Speed

Victor Couture[*†]
University of California, Berkeley

Gilles Duranton[*‡]
University of Pennsylvania

Matthew A. Turner[*§]
Brown University

19 December 2016

ABSTRACT: We investigate the determinants of driving speed in large US cities. We first estimate city level supply functions for travel in an econometric framework where both the supply and demand for travel are explicit. These estimations allow us to calculate a city level index of driving speed and to rank cities by driving speed. Our data suggest that a congestion tax of, on average, about 3.5 cents per kilometer yields welfare gains of about 30 billion dollars per year, that centralized cities are slower, that cities with ring roads are faster, and that the provision of automobile travel in cities is subject to decreasing returns to scale.

Key words: roads, vehicle-kilometers traveled, public transport, congestion, travel time.

JEL classification: L91, R41

[*]We thank two anonymous referees, the editor, Vincent Breteau, Pierre-Philippe Combes, Kerem Cosar, Marc Gaudry, Ed Glaeser, Andy Haughwout, Diego Puga, Matti Sarvimäki, Ken Small, Will Strange, Francisco Trebbi, and conference and seminar participants for insightful comments. Financial support from the Canadian Social Science and Humanities Research Council is gratefully acknowledged.

[†]Haas School of Business, University of California, Berkeley, 2220 Piedmont Avenue, Berkeley, CA 94720, USA (e-mail: `couture@haas.berkeley.edu`).

[‡]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 Centre for Economic Policy Research.

[§]Department of Economics, Box B, Brown University, Providence, RI 02912 (e-mail: `matthew_turner@brown.edu`; website: `http://www.economics.utoronto.ca/mturner/index.htm`). The hospitality and support of the Enaudi Institute of Economics and Finance and the Property and Environmental Research Center are gratefully acknowledged.

1. Introduction

The average US driver spent about 72 minutes driving per day in 2008; the average household devoted nearly 9,000 dollars or about 18% of its expenditure to transportation, 95% of which went to buying, maintaining, or operating a private vehicle; in a typical year, the US spends about 150 billion dollars on road construction and maintenance; and the value of capital stock associated with road transportation in the US tops 7 trillion dollars (US BTS, 2013). In short, household road transportation is economically important.

The problem of understanding household transportation behavior is also conceptually difficult. Road transportation allows households to get to work, buy consumption goods, and enjoy leisure. Travel may also have independent consumption value. For households, the transportation problem involves decisions about the number, purpose, destinations, mode and the time of departure for their trips. These decisions in turn affect the amount and quality of household leisure, consumption of housing, choice of jobs, retail stores, amenities, and friends.

Underlying all of these choices is a road travel technology that governs the relationship between resources directed to the provision of transportation and road travel. We estimate this relationship. As we will see, the relationship between aggregate city travel time, road infrastructure and speed is formally equivalent to a production function for travel. Therefore, it implies marginal and average cost curves and, together with information about travel demand, describes the equilibrium provision of road travel in cities. This allows us to evaluate the welfare implications of counterfactual infrastructure policies and to arrive at estimates of the deadweight loss of congestion.

We estimate that the deadweight loss from congestion is about 30 billion dollars per year. We find that the high costs of expanding the roadway implies larger welfare gains from managing demand for travel than from expanding supply. In particular, our estimates suggest that a gasoline tax of about 60 cents to 1.6 dollars per gallon would be a welfare improving response to traffic congestion. These are precisely the sorts of calculations that economists have sought since Vickrey's pioneering characterization of traffic congestion (Vickrey, 1963).

The description of travel provision above implicitly assumes a scalar 'price' for travel. Reality is more complicated. Our data indicate that the speed of trips increases systematically with their distance. That is, cities do not offer a scalar cost of travel, but a menu of unit travel costs that vary with the distance of the trip to be undertaken. Given this, our exercise requires an important preliminary step. We must estimate the menu of speed and distance combinations on offer in each city and use this menu to calculate a scalar speed index to describe the cost of travel in a city. This calculation is analogous to the calculation of more conventional price indices, and our speed index describes the cost of road travel in a city in exactly the same sense that a city specific price index would describe the price of goods.

Complicating our exercise further, we expect individual drivers to adjust their behavior in response to the menu of trip distance and speed combinations they face. Estimating a city's speed distance menu requires an econometric response to this simultaneity problem. Our econometric model explicitly accounts for this simultaneity problem. We exploit demand variation arising from differences in trip distances across trips made for different purposes to identify the supply relationship. Our results suggest that this simultaneity problem is both economically and statistically important.

In addition to its importance to our subsequent investigation of the technology of travel provision, our speed index is of independent interest as a measure of city level the cost of travel. In particular, it provides an alternative to the Texas Transportation Institute's (TTI) widely cited congestion index (Schrank and Lomax, 2009, Schrank, Lomax, and Turner, 2010). However, unlike the TTI index, our index is grounded in economic theory and hence can be more easily interpreted. We find that Miami is the slowest city in the US and that it is 28% slower than Louisville, the fastest city in our sample.

Our investigation of the relationship between our speed index, aggregate vehicle travel time and roads is in the spirit of standard analyses of factor productivity. Since the supply of roads or aggregate travel time in a city may reflect unobserved determinants of speed, we account for the probable endogeneity of inputs in this estimation. Our main findings are that the elasticity of speed with respect to roads is about 0.09 whereas the elasticity of speed with respect to aggregate vehicle travel time is -0.13. As we will show below, that the sum of these two numbers is negative suggests that travel is produced with decreasing returns to scale in US cities. We also find suggestive evidence that more centralised cities are slower and that cities with more ring roads are faster.

Our investigation also addresses one of the central questions of transportation economics, 'what does the speed-flow curve look like?' The current literature on this question finds that speed decreases by between 50 to 60% in response to a doubling of the number of vehicles on a road. Our methodology allows us to estimate a similar elasticity describing the relationship between speed and the total time devoted to travel. Our estimates for this elasticity are on the order of 15%.

This reflects two important methodological differences. First, our unit of study is a city/year, while the existing literature typically considers particular roads or small areas at particular times. By considering averages over large areas and long time periods, we calculate a speed flow curve that implicitly reflects possible equilibrium responses to traffic, such as changes in routes or the timing of trips. While segment specific estimates are clearly of use, for the purpose of setting metropolitan or national transportation policy, our city level estimates seem more relevant. Second, extant estimates of speed flow curve have largely ignored the fact that observed travel behavior results from an equilibrium that depends upon both supply and demand conditions.[1] Our econometric strategy deals with this problem explicitly.

Our results are also broadly relevant to urban economics. In the ubiquitous monocentric models unit travel cost is usually the fundamental parameter that determines the location choices of households within cities, their consumption of housing, land use, and the population size of cities. Our work provides better estimates for this fundamental parameter, and how it varies with population and road infrastructure. We also refine the results in Duranton and Turner (2011). Where Duranton and Turner (2011) is primarily concerned with the relationship between the stock of roads in a city and the equilibrium level of traffic, we are here more interested in uncovering the underlying structure of the supply of travel. This, in turn, allows a more explicit evaluation of welfare.

Finally, while productivity in the manufacturing and service sectors is extensively studied, in spite of its size, the transportation sector has received much less attention.[2] The estimation of production functions is usually afflicted by serious issues of unobserved prices and the simultaneous determination of inputs and productivity. The first part of our methodology, which estimates the supply of travel in cities, allows us to recover appropriate prices for urban travel. The specific nature of the two main inputs into the production of travel also enables us to use plausible instruments to circumvent the problem of the simultaneous determination of inputs and productivity. Our estimated city-level supply functions allow us to investigate the cross-sectional

[1] In their authoritative book, Small and Verhoef (2007) exposit the supply and demand for travel separately, indeed, in different chapters. The fact that some variables may affect both supply and demand is recognised but only discussed in the context of car purchases. This absence of recognition of this simultaneity problem is all the more puzzling since transportation theory makes heavy use of supply and demand frameworks which constitute the starting point of all the economic calculations of the costs of congestion.

[2] Public transportation is an exception. Following Meyer, Kain, and Wohl's (1965) classic work, the estimation of the productivity of public transportation providers is a standard exercise (see Small and Verhoef, 2007, for a review).

determinants of efficiency in transportation. Consistent with the large extant literature investigating productivity in firms (e.g., Syverson, 2011), we find that some cities are dramatically more efficient than others. This suggests that there may be large gains if slow cities can emulate fast cities. As highlighted above, we also find evidence of mild decreasing returns to scale in the production of road transportation and of a low share for the fixed factor (roads). This is in contrast with standard findings for the production of consumption goods.

2. Data

Our data describe aggregate travel behavior in a set of large US cities and the individual driving trips taken by a sample of each city's residents. Our cities are mainly US (Consolidated) Metropolitan Statistical Areas (MSA) drawn to 1999 boundaries. MSAs are census reporting units and are aggregations of counties containing a major urban center and its surrounding region. To assess the robustness of our findings, we also sometimes use US Primary Metropolitan Statistical Areas (PMSAs). Our analysis relies heavily on household survey data. To ensure that we observe a sufficiently large number of households in each MSA, in most of our work we consider a sample of the 100 largest MSAs according to their census population in 2010. When our analysis requires a large sample of households in each city, we restrict attention to the 50 largest MSAs.

Data on individual travel behavior come from the 1995-1996 National Personal Transportation Survey and the 2001-2002 and 2008-2009 National Household Transportation Surveys. In a slight abuse of language, we to refer these surveys as the 1995, 2001 and 2008 NHTS. Each of the NHTS surveys reports household and individual demographics for a nationally representative sample of households.[3] More importantly, the 'travel day file' of each NHTS survey codifies a travel diary kept by every member of each sampled household. For each adult member of participating households we observe the distance, duration, mode, purpose, and start time for each trip taken on a randomly assigned travel day. See Online Appendix A for further details. We eliminate trips entered by non-drivers in order to focus our investigation on the movement of vehicles rather than the movement of people. In our sample of 100 MSAs, the NHTS describes 419,331 trips, 102,462 drivers and 71,287 households in 2008; 168,683 trips, 40,333 drivers and 27,574 households in 2001; and 152,512 trips, 33,860 drivers and 22,592 households in 1995.

[3] The sample may not be representative for smallers MSAs with fewer observations, but we control for individual and trip characteristics.

We aggregate to describe travel behavior at the MSA level. To estimate MSA vehicle kilometers traveled (VKT) and vehicle time traveled (VTT), we sum the time and distance of each trip over all of an individual's trips. We then compute the average distance and time driven by an individual in each MSA. We multiply this individual average by MSA adult population (from the US Census) to obtain total MSA VKT and VTT.

Our data on MSA road infrastructure are from the 1995, 1996, 2001, 2002, and 2008 Highway Performance and Monitoring System (HPMS) Universe and Sample data. The US federal government administers the HPMS through the Federal Highway Administration in the Department of Transportation. This annual survey, which is used for planning purposes and to apportion federal highway funding, collects data about the entire interstate highway system (HPMS Universe data) and a large sample of other roads in urbanized areas (HPMS Sample data).

The HPMS Universe data describe every segment of interstate highway (IH) and allow us to calculate the number of lane kilometers of IH in each MSA for each NHTS year. To calculate lane kilometers of major urban roads (MRU) in the urbanized parts of an MSA, we sum lane kilometers for four classes of roads reported in the HPMS Sample data; 'collector', 'minor arterial', 'principal arterial' and 'other highway'. We omit a residual class, 'local roads' because they are not systematically reported. To ensure that the resulting measures of road infrastructure are comparable to the NHTS surveys, which are collected over two years, we average each of the HPMS variables over the two relevant NHTS sampling years.

Table 1 contains summary statistics for our main variables in the 100 largest MSAs. Means and standard deviations for trip-level variables are reported in Panel A. Trip distance and trip duration increase from 1995 to 2001, from 12.5 to 13.2 km and from 15.1 to 17.6 minutes. Some of the increase in average trip duration is accounted for by a decrease in average trip speed (computed across trips) from 43 to 39 km/h. Average trip duration, distance, and speed are very similar in 2001 and 2008. The average number of trips decreases from 4.5 in 1995 to 4.1 in 2008. We note that 1995 NHTS survey asks respondents to report the time it took to get to their destination, while the 2001 and 2008 surveys ask respondents to report exact departure and arrival times. This slight difference in wording may partly explain the observed decrease in speed between 1995 and 2001.[4] We also

[4]To support this conjecture we note that our results below show that the drop in speed in 2001 is nearly entirely accounted by a small increase of slightly above 1 minute in the cost of the first kilometer of each trip. All comparisons between the 1995 NHTS and other years are subject to this caveat.

Table 1: Summary Statistics for the 100 largest MSAs

Variable	1995	2001	2008
Panel A. Trip-level data based on the NHTS			
Mean trip distance (km)	12.5	13.2	12.8
	(16.2)	(17.1)	(16.4)
Mean trip duration (min)	15.1	17.6	17.5
	(14.2)	(15.3)	(15.2)
Mean trip speed (km/h)	43.1	39.5	38.5
	(23.0)	(22.5)	(22.2)
Mean trip number (per driver)	4.5	4.2	4.1
	(2.6)	(2.4)	(2.3)
Total observed number of trips	152,512	168,683	419,331
Panel B. MSA-level data based on the HPMS and Census			
Mean daily vehicle kilometers traveled ('000,000 km)	51.3	59.7	64.2
	(74.7)	(85.1)	(90.9)
Mean daily vehicle travel time ('000,000 min)	62.1	79.2	87.3
	(91.4)	(114.6)	(126.2)
Mean lane km (interstate highways, '000 km)	2.1	2.3	2.4
	(2.3)	(2.4)	(2.4)
Mean lane km (major urban roads, '000 km)	10.5	11.9	14.4
	(13.5)	(16.1)	(18.2)
Mean MSA population ('000)	1,777	1,923	2,090
	(2,715)	(2,877)	(3,028)

Notes: Authors' computations using NHTS sampling weights to compute the means of Panel A. Standard deviations in parentheses. Total vehicle kilometers traveled and total vehicle time traveled are estimates for privately operated vehicles in MSAs. Interstate highways are for entire MSAs. Major urban roads are measured within the urbanized area of MSAs.

note that driving is sensitive to the business cycle, another reason to be cautious when comparing across years.

Panel B of table 1 reports means and standard deviations for MSA-level aggregates. Average VKT and VTT grow from 1995 to 2008, by 20% for VKT and by 29% for VTT, with much of the increase in VKT accounted for by the sample average MSA population growth of 10%. Lanes of interstate highway grow by 14% between 1995 and 2008 while lanes of major urban roads grow by 40%. While MSA boundaries are constant over time, urbanized area boundaries are not, so that some of the growth in lane kilometers of major urban roads reflects the expansion of urbanized areas.[5]

The NHTS data report the purpose of each trip using 10 consistently defined categories such as

[5]Schrank and Lomax (2007) argue that VKT grows more quickly than road capacity. While our data confirm this for Interstate Highways, this is not the case for major urban roads. Besides the issues of boundary changes, we also note that our NHTS based travel estimates capture more travel than do the HPMS based estimates on which Schrank and Lomax (2007) is based. See Duranton and Turner (2011) for a more extensive discussion of the differences between HPMS and NHTS data.

Table 2: Mean trip distance in kilometers, by trip purpose, for the 100 largest MSAs

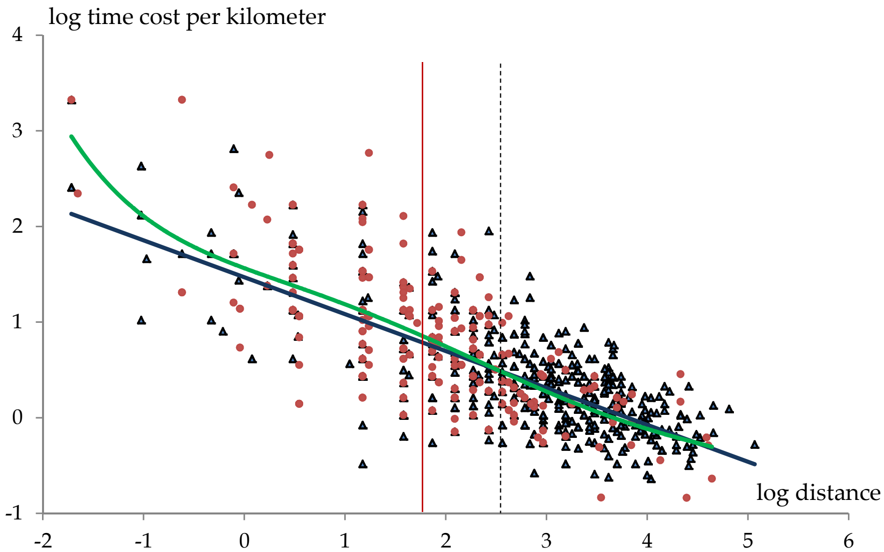
Trip purpose	Frequency (1995-2008)	km 1995	km 2001	km 2008
To/from Work	23.6%	18.6	18.8	19.1
		(19.0)	(18.8)	(19.1)
Work-related business	3.3%	17.6	20.9	18.4
		(21.0)	(23.4)	(21.5)
Shopping	21.8%	7.8	8.6	8.2
		(11.3)	(12.1)	(11.2)
Other family/personal business	24.3%	9.4	10.1	9.4
		(12.7)	(14.3)	(13.7)
School/church	4.6%	11.5	11.6	12.2
		(13.3)	(13.6)	(13.5)
Medical/dental	2.2%	13.3	12.8	13.0
		(14.9)	(13.2)	(13.5)
Vacation	0.3%	35.1	34.5	25.6
		(41.0)	(40.3)	(34.6)
Visit friends/relatives	5.7%	15.7	17.8	17.2
		(20.2)	(23.0)	(22.7)
Other social/recreational	13.8%	12.4	12.2	11.1
		(17.1)	(16.4)	(15.1)
Other	0.5%	13.4	20.3	22.4
		(18.6)	(25.4)	(23.9)

Notes: Authors' computations using NHTS sampling weights and all three years of data (pooled together to compute frequencies) by averaging across all trips. Standard deviations in parentheses.

'to/from work', 'shopping', or 'medical/dental'. Table 2 shows the mean and standard deviation of distance by trip purpose for the MSAs in our sample. There is significant and persistent variation in average trip distance across trip purposes. Shopping trips are shortest at about 8.2 km on average in 2008. Vacation trips are the longest and average 25.6 km. We note that 'vacation' and 'other' trips occur infrequently in the data and we sometimes exclude them from our analysis.

Figure 1 plots log distance and log (inverse) speed for two groups of trips in Chicago in 2008. The triangles represent commute trips and the circles represent trips taken for one of two other purposes, school/church or medical/dental. It is clear from the figure that for both groups, speed is higher for longer trips. In fact, this relationship between speed and trip distance is one of the most important features of our data. Consistent with sample averages reported in table 2, for Chicago in 2008 commute trips are about twice as long as the other class of trips described in figure 1 (20.3 km for commutes, 9.0 km for school/church trips, and 12.6 km for medical/dental trips). The figure also represents two trend lines: linear and 5th-order polynomial. The high-order polynomial stays remarkably close to the linear trend, deviating only for very short trips that

Figure 1: Speed and distance for some Chicago trips in 2008.



Commute trips are represented by triangles (mean log distance 2.52, plain line). Church, school, medical, and dental trips are represented by circles (mean log distance 1.81, dashed line). Linear trend line in black and 5th-order polynomial trend line in green (grey).

account for a small fraction of total travel. We see a similar pattern in other cities.

In addition to the NHTS and HPMS, we exploit several other sources of MSA level data as explanatory variables or as instrumental variables. Specifically, in our investigation of the determinants of MSA driving speed, we consider a number of geographical characteristics of cities: ruggedness, elevation range, and cooling and heating degree days. We also develop novel variables to measure urban form and the shape of the road network in cities. Finally, we use variables describing historical transportation networks (1947 interstate highway plan, 1898 railroads, and old exploration routes of the continent dating back to 1528) as instruments for the modern road network. Details about these variables are available in Online Appendix A and in Duranton and Turner (2011).

3. A theory of speed and the supply of travel

3.1 A city level model of the equilibrium provision of VKT

We begin by defining a production function for vehicle kilometers travelled. Let i index our sample of cities, R_i measure city i's stock of roads, VTT_i be aggregate vehicle travel time for the city, X_i be a set of other city characteristics and let ν_i be an error term. With this notation in place, we can define

$$\log \text{VKT}_i = \alpha \log R_i + (1-\theta) \log \text{VTT}_i + X_i \phi + \nu_i \,. \tag{1}$$

This is a standard Cobb-Douglas production function: vehicle kilometers traveled is our measure of output; roads and vehicle time traveled are factors of production; α is the share of roads in the production of travel; $1-\theta$ is the share of vehicle travel time; and finally, $X_i\phi + \nu_i$ is total factor productivity, the ability of a city to move its residents conditional on its stock of roads and aggregate time spent in cars. Some of the determinants of that productivity may be observed and included in X_i. Other determinants are unobserved and included in the residual ν_i.

Vehicle kilometers traveled is equal to vehicle time traveled multiplied by speed, S: $\text{VKT} \equiv \text{VTT} \times S$. Using this expression, we can rewrite equation (1) as a regression of speed on roads and vehicle travel time,

$$\log S_i = \alpha \log R_i - \theta \log \text{VTT}_i + X_i\phi + \nu_i \,. \tag{2}$$

Although equation (1) is equivalent to the travel production equation (2), we prefer to focus on equation (2) in the second of our two main empirical exercises for a number of reasons. First, the dependent variable, S_i – an index describing the speed of travel in an MSA, is a measure of travel efficiency that is easier to interpret than city aggregate VKT. Second, the exact definition of S_i is non-trivial as we discuss below. More generally, focusing on speed as dependent variable will make our discussion of identification issues clearer. Third, equation (2) maps more directly into our welfare analysis.

For later reference, note that we can determine the nature of returns to scale in the provision of automobile transportation from the production function (1) and estimates of α and θ. In particular, $\alpha - \theta$ is a measure of returns to scale and if $\alpha < \theta$ there are decreasing returns to scale in the production of VKT.

From equation (2) we can easily derive average and marginal cost curves for VKT in a city. To proceed, define $-\log \Omega_i \equiv \alpha \log R_i + X_i\phi + \nu_i$. Substituting into (2), rearranging and suppressing the i subscript for legibility gives,

$$C \equiv 1/S = \Omega \ \text{VTT}^{\theta} \,. \tag{3}$$

Equation (3) gives the average cost of a kilometer of travel as a function of MSA aggregate travel time. Using the fact that $\text{VTT} = \text{VKT}/S = \text{VKT} \times C$ and substituting into equation (3) implies,

$$C = AC(\text{VKT}) = \Omega^{\frac{1}{1-\theta}} \ \text{VKT}^{\frac{\theta}{1-\theta}} \,. \tag{4}$$

Roads are congestible and, in an equilibrium where access to the roads is not priced, drivers do not account for the costs their presence on the roads imposes on their fellow drivers. Hence, Equation

(4) is an aggregate inverse supply curve for automobile travel in an MSA as all drivers experience the prevailing average time cost of travel. Multiplying the average cost of travel in equation (4) by VKT and differentiating gives the marginal time cost of travel

$$MC(\text{VKT}) = \frac{C}{1-\theta}\,. \tag{5}$$

This marginal cost function reflects the private cost of a marginal kilometer of travel and also the extent to which this marginal kilometer slows down other drivers. That is, this marginal cost curve reflects the full social cost of travel and congestion. Thus, from regression equation (1) we derive the marginal and average cost curves for VKT.

Turning to the demand for VKT, we define it as $\text{VKT} = \Gamma C^{-\sigma}$, where Γ is a constant and σ is the price elasticity of the demand for VKT. We rearrange this expression to write it as an inverse demand curve,

$$C = \Gamma^{\frac{1}{\sigma}} \text{VKT}^{-\frac{1}{\sigma}}\,. \tag{6}$$

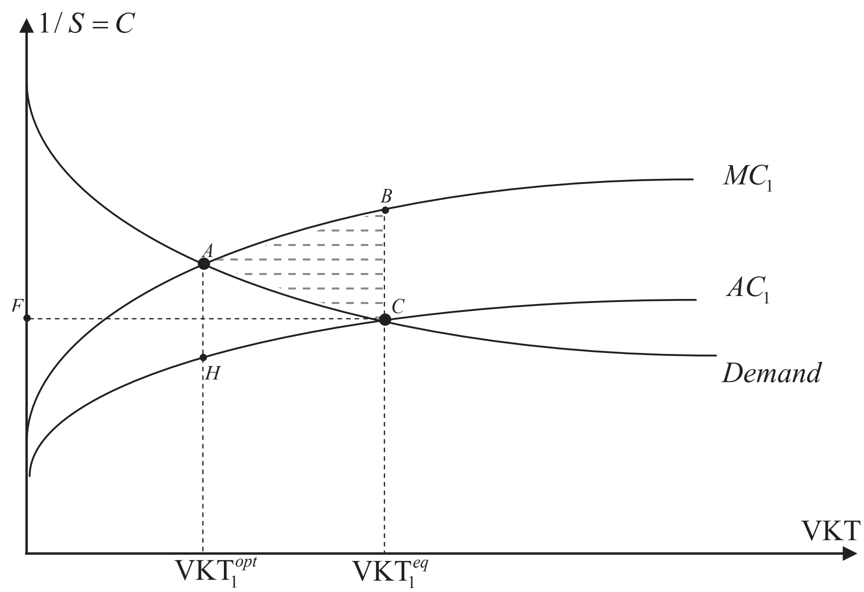
Following Vickrey (1963), figure 2 illustrates a simple partial equilibrium model of the provision of automobile travel. The vertical axis of this figure describes the cost of travel in minutes per kilometer, the inverse of speed. The horizontal axis describes vehicle kilometers travelled per year in the city. Equilibrium travel, VKT_1^{eq}, is determined by the intersection of demand and average cost, AC_1. The optimal level of travel, VKT_1^{opt}, is determined by the intersection of marginal cost, MC_1, and demand. The deadweight loss in equilibrium is given by the hatched region with vertices A, B, and C. This is the economic cost of excess equilibrium congestion. The optimal congestion tax is given by the height of the segment AH. With a tax of $|AH|$ per kilometer, we shift up the average cost curve AC_1 so that the intersection of $AC_1(\text{VKT}_1^{opt}) + |AH|$ is equal to the demand at VKT_1^{opt}. To calculate this tax, we must evaluate the difference between supply and marginal cost at optimal VKT.

Equating average cost (4) and demand (6) we can solve for equilibrium VKT. Optimal VKT results from equilibrating demand (6) and marginal cost (5). In Online Appendix B, we derive expressions for equilibrium and optimal travel as a function of parameters. In the same appendix, we also derive two expressions that will facilitate the welfare and policy analysis presented later.

The first of these is,

$$\Delta = \left[\left(1 - (1-\theta)^{\frac{\sigma}{1-\theta(1-\sigma)}}\right) - \frac{\sigma}{\sigma - 1}\left(1 - (1-\theta)^{\frac{(\sigma-1)(1-\theta)}{1-\theta(1-\sigma)}}\right) \right]. \tag{7}$$

Figure 2: Welfare analysis.



This expression gives the deadweight loss from congestion in a particular city as a proportion of the total travel time. It depends on just two parameters, the supply and demand elasticities, θ and σ. The second is an expression for τ^*, the optimal congestion tax as a function of parameters and observed quantities,

$$\tau^* = \theta(1-\theta)^{\frac{\sigma\theta}{1-\theta(1-\sigma)}-1} \frac{\text{VTT}^{eq}}{\text{VKT}^{eq}} \tag{8}$$

Note that the units of the tax in this calculation are minutes per kilometer. We will ultimately multiply by a cost of time to convert to cents per kilometer.

3.2 *A model of individual travel behavior and city level trip length supply schedules*

The preceding analysis describes the behavior of city level aggregates, including a city level index of the speed of travel. Such an aggregate measure of travel speed must derive from individual trips. However, 'the speed of travel in a city' is not well defined at the trip level. Cities do not offer a single speed of travel. They offer a menu of feasible speed and trip distance combinations. Therefore, we now turn to the problem of measuring the speed of travel in a city.

To describe a city's ability to supply road travel we will eventually calculate a speed index analogous to a Laspeyres price index. Loosely speaking, this index will indicate the time premium required to complete a standard bundle of trips in a given city relative to an average city. Before we compute this index, we need to estimate the menu of feasible speed and trip distance combinations in each city.

More specifically, we need to estimate the time cost of travel per unit distance for any trip of a given distance in each city. Although this is again a supply relationship that relates the time cost of travel to vehicle kilometers traveled, these quantities differ from those we discussed in the context of figure 2. We are here concerned with the behavior of an individual in a given city choosing how far to drive to accomplish a particular errand, taking as given the behavior of all other drivers and all other city level characteristics.

Our unit of observation is a trip, k, made by a particular driver, j, in a particular city, i. Let x_{ijk} denote distance for trip ijk in kilometers, and c_{ijk} the log time cost of the trip in minutes per kilometer. Note that c, the time cost of distance, is simply the inverse of speed. Let $\tau_{ijk} \in \{1,..T\}$ index the possible purposes for trip ijk and let χ^{τ}_{ijk} be an indicator variable that is one for trips of type τ and zero otherwise.[6] We are primarily concerned with variation in trip speed and distance within a city and therefore often omit the i subscript to increase legibility.

As figure 1 shows, the relationship between speed and distance is an important feature of our data. The unit price of travel is declining in trip length. Given this, define the inverse trip length supply schedule to be

$$c^s_{jk} \quad = \quad x_{jk}^{-\gamma} \exp(\bar{c} + \delta_j + \epsilon_{jk}). \tag{9}$$

Here, the locus c^s_{jk} describes technically feasible prices and trip lengths. It reports the average price in minutes per kilometer, on a particular trip of length x. The parameter δ_j measures drivers' abilities to drive fast on all trips and reflects characteristics such as the driver's skillfulness or the proximity of his or her home to a freeway. The parameter ϵ_{jk} measures a driver's ability to drive fast on a particular trip and reflects events such as stormy weather or road construction. The parameter $\bar{c}$ is the (log) time cost of a one kilometer trip when $\delta_j = 0$ and $\epsilon_{jk} = 0$. Finally, γ is the price elasticity of the supply of distance.

The trip length supply schedules in equation (9) are central to our analysis. They allow us to calculate the total time required to complete standardized trip bundles, components of our speed index. Estimating these curves, the parameters $\bar{c}$ and γ in particular, is the goal of our first empirical exercise.

[6] N.B.: We here use τ to index trip purposes, not as a tax, as in the preceding section.

Let c^d_{jk} be driver j's willingness to pay, in minutes per unit distance, for trip k. Define the driver's willingness to pay schedule as,

$$c^d_{jk} \quad = \quad x_{jk}^{-\beta} \exp(\Sigma_{\tau=1}^{T} A^{\tau} \chi^{\tau}_{jk} + \eta_j + \mu_{jk}), \tag{10}$$

Since χ^{τ}_{jk} is a trip purpose indicator for trip k, the summation $\Sigma_{\tau=1}^{T} A^{\tau} \chi^{\tau}_{jk}$ describes a trip purpose specific constant and allows the intercept of the driver's willingness to pay curve to vary with trip purpose. A^{τ} measures the log willingness to pay for a one kilometer trip of type τ when $\eta_j = 0$ and $\mu_{jk} = 0$. With $\beta > 0$, the unit cost of distance falls with trip length for all trip purposes. Hence, drivers are willing to drive longer distances to their preferred restaurant or supermarket as the time cost per unit of distance falls. Similarly, they may also choose to reside in more remote locations. The 'slope' parameter β is the price elasticity of the willingness to pay for trip distance and determines the rate of decline in a driver's log willingness to pay for a kilometer as trip distance increases. The parameter η_j describes a driver's idiosyncratic willingness to give up time for distance depending on his innate impatience or value of time. The parameter μ_{jk} reflects trip-specific factors which affect willingness to give up time for distance, for example, how busy a day the driver is having.

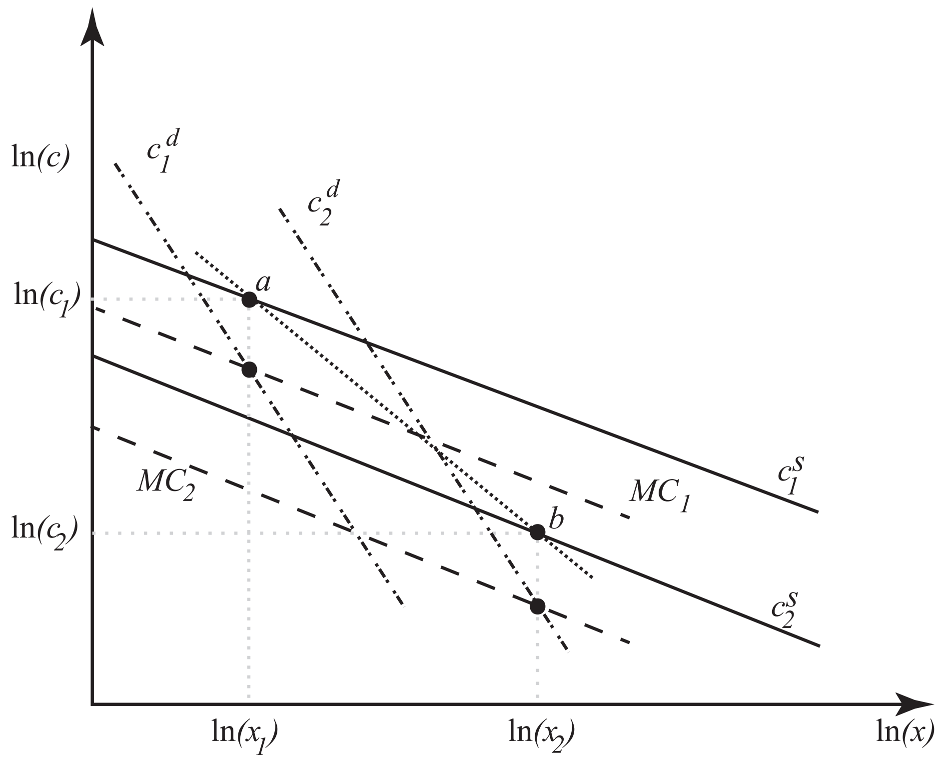
Since drivers recognize that their choice of distance affects the speed of travel, they choose trip distance to satisfy

$$c^d = MC(x) \equiv \frac{d(x\, c^s)}{dx} = (1-\gamma)c^s. \tag{11}$$

That is, the marginal willingness to pay for trip distance equals the marginal cost of trip distance. Note that despite affecting the unit price of distance with their choice of trip length, drivers are still price takers in the sense that they take the trip length supply schedule as given. In particular, they do not recognize that their driving behavior may contribute to congestion in the network and thus shift the whole supply schedule.

Figure 3 illustrates this equilibrium for the case when $\beta > \gamma$ and $cov[(\eta_j + \mu_{jk}),(\delta_j + \epsilon_{jk})] > 0$. The axes on this figure are identical to those of figure 1. In figure 3, c^s_1 describes our supply relationship for a particular realization of ϵ. The marginal cost curve associated with this average cost curve is MC_1, a dashed line. Because we have a logarithmic scale, the marginal cost curve is a vertical translation of the average cost curve, rather than a rotation. c^d_1 describes a driver's demand schedule for a particular realization of μ. An optimizing driver chooses trip distance, x_1,

Figure 3: Model of equilibrium trip distance.



to equalize marginal trip cost with its marginal value. The resulting unit price of distance for this trip is c_1, which is determined by the average cost curve. The curves c_2^s and c_2^d reflect different draws of ϵ and μ and give rise to different equilibrium trip distance and speed. Our data consist of equilibrium pairs of speeds and distances, e.g., the points a and b: this is exactly what is illustrated by figure 1. Our goal is to estimate the average cost functions, c_1^s and c_2^s. From the figure, it is clear that naively plotting the line of best fit for these equilibrium pairs, the dotted line connecting points a and b, need not accomplish this objective. To estimate average cost curves, we require variation in demand that is unrelated to variation in supply.[7]

More formally, using equations (9) and (10) in (11), and taking (natural) logarithms we arrive at the following system of equations,

$$\log x_{jk} \quad = \quad D_j + \Sigma_{\tau=1}^{T-1} \tilde{A}^{\tau} \chi_{jk}^{\tau} + \zeta_{jk} \tag{12}$$

[7]This is a complicated figure because it portrays two common problems at the same time. The first is the driver's optimization problem, which is formally equivalent to the problem of partial equilibrium with monopsony. The second is simultaneous equations bias. Thus, this figure illustrates the problem of simultaneous equations bias in the context of a monopsony equilibrium.

$$\log c_{jk} \quad = \quad \bar{c} + \delta_j - \gamma \log x_{jk} + \epsilon_{jk}\,, \tag{13}$$

where c is the observed equilibrium price, $D_j \equiv \frac{\log(1-\gamma)}{\gamma-\beta} + \frac{\bar{c}}{\gamma-\beta} - \frac{A^T}{\gamma-\beta} + \frac{\delta_j-\eta_j}{\gamma-\beta}$, $\tilde{A}^\tau \equiv \frac{A^T-A^\tau}{\gamma-\beta}$, $\tau \in \{1,..,T-1\}$, and $\zeta_{jk} \equiv \frac{\epsilon_{jk}-\mu_{jk}}{\gamma-\beta}$. Note that the equilibrium price is c^s, not c^d, since equation (11), requires the two quantities to diverge in equilibrium.

Inspection of equations (12) and (13) shows that A^τ, the willingness to pay for a trip of type τ, χ^τ_{jk}, the dummy for the trip being of type τ, and η_j (a component of D_j) the individual characteristics affecting the demand for trips of driver j, all appear in the distance equation (12), but not in the speed equation (13). It follows that variables measuring these quantities are candidate sources of exogenous variation in demand with which to resolve our simultaneity problem.

In practice, it is hard to think of individual characteristics that affect the demand for trips but not the ability to produce them. Educational attainment affects a driver's opportunity cost of time and hence demand for trip distance, but may also affect driving skills and thus the ability to drive at a high speed. This suggests that individual characteristics are unlikely to provide good sources of exogenous variation in demand.

Trip type indicators are more defensible sources of variation with which to identify the inverse supply curve described by equation (9). Trip type dummies occur explicitly in equation (12) and not in equation (13), so the rationale for using them as an instrument is transparent. Denote these instruments Z_{jk}. As made clear by the discussion above, valid instruments for trip distance must satisfy two conditions. First, they must predict trip distance conditional on the other controls: $cov(Z_{jk}, x_{jk}|.) \neq 0$ (relevance). We demonstrate that this condition holds below. Second, instruments must be uncorrelated with the error term of equation (13): $cov(Z_{jk}, \epsilon_{jk}|.) = 0$ (exogeneity).

If trip type dummies are orthogonal to ϵ_{jk} then we are not more (or less) likely to observe trips of type τ when such trips are particulary fast. In fact, we suspect that some trips (e.g., 'recreational' trips, to take an example from the data) might be taken with greater propensity when traffic conditions are good, i.e., when ϵ_{jk} is high. To understand why such a correlation might arise, assume there are only two types of trip: to the gym and to work. Also suppose that drivers stop going to the gym when there is more than 10 centimeters of snow on the ground and stop going to work when there is more than 30 centimeters of snow (and traffic gets even slower). In this case, trips to the gym will be positively correlated with the error term. This violates the exogeneity condition.

To circumvent this possible problem, we can restrict attention to trips which are not discretionary such as trips 'to and from work', 'work related business', 'school church', and 'medical/dental' trips.[8] Adding controls reduces the role of unobserved determinants of speed. In our case, we know trip characteristics like; month, day of week, and time of day. If adding these controls does not cause big changes in our estimations it suggests that our instruments are uncorrelated with ϵ_{jk}.

In addition to trip type dummies, we also rely on mean distance by trip type by city as instruments. The rationale for this instrument is somewhat different from that for trip type dummies and is described in Online Appendix C.

Apart from concerns about the validity of our instruments, we may worry about the sorting of drivers. As we see in equation (13), the constant term in the speed equation is the sum of the intercept of the inverse-supply curve $\bar{c}$ (the coefficient of interest), and driver characteristics affecting supply δ_j. Since we only observe drivers driving in one city, $\bar{c}$ and δ_j cannot be separately identified. Stated precisely, our concern is that fast drivers, those with high δ's, might systematically choose to locate in fast cities, those with high $\bar{c}$'s.

We have two responses to this problem. The first is to consider large areas, the largest US consolidated statistical metropolitan areas (MSAs), as our unit of observation. As long as the problematic sorting of drivers occurs at a smaller scale than our unit of observation, it will not lead to systematic differences between drivers in one MSA and another. Drivers with a desire to drive fast can always locate close to highways in a less densely populated part of nearly any large MSA in the US. Much the same logic is widely used to identify local peer group effects (e.g., Evans, Oates, and Schwab, 1992, for an early example). Our second response to the sorting problem is to parameterize individual effects as a function of observable driver characteristics. In particular, we expect that controls such as age, income, gender, and education are correlated with individual unobservables. Since only the residual ϵ_{ijk} will be confounded with the intercept, we use our controls to reduce this residual as much as possible.

3.3 *Calculation of the speed index*

After estimating a 'trip length supply schedule' relating trip speed to trip length in each city, we can now compute a speed index for each city.

[8] Actually, we only need to restrict attention to trips with the same level of discretion.

Let $\bar{c}_{US}$ denote our estimates of $\bar{c}$ from a particular regression specification for all MSAs. Let $\bar{c}_i$ be a corresponding estimate for a particular MSA, and let $\sum_{jk}$ be a sum over all individuals j and trips k (i.e., the universe of all trips taken in the data). Then, suppressing NHTS trip weights and year indices, the speed index for city i is

$$S_i = \frac{\Sigma_{jk} x_{jk} \exp\left(\bar{c}_{US} - \gamma_{US} \log x_{jk}\right)}{\Sigma_{jk} x_{jk} \exp\left(\bar{c}_i - \gamma_i \log x_{jk}\right)}. \tag{14}$$

That is, we compute the time that it would take to realize all (weighted) US trip distances in our data at the average estimated US speed relative to how much time it would take to realize the same trips at the estimated speed of a given MSA. Formally, this is the inverse of a time cost index or equivalently, a speed index.[9]

The speed index defined in equation (14) is analogous to a Laspeyres price index in the sense that we compare the speed of travel across US MSAs for the same (national) bundle of trips. A possible worry with this index is that the relative time cost of different types of trips may vary a lot across MSAs and these different trips may be highly substitutable. To assess this potential problem, we also compare the speed of travel across MSAs for the trips that actually occur in these MSAs in robustness checks below. This alternative speed index is analogous to a Paasche price index.

Finally, recall that in our analysis of the determinants of speed, we examine the relationship between our speed index and probable determinants of travel speed. In particular, the extent and configuration of the road network, and the physical geography and configuration of sample MSAs. By construction our index describes the cost of travel. This allows us to abstract from the changes in the value of travel often associated with changes in accessibility, although we allow for changes in the value of travel in our welfare analysis.

4. Estimation of the city level supply schedules

Our goal is to understand the production of travel in cities. We proceed as follows. In this section we estimate city level supply schedules. In the following section we use these supply schedules to calculate a speed index for each city. Finally, we use this speed index, together with city level measures of VKT and infrastructure to estimate the aggregate supply relationship described by equation (2).

[9]Formally, S is a scalar without units. However, we can easily interpret it as a speed. An MSA for where the national bundle of trips requires only half as much time the US average speed has an index of 2 and is twice as fast as an average MSA.

We now turn to the estimation of city level speed distance schedules. We start by estimating variants of the equation

$$\log c_{ijk} \quad = \quad \overline{c}_i + Y_j\,\delta - \gamma_i\,\log x_{jk} + T_{jk}\,\xi + \epsilon_{ijk}\,. \tag{15}$$

This equation differs from equation (13) in two regards. First, it includes a vector of trip attributes, T_{jk}, not present in (13). These trip attributes control for variation in traffic conditions by time of day, day of week, and month of year. Second, equation (15) includes a vector of individual control variables Y_j. This generalizes equation (13) which restricts attention to individual fixed effects.

We estimate equation (15) using NHTS trips by drivers residing in each of the MSAS in our sample. For each MSA, we thus estimate an intercept $\overline{c}_i$ and a slope γ_i. Because it is not enlightening to report a large number of coefficients, table 3 reports some summary results. In each panel, we report the mean values of the MSA intercept $\overline{c}_i$ and the MSA slope γ_i. For both variables, we also report the standard deviations of the mean of these coefficients in parenthesis and the mean of their standard errors in squared brackets. Panel A report results based on the 2008 NHTS for trips by drivers in the 100 largest MSAS. Panel B replicates panel A but restricts attention to the 50 largest MSAS. Panels C and D reproduce panel B but are based on the 2001 and 1995 NHTS.

In column 1, we estimate equation (15) without driver or trip controls. The mean value of $\overline{c}_i$ for the 100 largest MSAS in 2008 appears in the first row of panel A. Its value of 1.412 implies just above 4 minutes for a trip of one kilometer.[10] This is slightly less than 15 kilometers per hour. The second row of the same column reports the standard error of the mean of the intercepts *across* MSAS. Its value of 0.094 implies an $e^{0.094} \approx 10\%$ difference in speed for a trip of one kilometer. The third row reports that the mean of the standard error *within* MSAS is only 0.030. This suggests that the differences in intercepts across MSAS reflect mostly true differences in speed, not sampling error.

The fourth row of panel A reports the average of the coefficients for log distance. In column 1, its value of 0.426 implies that speed increases by about $2^{0.426} \approx 34\%$ when trip distance doubles. Of course, we cannot expect this relationship to scale up for extremely long trips. However, 99% of the trips we observe are between zero and 83 kilometers and this elasticity estimate applies in this range. The fifth row reports the average standard deviation for these estimates of γ across MSAS. It equals 0.034. Since this is more than twice as large as the mean standard error for γ within MSAS,

[10] Since this quantity is an exponential of an average of logs from which we omit the errors, strictly speaking, it is not predicted speed.

Table 3: Estimation of inverse-supply curves

	(1) OLS1	(2) OLS2	(3) OLS3	(4) FE	(5) IV1	(6) IV2	(7) IV3	(8) IV4	(9) IV FE
Panel A. 100 largest MSAs for 2008									
Mean $\bar{c}$	1.412	1.389	1.390	1.393	1.314	1.308	1.310	1.248	1.265
	(0.094)	(0.091)	(0.093)	(0.102)	(0.147)	(0.148)	(0.132)	(0.143)	(0.253)
	[0.030]	[0.002]	[0.001]	[0.033]	[0.051]	[0.050]	[0.046]	[0.052]	[0.147]
Mean γ	0.426	0.421	0.421	0.416	0.356	0.353	0.355	0.342	0.348
	(0.034)	(0.034)	(0.034)	(0.037)	(0.074)	(0.078)	(0.066)	(0.075)	(0.131)
	[0.014]	[0.001]	[0.001]	[0.024]	[0.035]	[0.034]	[0.031]	[0.036]	[0.105]
Panel B. 50 largest MSAs for 2008									
Mean $\bar{c}$	1.415	1.396	1.396	1.396	1.293	1.283	1.293	1.223	1.252
	(0.068)	(0.068)	(0.069)	(0.073)	(0.096)	(0.089)	(0.093)	(0.085)	(0.131)
	[0.020]	[0.002]	[0.001]	[0.022]	[0.036]	[0.036]	[0.034]	[0.039]	[0.067]
Mean γ	0.421	0.416	0.415	0.411	0.336	0.330	0.337	0.318	0.336
	(0.020)	(0.019)	(0.018)	(0.021)	(0.046)	(0.042)	(0.045)	(0.042)	(0.062)
	[0.010]	[0.001]	[0.001]	[0.016]	[0.025]	[0.025]	[0.023]	[0.026]	[0.047]
Panel C. 50 largest MSAs for 2001									
Mean $\bar{c}$	1.385	1.384	1.380	1.351	1.326	1.318	1.328	1.260	1.264
	(0.071)	(0.067)	(0.070)	(0.067)	(0.113)	(0.121)	(0.104)	(0.132)	(0.241)
	[0.025]	[0.002]	[0.001]	[0.029]	[0.048]	[0.048]	[0.045]	[0.050]	[0.098]
Mean γ	0.412	0.407	0.406	0.394	0.349	0.344	0.350	0.341	0.350
	(0.021)	(0.021)	(0.021)	(0.023)	(0.061)	(0.064)	(0.056)	(0.066)	(0.113)
	[0.011]	[0.001]	[0.001]	[0.020]	[0.032]	[0.031]	[0.029]	[0.033]	[0.066]
Panel D. 50 largest MSAs for 1995									
Mean $\bar{c}$	1.189	1.196	1.186	1.133	1.115	1.110	1.111	1.065	1.044
	(0.083)	(0.080)	(0.080)	(0.090)	(0.117)	(0.120)	(0.115)	(0.124)	(0.141)
	[0.027]	[0.006]	[0.002]	[0.029]	[0.050]	[0.049]	[0.046]	[0.051]	[0.075]
Mean γ	0.380	0.375	0.373	0.350	0.335	0.332	0.331	0.326	0.320
	(0.021)	(0.023)	(0.023)	(0.028)	(0.058)	(0.059)	(0.053)	(0.061)	(0.064)
	[0.013]	[0.001]	[0.001]	[0.022]	[0.035]	[0.034]	[0.032]	[0.036]	[0.054]

Notes: Mean of the coefficients across all cities. Standard deviation of city coefficients in parentheses. Mean of the standard deviation of city coefficients in squared parentheses. OLS estimations in columns 1-4 and IV in columns 5-9.
Dependent variable: log minutes per kilometer for individual trips.
Controls: No control in column 1. Controls for household income and its square, driver's education and its square, age, dummies for males, blacks, and workers, and a quartic for the time of departure in columns 2 and 5-8. 17 dummies for household income, four dummies for education, age, dummies for males, blacks, hispanics, and workers, 23 dummies for the hour of departure, 11 dummies for the month of departure, and a dummy for trip taken during the weekend in column 3. Driver fixed effects in columns 4 and 9.
Instruments: Mean trip distance for trips of the same purpose in the same MSAs in column 5. Same instrument but computed from the four most similar MSA in term of population in columns 6 and 9. Trip purpose in column 7 (8 categories) and column 8 (2 categories; commutes and other work related trips, shopping, medical and dental, and school and church). See the text for a discussion of instrument strength.

0.014, reported in the sixth row, this probably reflects again true heterogeneity across MSAs and not sampling error.

In column 2, we augment the regression of column 1 with several controls for driver characteristics; household income and its square, driver's education and its square, age, and dummies for males, blacks, and workers. We also include a quartic in departure hour and a weekend dummy as trip controls. In column 3, we include more exhaustive driver and trip controls. For drivers we include 17 dummies for household income, four dummies for education, age, and dummies for males, blacks, hispanics, and workers. For trips, we include 23 dummies for the hour of departure, 11 dummies for the month of departure, and a dummy for trips taken during the weekend.

We constrain the effect of driver and trip characteristics to be the same for all MSAs. This increases the efficiency of our estimations, increases the transparency of the speed indices calculated below, and eases the calculation of these indices. Moreover, since regressions with driver and trip characteristics give similar results to regressions with driver fixed effects, it seems unlikely that this simplifying assumption is important to our estimates of the speed distance schedules.[11]

Because our explanatory variables are centered, we can directly compare estimates of $\bar{c}_i$ across columns 1, 2, and 3. Their means are within 0.023 of each other or less than 2% apart. We can also compare estimates of the distance elasticity of speed, γ, across columns. These estimates are stable. The R^2s for the different specifications are also stable. The (adjusted) R^2 associated with column 1 when estimating an intercept ($\bar{c}$) and a slope (γ) for each city in a single regression is 56.7%. Adding driver and trip controls in column 2 raises this R^2 slightly, to 57.7%. The more exhaustive controls of column 3 also increase the R^2 slightly, this time to 57.8%. Controls for trip and driver characteristics do not affect our estimation of the distance elasticity.

This does not imply that driver and trip characteristics do not affect speed. They do. While we do not report these coefficients, several are interesting. Women are about 0.5% slower than men. Age is more important. A year of age is associated with 0.3% slower speed. Black drivers drive about 8% slower. Drivers with more education and drivers with higher income are faster, although in both cases the relationship tapers off after a threshold: drivers with a Bachelor degree are about 7% faster than workers with less than high school; drivers from households with annual income

[11] It would also be of interest to investigate whether the effect of driver and trip characteristics vary across MSAs, e.g., to check if 'peak' hours differ in intensity and duration across MSAs. However, given that our ultimate objective is an understanding of city level determinants of speed, we leave such an investigation for future research.

around \$60,000 are about 9% faster than drivers from the poorest households.[12] Our findings on the effect of trip characteristics are unsurprising: weekend trips are about 4% faster than week-day trips; trips departing during the morning peak are about 4% slower than trips in the middle of the night; trips departing during the evening peak are about 10% slower than trips in the middle of the night; there are small differences between months, Winter and Fall months are about 1% faster than Spring and Summer months.

In column 4, we return to the specification of column 1 and introduce driver fixed effects. The results for this column confirm that driver characteristics do not affect the estimation of our parameters of interest. With driver fixed effects, the mean of both intercepts and slopes are nearly unchanged from column 1.

Column 5 replicates the specification of column 2, but, to instrument for trip distance, uses the mean log distance of other trips with the same purpose in the same MSA. We note that this estimation raises a technical issue. We want to estimate a separate intercept and slope for each MSA. This implies estimating a separate IV regression for each MSA so that we instrument trip distance in a city by only the instruments for this city instead of the entire set of instruments. At the same time, we want to constrain the effect of driver and trip characteristics to be the same everywhere to remain consistent with the OLS estimations of column 2. This calls for a two-step approach where the effects of driver and trip characteristics on speed are estimated first from the cross-section of MSAs. We then take these coefficients as given (i.e., treat them as constraints) when estimating a separate TSLS regression for each MSA.

If drivers take longer trips when travel is faster, then OLS estimates of γ are biased upwards. Comparing the IV results in column 5 panel A with the corresponding OLS results in column 2 we see that, as expected, the IV estimates of γ are smaller than the OLS estimates. In column 5, the mean IV mean elasticity of speed with respect to distance is 0.356. The corresponding OLS value from column 2 is 0.421. Although modest, this 20% difference between the OLS and IV estimates is statistically significant for a large majority of cities. These elasticities imply that after controlling for simultaneity in the choice of trip distance and speed, speed increases by only about $2^{0.360} \approx 28\%$ when trip distance doubles as opposed to the 34% increase we observe in equilibrium speed.

We also observe that the estimates of $\bar{c}$ are lower with IV than OLS. From figure 3, the equilibrium

[12] This might obviously be related to the state of their vehicles which we control for indirectly with demographic characteristics.

relationship between the unit time cost of travel and trip distance is given by the line passing through the points a and b. On the other hand, the supply relationship is given by c^s and has a smaller slope and intercept. This corresponds exactly to the observed relationship between OLS and IV estimates. Consistent with this, the IV results of column 5 imply a speed of nearly 17 kilometers per hour for a trip of one kilometer. This is about two kilometers per hour faster than is implied by the OLS estimate of column 2 (recall that a smaller value of $\bar{c}$ implies a lower time cost of travel and hence a higher speed).

A possible worry is that these results are driven by weak instruments. For the 50 largest cities in 2009, the average first-stage F statistic for the excluded instrument is equal to 595. There is nonetheless a lot of heterogeneity across cities. At one extreme, we have nearly 33,000 trips for Dallas and commutes are about twice as long as shopping trips. With driver characteristics playing only a minor role, it is unsurprising that the F statistic is above 3,000 in this MSA. While no MSA in the top 50 in 2008 has an F statistic below 10, there is one below 20. This is Baton Rouge (LA) for which we observe only about 800 trips. For 2001, one MSA has an F statistic below 10 and three are below 20. For 1995, the numbers are respectively 5 and 9. Ignoring MSAs for which the instrument is weak makes no difference to our results. We also experimented with clustering our estimations at the household level for each MSA and with robust standard errors and this makes no perceptible difference either.

In column 6, we replicate the IV estimations of column 5, but instrument for trip distance using mean log distance for trips of the same type in the four MSAs with most nearly the same population. The results are close to those of column 5. Mean log distance for trips of the same purpose in MSAs with similar population is a marginally stronger instrument than mean log distance by trip purpose in the same MSA because the instrument is measured with more precision, being computed using more observations (i.e., trips from four MSAs instead of only one). As a result, no MSA among the largest 50 has an F test below 20 in 2008, only one in 2001, and five in 1995. For this reason and because mean trip distance in other cities is arguably more likely to satisfy the exclusion restriction we prefer the specification of column 6 and we use these regressions to compute our benchmark speed index.

In column 7, we use seven trip purpose dummies as instruments. Despite the different rationales for the validity of the instruments, this yields mean estimates close to those of columns 5 and 6. In column 8, we restrict our sample of trips to less discretionary trips: commutes, work-related trips, school and church, and medical/dental and use a dummy for commutes and work-related trips as instrument. The slopes we estimate for the inverse-supply curves are close to those from the other IV estimates. We note that the mean intercept is lower than for previous IV estimations since less discretionary trip are often taken at busier hours and tend to be slower.

Finally, in column 9 we return to the same instruments as in column 6 but apply them to the fixed effect estimation of column 4. This is a very demanding estimation strategy since the effect of distance on speed is identified within driver from the speed differences between long work trips and shorter trips. We can see that the estimates of the slopes of the inverse supply curves nonetheless remain close to previous IV estimations but have larger standard errors.[13]

Panel B replicates panel A for our sample of the 50 largest MSAs. For more demanding estimations like column 9, the coefficients of interest are more precisely estimated than when we consider the 100 largest MSAs. Even for the specification of column 1, the mean of the standard error on $\bar{c}$ and on γ is about twice as large for MSAs ranked between 51 and 100 in terms of population as for the largest 50.

Panels C and D replicate panel B for 2001, and 1995, respectively. The results for 2001 are similar to those for 2008. This is consistent with mean trip distances and speeds being essentially the same in 2001 and 2008. We note nonetheless that the variances of the city intercepts and distance elasticities are larger in 2001 than in 2008. This probably reflects the larger sample drawn by the 2008 NHTS. For 1995, the estimated distance elasticities of speed are close to but smaller than those for 2001 and 2008. The city intercepts are also smaller. This is consistent with the observed reduction in mean speed after 1995.

To assess the robustness of our findings further, we perform a variety of estimations using alternative samples and geography. The results are reported in table 9 in Online Appendix D. A first worry is that we may identify the speed distance schedules mostly from long trips which are relatively rare. In addition, the variance for short trips is larger as made clear by the data for Chicago represented in figure 1. To tackle these two issues, we replicated the OLS estimates of column 3 of table 3 and the IV estimates of column 5 of the same table excluding all trips in the top and bottom distance quartile in each city. Although the IV estimates are imprecise, the average OLS elasticity of speed with respect to distance remains close to those based on the whole sample.

[13]This is to some extent driven by weaker instruments. The mean F is less than half relative to column 5 or 6. Among the largest 50 MSAs in 2009, one has an F below 10 and 5 are below 50. Among the largest 100 MSAs, 21 are below 10 and 30 are below 20.

A second worry is that the trip distance instrument we use above may be affected by the times at which people travel. These times of travel differ across types of trips. To assess this potential problem, we again replicated our preferred OLS and IV estimates of table 3 restricting our sample to peak-hour trips. The OLS and IV elasticities of the speed of travel with respect to distance are very similar to those of table 3. The intercepts are slightly lower given that travel is generally slower at peak hour. Related to this, one may also worry that commutes, which are typically longer trips, may also differ in other respects including for instance the direction of travel. To assess whether commutes affect our results, we also performed separate estimations that restrict our sample to commuting and work-related trips. The OLS results are very close to those for all trips. The IV results are less conclusive since the dramatic reduction in sample size when we only consider commutes and work-related trips implies that IV regressions can be meaningfully estimated for only the largest cities in 2008.

In addition, our estimations rely on (consolidated) metropolitan statistical areas. These are geographically large units that will group, for instance, Baltimore with Washington DC, Gary with Chicago, Fort Lauderdale with Miami, and Northern New Jersey with New York City. While the congestion of Miami and New York City may spill over to Fort Lauderdale or to Northern New Jersey, it is much less clear whether Baltimore and Gary (IN) are part of the same transportation equilibrium as Washington or Chicago. For 1995 and 2001, we only know household location for MSAs. For 2009, we have more precision for household location and can re-estimate the speed distance schedules for primary metropolitan statistical areas. As reported in table 9, this makes virtually no difference to our estimates of the average slopes and intercepts of the inverse supply schedules.

As a final robustness check, we experiment with alternative functional forms. First, we estimate polynomial regressions, and add non-linear terms of log trip distance to the equation in column 1 of Table 3. The polynomial regressions only marginally improve the fit of the regressions. The mean (non-adjusted) R^2 computed across the 100 largest cities in 2008 increases from 0.571 in the linear regression to 0.595 in the fifth-order polynomial regression. While the coefficients associated with the non-linear terms are in most cases statistically significant, they make little economic difference. The linear specification implies an elasticity of speed with respect to trip distance of 0.417 in 2009. The fifth-order polynomial specification implies an elasticity of 0.444 for a three-kilometer trip (at the first decile of trip distance) and an elasticity of 0.325 for a much longer trip of 31 kilometers (at the ninth decile of trip distance). In comparison, the largest corresponding elasticity reported in the top panel of table 3 is 0.426 and the smallest is 0.342, so this range is not large relative that associated with other variations in specification. As we report below, we also experimented with local polynomial smoothing regressions and show that taking these non-linearities into account makes close to no difference to our estimation of the speed-distance schedules and the speed indices we derive from them.

In sum, we draw four conclusions from our estimations of the speed distance schedules. First, from the differences between our OLS and IV estimates, there is evidence of the simultaneous determination of speed and distance. This modestly affects our estimates of the elasticity of speed with respect to distance. Second, different IV strategies yield similar estimates for $\bar{c}$ and γ. Third, our estimates are robust to the inclusion of other trip and driver controls. Fourth, our estimates are also robust to considering a different subsamples of trips, departure time, and geographical units.

5. Speed index

5.1 Calculation of speed indices

Table 4 reports our preferred speed index for the largest 50 US MSAs in 1995, 2001, and 2008. This index is based on our preferred estimate of equation (3) from column 6 of table 3, where we instrument for trip distance with mean distance for trips of the same type in the four cities most nearly the same size.

For 2008, we find that the speed of driving in the slowest MSA, Miami, is 28% lower than in the fastest, Louisville. Despite its name, Grand Rapids is only the second fastest MSA. More generally, among the 10 slowest MSAs, we find the four largest (New York, Los Angeles, Chicago, and Washington), another from the top 10 largest (Boston), four large cities with a difficult geography (Miami, Seattle, New Orleans, and Pittsburgh) and one city with famously stringent zoning regulations (Portland).

There are some changes in ranking between 1995 and 2008. The Spearman rank correlation between the 2008 and 2001 ranking is 0.82 while that between the 2008 and 1995 ranking is 0.59. These correlations are high but far from perfect. In part at least, these changes in rank reflect changes in city level fundamentals: in a regression of changes in the speed index from 1995 to 2008 against population changes over the same period and the 1995 value of the same index, we

Table 4: Ranking of the 50 largest MSAs, slowest at the top

	2008 Index	2008 Rank	2001 Index	2001 Rank	1995 Index	1995 Rank	Population rank
Miami-Fort Lauderdale, FL	0.88	1	0.87	1	0.92	2	12
Chicago-Gary-Kenosha, IL-IN-WI	0.91	2	0.94	2	0.90	1	3
Portland-Salem, OR-WA	0.94	3	1.04	18	1.09	27	21
Seattle-Tacoma-Bremerton, WA	0.94	4	0.95	3	0.99	8	14
Los Angeles-Riverside-Orange County, CA	0.95	5	0.98	7	1.00	11	2
New York-Northern NJ-Long Isl., NY-NJ-CT-PA	0.95	6	0.95	4	0.94	3	1
New Orleans, LA	0.96	7	0.98	9	1.04	14	44
Washington-Baltimore, DC-MD-VA-WV	0.96	8	0.98	9	0.98	6	4
Boston-Worcester-Lawrence-Low.-Brock., MA-NH	0.96	9	0.98	11	0.99	9	8
San Francisco-Oakland-San Jose, CA	0.96	10	1.01	13	1.02	13	5
Pittsburgh, PA	0.96	11	0.98	8	1.02	12	22
Houston-Galveston-Brazoria, TX	0.97	12	1.06	21	1.11	30	9
Sacramento-Yolo, CA	0.98	13	1.03	17	1.22	47	24
Philadelphia-Wilmington-Atl. City, PA-NJ-DE-MD	0.98	14	0.97	5	0.97	5	6
Tampa-St. Petersburg-Clearwater, FL	0.98	15	1.03	15	1.09	26	19
Orlando, FL	0.98	16	1.03	16	1.05	15	19
Baton Rouge, LA	0.98	17	1.12	35	1.14	38	46
Las Vegas, NV-AZ	0.98	18	0.98	10	1.16	42	23
Norfolk-Virginia Beach-Newport News, VA-NC	0.99	19	1.12	37	1.09	23	34
Phoenix-Mesa, AZ	1.00	20	1.04	19	1.10	29	13
Cleveland-Akron, OH	1.01	21	1.12	36	0.96	4	18
Austin-San Marcos, TX	1.03	22	1.07	22	1.08	21	33
Atlanta, GA	1.03	23	1.08	24	1.09	24	11
San Diego, CA	1.04	24	1.09	26	1.11	31	16
St. Louis, MO-IL	1.04	25	1.09	29	1.20	44	20
Detroit-Ann Arbor-Flint, MI	1.04	26	1.08	23	1.06	18	10
Dallas-Fort Worth, TX	1.04	27	1.08	25	1.12	33	7
Salt Lake City-Ogden, UT	1.05	28	1.10	30	1.09	25	36
Denver-Boulder-Greeley, CO	1.06	29	0.99	12	0.99	10	17
San Antonio, TX	1.06	30	1.11	32	1.15	40	27
Indianapolis, IN	1.06	31	1.04	20	1.12	34	30
Jacksonville, FL	1.07	32	1.15	42	0.99	7	40
Hartford, CT	1.07	33	1.14	41	1.26	49	43
Charlotte-Gastonia-Rock Hill, NC-SC	1.07	34	1.13	39	1.15	41	29
West Palm Beach-Boca Raton, FL	1.09	35	1.11	33	1.07	20	39
Columbus, OH	1.09	36	1.16	43	1.07	19	32
Minneapolis-St. Paul, MN-WI	1.10	37	1.11	34	1.12	32	15
Memphis, TN-AR-MS	1.10	38	1.25	48	1.14	36	41
Milwaukee-Racine, WI	1.10	39	1.09	28	1.06	16	31
Cincinnati-Hamilton, OK-KY-IN	1.10	40	1.01	14	1.05	17	26
Richmond-Petersburg, VA	1.11	41	1.19	45	1.18	43	47
Nashville, TN	1.12	42	1.10	31	1.22	46	37
Buffalo-Niagara Falls, NY	1.12	43	1.19	44	1.09	28	48
Raleigh-Durham-Chapel Hill, NC	1.12	44	1.21	46	1.22	45	35
Oklahoma City, OK	1.15	45	1.12	38	1.14	37	42
Rochester, NY	1.16	46	1.14	40	1.15	39	50
Kansas City, MO-KS	1.18	47	1.29	49	1.08	22	28
Greensboro–Winston-Salem–High Point,NC	1.19	48	1.24	47	1.25	48	38
Grand Rapids-Muskegon-Holland, MI	1.22	49	1.29	50	1.13	35	45
Louisville, KY-IN	1.22	50	1.09	27	1.29	50	49

Notes: Speed index constructed from the estimations reported in column 6 of table 3.

find a coefficient -0.078 (significant at 10%) for the 50 largest MSAs and -0.132 (significant at 5%) for the 100 largest MSAs. With this said, some changes in ranking across years are probably due to sampling error.

5.2 *Robustness checks*

To assess the robustness of our preferred ranking, we compare it to alternative rankings based on the same data but different aggregation methods to construct the speed index, different estimation strategies, or different geography.

First, the exact construction of our index does not matter. Our speed index compares how much time it would take to complete all US trips at the average estimated US speed with how much time it would take to complete all US trips at the estimated MSA speed. As an alternative, we can measure how much time it would take to complete all trips in an MSA at this MSA's speed relative to average US speed. This would be a Paasche index instead of a Laspeyres index. This alternative index allows us to compare MSAs for the trips drivers actually take in those cities. This is of particular importance when trips of different types or different distances are easily substitutable. However, this alternative index is more difficult to interpret since speed differences across MSAs can now be caused by both the speed and composition of trips.

Empirically, allowing for differences in the composition of trips is not important. Using our preferred estimation strategy and our sample of the 50 largest MSAs, the rank correlation between our preferred (Laspeyres) ranking and the alternative (Paasche) index just described is above 0.99 for 1995, 2001, and 2008. If we consider the 100 largest MSAs the corresponding correlations are all above 0.96. These findings are not specific to our choice of estimation. We find similarly high correlations for the Laspeyres and Paasche rankings constructed from the output of OLS estimation for column 2 of table 3 (i.e., the OLS estimation that corresponds to our preferred IV).

Next, we compare our preferred ranking, calculated from column 6 of table 3, with alternative rankings calculated from other columns of the same table and with average speed calculated directly from the data. Starting with the latter, the rank correlation between our preferred ranking and one obtained based on MSA average speed is 0.68 for 50 MSAs in 2008. The rank correlations between our preferred ranking and alternative rankings obtained from the OLS estimates of columns 1 to 3 of table 3 are between 0.89 and 0.94 for 50 MSAs in 2008. For the fixed-effect estimation of column 4, the correlation is slightly lower at 0.87. For the IV estimations of columns 5, 7, 8, and 9, the correlations are 0.98, 0.97, 0.97, and 0.68, respectively. This last correlation is lower because it is based on the noisier estimates of column 9 (our most demanding estimation, with driver fixed effects in an IV regression). For 1995 and 2001, correlations across rankings are slightly lower.[14]

We also experimented with nonparametric specifications of the speed-distance schedules using local polynomial smoothing regressions.[15] Comparing the ranking obtained from this nonparametric estimation for the largest 50 MSAs in 2008 with that obtained from the corresponding linear OLS estimation of column 1 of table 3, the rank correlation is 0.88. The rank correlations for 2001 and 1995 are even higher at 0.94 and 0.91. In other words, a nonparametric specification generates a ranking of cities by speed that closely matches what we derive from a linear specification.

We draw a number of conclusions from these correlations. First, the relatively low correlation between our preferred ranking and raw measures of speed underscores the importance of controlling for trip distance. Second, the high correlations between the indices derived from our various IV estimations suggest that our preferred ranking is not sensitive to the details of our instrumentation strategy provided the relationship between speed and distance is precisely estimated. Third, the relatively high correlations between our preferred ranking and the rankings derived from OLS estimates suggest that controlling econometrically for the simultaneous determination of speed and distance has only a small effect on the final ranking of MSAs. Finally, the high correlations between our parametric non-parametric indices indicates that the exact functional form chosen to control for distance does not have big effects on the resulting speed index.

5.3 *Comparisons with TTI's index and a PMSA speed index*

The Texas Transportation Institute produces the best known and most widely reported indices of travel speed. We here investigate the differences between our index and the 2008 and 2009 TTI indices (Schrank and Lomax, 2009, Schrank *et al.*, 2010).

While TTI reports their index annually, their methods changed from 2008 to 2009 (Schrank and Lomax, 2009, Schrank *et al.*, 2010). The 2008 TTI index is based on an estimate of the time cost of travel constructed from data in the HPMS describing road characteristics and traffic levels on

[14]In our sample of the 50 largest MSAs, the rank correlation between our preferred ranking and the alternative ranking obtained from the OLS estimates of column 2 of table 3 is 0.85 in 2001 and 0.79 in 1995 instead of 0.94 in 2008. For the 100 largest MSAs, the rank correlation between our preferred ranking and the same alternative based on OLS estimates is 0.81 in 2008. More generally, correlations drop when we use 100 MSAs instead 50.

[15]We use a standard Epanechnikov kernel function to weight the local polynomial, a rule-of-thumb bandwidth estimator, and local-mean smoothing (i.e., polynomial of degree 0 used in the smoothing).

interstate highways and other federally funded roads. The 2009 TTI, on the other hand, is based on directly observed time costs of travel on highways and major arterial roads for a self selected sample of commercial vehicles and the drivers of privately-owned vehicles. Unlike the TTI index, our index is based on a sample constructed to be broadly representative of all trips in privately-owned vehicles. In addition, the TTI indices do not attempt to control for the simultaneity in the choice of distance and speed.[16]

Aside from differences in the quality of the underlying data and methodology, the TTI indices are based on a different geography than we use (see Online Appendix A). While neither geography is intrinsically preferable, this difference complicates the comparisons of indices. For the 47 cities reported by TTI also in our sample of the 50 largest MSAs, the rank correlation of our preferred speed index is 0.69 for the inverse of the 2008 TTI travel cost index and 0.74 for 2009. For the 71 cities reported by TTI also in our sample of the 100 largest MSAs, those correlations are 0.61 and 0.63 respectively.

As mentioned above, we also duplicate our analysis for primary statistical metropolitan areas. Among the 50 largest PMSAs, 17 are the core regions of larger MSAs, 11 are secondary centers in larger MSAs, and 22 contain and coincide with their entire MSA. Changing the unit of observation makes it difficult to provide direct comparisons with our MSA based results. From the ranking of the 50 largest PMSAs (not reported here) we note the following features. Excluding New Orleans which is not among the largest 50 PMSAs and thus no longer considered, the 10 slowest MSAs are all among the 14 slowest PMSAs. The four 'entrants' in this group are Nassau-Suffolk (NY), Bergen-Passaic (NJ), Fort Lauderdale (FL), and Oakland (CA), all secondary centers from slow MSAs: New York, Miami, and San Francisco. More generally, secondary centers of slow MSAs tend to be slow as well, much slower than their population size alone would suggest.[17] This explains why the rankings of the main urban centers are very close whether we use MSAs or PMSAs. The rule is not absolute though. A few secondary centres are much faster than their main urban core. This is for instance the case of Gary (IN) and Chicago. As a result, after taking away Gary, Chicago

[16]The TTI indices are 'linked-based' as they measure the time cost of travel for specific road segments. This makes the simultaneity slightly different relative to trip-based measures like ours. Still, the major roads and interstate highways over which the time cost of travel is measured are relatively more likely to be used when travel is fast. Also, the TTI indices ignore local roads over which a significant fraction of travel occurs and for which speed may vary differently across cities.

[17]The Jersey City (NJ) PMSA is a case in point. Although too small to enter the group of the 50 largest PMSAs, it is actually the slowest PMSA in the entire country. This is unsurprising. Large parts of this PMSA have a density higher than Manhattan without its wide avenues, not to mention their heavily congested accesses to Manhattan by bridge or by tunnel.

Table 5: The determinants of speed, 100 MSAs in 2008

	(0) S_{raw}	(1) S_{OLS1}	(2) S_{OLS2}	(3) S_{OLS3}	(4) S_{FE}	(5) S_{IV1}	(6) S_{IV2}	(7) S_{IV3}	(8) S_{IV4}	(9) $S_{\text{IV FE}}$
log lane	0.073^c	0.073^a	0.083^a	0.085^a	0.064^a	0.084^b	0.090^b	0.084^b	0.096^a	0.037
	(0.037)	(0.024)	(0.023)	(0.023)	(0.023)	(0.038)	(0.040)	(0.033)	(0.036)	(0.038)
log VTT	-0.094^b	-0.094^a	-0.11^a	-0.11^a	-0.085^a	-0.13^a	-0.13^a	-0.13^a	-0.14^a	-0.064^c
	(0.036)	(0.020)	(0.020)	(0.020)	(0.020)	(0.034)	(0.035)	(0.029)	(0.033)	(0.033)
R^2	0.12	0.35	0.46	0.47	0.31	0.37	0.38	0.41	0.39	0.06

Notes: OLS regressions with a constant in all columns. Robust standard errors in parentheses. *a*, *b*, *c*: significant at 1%, 5%, 10%. 100 observations per column. In column 0, the dependent variable is average trip speed. In columns 1-9, the dependent variable is the log of the index computed from the results of the regressions reported in the corresponding column of table 3. An F-test in all columns rejects that the sum of the two coefficients is zero.

turns out to be the slowest PMSA overtaking Miami. The absence of exurban counties in large PMSAs also implies that the slowest PMSAs are slower than their corresponding MSA. For instance the speed index of Chicago PMSA is 0.83 instead of 0.91 for the Chicago MSA. Nonetheless the high correlation of the speed index between the PMSAs that compose the largest MSAS suggests that MSAS are a relevant level of analysis.

6. Determinants of speed

6.1 Determinants of speed, empirical results

Table 5 reports OLS estimations of a simple version of equation (2) for our sample of the 100 largest MSAs. In all specifications we regress a measure of travel speed on the log of MSA lane kilometers of interstate highways and major urban roads and the log of MSA vehicle travel time. In column 0, the dependent variable is a measure of log average speed taken directly from the data. In columns 1 to 9, the dependent variable is the log of the speed index computed from the results of the regressions reported in the corresponding column(s) of table 3.

Column 6 reports our preferred regression in table 5. In this regression the dependent variable is our preferred speed index calculated from the results of column 6 in table 3. This regression implies an elasticity of travel speed with respect to lane kilometers of road of 0.09. This quantity also corresponds to the share of roads in the production of travel. The coefficient for log VTT implies a negative elasticity of speed with respect to aggregate vehicle travel time of -0.13. This value of

$\theta = 0.13$ also implies that the elasticity of vehicle kilometers traveled with respect to travel time is $1 - 0.13 = 0.87$.

The estimated coefficients for the log of lane kilometers and vehicle travel time are very similar in columns 0-8 of table 5. In all specifications, the coefficient on log lane kilometers remains between 0.06 and 0.10, while that on log vehicle travel time remains between -0.09 and -0.14. Column 9 uses a speed index based on the least precisely estimated speed-distance relationship, which may create to an attenuation bias.

Online Appendix E describes extensive robustness tests for the results reported in table 5. We show that the coefficients we find on roads, on aggregate travel time, and the existence of modest decreasing returns are also robust to our choice of sample, year of data, definition of roads, and measure of aggregate travel time. In this appendix we also tackle the endogeneity of both roads and VTT.[18] In a first exercise, we instrument roads and/or VTT. In a second exercise, we implement the methodology developed by Levinsohn and Petrin (2003). Both exercises largely confirm the results of table 5.

For 2008, we also experimented with similar regressions for primary metropolitan statistical areas instead of consolidated metropolitan areas. When regressing our preferred speed index on roads and vehicle travel time, we estimate elasticities of 0.15 for roads and -0.20 for vehicle travel time. These are slightly larger magnitudes than for MSAs where the analogous elasticities are 0.09 for roads and -0.13 for vehicle travel time.

We do not know of estimates directly comparable to ours in the literature. Bombardini and Trebbi (2012) regress the TTI time cost of travel index and another TTI-generated measure of travel delay on population. They report an unconditional elasticity of speed with respect to population of -7.5% and a corresponding elasticity of 10% for travel delay. Using slightly different population data and a slightly different sample, we obtain a similar unconditional elasticity of speed with respect to population of -7.2%. Further estimates using TTI data are reported in Online Appendix E. Using a very different approach and further away from our focus, Combes and Lafourcade (2005) decompose the decline in generalized transportation costs for trucks in France over 1978-1998 and find that changes in the road infrastructure only accounts for 8% of this decline.

[18]The econometric problems with estimating production functions are well-known and have received considerable attention (e.g., Ackerberg, Benkard, Berry, and Pakes, 2007, Syverson, 2011). Very much the same issues apply here except that we observe the true cost of travel. Hence, our analysis is not subject to problems associated with unobserved prices.

Our results also relate to the large literature estimating speed-flow curves in transportation economics and engineering. Estimates for typical road segments (or for small sets of road segments) typically suggest elasticities of speed with respect to the number of vehicles of around -50 to -60% (see for instance the speed-flow curves reported in Small and Verhoef, 2007).[19] In their area study for central Yokohama, Geroliminis and Daganzo (2008) report a speed-flow curve that implies similar speed elasticities. These magnitudes are much larger than those reported in table 5 for the coefficient on vehicle travel time.

To understand these differences, it is best to think of standard speed-flow elasticities as 'micro-elasticities' whereas our estimates are 'macro-elasticities'. The first difference is that with standard speed-flow curves each observation considers the actual number of vehicles on the road at one point in time whereas our measure of vehicle time travel is averaged over time. By re-scheduling their trips drivers can avoid the worst of traffic. Consistent with this conjecture, our data indicate that peak-hours last longer in larger cities. The second difference is that drivers can also choose alternative routes when traffic gets too dense somewhere. Finally, as we emphasize above, extant estimates of speed-flow curves are subject to a simultaneity problem.

Given the analogy between our estimating equation (2) and a firm level production function, it is also interesting to compare our results about the production of travel with what is known about the production of other goods. We find that the share of roads in the production of travel is around 0.10, whereas typical estimates regarding the share of capital in conventional sectors tend to be around one third. Notwithstanding the fact that standard production function estimation ignores consumer time as an input, this suggests that the production of travel is an extremely labour (or time) intensive activity.

We also find evidence of slightly decreasing returns. In all columns of table 5 except for column 0 and column 9 (for which the coefficients are perhaps less reliably estimated) an F-test soundly rejects that the sum of the two coefficients is zero. We note that these calculations of decreasing returns do not take into account the fact that building a lane kilometer of interstate highways of major urban roads in larger cities is considerably more expensive (Ng and Small, 2012). Using a measure of cost of capital instead of a measure of roads suggests that construction costs may

[19]Strictly speaking, traditional estimates of the elasticity of speed with respect to the number of vehicles estimate $\theta/(1-\theta)$. Recall that we regress speed on vehicle travel time to estimate θ. In turn, vehicle travel time is proportional to the number of vehicles and their speed (keeping trip length constant). Hence, the traditional speed flow regressions of travel speed on the number of vehicles estimate $\theta/(1-\theta)$. This makes little difference in our context given that our preferred value of θ is 0.13.

increase more rapidly with scale than productivity decreases.[20] Theoretically, one may think of a variety of reasons why there could be decreasing returns to scale in the provision of road travel. As suggested previously by Strotz (1965), Mohring (1976) and Kraus (1981), a greater density of roads implies a more than proportional increase in intersections, arguably a source of decreasing returns. One may also imagine that bottlenecks in the more central part of cities are worse in larger cities.[21] By contrast, the literature that estimates the production function for firms contains a variety of results but micro-data estimates are often suggestive of constant returns.

Decreasing returns in the provision of travel has important implications for the financing of highways, and more specifically to the debate about whether optimally priced highways are self-financing (Mohring and Harwitz, 1962). Although existing studies suggest mild economies of scale in road construction, which almost certainly disappear for the largest, densest cities, we are the first to estimate returns to scale in the congestion technology for a metropolitan area wide network. If decreasing returns in congestion dominate economies of scale in construction, then optimally priced networks would be revenue generating.

Finally, consistent with extant research on productivity in firms, we find considerable dispersion in productivity across cities. However, there is much less dispersion in the ability of US cities to produce travel out of roads and vehicle travel time than in the ability of firms to produce output from capital and labour. A city at the 90th percentile in our preferred estimation (column 6 of table 5) produces around 20% more travel from the same inputs than a city at the 10th percentile. Firm level data usually imply that a firm at the 90th percentile of productivity produces 100 to 200% more output from the same inputs than a firm at the 10th percentile (Syverson, 2011, Fox and Smeets, 2011).

Table 6: The determinants of speed, further explanatory variables

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Added:	Emp. central.	Pop. central.	Job/resid. mismatch	$\mathbb{E}$ pop. growth	log pop. 1920	s. manuf. emp.	Cooling deg. days	Heating deg. days
log lane (total)	0.085^b	0.077^b	0.098^b	0.071^c	0.072^c	0.070^c	0.087^b	0.080^c
	(0.039)	(0.036)	(0.039)	(0.042)	(0.040)	(0.042)	(0.040)	(0.041)
log VTT	-0.15^a	-0.14^a	-0.14^a	-0.11^a	-0.14^a	-0.11^a	-0.13^a	-0.13^a
	(0.033)	(0.031)	(0.035)	(0.037)	(0.035)	(0.037)	(0.035)	(0.036)
Added variable	-0.23^b	-0.31^a	-0.15^b	-0.20^a	0.29^a	0.25^a	-0.18^a	0.17^b
	(0.107)	(0.113)	(0.070)	(0.058)	(0.079)	(0.066)	(0.067)	(0.069)
R^2	0.40	0.42	0.41	0.42	0.43	0.45	0.42	0.41

Notes: OLS regressions with a constant in all columns. Standardized coefficient reported for the added variable. Robust standard errors in parentheses. *a*, *b*, *c*: significant at 1%, 5%, 10%. 100 observations per column. Dependent variables is log S_{IV2} in all columns.

6.2 *Other determinants of speed*

In table 6 we consider a broader set of determinants of travel speed. In column 1, we introduce a measure of employment centralization, the share of employment within 20 kilometers of the employment weighted centroid of the MSA in 1992.[22] In column 2, we use instead the corresponding measure of centralization for population. The results for both columns indicate that more centralized cities are slower. These findings should be regarded as suggestive since the R^2 increases only marginally relative to the benchmark estimation without these additional variables and the significance of these coefficients often disappears when further controls are added. In column 3, we use a measure of mismatch between employment and residents and find again that a greater mismatch is associated with slower traffic speeds. We have experimented more broadly with 'urban form' variables than we report here. Consistent with the results reported in table 6, we find that conditional associations with various measures of density, physical area and employment

[20] Using data from Duranton and Turner (2012) combined with highway construction costs from Ng and Small (2012), we obtain 0.37 as a rough estimate elasticity of the cost of construction of a lane kilometer of highway with respect to city population. We can then calculate a measure of highway capital cost for each MSA by multiplying lane kilometers by population elevated to the power 0.37. When we re-estimate the regressions of table 5 with this measure of road capital cost instead of lane kilometers, we obtain marginally lower coefficients for roads and for vehicle travel time. As a result, decreasing returns go from about -4% in table 5 to about -6% to -7%. More precise estimates of highway costs are likely to make returns even more strongly decreasing. These two forms of decreasing returns are conceptually different. Our main exercise is concerned with measuring how output varies with the quantity of inputs whereas this second exercise is about costs.

[21] Be it only because the central part of larger cities is denser. This prediction of standard urban models (Alonso, 1964, Mills, 1967, Muth, 1969) is strongly supported in our data.

[22] We prefer to use lagged variables to minimize endogeneity problems. These lags typically reduce the significance of the coefficients.

concentration, occur routinely in our results. Consistent with earlier results in the literature, (e.g., Glaeser and Kahn, 2004), these findings suggest that more compact, centralized cities are slower. Of course, compact centralized cities could still be desirable even if travel within them is more costly. Speed is not synonymous with optimality.

In column 4 of table 6, we turn to a different type of variable, population growth. We find a strong association between slow travel speeds and higher expected growth between 1980 and 2000.[23] A similar but weaker association is found with actual population growth between 1980 and 2008. That traffic should be slower in cities that have grown fast even after controlling for their current roadway and vehicle travel time is perhaps unsurprising. Roadway expansion often takes place at the urban fringe while bottlenecks worsen in the more central parts. In column 5, we replace expected population growth with 1920 population. Given that we also condition for current vehicle time travel, which is highly correlated with current population, our negative coefficient implies that cities that have grown less since 1920 are faster. This confirms the finding of column 4. In column 6, we turn to the share of manufacturing in employment in 1983 and find that cities more specialized in manufacturing are much faster. Given that cities with a greater share of manufacturing employment have grown less in population and tend to be more decentralized (Glaeser and Kahn, 2001), the positive sign on manufacturing employment is consistent with our two main findings so far.

In columns 7 and 8, we introduce two measures of temperature, cooling and heating degree days, and uncover a weak association between slower traffic and more extreme weather conditions. We also experimented with other geographic characteristics of cities such as their elevation range or the ruggedness of their terrain but found nothing. We also found no result for a broad range of socioeconomic characteristics of cities such as their income, education, etc. Finally, we note that introducing all these supplementary explanatory variables has little effect on the coefficients of our two main regressors, roads and vehicle travel time, both of which remain significant and nearly constant in all the columns of table 6.

In table 7, we investigate the relationship between road network configuration and speed. Following Baum-Snow, Brandt, Henderson, Turner, and Zhang (2012) we construct measures of ring road and radial road capacity of two networks for each MSA, the interstate highway network

[23]We use growth predicted by the composition of economic activity of cities in 1980 and the subsequent changes in employment by sector.

Table 7: The determinants of speed, road network variables

	(1) Rings IH	(2) Rays IH	(3) Both IH	(4) Both IH+MR	(5) IH per km	(6) IH in log	(7) + pop central.	(8) + s. manuf. emp.
log lane (total)	0.072^c	0.079^c	0.071^c	0.081^b	0.088^b	0.071^c	0.059	0.073^c
	(0.042)	(0.041)	(0.042)	(0.040)	(0.043)	(0.042)	(0.039)	(0.043)
log VTT	-0.13^a	-0.13^a	-0.13^a	-0.13^a	-0.13^a	-0.13^a	-0.14^a	-0.11^a
	(0.036)	(0.036)	(0.036)	(0.035)	(0.037)	(0.036)	(0.032)	(0.037)
Ring index	0.019^a		0.016^b	0.0029^c	0.12^b	0.039^b	0.036^b	0.035^b
	(0.0068)		(0.0076)	(0.0017)	(0.058)	(0.016)	(0.015)	(0.014)
Rays index		0.0081^c	0.0046	-0.0018	0.014	-0.0026	0.0012	-0.0017
		(0.0046)	(0.0051)	(0.0014)	(0.043)	(0.016)	(0.014)	(0.013)
Added variable							-0.17^b	0.25^a
							(0.066)	(0.071)
R^2	0.42	0.40	0.43	0.40	0.41	0.42	0.46	0.48

Notes: OLS regressions with a constant in all columns. Robust standard errors in parentheses. *a*, *b*, *c*: significant at 1%, 5%, 10%. 100 observations per column. Dependent variables is log S_{IV2} in all columns. In columns 1, 2, and 3, the ring and rays indices are computed as described in Online Appendix A from the 2005 interstate network. In column 4, the ring and rays indices are computed from the 2005 network of interstate highways and major urban roads. In column 5, we normalize the ring and rays indices (from the 2005 interstate network) by lane kilometers of interstate highways and major urban roads (measured in thousands). In column 6, we take the log of one plus the ring and rays indices (from the 2005 interstate network). In columns 7 and 8, we enrich the specification of column 3 with an index of population centralisation and the share of manufacturing employment, respectively.

and the union of this road network with all major urban roads, both for 2005. Online Appendix A provides details about the construction of these variables. In column 1, we add a ring road index computed from the 2005 interstate highway network to our preferred specification of table 5, column 6. In column 2, we use instead an index of 2005 interstate highway rays. In column 3, we use both indices at the same time. The rays index becomes insignificant. In contrast, the ring road index remains significant and indicates that speed is about 1.5% higher in cities where the ring road index is one unit higher. Since a unit increase in the ring road index reflects a road traveling about one fourth of the way around a city, the addition of a complete ring road is associated with about a 6% increase in speed. We note that the coefficient on vehicle travel time is virtually unchanged and the coefficient on lane kilometers is marginally lower.

In column 4, we consider the same two indices but compute them for the entire network of interstate highways and major urban roads. The coefficient on ring roads becomes much smaller. Next, we return to the rays and ring roads indices computed solely from the interstate highway network but normalise them by kilometers of interstate highways in column 5 or take them in logs in columns 6. In both cases, the results are unchanged and the ring road index remains significant.

That ring roads appear to matter in explaining traffic speed is reminiscent of the results of table 6 that show the importance of characteristics describing urban decentralisation. To compare both network and urban form variables, in column 7 we add a measure of centralization for population to the specification of column 3, while in column 8 we add the 1983 share of manufacturing employment. Interestingly the coefficients on these two variables are very close to those of table 6 while the coefficient on the ring road index more than doubles relative to column 3. This last set of results suggests that both city characteristics and road network characteristics play a role in explaining traffic speed.

Overall, the results of table 7 suggest an important role for highway network characteristics, and ring roads in particular, in explaining speed. Some caution is nonetheless needed because we cannot establish causality. We note nonetheless from results not reported here that using rays computed from planned 1947 interstate highways yields similar estimates to those reported in table 7.[24]

7. The value of speed

Section 3 presents expressions for the deadweight loss from congestion and for the optimal congestion tax. We now turn our attention to evaluating these expressions.

Our preferred estimate of the average supply elasticity θ is 0.13 in section 6. Using individual data, Duranton and Turner (2011) suggests that the demand for VKT is highly elastic and provides a point estimate of 16.[25] We note that this number describes driving on interstate highways. For other roads this elasticity may be lower. On the basis of case studies of two particular cities (Small and Verhoef, 2007, , p.11) suggest values for σ a bit less than 1.

Table 8 reports calculated values of equation (7) using different values of θ and σ. In the top three rows, we use our preferred value of $\theta = 0.13$ and range of values for σ that reflects our uncertainty over this value; our preferred value of 16 from Duranton and Turner (2009), the much smaller value suggested by case studies,[26] and to check robustness, double the Duranton and

[24]There are no ring roads in the 1947 highway plan.

[25]The working paper version of that work (table 6, columns 4 and 8 in Duranton and Turner, 2009) contains an explicit estimation of the demand for VKT which yields $\sigma \approx 16$.

[26]Actually, Δ is undefined for $\sigma = 1$. We report calculations based on $\sigma = 1.01$.

Table 8: Estimates of the value of congestion and the optimal congestion tax

θ	σ	Δ	DWL ($\$\times 10^9$)	DWL (hours/person)	τ^* ($/km)
0.13	16	0.040	29.8	12.4	0.035
0.13	1	0.009	6.7	2.8	0.038
0.13	32	0.045	33.6	14.0	0.035
0.07	16	0.017	12.6	5.2	0.019
0.19	16	0.065	49.2	20.5	0.052

Turner (2009) value. When $\theta = 1$ and $\sigma = 16$, Δ is about 0.040. That is, the value of deadweight loss from congestion is equal to about 4% of total travel time. Taking the aggregate travel time of 87.3m minutes for 2008 from panel B of table 1, converting from minutes per day to hours per year, and adjusting upwards by 1.25 persons per car, we find 26 million hours per year of deadweight loss due to congestion in an average MSA in our sample, or about 12.4 hours per person. If, following common practice in the transportation economics literature, we value these hours at half the average wage of 23 dollars per hour, this is 142 dollars per person per year or 29.8 billion dollars for our entire sample of 100 MSAs.[27].

In the second row, we consider our preferred estimate of θ and less elastic demand, $\sigma = 1$. This reduces our estimate of the cost of congestion to about 1% of travel time and deadweight loss to 6.7 billion dollars. In the third row, we consider an even more elastic travel demand, $\sigma = 32$. This slightly increases our estimate of the cost of congestion to 4.5% of travel time. In rows 4 and 5 of table 8 we consider $\sigma = 16$ and consider an estimates of the supply elasticity θ about two standard errors above and below our preferred estimate. Unsurprisingly, deadweight loss is larger as supply is less elastic. In sum, this table suggests that the value of congestion is on the order of tens of billions of dollars per year, though uncertainty about the magnitude of the demand elasticity implies precludes much more precision.

Our estimates of annual deadweight loss are conservative in that we only consider the time cost of travel and ignore fuel and other car usage costs, e.g., maintenance, pollution and collisions for which estimates are reviewed in Parry, Walls, and Harrington (2007). We also consider only travel by residents and ignore commercial traffic for which the time and fuel costs of congestion

[27]Existing estimates of value of time traveled for commuters generally center around 50% of an individual's hourly wage (Small and Verhoef, 2007). We note that the Texas Transportation Institute uses a much higher number of nearly 24 dollars per hour per person (Schrank *et al.*, 2010)

are arguably higher.[28] In addition, our estimates are based only on the population of the 100 largest MSAS. While these MSAS account for about two thirds of the US population, and presumably, most of the traffic congestion, to the extent that congestion occurs outside of these 100 MSAS, our estimates are too low. Finally, we expect drivers to change their time of departure to avoid congestion.[29] Such re-scheduling is costly and is not reflected in our calculations.

We can also evaluate the optimal congestion tax that we calculate in equation (8). Since equation (8), gives τ^* in minutes per kilometer, while congestion taxes are usually stated in cents per kilometer, we convert from minutes to dollars using the same 11.5 dollars per hour cost of time that we used above. We see from the top row of table 8 that evaluating this tax at equilibrium quantities with $\theta = 0.13$ and $\sigma = 16$ gives $\tau^* = 0.035$ dollars, or 3.5 cents, per kilometer. This tax is almost unchanged in response to changes in the demand elasticity, but more dramatically in response to large changes in the supply elasticity θ.

In their review of the literature, Parry *et al.* (2007) recommend a value of 3.2 cents per kilometer on the basis of 'back-of-the-envelope calculations by the US Federal Highway Administration. Parry and Small (2005) propose using a range from of 1.7 to 3.2 cents per kilometer with, as they mention, "a considerable range of uncertainty".[30] Morhing (1999) suggests a value of 2.5 cents per kilometer for Saint Paul and Minneapolis. Studies with a more limited focus often suggest larger congestion taxes. For instance Keeler and Small (1977) find a range of values between 6 and 22 cents per kilometer for rural highways and between 87 and 112 cents per kilometer for city highways.

The congestion tax estimates presented in table 8 range from about 2 to 5 cents per kilometer, and are always close 3.5 cents per kilometer for our preferred value of the supply elasticity, $\theta = 0.13$. Three comments about this finding are required. First, we calculate an average congestion price over roads and time. Our relatively small congestion tax reflects the fact, suggested by the average travel speeds of table 1, that most trips occur at uncongested times and places. Second, our methodology fully reflects responses to congestion that are not permitted in other estimates. In

[28] To fix ideas, according to Duranton and Turner (2011), the share of trucks traffic for interstate highways in 228 large US MSAS was 13% in 2003. In the calculations of the Texas Transportation Institute the cost of truck congestion represents slightly more than a quarter of the total cost of congestion. As for fuel losses, they represent only a fraction of the time cost of congestion. To see this, consider an hour lost to drive 10 kilometers. With a car consuming 15 liters per 100 kilometers, this is only 1.5 liters of fuel or about 1.5 dollars at one dollar per liter. This is small relative to a time cost of 11.5 dollars per hour.

[29] Trip scheduling is the subject of a large literature, e.g., Vickrey (1969), Noland and Small (1995).

[30] The Parry and Small gas tax estimate is adjusted downward to reflect the fact that part of the response to the tax involves increased fuel economy. Absent this adjustment, their estimates would be somewhat larger.

particular, we calculate the cost of congestion conditional on drivers optimizing against all possible reroutings and changes in trip times. Ours is a congestion price to be applied to all roads at all times. That this quantity is smaller than taxes we might apply to important roads at busy times is not a surprise. On the other hand, that such a crude tax should yield a welfare gain measured in tens of billions of dollars per year is more surprising and indicates the economic importance of traffic congestion.

It is also of interest to compare the magnitude of our optimal congestion tax with the magnitude of the gasoline tax. In 2008, gasoline taxes in the US were about 40 cents per gallon.[31] At 20 miles (or 32 kilometers) per gallon, our 2 to 5 cents per kilometer congestion price suggests a gasoline tax of 62 to 160 cents per gallon. Since the current federal gasoline tax is a user fee intended to pay for the construction and maintenance of the highway system, in order to provide the same level of highway funding and address congestion, albeit very crudely, gasoline taxes in the US would need to increase by a factor of 2.5 to 4.

Expansions of the road network are often proposed as a response to congestion. To assess this policy we compare the change in social surplus with construction costs. In figure 2, social surplus at VKT_1^{eq} is the integral from zero to VKT_1^{eq} of the difference between the demand and marginal cost curves. As a consequence of the fact that access to the road network is unpriced, $C^{eq} = AC(\text{VKT}^{eq})$ in equilibrium. Because the total cost of VKT^{eq} is $\text{VKT}^{eq} \times AC(\text{VKT}^{eq})$, it follows that in any equilibrium where access to the road network is unpriced, the value of social surplus is simply consumers surplus, the area under the demand curve but above the price line. Thus, the equilibrium change in social surplus resulting from a decrease in the cost of transportation is simply the corresponding change in consumers surplus.

For small changes in price, we can approximate the resulting change in consumers surplus with the product of the price change and the ex ante equilibrium quantity. To calculate the roads elasticity of equilibrium speed, ϵ_{SR}, we use the demand equation (10) and $\text{VKT} = S \times \text{VTT}$ to solve for $\epsilon_{SR} = \alpha/(1+\theta(\sigma-1))$. Evaluating at $\theta = 0.13$, $\alpha = 0.09$ and $\sigma = 16$, we have $\epsilon_{SR} \approx 0.03$. That is, a 1% increase in a city's stock of road causes about 0.03% increase in equilibrium speed.

With a mean daily MSA VKT of 64.2 million, this increase in speed implies an increase in consumers surplus of about 162,000 hours annually. Valuing this time at 11.5 dollars per hour and

[31] Federal gasoline taxes in the US were 18.4 cents per gallon in 2008 and the average state gasoline tax was about 20 cents per gallon (`http://www.fhwa.dot.gov/policyinformation/statistics/2014/mf205.cfm`, February 2016) for a total of 38.4 cents per gallon.

scaling by 100 to calculate a total for our whole sample, gives an increase in consumers surplus of about 186 million dollars per year.[32] When $\sigma = 1$, the corresponding increase in consumer surplus is 550 million dollars per year, about three times as large.

Using data from Duranton and Turner (2012) based on estimates of road construction costs by Ng and Small (2012) and maintenance costs from the US Bureau of Transportation Statistics (2007), increasing the supply of interstate highways by 1% in all of the 100 largest MSAs has an annual cost of 1.45 billion dollars. It is much harder to know the cost of expanding other major urban roads by 1%. Panel A of table 1 indicates that they represent five to six times as many lane kilometers. Even if major urban roads are much cheaper to build, a total annual cost of 3 billion dollars for a 1% expansion of the road network is most likely a lower bound. This is many times larger than the estimated benefits.

Comparing this calculation of costs and benefits suggests that the alleviation of congestion is probably not a sufficient justification for wholesale expansions of the road network. Speed is simply not sensitive enough to changes in the road network. With this said, there are other reasons to build roads and the alleviation of congestion might help to justify road construction in some cases, particularly for specific bottlenecks or in places less well provided with roads than our sample of major US cities. We note that this calculation confirms the conclusion of a similar calculation performed in Wheaton (1978).

8. Conclusion

Road transportation accounts for 18% of the budget for an average US household and about 72 minutes per day for an average US driver in 2008. In a typical year, the US spends about 150 billion dollars on road construction and maintenance. Congestion naturally arises in the course of turning these resources into travel. Despite a distinguished history of research in transportation economics, the extant literature has struggled to measure the value of investments in road transportation or the social cost of road congestion.

We make three advances in this agenda. First, we develop an econometric methodology and data to identify city level supply curves for trip travel. With these supply curves, we construct an index of travel speed for each of the largest US cities. This index number, for the first time,

[32] With $\epsilon_{SR} = 0.03$, at 87.3 m minutes per msa-day, an increase in speed of 0.03% implies a reduction in time of $87.3\text{m} \times \frac{0.03}{100} \times \frac{365}{60} \approx 159{,}000$ hours per MSA/year. Multiplying this by 11.5 \$/hour and by 100 MSAs gives the final number.

provides a theoretically founded measure of the efficiency with which different locations produce transportation.

Our investigation of the determinants of our speed index suggests that the production of transportation at the city level is subject to slight decreasing returns to scale. This finding provides an empirical basis for the positive relationship between city size and congestion costs that is axiomatic in nearly all extant theoretical models of city size. It also reveals the role of roads, travel time, and unobserved productivity. The presence of economically important variation in the unobserved productivity of cities at producing transportation suggests the possibility of large gains in efficiency if slow cities are able to emulate fast cities. We conduct a rudimentary investigation of city characteristics associated with efficiency. Our findings are unsurprising: dense centralized cities do not allow automobiles to travel at a high speed, while cities with ring roads do allow automobiles to travel at high speeds. Refining this investigation is an important topic for further research and may provide an empirical basis for the design of cities where travel is provided more efficiently.

Our investigation of the efficiency of travel amounts to the comparison of equilibrium speed with counterfactual, out-of-equilibrium scenarios. In this, it resembles the widely known travel cost indices published by TTI. We note however that the congestion costs estimated by TTI do not evaluate an economically meaningful definition of congestion. Such a measure of congestion is intrinsically subtle, and requires the calculation of the deadweight loss incurred at equilibrium levels of travel as opposed to optimum. Our estimates allow us a rough, and highly aggregated calculation of the deadweight loss from congestion. This calculation indicates that the losses from congestion are probably even larger than those suggested by counterfactual supply improvements. Since equilibrium responses dissipate much of the benefit from supply improvements, our results suggest that the largest gains in transportation policy can be obtained by managing demand.

References

Ackerberg, Daniel, C. Lanier Benkard, Steven Berry, and Ariel Pakes. 2007. Econometric tools for analyzing market outcomes. In James J. Heckman and Edward E. Leamer (eds.) *Handbook of Econometrics*, volume 6A. Amsterdam: North-Holland, 4271–4276.

Alonso, William. 1964. *Location and Land Use; Toward a General Theory of Land Rent*. Cambridge, MA: Harvard University Press.

Bartik, Timothy. 1991. *Who Benefits from State and Local Economic Development Policies?* Kalamazoo (MI): W.E. Upjohn Institute for Employment Research.

Baum-Snow, Nathaniel. 2007. Did highways cause suburbanization? *Quarterly Journal of Economics* 122(2):775–805.

Baum-Snow, Nathaniel, Loren Brandt, J. Vernon Henderson, Matthew A. Turner, and Qinghua Zhang. 2012. Roads, railroads and decentralization of Chinese cities. Processed, University of Toronto.

Bombardini, Matilde and Francesco Trebbi. 2012. Gains from distortions in congested city networks. Processed, University of British Columbia.

Burchfield, Marcy, Henry G. Overman, Diego Puga, and Matthew A. Turner. 2006. Causes of sprawl: A portrait from space. *Quarterly Journal of Economics* 121(2):587–633.

Combes, Pierre-Philippe and Miren Lafourcade. 2005. Transport costs: Measures, determinants, and regional policy implications for France. *Journal of Economic Geography* 5(3):319–349.

Duranton, Gilles and Matthew A. Turner. 2009. The fundamental law of road congestion: Evidence from US cities. NBER Working Paper 15376.

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. 2012. Urban growth and transportation. *Review of Economic Studies* 79(4):1407–1440.

Evans, William N., Wallace E. Oates, and Robert M. Schwab. 1992. Measuring peer group effects: A study of teenage behavior. *Journal of Political Economy* 100(5):966–991.

Fox, Jeremy T. and Valérie Smeets. 2011. Does input quality drive measured differences in firm productivity? *International Economic Review* 52(4):961–989.

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.

Glaeser, Edward L. and Matthew Kahn. 2001. Decentralized employment and the transformation of the American city. *Brookings-Wharton Papers on Urban Affairs* :1–47.

Glaeser, Edward L. and Matthew E. Kahn. 2004. Sprawl and urban growth. In Vernon Henderson and Jacques-François Thisse (eds.) *Handbook of Regional and Urban Economics*, volume 4. Amsterdam: North-Holland, 2481–2527.

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.

Kraus, Marvin. 1981. Scale economies analysis for urban highway networks. *Journal of Urban Economics* 9(1):1–22.

Levinsohn, James and Amil Petrin. 2003. Estimating production functions using inputs to control for unobservables. *Review of Economic Studies* 70(2):317–342.

Meyer, John R., John F. Kain, and Martin Wohl. 1965. *The Urban Transportation Problem*. Cambridge (MA): Harvard University Press.

Mills, Edwin S. 1967. An aggregative model of resource allocation in a metropolitan area. *American Economic Review (Papers and Proceedings)* 57(2):197–210.

Mohring, Herbert. 1976. *Transportation Economics*. Cambridge (MA): Ballinger Pub. Co.

Mohring, Herbert and Mitchell Harwitz. 1962. *Highway Benefits: An Analytical Framework*. Evanston (IL): Northwestern University Press.

Morhing, Herbert. 1999. Congestion. In José A. Gómez-Ibáñez, William B. Tye, and Clifford Winston (eds.) *Essays in Transportation Economics and Policy: A Handbook in Honor of John R. Meyer*. Washington DC: Brookings Institution Press, 181–222.

Muth, Richard F. 1969. *Cities and Housing*. Chicago: University of Chicago Press.

Ng, Chen Feng and Kenneth A. Small. 2012. Tradeoffs among free-flow speed, capacity, cost, and environmental footprint in highway designs. *Transportation* 39(6):1259–1280.

Noland, Robert B. and Kenneth A. Small. 1995. Travel-time uncertianty, departure time choice, and the cost of the morning commute. Processed, University of Californa at Irvine.

Parry, Ian W. H. and Kenneth A. Small. 2005. Does Britain or the United States have the right gasoline tax? *American Economic Review* 95(4):1276–1289.

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

Schrank, David and Tim Lomax. 2007. Urban mobility report 2007. Technical report, Texas Transportation Institute.

Schrank, David and Tim Lomax. 2009. Urban mobility report 2009. Technical report, Texas Transportation Institute.

Schrank, David, Tim Lomax, and Shawn Turner. 2010. Urban mobility report 2010. Technical report, Texas Transportation Institute.

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

Strotz, Robert H. 1965. Urban transportation parables. In Julius Margolis (ed.) *The Public Economy of Urban Communities*. Washington (DC): Resources for the Future, 127–169.

Syverson, Chad. 2011. What determines productivity. *Journal of Economic Literature* 49(2):326–365.

United States Bureau of Transportation Statistics. 2007. *Transportation Statistics Annual Report 2007*. Washington, DC: US Government printing office.

United States Bureau of Transportation Statistics. 2013. *Transportation Statistics Annual Report 2013*. Washington, DC: US Government printing office.

United States Department of Transportation, Federal Highway Administration. 2003*a*. *Archive Highway Performance Monitoring System (HPMS) Data Item Descriptions 1993 - 1998*. Washington, DC.

United States Department of Transportation, Federal Highway Administration. 2003*b*. *Archive Highway Performance Monitoring System (HPMS) Data Item Descriptions 1982-1987*. Washington, DC.

United States Department of Transportation, Federal Highway Administration. 2005. HPMS *Field Manual*. Washington, DC.

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.

Wheaton, William C. 1978. Price-induced distortions in urban highway investment. *Bell Journal of Economics* 9(2):622–632.

Speed: Online appendix

19 December 2016

Victor Couture[ab]

University of California, Berkeley

Gilles Duranton[1c]

University of Pennsylvania

Matthew A. Turner[1d]

Brown University

[a]We thank two anonymous referees, the editor, Vincent Breteau, Pierre-Philippe Combes, Kerem Cosar, Marc Gaudry, Ed Glaeser, Andy Haughwout, Diego Puga, Matti Sarvimäki, Ken Small, Will Strange, Francisco Trebbi, and conference and seminar participants for insightful comments. Financial support from the Canadian Social Science and Humanities Research Council is gratefully acknowledged.

[b]Haas School of Business, University of California, Berkeley, 2220 Piedmont Avenue, Berkeley, CA 94720, USA (e-mail: `couture@haas.berkeley.edu`).

[c]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 Centre for Economic Policy Research.

[d]Department of Economics, Box B, Brown University, Providence, RI 02912 (e-mail: `matthew_turner@brown.edu`; website: `http://www.economics.utoronto.ca/mturner/index.htm`). The hospitality and support of the Enaudi Institute of Economics and Finance and the Property and Environmental Research Center are gratefully acknowledged.

Appendix A. Data

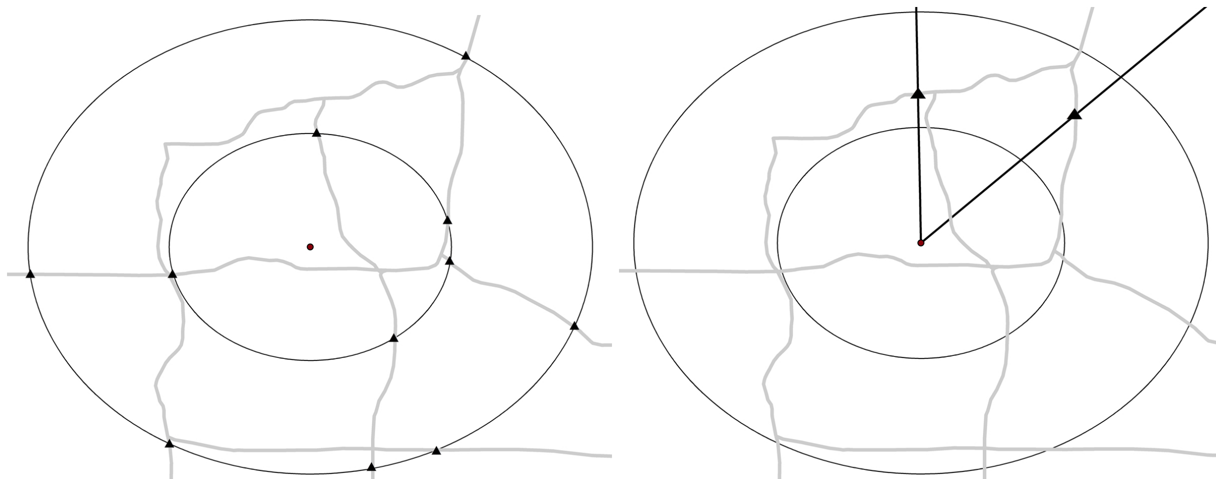
Consistent MSA *definition:* MSAS are defined as aggregations of counties. We use the 1999 definition of consolidated metropolitan statistical areas (MSA) and, in some robustness checks, primary metropolitan statistical areas (PMSAS). For 2008-2009, NHTS data contain a county identifier that allows us to identify the corresponding MSA and PMSA. For 1995-1996 and 2001-2002, the NPTS/NHTS data only contain an MSA identifier consistent with our MSA definition. Other MSA level variables are constructed either from administrative data reported at the county level, or from GIS data assigned to MSAS on the basis of an electronic map of 1999 county or MSA boundaries.

NHTS *data (trip-level data):* Our trip-level data are from the 1995-1996 National Personal Transportation Survey (NPTS) and its successors, the 2001-2002 and 2008-2009 National Household Transportation Surveys (NHTS). The surveys are sponsored by various agencies at the US Department of Transportation. Detailed documentation on the NHTS sample design and data content is available on the NHTS website (`http://nhts.ornl.gov/`). The data are intended for use by transportation planners in governmental agencies and aim at providing reliable, representative and comprehensive micro-data about the daily travel of Americans.

NHTS *aggregate data:* As explained in the text, the general idea behind our method for computing MSA totals (VKT and VTT) from trip level data is to multiply averages for individual travel in each MSA by population size. In practice however, this simple multiplication could lead to biased estimates, because of how we remove outliers.

We therefore proceed as follow. Our sample consists of all trips entered by the driver of a car, van, SUV, pick-up or other truck in the 100 largest MSAS. We identify the 0.5% of trips with the longest distance and the 0.5% of trips with the shortest distance, and eliminate all trips by an individual who entered any of these extreme values. This is because including individuals with an incomplete trip schedule would bias any averages that we compute later. We use a similar procedure to remove outliers for trip time and speed. The individuals that we drop from the sample are more likely to be drivers, so to obtain population totals out of average distance and time in this clean sample, we need to compute, prior to eliminating outliers, the share of individuals who drive at least one trip in each MSA. Then, we multiply this share by MSA adult population to obtain an estimate of the total number of people who drive at least one trip in each MSA. Finally, to

Figure 4: Illustrations of ring and radial road algorithms for Oklahoma City in 2005.



Left panel illustrates radial road algorithm. Right panel illustrates ring road algorithm.

obtain an estimate of VTT and VKT in each MSA, we just take the average time and distance traveled from the clean sample of drivers, and multiply it by the population of drivers.

As an example, suppose that in a given MSA, the NHTS contains 5 individuals driving at least one trip, out of 10 individuals in our sample for that MSA. If MSA adult population is equal to 100, then our estimate of the number of drivers in this MSA is $(5/10) \times 100 = 50$. Now suppose that for the drivers who remain in the sample after removing outliers, average distance driven is equal to 20 kilometers. Then our estimate of VKT in this MSA is $50 \times 20 = 1000$ vehicle kilometers traveled.

HPMS *data (road infrastructure):* To estimate our travel production function, we rely on the Highway Performance Monitoring System (HPMS) data for 1995, 1996, 2001, 2002, and 2008. These data are collected and maintained by the US Federal Highway Administration in cooperation with many sub-national government agencies. Documentation is available in several reports from the Federal Highway Adminstration (DOT, 2003*a*, 2003*b*, and 2005). See also Duranton and Turner (2011).

For the interstate highway system, the HPMS records number of lanes, length, average annual daily traffic (AADT), and county. By construction, road segments do not cross county borders. For segments in urbanized areas, the HPMS also provides an urbanized area code. Since MSAs are county based units, these data allow us to calculate lane kilometers for the urbanized and non urbanized area interstate systems by MSA.

Road network measures: Our ring road and radial road index variables are based on the 2005

National Highway Performance Network (NHPN) map of the US road network. This map, which complements the entirely tabular HPMS data, describes all interstate highways in the US and a subset of other major roads. From the source (NHPN) we construct two subnetworks. The first consists of all open segments of the interstate highway network, the second consists of all open segments of interstate highway, along with all principal arterial roads and urban freeways or expressways. These two networks correspond closely to the measures of interstate highways and major urban roads (including interstate highways) that we construct from the HPMS.

For each of these networks we calculate a measure of ring road capacity and a measure of radial road capacity. As their names suggest, these indices measure the ability of the network to carry traffic radially from the center and circumferentially around the center. Since both concepts fundamentally rely on knowing where the center of each city is, our first step is to locate the centers of our MSAs. To accomplish this, we use the 2000 zipcode business patterns data and the 2000 census zipcode boundaries files to identify the zipcode in each MSA with the highest employment density.[33]

Once we have identified the center of each MSA we are able to calculate our ring and radial road indexes. The left panel of figure 4 illustrates the calculation of our radial road index for the Oklahoma City MSA and the interstate highway network. To calculate the radial road index, we first draw two circles centered on the CBD, the first with radius 5 kilometers and the second with radius 10 kilometers. We then calculate the number of times the road network intersects each ring. The minimum of these two values is our radial road index, which in this case is 5. If doing this calculation by 'eyeball' one would probably struggle over whether there were 4 or 5 radial roads. Calculating the ring road index is more involved, and is illustrated in the right panel of figure 4. To begin, we draw two rays out from the center, one due North and the other Northwest. We next calculate the number of times the road network intersects each ray in the interval between 5 and 9 kilometers from the center. The minimum of the these two is the radial road index for this quadrant in the 5 to 9 kilometers donut. In the right panel of figure 4, this value is one. We then replicate this exercise for larger donuts, 9-15 kilometers and 15-25 kilometers, and sum over all three donuts to get the radial road index for the first quadrant. The size of the three disks is chosen to preserve proportions as we scale up the size of the disks. To calculate our ring road index, we

[33]The census maintains 'zipcode' boundary files based on census blocks. They do not exactly correspond to the zip codes boundaries used by the post office and are more stable over time.

replicate this procedure in each quadrant and sum over quadrants. We note that these indexes are based on logic first developed in Baum-Snow *et al.* (2012).

Population and employment data: Population data for 1995-1996, 2001-2002, and 2008-2009 is obtained from annual population estimates provided by the US census for these years. These estimates are themselves based on interpolations and extrapolations of population counts made by the US census. In some regressions we also use population data dating back to 1920, the first census year which allows us to retain our samples of MSAs. We also use employment data from the County Business Patterns to build two variables for our exploration of the determinants of speed. The first is that of manufacturing employment. The second is an exogenous measure of MSA employment growth which interacts the sectoral composition of economic activity in an MSA in 1980 with the national growth of these sectors between 1980 and 2000.

Climate: We use two measures of temperature in US MSAs taken from the data used by Burchfield, Overman, Puga, and Turner (2006).

Urban form: We have available the data underlying Burchfield *et al.* (2006). These data provide fine scale employment data from 1994 zipcode business patterns and from 1990 tract level census data: for every 990 meter $\times$ 990 meter cell in a regular grid covering the whole of the continental US, these data report imputed 1990 population and 1994 employment. These data allow us to calculate the measures of urban form used in the first three columns of table 6; employment centrality, population centrality, and employment-residence mismatch.

The population centrality measure reports the largest share of population that occurs in any ring of radius 20 kilometers whose center lies in the MSA. The employment centrality measure is identical, but is based on employment rather than population.

The mismatch variable provides an aggregate measure of the extent to which people do not live where jobs are (though not of the extent to which people do not live where *their* jobs are). Specifically, let e_i be employment in cell i and p_i be population. Our mismatch measure is

$$\frac{\sum_{i \in \text{MSA}_j} |p_i - \alpha e_i|}{\sum_{i \in \text{MSA}_j} p_i},$$

where α is the inverse share of employment in total population $\left(\alpha \sum_{i \in \text{MSA}_j} e_i \equiv \sum_{i \in \text{MSA}_j} p_i\right)$. Since we are normalizing by population, this is a per capita measure of mismatch.

Instruments for contemporary interstate highways: Our measures of the 1947 interstate highway plan and the 1898 railroad network are taken from Duranton and Turner (2011) and are documented there. Further discussion of the 1947 highway plan is available in Baum-Snow (2007) and Duranton and Turner (2012).

Texas Transportation Institute travel time indices: Until 2009 (using 2008 data), the annual TTI travel time indices were constructed using counts of vehicles on interstate highways and major urban roads from the average annual daily traffic (AADT) item measured in the HPMS 'Sample' data. More specifically, the TTI methodology subjected the AADT variable to a series of transformations using external information and turned it into a measure of speed. This measure of speed was then 'operationally' adjusted for several factors (e.g., the existence of a ramp metering system to access interstate highways) to obtain a final measure of speed by city. These quantities were then compared to a measure of free flow traffic to infer a travel cost index by city and their travel cost index for the US.

In 2010 (to exploit 2009 data), the TTI paired with INRIX, a leading provider of traffic information, directions, and driver services. Real speed data is now measured using information provided by location devices from vehicles operated by various fleet operators and by the smart phones of voluntary individuals for a subsample of segments of roads covered by the HPMS 'Sample' data.

Importantly, the TTI data cover only urbanized areas. This implies that areas for which the TTI computes its index are smaller than the MSAs that we use. This said, the 'urbanized' part of MSAs will host the vast majority of their residents and jobs. In order to compare our MSA based index with TTI, we merged several TTI urbanized areas to approximate MSAs more closely. To do this, we weighted urbanized areas by their population. The full list of merged TTI urbanized areas is: Washington DC and Baltimore, San Francisco-Oakland and San Jose, Cleveland and Akron, Los Angeles and Riverside-San Bernardino, Denver and Boulder, Greensboro and Winston-Salem, Boston and Worcester. Finally, for Fort Myer-Cape Corral MSA, the only TTI data point is for Cape Corral and for Norfolk-Virginia Beach-Newport News MSA, the only TTI data point is for Virginia Beach. Overall we can match 47 of the 50 largest MSAs and 71 of the 100 largest using data for the 90 urban areas for which the TTI reports an index for 2008.

Appendix B. Derivation of VKT^{opt}, VKT^{eq}, τ^* and Δ

To solve for equilibrium VKT, we set average cost (4) equal to the demand for VKT (6) and solve to get,

$$\text{VKT}^{eq} = \left(\Gamma^{1-\theta}\Omega^{-\sigma}\right)^{\frac{1}{1-\theta(1-\sigma)}} . \tag{B1}$$

Optimal VKT results from equating demand (6) with marginal cost (5)

$$\text{VKT}^{opt} = (1-\theta)^{\frac{\sigma(1-\theta)}{1-\theta(1-\sigma)}} \text{VKT}^{eq}. \tag{B2}$$

We note that we can use these expressions to evaluate the amount of 'overdrive', that is, the difference between equilibrium VKT and optimal VKT.

Equilibrium deadweight loss is the area of the approximately triangular region of figure 2 with vertices at points A, B and C. More precisely, this area is,

$$\text{DWL} = \int_{\text{VKT}^{opt}}^{\text{VKT}^{eq}} \left[\frac{1}{1-\theta}\Omega^{\frac{1}{1-\theta}}\text{VKT}^{\frac{\theta}{1-\theta}} - \Gamma^{\frac{1}{\sigma}}\text{VKT}^{-\frac{1}{\sigma}}\right] d\text{VKT} . \tag{B3}$$

Evaluating this integral, we have

$$\text{DWL} = \Omega^{\frac{1}{1-\theta}} \left[\left(\text{VKT}^{eq}\right)^{\frac{1}{1-\theta}} - \left(\text{VKT}^{opt}\right)^{\frac{1}{1-\theta}}\right] - \frac{\sigma}{\sigma-1}\Gamma^{\frac{1}{\sigma}} \left[\left(\text{VKT}^{eq}\right)^{\frac{\sigma-1}{\sigma}} - \left(\text{VKT}^{opt}\right)^{\frac{\sigma-1}{\sigma}}\right] . \tag{B4}$$

Using equation (B2), this becomes

$$\text{DWL} = \left(1-(1-\theta)^{\frac{\sigma}{1-\theta(1-\sigma)}}\right)\Omega^{\frac{1}{1-\theta}}\left(\text{VKT}^{eq}\right)^{\frac{1}{1-\theta}} - \left(1-(1-\theta)^{\frac{(\sigma-1)(1-\theta)}{1-\theta(1-\sigma)}}\right)\frac{\sigma}{\sigma-1}\Gamma^{\frac{1}{\sigma}}\left(\text{VKT}^{eq}\right)^{\frac{\sigma-1}{\sigma}} . \tag{B5}$$

In equilibrium, we can use the supply curve (4) and the demand curve (6) to write

$$\Gamma^{\frac{1}{\sigma}} = C^{eq}\left(\text{VKT}^{eq}\right)^{\frac{1}{\sigma}} \tag{B6}$$

and

$$\Omega^{\frac{1}{1-\theta}} = C^{eq}\left(\text{VKT}^{eq}\right)^{-\frac{\theta}{1-\theta}} . \tag{B7}$$

Using these two identities, along with the fact that $\text{VTT}^{eq} = C^{eq} \times \text{VKT}^{eq}$, we can write deadweight loss entirely in terms of total travel time and parameters,

$$\text{DWL} = \left[\left(1-(1-\theta)^{\frac{\sigma}{1-\theta(1-\sigma)}}\right) - \frac{\sigma}{\sigma-1}\left(1-(1-\theta)^{\frac{(\sigma-1)(1-\theta)}{1-\theta(1-\sigma)}}\right)\right]\text{VTT}^{eq} . \tag{B8}$$

Dividing by VTT^{eq} gives Δ.

The magnitude of the optimal Pigovian congestion tax is the difference between the marginal and average cost of VKT at the optimum. Using equation (5), we have

$$\tau^* = \frac{\theta}{1-\theta} AC(\text{VKT}^{opt}). \tag{B9}$$

Together with equation (4), this gives

$$\tau^* = \frac{\theta}{1-\theta} \Omega^{\frac{1}{1-\theta}} (\text{VKT}^{opt})^{\frac{\theta}{1-\theta}}. \tag{B10}$$

To proceed, note that we can use equation (B7) to write $\Omega^{\frac{1}{1-\theta}}$ in terms of VKT^{eq} and equation (B2) to write VKT^{opt} in terms of VKT^{eq}. Substituting and simplifying gives,

$$\tau^* = \theta(1-\theta)^{\frac{\sigma\theta}{1-\theta(1-\sigma)}-1} C^{eq}. \tag{B11}$$

Recalling that $\text{VTT}^{eq} = C^{eq}\text{VKT}^{eq}$ gives the desired result.

Appendix C. Discussion of trip length instrument

While mean distance by trip type does not appear explicitly in our the system of equations (12)-(13), it is arguably a good proxy for the willingness to pay for a trip of a given type. Just as trip type dummies vary with the intercept of the demand curve, so does mean trip distance by type. However, the arguments for the validity of trip type dummies and the validity of mean distance by trip type are different.

Trips to the gym may be observed on average under better weather (and traffic) conditions than trips to work. This is not an issue as long as this differential selection of trips does not affect the distance of trips to the gym relative to the distance of trips to work. However, suppose that when the weather is worse drivers go to a closer gym and to a closer workplace. If weather conditions affect distances equally for all types of trips, then unobserved weather introduces noise but does not lead to a violation of the exogeneity condition. On the other hand, if more than 10 centimeters of snow causes drivers to choose a closer gym but does not affect the choice of workplace, then mean trip distance by type is correlated with the error term in the speed regressions. More generally, average distances by trip type do not satisfy the exogeneity condition when the unobserved state of traffic differentially affects the distance of different types of trips depending on mean distance.

Whether the length of shorter trip types should be more (or less) sensitive to traffic conditions than distance for longer trips types is not obvious. While (on average) distance for short shopping trips may be sensitive to traffic conditions, that of longer recreational trips might be as well. On the other hand, distance for short trips to school may be insensitive to traffic conditions and long trips to work may be similarly insensitive. This said, to avoid a possible correlation between average trip distance and unobserved traffic conditions, we can again restrict attention to trips with low discretion. We can also use extensive controls for trip characteristics, as we do when using trip type dummies as instruments.

We note that if trip type dummies and mean distance by trip type fail their respective exogeneity conditions, they do so for different reasons. Trip type dummies may be subject to a selection bias where different types of trips are observed only under systematically different unobserved traffic conditions. Mean trip distance may be subject to a simultaneity problem where mean distance for trips of different types is affected differently under different unobserved traffic conditions in a way that is correlated with this speed. Therefore, if both types of instruments lead to similar estimates this indicates that they are either both valid or, improbably, that the selection problem for trip type dummies has the same bias on our estimates as the hypothetical simultaneity problem for average distance by trip types.

Appendix D. Further robustness tests for the speed-distance relationship

Table 9 reports further results which are commented in the text.

Appendix E. Robustness tests for determinants of speed

Table 10 confirms our findings using different samples of MSAs, years of data, and variables. In column 1 we duplicate our preferred estimation of column 6 of table 5 using only the 50 largest MSAs and estimate slightly larger magnitudes for both the coefficient on lanes and vehicle travel time. The R^2 is also higher given the greater precision of our index in larger MSAs. In columns 2 and 3, we also duplicate our preferred regression from table 5 but use 100 MSAs for 2001 and 1995, respectively. While the results for 2001 are close to those for 2008, the coefficient on lanes for 1995 is insignificant and the R^2 is much lower. In columns 4 and 5, we use a log speed index computed from the 2008 and 2009 TTI travel cost indices, respectively, as dependent variables. The results for the 2008 TTI index are close to those of our preferred estimation. With the 2009 TTI index, the

Table 9: Robustness checks of the estimation of inverse-supply curves

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	mid. dist. 2 quartiles		peak hours		commutes		PMSA	
	OLS	IV	OLS	IV	OLS	IV	OLS	IV
Panel A. 100 largest MSAs for 2008								
Mean $\bar{c}$	1.258	1.022	1.390	1.190	1.352	-	1.413	1.331
	(0.144)	(0.466)	(0.107)	(0.182)	(0.166)	-	(0.090)	(0.142)
	[0.003]	[0.171]	[0.003]	[0.068]	[0.004]	-	[0.001]	[0.047]
Mean γ	0.372	0.251	0.413	0.352	0.394	-	0.420	0.350
	(0.063)	(0.243)	(0.043)	(0.089)	(0.060)	-	(0.028)	(0.071)
	[0.001]	[0.118]	[0.001]	[0.044]	[0.001]	-	[0.001]	[0.032]
Panel B. 50 largest MSAs for 2008								
Mean $\bar{c}$	1.269	1.032	1.386	1.180	1.381	1.101	1.417	1.303
	(0.108)	(0.433)	(0.075)	(0.123)	(0.142)	(0.831)	(0.075)	(0.094)
	[0.002]	[0.144]	[0.002]	[0.053]	[0.003]	[0.069]	[0.001]	[0.034]
Mean γ	0.371	0.196	0.402	0.329	0.393	0.291	0.411	0.323
	(0.039)	(0.228)	(0.024)	(0.061)	(0.047)	(0.283)	(0.020)	(0.041)
	[0.001]	[0.101]	[0.001]	[0.034]	[0.001]	[0.046]	[0.001]	[0.023]
Panel C. 50 largest MSAs for 2001								
Mean $\bar{c}$	1.311	1.093	1.372	1.262	1.349	-	-	-
	(0.109)	(0.355)	(0.083)	(0.116)	(0.125)	-	-	-
	[0.004]	[0.180]	[0.003]	[0.067]	[0.005]	-	-	-
Mean γ	0.388	0.198	0.395	0.334	0.385	-	-	-
	(0.045)	(0.191)	(0.028)	(0.058)	(0.037)	-	-	-
	[0.002]	[0.125]	[0.001]	[0.042]	[0.001]	-	-	-
Panel D. 50 largest MSAs for 1995								
Mean $\bar{c}$	1.042	0.890	1.190	0.992	1.162	-	-	-
	(0.144)	(0.358)	(0.090)	(0.155)	(0.140)	-	-	-
	[0.008]	[0.173]	[0.005]	[0.070]	[0.009]	-	-	-
Mean γ	0.312	0.208	0.363	0.323	0.347	-	-	-
	(0.056)	(0.186)	(0.023)	(0.073)	(0.039)	-	-	-
	[0.002]	[0.121]	[0.001]	[0.042]	[0.002]	-	-	-

Notes: Columns 1 and 2 censor the bottom and top quartile of distance per MSA and year. Columns 3 and 4 only consider trips taken at peak hours. Columns 5 and 6 consider only commutes and work-related trips. Columns 7 and 8 use primary instead of consolidated metropolitan areas. Instruments in columns 6 are weak except for the largest cities in 2009. Household location at the PMSA level is known only for 2009. Mean of the coefficients across all cities. Standard deviation of city coefficients in parentheses. Mean of the standard deviation of city coefficients in squared parentheses.
Dependent variable: minutes per kilometer for individual trips.
Controls: Controls for household income and its square, driver's education and its square, age, dummies for males, blacks, and workers, and a quartic for the time of departure in all columns.
Instruments: Mean trip distance for trips of the same purpose computed from the four most similar MSA in term of population in columns 2, 4, 6, and 8.

Table 10: The determinants of speed, alternative years, samples, and variables

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	S_{IV2} 2008	S_{IV2} 2001	S_{IV2} 1995	TTI_{2008} 2008	TTI_{2009} 2008	S_{IV2} 2008	S_{IV2} 2008	S_{IV2} 2008
log lane (total)	0.15^a	0.066^c	0.022	0.11^a	0.032	0.066^c		
	(0.037)	(0.034)	(0.049)	(0.023)	(0.022)	(0.038)		
log lane (interstate highways)							0.071^a	0.069^a
							(0.014)	(0.013)
log lane (major urban roads)								0.066^b
								(0.030)
log VTT	-0.19^a	-0.12^a	-0.074	-0.17^a	-0.077^a		-0.11^a	-0.17^a
	(0.031)	(0.032)	(0.046)	(0.019)	(0.018)		(0.012)	(0.031)
log population						-0.12^a		
						(0.035)		
Observations	50	100	100	71	78	100	98	98
R^2	0.52	0.35	0.24	0.70	0.65	0.34	0.46	0.49

Notes: OLS regressions with a constant in all columns. Robust standard errors in parentheses. *a*, *b*, *c*: significant at 1%, 5%, 10%. Dependent variables is in log in all columns. In columns (4) and (5), we use the inverse of the TTI indices to be able to interpret them as speed.

coefficient on lanes is small and insignificant. Column 6 substitutes population for VTT finds results similar to the other specifications. In column 7, we use log interstate highway lane kilometers as a measure of roads. Excluding smaller roads leads to a smaller coefficient. Finally in column 8 we control separately for log lane kilometers of interstate highways and major urban roads. We find a marginally higher coefficient for interstate highways than for major urban roads which is consistent with the more important capacities and the greater speed of interstate highway lanes relative to lanes of major urban roads. Importantly, the sum of these two coefficients is very close to the coefficient for both types of roads considered jointly in our preferred regression.

Table 11 further confirms our findings using alternative estimation techniques. As argued above, vehicle travel time is expected to be determined simultaneously with travel speed. To deal with this problem, in column 1 we instrument for VTT using population. As might be expected, population is a strong predictor of VTT. The estimated coefficient on VTT is -0.12, only marginally smaller in magnitude than (and statistically undistinguishable from) its OLS counterpart. One may worry that the exclusion restriction associated with this regression is not satisfied since population in 2008 may be correlated with traffic speed through variables other than roads and VTT. This worry may not be as important as it seems. First, Duranton and Turner (2012) document that interstate

Table 11: The determinants of speed, 100 MSAs in 2008 IV regressions

	(1) TSLS	(2) TSLS	(3) TSLS	(4) TSLS	(5) TSLS	(6) TSLS	(7) TSLS	(8) TSLS	(9) TSLS	(10) LEVPET
log lane	0.075^{a}	0.081^{a}	0.077^{b}							0.090^{c}
	(0.040)	(0.041)	(0.039)							(0.048)
log lane (IH)				0.086^{a}	0.090^{a}	0.084^{a}	0.082^{a}	0.086^{a}	0.079^{a}	
				(0.024)	(0.024)	(0.025)	(0.024)	(0.023)	(0.025)	
log lane (MRU)					0.066^{b}	0.057^{c}		0.056^{c}	0.044	
					(0.030)	(0.032)		(0.029)	(0.032)	
log VTT	-0.12^{a}	-0.11^{a}	-0.096^{b}	-0.12^{a}	-0.18^{a}	-0.16^{a}	-0.11^{a}	-0.17^{a}	-0.14^{a}	-0.11^{a}
	(0.035)	(0.036)	(0.038)	(0.020)	(0.035)	(0.039)	(0.019)	(0.033)	(0.040)	(0.025)
Instrumented										
log VTT	X	X	X				X	X	X	
log lane (IH)				X	X	X	X	X	X	
Instruments										
log pop. 2008	X	X	X				X	X	X	
log highways 1947				X	X	X	X	X	X	
log railroads 1898				X	X	X	X	X	X	
Controls										
$\Delta_{1980-2008}$ log pop.		X								
$\mathbb{E}\Delta_{1980-2000}$ log pop.			X			X			X	
log pop. 1950						X			X	
log pop. 1920						X			X	
Observations	100	100	100	94	94	94	94	94	94	300
Overid. p-value	-	-	-	0.20	0.26	0.43	0.20	0.24	0.46	-
First-stage stat.	567	494	450	32.3	32.4	28.0	21.4	21.1	18.5	-

Notes: Regressions with a constant in all columns. Robust standard errors in parentheses. *a*, *b*, *c*: significant at 1%, 5%, 10%. Dependent variables is log S_{IV2} in all columns. The first stage statistic is a Kleibergen-Paap F-statistic for robust estimation. They are always above the Stock-Yogo critical threshold at 10%.

highways have only a modest effect on urban growth. Second, to explore this question further, in column 2 we duplicate the regression of column 1 but also control for population growth between 1980 and 2008. The results remain the same. To confirm this finding, in column 3 we use a more exogenous proxy for urban growth, namely expected population growth between 1980 and 2000. This proxy is computed by interacting the sectoral composition of employment in MSAs in 1980 and employment growth for these sectors between 1980 and 2000.[34] The results are still very close to the previous estimates.[35]

[34]This variable is inspired by Bartik (1991) and further details about its construction can be found in Duranton and Turner (2011).

[35]Overall the finding of a slightly less negative effect of travel time on speed in TSLS relative to OLS after controlling for population growth is consistent with higher travel speed having a modest positive effect on population.

In column 4, we turn to the simultaneous determination of roads and travel speed. Unfortunately, we do not know of any good exogenous predictor of lane kilometers of both interstate highways and major urban roads. However, we can follow Duranton and Turner (2011, 2012) and use the 1947 plan of interstate highways and a map of 1898 railroads to predict contemporaneous lane kilometers of interstate highways in US MSAs. Online Appendix A provides further details about the construction of these two instruments. As can be seen from the results, these two instruments are good predictors of contemporaneous lane kilometers of interstate highways and pass the appropriate overidentification test. The coefficient on lane kilometers of interstate highways is 0.086. This is slightly above its corresponding OLS estimate of 0.071 in column 7 of table 10. In column 5, we estimate the same regression but add lane kilometers of major urban roads as control. The sum of the two lane coefficients is 0.156, slightly above the sum of the corresponding OLS estimates of 0.135 (in column 8 of table 10). Again, our instruments could be correlated with travel speed through the error term. Cities that received more railroads during the 19th century or cities that were allocated more roads in 1947 may be different in systematic ways. In particular, they were bigger at the time and thus may be spatially organised in a different way and may have a different transportation network relative to more recent cities. In turn, that might affect travel speed. To preclude these correlations, in column 6 we introduce 1920 population (the closest year to 1898 for which we can get population estimates), 1950 population, and expected population growth between 1980 and 2000 as before. Introducing these controls in column 6 changes close to nothing to the estimates of column 6. The marginally significant coefficient on major urban roads is now marginally insignificant.

In columns 7 to 9, we repeat the same strategy as in columns 4 to 6 but now instrument for both lane kilometers of highways and vehicle travel time. The results remain the same as before. Finally in column 10, we use a completely different instrumenting approach which makes use of all three cross-sections of data (and changes in the roadway prior to 1995) and implement the estimation technique for productivity suggested by Levinsohn and Petrin (2003). It is true that, unlike firms, cities do not invest in roads in response to positive demand shocks to maximize profit (be it only because roads are funded mainly by the federal government). Nonetheless we expect road provision to respond to changes in travel conditions. While suggestive, our Levinsohn-Petrin results are very close to our OLS and TSLS estimates.