, 10%. There are 18 indicators for each fortnight of our data collection period. We also have eight weather variables (temperature, humidity, atmospheric pressure, visibility, wind speed, fog, rain, and thunderstorm.

than in table 7 and we no longer estimate negative mean marginal elasticities. All the regressions involving a cubic term for the log number of travellers estimate a mean marginal elasticity of about 0.06 and a maximal marginal elasticity of about 0.2. This suggests that unobserved trip characteristics play a role in determining the results of table 6 but not for table 7.

To confirm this feature, panel (A) of table 8 reports the results of five regressions that include trip fixed effects. These fixed-effect regressions estimate elasticities of the time cost of travel which are essentially identical to the corresponding elasticities of table 7. Panel (B) of table 8 duplicates these five regressions but also includes an hour fixed effect. This panel relies only on variations in traffic speed and number of travellers within each hour and within each trip. Because variations in the number of travellers are arguably not very well measured over a short time frame, these regressions estimate a somewhat lower elasticity. These results are nonetheless reassuring since they show that the low elasticities estimated previously are not driven by unobserved hour effects.

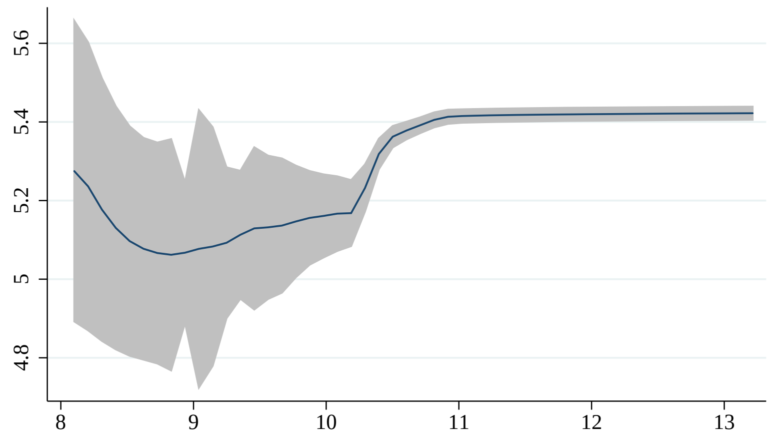
To visualise these results, figure 4 presents three non-parametric estimates of the relationship between the log time cost of travel and the log number of travellers. The figure in the first panel corresponds to the parametric estimates of table 5 which use actual trip durations to compute both the time cost of travel for each trip and the number of travellers. For this figure, we note the imprecision of the estimates for the times when very few trips occur and the initial region where the time cost of travel decreases as the number of travellers increases. The figure in the second panel corresponds to the parametric estimates of table 6 which use counterfactual GM trip durations for the times at which the actual trips were taken to compute both the time cost of travel for each trip and the log number of travellers. This figure exhibits again an initial region of declining time cost of travel. Finally, the figure in the third panel corresponds to the parametric estimates of table 7 which use counterfactual GM trip durations at all times to compute both the time cost of travel for each trip and the log number of travellers. In this figure, there is no region of decreasing time cost of travel. This is consistent with the notion that slower trips are taken at times when the number of travellers is low. This would be the case if, for instance, more central trips, which are normally slower, are over-represented late at night.

The defining feature of the third panel of figure 4 is that the time cost of travel is essentially flat at low levels of traffic and then increases in an intermediate region before reaching a high plateau. As the log number of travellers increases from about 8 to about 13, the log time cost of travel increases from about 4.65 to 5.1. This suggests an average elasticity of the time cost of travel with respect to the number of travellers of about 0.09. This is close to the estimated mean elasticity of about 0.06 estimated in tables 6 and 7. At its steepest, the slope of the curve is slightly above 0.20. This figure is also consistent with estimates of the maximum marginal effects reported in tables 6 and 7.

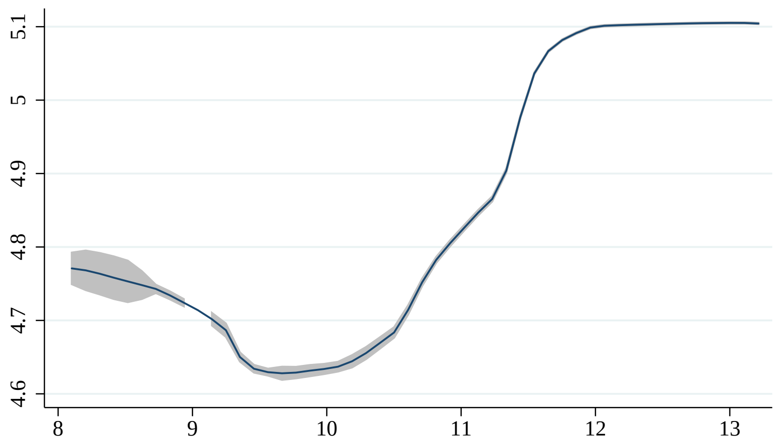
To appreciate the reality underlying these figures, we note that the GM number of travellers is multiplied by 17 between the least busy hour of the day at 2 am and 4 am.[32] At the same time, the

[32]Our analysis uses five-minute intervals. We aggregate to hourly figures for illustration purpose.

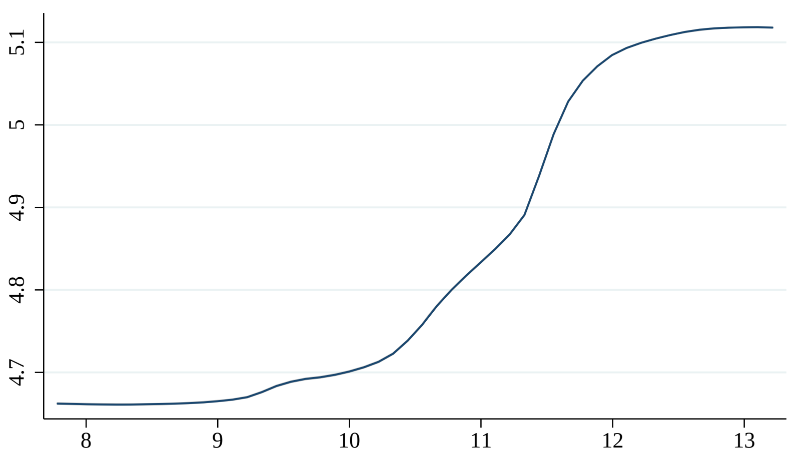
Figure 4: Non parametric estimates of the supply curve



Panel A: Actual survey trips



Panel B: Counterfactual GM trips at actual survey times



Panel C: counterfactual GM trips at all times

Note: Each graph reports the log number of travellers on the horizontal axis and log time cost of travel on the vertical axis. The time cost of travel is in seconds per kilometre. The log number of travellers is computed from our sample and extrapolated to the population of Bogotá. 95% confidence intervals in grey.

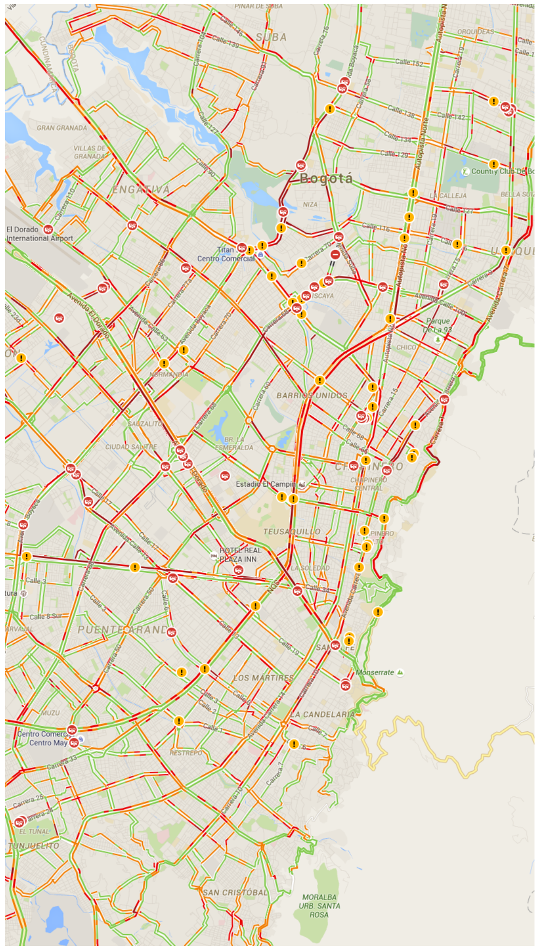
time cost of travel only increases by 4%. The arc elasticity of the time cost of travel with respect to the number of travellers implied by those figures is only 0.01. This corresponds to the initial flat in the curve of the third panel of figure 4. Then, between between the hours of 4 am and 6 am, the number of travellers is multiplied by about 7 while the time cost of travel increases by 40%. The arc elasticity implied by those figures is 0.17. Similarly, at 9 pm, the number of travellers is only about a quarter of what it is at 6 pm and the time cost of travel is about 25% less, implying an arc elasticity of 0.21. These two changes take place along the steep part of the curve of the third panel of figure 4. To understand the second flat of the curve for high numbers of travellers, we note that the number of travellers declines by about 10% between the hours of 5 pm and 6 pm while the time cost of travel actually increases slightly.[33]

The maximum elasticity of the time cost of travel that we estimate at about 0.20 is well below extant estimates from the literature for a single road or a set of roads (Small and Verhoef, 2007) and much below the unit elasticity needed to generate hypercongestion. That the time cost of travel plateaus above a certain threshold is also in sharp contrast with results from a single road indicating that traffic will reach a complete standstill beyond a threshold. Since the pattern in the first panel of figure 4 which uses survey data directly is broadly the same as in the last panel, these two differences are not driven by our identification strategy that relies on the use of counterfactual travel data. Instead, we attribute these two differences to the fact that we consider a broad area instead of a single road or a small group of roads. Simply put, there is always a local street that can be taken when the highway is congested. As a result, the time cost of travel never becomes infinite for the numbers of travellers that we observe.

To illustrate this argument, figure 5 shows a GM representation of traffic conditions in Bogotá on Monday, 2nd of May 2016 at 6 pm. This is a fairly low resolution picture showing the main arteries for a large part of the city. The prevalence of red on these arteries indicates heavy traffic. The two panels of figure 6 zoom onto the two main centres of Bogotá which are arguably highly congested. Interestingly, the higher resolution of these pictures reveals that, while the main arteries are extremely busy, lots of local streets are still green or light orange indicating lower levels of traffic. Put differently, the time cost of travel is bounded from above by the existence of local

[33] There is also a small increase in the time cost of travel between 6 am and 7 am despite a small decrease in the number of travellers. The hour of 6 am (from 6.00 to 6.59) is the busiest since the work day often starts at 7 am.

Figure 5: Traffic at 6 pm, low resolution

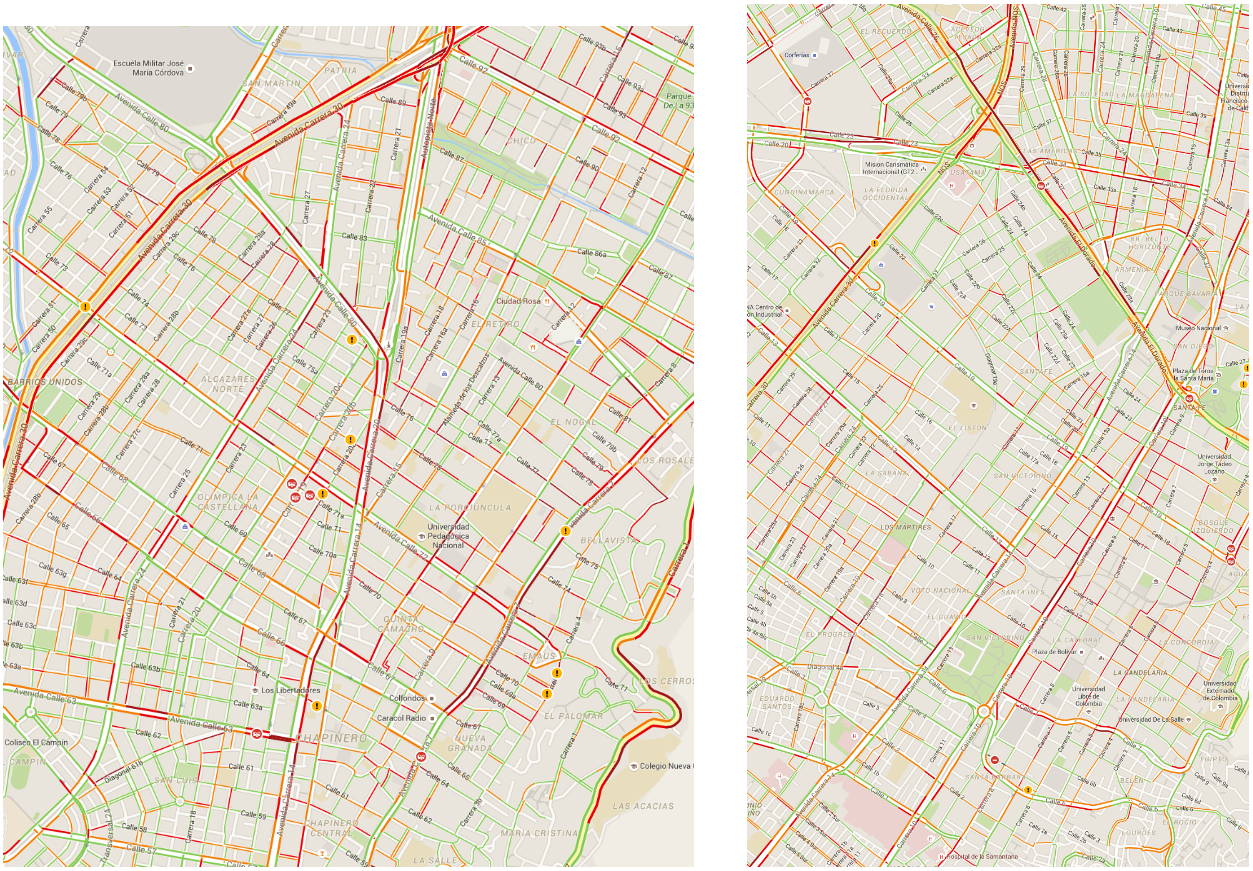


Notes: Google Maps screen capture taken on Monday, 2nd of May 2016 at 6 pm. Higher levels of congestion as the colour ranges from green to amber, to red. The screen capture represents most of the city of Bogotá.

streets which do not get clogged for extant levels of traffic.[34]

Finally, we can assess how the time cost of travel varies with the number of travellers in different parts of Bogotá. We use the conventional division of Bogotá into 20 districts but group the smaller districts together and end up with 12 areas. To compute the number of travellers in each area, we use detailed driving instructions obtained from GM. For the first itinerary returned by each query, we computed the fraction of each trip taking place in each area. The main limitation to our data construction here is that we attribute the average time cost of travel of the entire trip to each area. We use this average that may straddle several areas to compute time cost of travel and the number of travellers in each area. This affects our dependent variable, the time cost of travel for each trip

Figure 6: Traffic at 6 pm, high resolution



Panel A: Historical centre

Panel B: Northern CBD

Notes: Google Maps screen capture taken on Monday, 2nd of May 2016 at 6 pm. Higher levels of congestion as the color ranges from green to amber, to red.

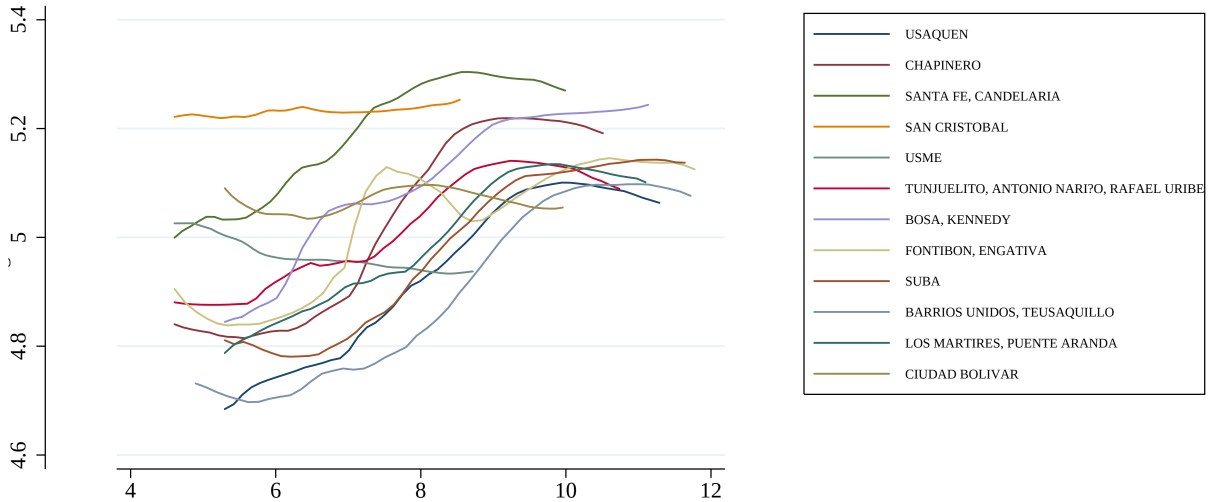
in a given area, as well as the dependent variable of interest, the number of travellers.

With these caveats in mind, the non-parametric estimate of relationship between the log time cost of travel and the log number of travellers for each area is represented in figure 7. This figure is analogous to figure 4 panel C for each area of Bogotá. While figure 7 is more noisy, the patterns are generally similar with the same S-shape evolution of the time cost of travel as the number of travellers increases. Although there are differences in the mean time cost of travel across areas, the elasticity of the time costs of travel with respect to the number of travellers for most areas is of the same magnitude as for the entire city.

The differences across areas are also interesting. The two areas for which we observe a flat or even a decreasing pattern are Usme and San Cristobal. These are two sparsely populated districts

[34] Arguably, increasing traffic density well beyond the highest levels we observe should lead to further increases in the time cost of travel.

Figure 7: Non-parametric estimates of the supply curve by district



The graph reports the log number of travellers on the horizontal axis and log time cost of travel on the vertical axis. Both are computed using counterfactual travel durations from GM. The time cost of travel is in seconds per kilometre.

at the extreme South of Bogotá where very few people live. We also note that the slowest area (Santa Fe, Candelaria) is the union of the two localities that form the historical centre of Bogotá with extremely narrow streets and short blocks. On the other hand, the northern districts of Suba, Usaquén, and Barrios Unidos, for which the residential density is lower and roads wider, are also the districts where travel is the fastest.

5.2 *Demand*

Table 9 reports our main estimation results for the demand equations (4) and (5). An observation is a counterfactual instance of a trip. At the time of this counterfactual, we measure whether the traveller surveyed by the BTS was travelling or not at the same time on the day of the survey. We use the resulting indicator variable as the dependent variable. We regress it on the log time cost of travel at the time of the query, an indicator variable for each hour of the day, and a set of control variables that varies across columns. As previously the time cost of travel is per unit of distance.

In panel A of table 9, we use all 24 hours of the day and estimate a different intercept for each hour. We do not report the hour indicators in the table. We instead provide a graph below. The estimated coefficient for the log time cost of travel in column 1 is -0.30. It implies that a traveller is 0.3% less likely to travel when the time cost of travel is 1% higher. Thus, this traveller will travel 1.3% fewer kilometres. This first specification also controls for log trip distance. This is perhaps

Table 9: Demand estimations, Poisson regressions

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel (A): All hours of the day							
log time cost	-0.30^a	-0.30^a	-0.29^a	-0.30^a	-0.29^a	-0.30^a	-0.41^a
	(0.078)	(0.079)	(0.079)	(0.083)	(0.082)	(0.091)	(0.097)
log trip distance	-0.17^a	-0.17^a	-0.15^a	-0.17^a	-0.15^a	-0.16^a	-
	(0.029)	(0.028)	(0.026)	(0.028)	(0.026)	(0.032)	
Hour indicators	Y	Y	Y	Y	Y	Y	Y
Mode	N	Y	Y	Y	Y	Y	-
Purpose / # trip travellers	N	N	Y	N	Y	Y	-
Traveller demographics	N	N	N	Y	Y	Y	-
Industry	N	N	N	N	N	Y	-
Trip fixed effect	N	N	N	N	N	N	Y
Observations	8,530,445	8,530,445	8,530,445	7,701,122	7,701,122	4,590,469	8,255,078
Panel (B): Day-time, from 7 am to 7 pm							
log time cost	-0.35^a	-0.35^a	-0.34^a	-0.34^a	-0.33^a	-0.34^a	-0.44^a
	(0.062)	(0.062)	(0.062)	(0.061)	(0.061)	(0.071)	(0.061)
log trip distance	-0.15^a	-0.15^a	-0.14^a	-0.16^a	-0.15^a	-0.15^a	-
	(0.028)	(0.027)	(0.027)	(0.028)	(0.027)	(0.035)	
Observations	3,771,485	3,771,485	3,771,485	3,375,751	3,375,751	2,027,162	3,774,458
Panel (C): Night-time, from 7 pm to 7 am							
log time cost	-0.20	-0.18	-0.17	-0.19	-0.19	-0.24	-0.42
	(0.21)	(0.22)	(0.23)	(0.26)	(0.26)	(0.30)	(0.30)
log trip distance	-0.12^b	-0.11^b	-0.085^b	-0.11^b	-0.086^b	-0.095^c	-
	(0.051)	(0.048)	(0.037)	(0.047)	(0.040)	(0.054)	
Observations	1,732,859	1,732,859	1,732,859	1,591,643	1,591,643	950,241	1,632,402
Panel (D): All hours of the day, time cost of travel in level							
Time cost	-0.0014^a	-0.0014^a	-0.0014^a	-0.0015^a	-0.0014^a	-0.0014^a	-0.0018^a
	(0.00039)	(0.00039)	(0.00039)	(0.00040)	(0.00040)	(0.00045)	(0.00048)
log trip distance	-0.16^a	-0.16^a	-0.14^a	-0.17^a	-0.15^a	-0.15^a	-
	(0.029)	(0.028)	(0.026)	(0.028)	(0.026)	(0.032)	
Observations	8,530,445	8,530,445	8,530,445	7,701,122	7,701,122	4,590,469	8,255,078
Mean elasticity	-0.23	-0.23	-0.22	-0.23	-0.22	-0.22	-0.32

Notes: Poisson regressions with a constant in all columns. Standard errors clustered by time intervals in parentheses. *a*, *b*, *c*: significant at 1%, 5%, 10%. An indicator for the travelling status is the dependent variable. The mode indicators are for taxi and road transit (default private vehicle). We have three purpose indicators: work, study, health (default is everything else). Traveller demographics include age, education, gender, and seven household income category indicators. We also have 17 industry indicators.

the most important characteristic of a trip that is likely to be correlated with the time at which travellers travel and thus with the time cost of travel.[35]

[35] In the US, commutes are longer than most other categories of trips and bunched at certain hours of the day (Couture *et al.*, 2016).

Column 2 of table 9 duplicates the specification of column 1 adding indicators for different modes of road travel. The coefficients for taxi and road transit are not reported in the table. They are tiny at about 0.01 and insignificant in this specification so that they do not affect our conclusions. Columns 3 to 6 further add the purpose of the trip, the number of travellers on this trip, a range of traveller demographics, and the industry of employment of the traveller. Throughout these columns, the estimated coefficient for the log time cost of travel is extremely stable and only varies between -0.29 and -0.30 despite the addition of a wide array of controls. Column 7 duplicates again the same regression, including this time a trip fixed effect. The estimated coefficient on the time cost of travel is now slightly larger in magnitude at -0.41. This is our preferred estimate.

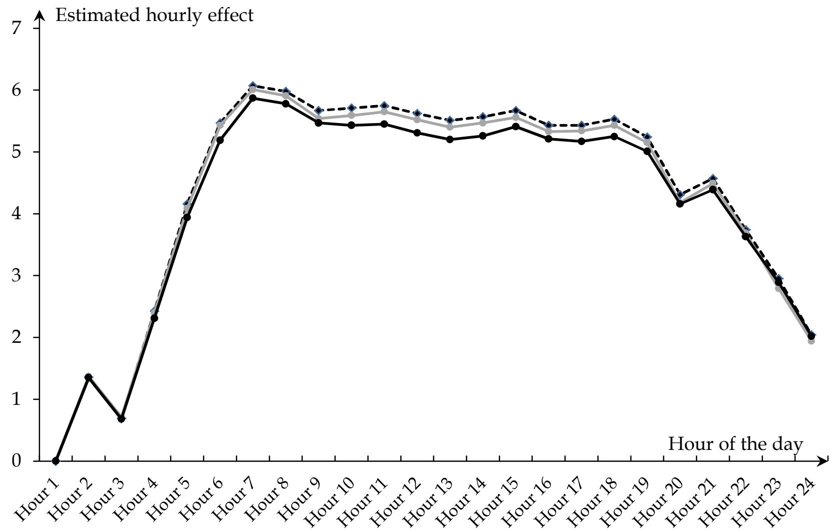
Panel (A) of table 9 also reports a coefficient on log trip distance. Again, this coefficient is extremely stable and only varies between -0.15 and -0.17. Although longer trips take proportionately longer (after conditioning for the time cost of travel), travellers at any point in time are less likely to be in such a trip. The coefficients on log trip distance in this panel imply an elasticity of the number of trips with respect to trip distance between -1.15 and -1.17. This is consistent with the gravity patterns of urban travel which have been evidenced in other contexts (e.g., Ahlfeldt, Redding, Sturm, and Wolf, 2015) and are heavily relied upon by traffic generation modelling (Erlander and Stewart, 1990).

Figure 8 plots the estimated hour effects for three columns of panel (A) of table 9. Again, the results are extremely stable across the three sets of estimates that are plotted. They indicate an early morning peak in demand followed by a high plateau throughout the day followed by a steep decline after 7 pm.

The elasticity of the demand for travel may vary across the day and this would affect the estimates of the hour intercepts. Panels (B) and (C) of table 9 explore this possibility. The cross-sectional estimates of columns 1 to 6 indicate an elasticity slightly larger in magnitude for daytime travel and somewhat smaller for nighttime travel. Because of fewer trips and less variation in the time cost of travel, the night coefficients estimated in panel (C) are insignificant.

Panel (D) of table 9 duplicates again panel (A) but includes the time cost of travel in level instead of in log. Column 1 estimates a coefficient of -0.0014 for the time cost of travel. This figure corresponds to a mean elasticity of -0.23 instead of -0.30 when the time cost of travel is included in log in the corresponding specification of panel (A). The mean elasticities estimated in columns

Figure 8: Estimated hour effects for weekday travel



The dashed black line represents the hour effects estimated in column 1 of panel (A) of table 9. The plain grey line and the plain black line represent the same hours effects estimated from columns 5 and 7 of the same panel, respectively. The first hour of the day is used as a reference and the third hour of the day is obtained by interpolation. The standard errors are not represented but all vary between 0.64 and 0.84.

Table 10: Demand estimation, robustness checks, Poisson regressions

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
log time cost	-0.29[a]	-0.26[a]	-0.083[a]	-0.15[c]	-0.17[a]	-0.15[a]	-0.26[a]	-0.33[a]
	(0.083)	(0.051)	(0.054)	(0.083)	(0.041)	(0.048)	(0.054)	(0.055)
Hour indicators	1 hour	-	30mn	30mn	30mn	30mn	30mn	30mn
Time window	day	day	w. 30mn	w. 60mn	w. 120mn	w. 30mn	w. 60mn	w. 120mn
District indicator	Y	by 30mn	N	N	N	N	N	N
Trip fixed effect	N	N	N	N	N	Y	Y	Y
Observations	7,453,362	7,701,068	1,979,261	2,219,688	2,695,049	2,187,332	2,453,277	2,968,248

Notes: Standard errors clustered by time intervals in parentheses. *a*, *b*, *c*: significant at 1%, 5%, 10%. An indicator for the travelling status is the dependent variable. All regressions include log trip distance, traveller demographics, number of travellers on the trip, and indicators for mode and trip purpose. The mode indicators are for taxi and road transit (default private vehicle). We have three purpose indicators: work, study, health (default is everything else). Traveller demographics include age, education, gender, and seven household income category indicators.

2 to 6 are very similar while, as previously, the fixed effect estimation of column 7 yields modestly different results.

Table 10 reports results for a number of robustness checks. Column 1 duplicates the estimation of column 5 of table 9 and adds a specific intercept for each district within which a trip primarily takes place. The coefficient on the log time cost is unchanged. Column 2 allows for the time indicators to vary across districts for every 30 minute period. This only leads to a coefficients

Table 11: Demand estimation, robustness checks, logit and negative binomial regressions

	(1) All day	(2) All day	(3) 7am-7pm	(4) All day	(5) All day	(6) All day
Panel (A): Logit estimations						
log time cost	-0.31^{a}	-0.31^{a}	-0.35^{a}	-0.27^{a}		
	(0.083)	(0.083)	(0.065)	(0.12)		
Time cost					-0.0015^{a}	-0.0015^{a}
					(0.00041)	(0.00047)
log trip distance	-0.18^{a}	-0.16^{a}	-0.16^{a}	-	-0.17^{a}	-0.16^{a}
	(0.030)	(0.027)	(0.029)		(0.030)	(0.033)
Hour indicators	Y	Y	Y	Y	Y	Y
Mode	N	Y	Y	-	N	Y
Purpose / # trip travellers	N	Y	Y	-	N	Y
Traveller demographics	N	Y	Y	-	N	Y
Trip fixed effect	N	N	N	Y	N	N
Observations	8,255,144	8,255,144	3,375,708	6,509,121	8,255,144	8,255,144
Mean elasticity	-0.31	-0.31	-0.35	-0.27	-0.23	-0.23
Panel (B): Negative binomial estimations						
log time cost	-0.30^{a}	-0.29^{a}	-0.33^{a}	-0.25^{a}		
	(0.078)	(0.079)	(0.061)	(0.11)		
Time cost					-0.0014^{a}	-0.0014^{a}
					(0.00039)	(0.00039)
Mean elasticity	-0.30	-0.29	-0.33	-0.25	-0.23	-0.22

Notes: Standard errors clustered by time intervals in parentheses. *a*, *b*, *c*: significant at 1%, 5%, 10%. An indicator for the travelling status is the dependent variable. The mode indicators are for taxi and road transit (default private vehicle). We have three purpose indicators: work, study, health (default is everything else). Traveller demographics include age, education, gender, and seven household income category indicators.

modestly smaller in magnitude. Rather than allowing the time cost of a trip across the entire day, columns 3 to 5 restrict the window of observation to 30, 60, and 120 minutes around the actual time of the trip. The coefficients on the time cost are now smaller in magnitude because the observed variation in the time cost of travel is arguably more noisy over a short time window. Columns 6 to 8 repeat the same exercise but also impose a trip fixed effect. We note that considering a two-hour window leads to estimates close to those obtained from observations taken over the entire day.

Finally, table 11 duplicates some of our previous specification but uses a logit estimator in panel (A) and a negative binomial estimator in panel (B). A comparison with the results of the Poisson estimations of table 9 indicates that the exact econometric estimation technique used to estimate the effects of the time cost of travel on travel status does not matter.

6. The deadweight loss of congestion

To compute the deadweight loss of congestion and the optimal congestion tax, we first provide a calculation under simplified functional forms which assume constant elasticities for both demand and supply and, second, a calculation which uses the more general functional forms that we estimate in some of our specifications. This latter calculation is more precise but less transparent black-box. Constant elasticities yield instead closed-form solutions and provide straightforward intuitions.

Returning to the framework of section 3, on the supply side, the time cost of travel per unit of distance at time t is given by:

$$C_t^S = \underline{C} \times (N_t^S)^{\sigma}\,, \tag{6}$$

where σ is the elasticity of the time cost of travel with respect to the number of travellers estimated above and $\underline{C}$ is the time cost of travel when no one travel, that is under 'free-flow' or uncongested travel.

On the demand side, we have:

$$N_t^D = \overline{B}_t \times (C_t^D)^{-\eta}\,, \tag{7}$$

where minus η is the demand elasticity of the number of travellers with respect to the time cost of travel and $\overline{B}_t$ is a time-varying demand shifter.

Distance travelled at time t (i.e., the flow of travel) is equal to $V_t = N_t/C_t^S$. In equilibrium, we have $C_t^S = C_t^D$ and $N_t^S = N_t^D$. Using equations (6) and (7), we find:

$$C_t^{eq} = \overline{B}_t^{\frac{\sigma}{1+\eta\sigma}} \underline{C}^{\frac{1}{1+\eta\sigma}}\,, \qquad N_t^{eq} = \overline{B}_t^{\frac{1}{1+\eta\sigma}} \underline{C}^{\frac{-\eta}{1+\eta\sigma}}\,, \text{ and } \qquad V_t^{eq} = \overline{B}_t^{\frac{1-\sigma}{1+\eta\sigma}} \underline{C}^{-\frac{1+\eta}{1+\eta\sigma}}\,. \tag{8}$$

At the equilibrium, N_t^{eq} travellers face a time cost of travel C_t^{eq} and the flow is given by V_t^{eq}.

To compute the optimum, we need to use the marginal cost curve associated with the (average) cost curve represented by equation (6). As argued in section 3, $MC_t^S = (1+\sigma)C_t^S$. Straightforward calculations using equation (8) imply that the optimum is described by:

$$MC_t^{opt} = (1+\sigma)C^{opt} = (1+\sigma)^{\frac{1}{1+\eta\sigma}} C_t^{eq}\,, \qquad N_t^{opt} = (1+\sigma)^{\frac{-\eta}{1+\eta\sigma}} N_t^{eq}\,, \quad \text{and} \tag{9}$$

$$V_t^{opt} = \frac{N_t^{opt}}{C_t^{opt}} = (1+\sigma)^{-\frac{\eta(1-\sigma)}{1+\eta\sigma}} V_t^{eq}\,. \tag{10}$$

At the optimum, N_t^{opt} travellers face a time cost of travel MC_t^{opt} and the flow is given by $V_t^{opt} = \frac{N_t^{opt}}{C_t^{opt}}$. The difference between the marginal time cost of travel and its average at the optimum is the optimal Pigovian tax τ:

$$\tau_t = MC^{opt} - C^{opt} = \sigma C^{opt} = \sigma(1+\sigma)^{\frac{-\eta\sigma}{1+\eta\sigma}} C_t^{eq} \,. \tag{11}$$

We value time travelling at half the travellers' wage as is customary in transportation economics (Small, 2012), assume $\sigma = 0.06$ and $\eta = 0.40$ for our baseline case, and consider a speed of 20 kilometres per hour (i.e., $C^{eq} = 0.05$). $\sigma = 0.06$ corresponds to our preferred mean supply elasticity in columns 3-5 of panel (A) of table 8. This said, we can consider, elasticities as high as $\sigma = 0.20$ to match the maximum elasticity of the supply curve in figure 4 panel (C). On the demand side, $\eta = 0.40$ corresponds approximately to our preferred elasticity estimated over the entire day. Because we worry that our approach may underestimate the magnitude of the true elasticity of demand, we also consider higher values of η. A travel speed of 20 kilometers per hour corresponds approximately to the travel speed that occurs in Bogotá during the morning and evening rushes.

For a Colombian wage of 3 US$ per hour, the values of $\sigma = 0.06$, $\eta = 0.40$, and $C = 0.05$ imply an optimal tax of slightly less than a half a penny per kilometre.[36] In the US, for a typical wage of 24 US$ per hour, the optimal congestion tax would be about 3.6 cents per kilometre. This is close to the optimal congestion tax suggested by Couture *et al.* (2016) who use cross-city variation. As can be seen by inspection from equation (11), this tax is fairly insensitive to η and, to a first approximation, proportional to σ. So even with $\sigma = 0.20$, the optimal congestion tax is only about 1.5 cent per kilometre for a Colombian wage of 3 US$ per hour with $\eta = 0.40$ and 1.25 cent per kilometer with a fully elastic demand($\eta \to +\infty$).

Next, we can compute the deadweight loss of congestion at time t. While equations (6) and (7) express C and N as a function of each other, the deadweight loss of congestion must be computed in the (V,C) plane as made clear by figure 3. With a slight abuse of notations, the deadweight loss of congestion is given by:

$$DWL_t = \int_{V_t^{opt}}^{V_t^{eq}} \left(MC_t^S(V) - C_t^D(V) \right) \mathrm{d}V \,. \tag{12}$$

Given that this deadweight loss is already an approximation using constant elasticities and that the objective is to provide some clear intuitions about what determines the magnitude of the deadweight loss, we start by approximating the deadweight loss of congestion in figure 3 with

[36] According to the Colombian statistical institute (DANE), the average wage in Bogotá in 2014 was 1.34 million pesos per month. At 2,900 pesos per dollar, 4.5 weeks per month, and 48 hours of work per week, that implies 2.15 dollars per hour.

the triangle ABE. Using equations (9)-(10) and the definition of V, we can write:

$$DWL_t \approx \frac{\sigma C_t^{eq}}{2} \times (V_t^{eq} - V_t^{opt}) = \frac{\sigma}{2}(1-\xi)\, N_t^{eq}\,, \tag{13}$$

where $\xi \equiv (1+\sigma)^{-\frac{\eta(1-\sigma)}{1+\eta\sigma}}$.

For our benchmark values of $\sigma = 0.06$. $\eta = 0.4$, we find that the deadweight loss of congestion represents about 0.06% of travel time. This figure seems extremely small. The formula in equation (13) shows clearly that the triangle ABE is given by the wedge σ multiplied by the proportion of excessive travel $(1-\xi)$ and the number of travellers divided by 2. For 'small' values of $\eta \times \sigma$, excessive travel is approximated by $\eta\sigma$ so that the deadweight loss is approximately a proportion $\eta\sigma^2/2$ of travel time. From the data, an average person in the BTS takes 1.15 motorised trips per weekday of an average duration of 54 minutes per trip. Ignoring that a good part of these 54 minutes per trip may not be spent in traffic, the deadweight loss associated with $\sigma = 0.06$. $\eta = 0.40$ represents only about a few seconds daily.

If instead of taking a triangular approximation of the deadweight loss as in equation (13) we compute the exact integral, simple but tedious calculations lead to:

$$DWL_t = \left[(1+\sigma)(1-\sigma)\left(1-\xi^{\frac{1}{1-\sigma}}\right) - \frac{\eta(1-\sigma)}{\eta-1-\eta\sigma}\left(1-\xi^{\frac{\eta-1-\eta\sigma}{\eta(1-\sigma)}}\right)\right] N_t^{eq}\,. \tag{14}$$

For our benchmark values of $\sigma = 0.06$. $\eta = 0.4$, the exact value of the deadweight loss is again 0.06%. The true deadweight loss is larger than the approximation given equation (13) by about 1% due to the concavity of the demand curve and the convexity of the supply curve. This difference is economically negligible.

An alternative to our calculation, which assumes a constant supply elasticity, is to compute the deadweight loss of congestion for each hour of the day using the non-linear supply estimates of column 5 of panel (A) of table 8. With that more exact approximation, the loss is now less than 0.1%. The reason is that most of the travel in Bogotá occurs at busy hours for which the supply elasticity σ is close zero.

We now consider the deadweight loss of congestion under alternative values of σ and η. If we retain $\sigma = 0.06$, but consider $\eta = 1$ instead of 0.4, the deadweight loss is now 0.15% instead of 0.06%. This is unsurprising since the deadweight loss of congestion is roughly proportional to η for 'small' values of $\eta\,\sigma$. Assuming a much more elastic demand with $\eta = 10$ implies a deadweight loss of 0.92%. As η becomes large, excessive travel can be approximated by σ (provided this latter quantity remains small). This approximation implies a deadweight loss of congestion of about

$\sigma^2/2$ or 1.8% for $\sigma = 0.06$. The exact integral implies a slightly larger value of 2.05% because of the concavity of the supply curve. With the same amount of travel as before, this represents slightly more than a minute every day.

Relying again on the intuition given above, the deadweight loss of congestion is more sensitive to σ. With $\sigma = 0.2$ and returning to $\eta = 0.4$, the deadweight loss is now 0.54% of the time spent in travel.[37] This is much more than 0.06% of course but this still only about 20 seconds per day for an average traveller. If we keep $\sigma = 0.2$ and assume a perfectly elastic demand, the deadweight loss now represents 5.65% of travel time. This is now 3 minutes and 30 seconds per day for a typical traveller. If we value travel time at half the wage again, this loss represents only about 0.37% of a daily wage. Even under those arguably extreme assumptions, the deadweight loss still remains small.

Making even extreme assumptions, we now impose $\sigma = 0.6$ and a perfectly elastic demand. The deadweight loss now grows to 7.9% of travel time. However, this only represents about 0.5% of a daily wage. While these assumptions still only involve a small loss, they are implausible. The demand for travel is arguably far from being perfectly elastic and $\sigma = 0.6$ would imply a fifteenfold increase of the time cost of travel when the number of travellers is multiplied by 100 between the least and most travelled hours of the day as it is the case in Bogotá. In the data, we only measure a 70% difference in the time cost of travel between the fastest and slowest hours.

7. Conclusion

This paper proposes a novel approach to assess the deadweight loss of congestion in a city. We implement this approach for road travel in the city of Bogotá using information from a local travel survey and counterfactual travel data generated from Google Maps. For the supply of travel, we find that the elasticity of the time cost of travel per unit distance with respect to the number of travellers is on average about 0.06. This elasticity is close to zero at low levels of traffic, then reaches a maximum magnitude of about 0.20 as traffic builds up and becomes small again at high levels of traffic. While this finding for an entire area is in sharp contrast with extant results for specific

[37] This type of result and the underlying quantitative intuition are not very different from the traditional results obtained when estimating the deadweight loss of monopoly. Since Harberger (1954), economists have consistently estimated deadweight losses from monopolies of the order of 1% to 2% of GDP and often much less. For instance, a monopoly facing a price elasticity of -6 optimally prices 20% above marginal cost to maximise its profits. This leads to a reduction in sales of about 66% relative to marginal cost pricing. The welfare loss is then approximately (20%×66%/2=) 6.6%. In our case, we obtain a similar loss with a wedge of 20% between the average cost of travel that drivers pay and the marginal cost of travel for a demand elasticity η of about 4.

road segments, we explain it by the existence of local streets which remain relatively uncongested even at peak hours. We also estimate an elasticity of the number of travellers with respect to the time cost of travel of about -0.40. Although road travel is costly in Bogotá and Bogotá is a 'highly congested' city following a popular terminology, our findings point to a small deadweight loss from congestion. To put it simply, w hen demand is strong and supply is limited, the price is high. That does not mean that the allocation is deeply inefficient.

We make several contributions to the literature. First, to provide a plausible estimate of the speed-density curve for urban travel, we tackle both the issue of trip selection and the simultaneous determination of the time cost of travel and the number of travellers. Second, our investigation of the effect of the time cost of travel on the demand for travel is also largely novel. Third, our approach also innovates by harnessing the power of 'big data' to explore a core economic question.

Beyond these academic contributions, we also provide important practical guidance for urban transportation policy. Our approach delivers the optimal congestion tax per unit of distance for each hour of the day in Bogotá. Our estimate of the deadweight loss of congestion is also indicative of the maximal gain that could be expected from curing the city from congestion. This said, given our low estimates of the deadweight loss of congestion, one might be tempted to turn away from traffic congestion and the regulation of the demand for traffic.

That would be wrong on three counts. First, we observe excess travel. Although the deadweight loss of congestion is fairly small, the demand for urban travel appears somewhat elastic. As a result, only a fairly low congestion tax is needed to achieve optimality. Second, our analysis examines only one externality associated with congestion: slower travel. In addition, congestion is by no means the only negative external cost associated with road travel (Parry, Walls, and Harrington, 2007). Importantly, we expect motorised traffic to increase pollution. This is a crucial issue for Bogotá. Excessively high levels of traffic also have other social costs such as noise, accidents, etc. We hope that future research will provide a more complete assessment of these costs.

Finally, travel in Bogotá is costly and will remain so even under optimal congestion pricing. There are potentially large gains in consumer surplus from policies that lower the cost of travel on the supply side. Such policies may involve a better traffic management, the development of the roadway, a de-concentration of the city, other land-use planning interventions, the development of non-roadway forms of transportation such as subways, etc. A detailed examination of the merits of these policies should also be a priority for future research.

References

Adler, Martin, Federica Liberini, Antonio Russo, and Jos van Ommeren. 2016. Road congestion and public transit: Evidence from Rome. Processed, Free University of Amsterdam.

Ahlfeldt, Gabriel M., Stephen J. Redding, Daniel M. Sturm, and Nikolaus Wolf. 2015. The economics of density: Evidence from the Berlin Wall. *Econometrica* (83):2127–2189.

Anderson, Michael L. 2014. Subways, strikes, and slowdowns: The impacts of public transit on traffic congestion. *American Economic Review* 104(9):2763–2796.

Ardekani, Siamak and Robert Herman. 1987. Urban network-wide traffic variables and their relations. *Transportation Science* 21(1):1–16.

Arnott, Richard, André De Palma, and Robin Lindsey. 1993. A structural model of peak-period congestion: A traffic bottleneck with elastic demand. *American Economic Review* 83(1):161–179.

Arnott, Richard and Eren Inci. 2010. The stability of downtown parking and traffic congestion. *Journal of Urban Economics* 68(3):260–276.

Beckmann, Martin, C. B. McGuire, and Christopher B. Winsten. 1956. *Studies in the Economics of Transportation*. New Haven, CT: Yale University Press for the Cowles Commission for Research in Economics.

Campbell, John Y., Stefano Giglio, and Parag Pathak. 2011. Forced sales and house prices. *American Economic Review* 101(5):2108–2131.

Chay, Kenneth Y. and Michael Greenstone. 2005. Does air quality matter? Evidence from the housing market. *Journal of Political Economy* 113(2):376–424.

Combes, Pierre-Philippe and Laurent Gobillon. 2015. The empirics of agglomeration economies. In Gilles Duranton, Vernon Henderson, and William Strange (eds.) *Handbook of Regional and Urban Economics*, volume 5A. Amsterdam: Elsevier, 247–348.

Coscia, Michele, Frank Neffke, and Eduardo Lora. 2016. Report on the poblacion flotante of Bogota (D.C.). Processed, Center for International Development, Harvard University.

Couture, Victor, Gilles Duranton, and Matthew A. Turner. 2016. Speed. *Review of Economics and Statistics* Accepted for publication.

Davis, Lucas W. 2008. The effect of driving restrictions on air quality in Mexico City. *Journal of Political Economy* 116(1):38–81.

Dewees, Donald N. 1979. Estimating the time costs of highway congestion. *Econometrica* 47(6):1499–1512.

Duranton, Gilles. 2015. A proposal to delineate metropolitan areas in Colombia. *Economia & Desarrollo* 75(0):169–210.

Duranton, Gilles and Matthew A. Turner. 2011. The fundamental law of road congestion: Evidence from US cities. *American Economic Review* 101(6):2616–2652.

Duranton, Gilles and Matthew A. Turner. 2016. Urban form and driving:Evidence from US cities. Processed, Wharton School, University of Pennsylvania.

Erlander, Sven and Neil F. Stewart. 1990. *The Gravity Model in Transportation Analysis: Theory and Extensions*. Utrecht: VSP.

Ewing, Reid and Robert Cervero. 2010. Travel and the built environment: A meta-analysis. *Journal of the American Planning Association* 76(3):265–294.

Fosgerau, Mogens and Kenneth A. Small. 2013. Hypercongestion in downtown metropolis. *Journal of Urban Economics* 76(0):122–134.

Geroliminis, Nikolas and Carlos F. Daganzo. 2008. Existence of urban-scale macroscopic fundamental diagrams: Some experimental findings. *Transportation Research Part B: Methodological* 42(9):759–770.

Gomez-Ibanez, Jose A. 1996. Big-city transit rider snip, deficits, and politics: Avoiding reality in Boston. *Journal of the American Planning Association* 62(1):30–50.

Graham, Daniel J. and Stephen Glaister. 2002. The demand for automobile fuel: A survey of elasticities. *Journal of Transport Economics and Policy* 36(1):1–25.

Haight, Frank A. 1963. *Mathematical Theories of Traffic Flow*. New York (NY): Academic Press.

Harberger, Arnold C. 1954. Monopoly and resource allocation. *American Economic Review* 44(2):77–87.

Ioannides, Yannis M. and Giorgio Topa. 2010. Neighborhood effects: Accomplishments and looking beyond them. *Journal of Regional Science* 50(1):343–362.

Keeler, Theodore E. and Kenneth A. Small. 1977. Optimal peak-load pricing, investment, and service level on urban expressways. *Journal of Political Economy* 85(1):1–25.

Parry, Ian W. H. and Kenneth A. Small. 2009. Should urban transit subsidies be reduced? *American Economic Review* 99(3):700–724.

Parry, Ian W.H., Margaret Walls, and Winston Harrington. 2007. Automobile externalities and policies. *Journal of Economic Literature* 45(2):1335–1353.

Pigou, Arthur C. 1920. *The Economics of Welfare*. London (UK): Macmillan and Company.

Rossi-Hansberg, Esteban, Pierre-Daniel Sarte, and Raymond Owens III. 2010. Housing externalities. *Journal of Political Economy* 118(3):485–535.

Safirova, Elena, Kenneth Gillingham, and Sébastien Houde. 2007. Measuring marginal congestion costs of urban transportation: Do networks matter? *Transportation Research Part A: Policy and Practice* 41(8):734–749.

Samad, Taimur, Nancy Lozano-Gracia, and Alexandra Panman. 2012. *Colombia Urbanization Review. Amplifying the Gains from the Urban Transition*. Washington (DC): World Bank.

Schrank, David, Bill Eisele, Tim Lomax, and Jim Bak. 2015. Urban mobility scorecard 2015. Technical report, Texas Transportation Institute.

Secretaría Distrital de Movilidad. 2013. Informe de indicadores. Encuesta de movilidad de Bogotá 2011. Report, Alcalía Mayor de Bogotá (DC).

Small, Kenneth A. 2012. Valuation of travel time. *Economics of transportation* 1(1):2–14.

Small, Kenneth A. and Xuehao Chu. 2003. Hypercongestion. *Journal of Transport Economics and Policy* 37(3):319–352.

Small, Kenneth A. and Erik T. Verhoef. 2007. *The Economics of Urban Transportation*. New York (NY): Routledge.

Vickrey, William S. 1963. Pricing in urban and suburban transport. *American Economic Review (Papers and Proceedings)* 53(2):452–465.

Vickrey, William S. 1969. Congestion theory and transport investment. *American Economic Review (Papers and Proceedings)* 59(2):251–260.

Walters, Alan A. 1961. The theory and measurement of private and social cost of highway congestion. *Econometrica* 29(4):676–699.