Gains From Distortions in Congested City Networks.

Francesco Trebbi and Matilde Bombardini*

November 2012

Abstract

This paper presents a model and an automated methodology for assessing gains from network distortions in cities. Distortions arise from excluding traffic along certain routes. Distortions degrade network connectivity, but can be paradoxically useful for congestion amelioration. We show that such distortions are quantitatively large, increasingly pervasive in larger cities, and potentially very valuable. The results ultimately support the view that Braess' (1968) paradox is not just a theoretically interesting possibility, but a widespread feature of city road networks.

*University of British Columbia and National Bureau of Economic Research, ftrebbi@mail.ubc.ca; University of British Columbia and National Bureau of Economic Research, matildeb@mail.ubc.ca, respectively. The authors would like to thank Daron Acemoglu, Richard Arnott, Gilles Duranton, David Lucca, Andrea Mattozzi, Tim Roughgarden, Jesse Shapiro, Ken Small, Adam Szeidl, Matthew Turner, and Erik Verhoef for useful discussion and Luigi Guiso, Paola Sapienza and Luigi Zingales for sharing their historical data on Italian cities. Kareem Carr, Seyed Ali Madani zadeh, Isaac Laughlin, and Carlos Sánchez-Martínez provided outstanding research assistance. Laughlin also contributed to the data description section. Financial support from the Initiative on Global Markets at Chicago Booth is gratefully acknowledged. An earlier version of the paper circulated under the title "City Structure and Congestion Costs".

I. Introduction

This paper offers two main contributions. First, we present an empirical methodology based on network theoretic foundations for studying the microstructure of city road networks. Second, we find strong empirical relationships between population, network structure, and congestion costs. These findings provide guidance in restricting the set of ambiguous theoretical predictions that we report.

Concerning the network structure, the paper shows that there is an intuitive, but overlooked connection between the way city traffic plans are organized and Braess' (1968) seminal counterintuitive observation that reducing the number of links in a network may actually ameliorate its equilibrium level of congestion[1]. City planners systematically close off roads and directions of flow to abate congestion at the cost of lower network connectivity. By any standard such approach is more frequently employed than congestion pricing and more readily implementable than changes in the provision of lane kilometers for any type of roads. The paper assesses the empirical prevalence of such distortions and binds their economic value in transportation networks.

Congestion costs are economically large. The Texas Transportation Institute assesses total congestion costs in the United States at $115 billion in 2009 (based on wasted time and fuel, excluding the environmental externalities of the 3.9 billion gallons of wasted fuel), 58 percent of which stem from very large urban areas. Couture, Duranton and Turner (2012) provide an even larger economic assessments ($141 billion in 2008). This motivates our interest.

We model a city as a directed graph and describe the time delay of going through the network (latency) as a function of the flow of agents routing through its arrows. We define a (city) planner's problem as finding a subgraph minimizing the aggregate congestion costs of

[1] Braess (1968) presents an example of how an increase in the connectivity of a network (e.g. adding a new link) may increase total traffic congestion in equilibrium. This counterintuitive result is commonly referred to as Braess' paradox. This paradoxical result occasionally also surfaces in other forms, such as the "fundamental law of road congestion" (i.e. increasing road supply does not ameliorate congestion levels). See Duranton and Turner (2011).

selfishly-routing agents moving through the network. This matches the empirical regularity that not all sites within a city are accessible by traffic and that the accessible vertex set is a very deliberate subset of the overall city network. We study the equilibrium properties of such optimal subgraph. Within our framework, traffic flow restrictions in the subgraph (e.g. one-way traffic and other forms of path distortion) are a direct expression of Braess' paradox in networks. In other words, traffic restrictions are interpretable as pruning arrows from a transportation network to abate latency.[2]

Placing theoretical structure on the data, we quantify the endogenous congestion-minimizing distortions in city transit networks in a large sample of United States and Italian cities, to show the increasing pervasiveness of distortions in larger cities and to place reasonable bounds on their economic value. Importantly, all this is achieved sidestepping the need for detailed micro traffic and speed data (hundreds of millions of roads/time data points, often noisily measured), a major obstacle for the empirical analysis of congestion beyond samples of very limited representativeness.[3] To this goal, we introduce a novel automated search algorithm on a widely available Internet mapping application[4]. The algorithm first generates source-target routes within the network and then samples from the empirically implemented viable subgraph (i.e. the actual traffic plan). For a large sample of US and Italian cities we estimate the endogenous congestion-minimizing distortions of transit networks by comparing restricted versus unrestricted routes within the city. An average US daily commute is made about 10 percent longer in terms of its distance as a result of network pruning. We also show empirically that, as population increases, cities display larger distortions within their networks. To the best of our knowledge, these stylized facts are new.

From a theoretical standpoint, the paper is related to a growing computer science literature focused on inefficiencies arising within selfishly routing networks. Our model builds heavily on the work of Roughgarden and Tardos (2002) and Roughgarden (2002), which in

[2]See Section 2 for further discussion.

[3]For a critical discussion, see Fosgerau and Small (2010). The authors also present an application to Danish freeways with (exceptionally) clean data.

[4]Our application of choice is Google Maps, available at http://maps.google.com/.

turn leverages on the voluminous transportation science literature on traffic equilibria stemming from Wardrop (1952). Our contribution should not be seen as theoretical though, but in tightly mapping such abstract models into data. The paper is related to the vast literature on Braess' paradox in networks[5] and proposes one of its few applications to real road networks in a large sample of cities. Indeed, the prevalence of Braess' paradox in transportation networks is neither theoretically obvious [6] nor one that has little economic implications, given its potentially large economic value.

From an empirical standpoint, we follow the same path as Lucca and Trebbi (2009) in presenting theoretically-grounded information retrieval algorithms from web sources designed to quantify hard-to-measure concepts in an automated and undirected fashion. This paper is related to a vast literature on agglomeration, spanning industrial organization, international trade, urban geography, urban planning, regional science, and urban economics. The number of important contributions in this area is too large for a comprehensive review here. Fujita, Krugman, and Venables (1999) and Fujita and Thisse (2002) provide excellent overviews of the economic geography literature, while Anas, Arnott, and Small (1998) and Small and Verhoef (2007) provide accurate reviews of the economics of urban spatial structure and transportation. More closely related is the economic literature on congestion and hypercongestion costs (Dewees, 1977; Arnott and Small, 1994; Verhoef, 2001; Small and Chu, 2003; Arnott, Rave and Schöb, 2005; Duranton and Turner, 2011; Couture, Duranton and Turner, 2012) to which this paper contributes by presenting a model of urban networks matched to micro data. The paper is also related to the literature on quantifying urban network structure and sprawl.[7] Relative to this line of research, mostly focused on entropy and dispersion measures, our approach differs in that we focus on endogenous distortions imposed on the network structure. Our contribution to this literature is dual, as we present

[5] Bounding the pervasiveness of Braess paradox in networks is studied, among others, by Lin, Roughgarden, and Tardos (2004) and Lin, Roughgarden, Tardos, and Walkover (2009).

[6] For a discussion see Frank (1981); Valiant and Roughgarden (2010); Chung and Young (2010).

[7] Examples include Gordon, Kumar, and Richardson (1989), Tsai (2005), Xie and Levinson (2006), Burchfield et al. (2006).

both new automated measures targeted at quantifying city structure and an algorithm which substantially improves on the small scale limiting previous studies.[8]

The work is organized as follows. Section 2 presents a model describing the city and the main theoretical results in terms of congestion and agglomeration costs. Section 3 shows how to link the theoretical representation of the network and the empirical city structure. Section 4 presents the mapping algorithm and data. Section 5 reports our main empirical results. Section 6 concludes and discusses future applications.

II. Model

We model a city as a directed acyclic graph $C = (V, A)$ with vertex set V, indicating sites within the city, and arrow set A. The graph displays $i = 1, ..., k$ source-target vertex pairs $s_i - t_i$ and multiple paths through C connecting each pair. We do not impose restrictions on C, so the model caters to monocentric or multicentric, sprawling or compact macro urban structures. This particular network representation and the analysis for selfish-routing flows are due to Roughgarden and Tardos (2002) and Roughgarden (2002, 2006).

We assume that the city population, N, is located at the sources $s_1, ..., s_k$ each with subpopulation shares $w_1, ..., w_k$, $\sum_i w_i = 1$.[9] We do not focus on the endogenous spatial equilibrium and take the location decision as predetermined. While theoretically possible to relax this assumption, unavailability of systematic data on the value of targets t_i for different city population subgroups (necessary to compute payoffs in location trade off) would make any empirical validation unfeasible.

Our location assumption is realistic in the case of short-term considerations (hence, our model could be considered complementary to long-run models of endogenous residential choice) or in instances where location decisions are driven more by amenities at s_i (school quality, local public goods, etc.) than by considerations on a specific commute $s_i - t_i$. In

[8] For instance, Youn, Gastner, and Jeong, (2008) present an empirical analysis of Braess paradox in a limited sample of 246 roads in Boston, Massachusetts.

[9] For tractability individuals do not locate on a non-source, non-sink vertex in the network.

this latter sense, Quinet and Vickerman (2004, p.50) conclude that "empirical studies shed further light on the theoretical analyses in suggesting that transport constitutes an important factor, but one which is far from the sole determinant, in the location decision of activities."[10]

Let us define r_i the total traffic traveling on $s_i - t_i$, that is the expected number of agents making the $s_i - t_i$ trip per time period[11], the input traffic vector $\overrightarrow{r} = [r_1, ..., r_k]'$, and $\sum_i r_i = r$ the total traffic rate in C. For example, we may consider s_i being a residential suburb, t_i a location in the central business district of the city, and $s_i - t_i$ the commute to work for r_i individuals out of the total $n_i = w_i N$ residing at s_i. Define p_i the per period probability that an individual in n_i has to undertake trip $s_i - t_i$.[12] For tractability we will assume that these events are independent across individuals, so that $r_i = p_i n_i > 0$.[13] Input traffic rates are taken as given in what follows, while Appendix B presents a generalization of the model allowing endogenous traffic demand $\overrightarrow{r}$.

For given origin-destination pair $s_i - t_i$, also called a commodity, we allow individuals to sort through different paths connecting s_i and t_i. Call Π_i^C the (non-empty) set of paths connecting s_i and t_i in C, and f_π the flow on path $\pi \in \Pi_i^C$. Define $\Pi^C = \cup_i \Pi_i^C$. A flow $f : \Pi^C \to \mathbb{R}^+$ is called feasible if $\sum_{\pi \in \Pi_i^C} f_\pi = r_i$ for all i. Let us indicate with f_a the flow on arrow $a \in A$, so that $f_a = \sum_{\pi : a \in \pi} f_\pi$.

We assume that the time necessary to travel along arrow a is described by a latency function $l_a(f_a)$ with $l_a : \mathbb{R}^+ \to \mathbb{R}^+$, nonnegative, non-decreasing, continuous, and semi-convex in the traffic flow along arrow a.[14] Non-decreasing and continuous latency functions $l_a(.)$

[10] For the case of highways, the FHWA states that "Recent empirical studies conducted in Ohio and North Carolina indicate that local patterns of population growth (measured at the Census Tract level) are not highly correlated with increases in highway capacity." See Hartgen, D.T. (2003a, 2003b).

[11] For the following theoretical analysis a specific assumption for time interval length is not necessary. However one can consider a day a reasonable reference period. In the empirical analysis, daily trips are sampled over one year periods.

[12] See Arnott, De Palma, and Lindsey (1993) for a structural model were endogenous departure decisions are modeled explicitly and Vickrey (1969) for an earlier seminal contribution. However, the network considered by Arnott, De Palma, and Lindsey (1993) is stark, as it includes a two vertex graph with a single bottleneck.

[13] For the following theoretical analysis this assumption is immaterial. For the empirical application the linear dependence of traffic rates on population within the network should be considered a special case. We discuss it within the empirical methodology section.

[14] Semi-convexity is a weaker assumption than convexity, in particular if $l_a(x)$ is semi-convex, then $x l_a(x)$ is convex.

are standard in the traffic routing literature (Roughgarden, 2002) and convexity is routinely employed in empirical analysis by the US Federal Highway Administration (FHWA) as well as in the theoretical literature (e.g. ; Chen and Kempe, 2007; Acemoglu and Ozdaglar, 2007).

Given a network C, a traffic rate through the network $\overrightarrow{r}$ and set of latency functions l, we indicate the total latency costs of a flow f in C as $L(f) = \sum_{a \in A} l_a(f_a) f_a = \sum_{\pi \in \Pi^C} l_\pi(f) f_\pi$, where the latency on path π is $l_\pi(f) = \sum_{a \in \pi} l_a(f_a)$.

A. *Optimal and Nash Flows*

a. Optimal Flows For graph C, traffic rate $\overrightarrow{r}$, and latency functions l, Roughgarden and Tardos (2002) define a flow f^* that minimizes total network latency an optimal flow. Particularly, they show that an optimal flow solves the following nonlinear program:

$$
\begin{aligned}
& & Min \sum_{a \in A} l_a(f_a) f_a & \\
(1) \qquad s.t. \sum_{\pi \in \Pi_i} f_\pi &= r_i & & \forall i \\
f_a &= \sum_{\pi \in \Pi : a \in \pi} f_\pi & & \forall a \in A \\
f_\pi &\geq 0 & & \forall \pi \in \Pi^C.
\end{aligned}
$$

Under our assumptions on l, program (1) is convex and admits one feasible global optimum f^*. The minimum-latency flow exists and is unique. We can think of f^* as the solution that a city planner able to fully and costlessly direct traffic flows through C would implement.

Intuition would suggest $L(f^*)$ be monotonically non-decreasing in traffic rates, which the following proposition confirms.

propositionf f^* and $f^{*\prime}$ are feasible optimal flows for $(C, \overrightarrow{r}, l)$ and $(C, (1+\gamma)\overrightarrow{r}, l)$ respectively, with $\gamma > 0$, then $L(f^*) \leq L(f^{*\prime})$.

Proof. In Appendix.

Note that in Proposition 1 $(1+\gamma)\overrightarrow{r}$ indicates a scalar multiplication, that is we are increasing proportionally traffic inputs at every source s_i. Although restrictive, it will become clear below that focusing on this type of traffic increments substantially facilitates the tractability of the analysis.

b. Nash Flows It is possible to extend the analysis to decentralized solutions, and characterize the flow f at a Nash equilibrium in the triple $(C, \overrightarrow{r}, l)$. Nash equilibria for this problem are reasonable approximations to individual behavior in presence of atomistic, selfish agents endogenously routing through the network (Roughgarden, 2002). Such agents fully incorporate in their decisions the latency of each arrow when choosing a path (picking the minimum-latency path), but will not incorporate in their decisions the additional congestion externality they bring to that specific flow. In transportation science this is commonly referred to as the Wardrop (1952) equilibrium problem for urban road networks.

definitionflow f is a Nash equilibrium flow if and only if $l_{\pi_1}(f) \leq l_{\pi_2}(f)$, for every flow-carrying path $\pi_1 \in \Pi_i^C$ and every path $\pi_2 \in \Pi_i^C$, for every $i = 1, ..., k$.

If a flow f is a feasible Nash flow, then any two $s_i - t_i$ paths $\pi_1, \pi_2 \in \Pi_i^C$ carrying positive flow must have the same latency $l_{\pi_1}(f) = \sum_{a \in \pi_1} l_a(f_a) = l_{\pi_2}(f) = \ell_i$. This further implies $L(f) = \sum_i \ell_i r_i$.

Beckmann, McGuire, and Winsten (1956) and Roughgarden and Tardos (2002) show the following existence and uniqueness result, whose proof we omit for brevity:

propositiononsider a triple $(C, \overrightarrow{r}, l)$ with l continuous and non-decreasing: (a) $(C, \overrightarrow{r}, l)$ admits a feasible Nash equilibrium flow f. (b) If f and $\bar{f}$ are feasible Nash flows, then $L(f) = L(\bar{f})$.

For given triple $(C, \overrightarrow{r}, l)$, a Nash flow f will generally not achieve minimum latency, that is $L(f) \geq L(f^*)$. Agents selfishly routing through the network fail to incorporate the negative externality they induce upon each other[15]. This results in the possibility of some

[15] For an early discussion see also Walters (1961).

counterintuitive outcomes in traffic networks which we now explore.

We begin by deriving a useful monotonicity result for Nash flows with respect to traffic inputs. We show that proportional traffic increases in multicommodity networks ($k \geq 1$) guarantee higher latency:

propositionf f and f' are feasible Nash flows for $(C, \overrightarrow{r}, l)$ and $(C, (1+\gamma)\overrightarrow{r}, l)$ respectively, with $\gamma > 0$, then $L(f) \leq L(f')$.

Proof. In Appendix.

The result that increased traffic should increase total congestion at Nash equilibrium may appear obvious, but it is not. To the contrary, equilibrium congestion may actually decrease when the overall traffic input in a network increases. Fisk (1979) provides an example where a traffic demand's increase along one origin-destination (O-D) pair reduces total latency in the network. Figure 1 modifies such example allowing input traffic increases in *all* O-D's. Total congestion falls in this case as well. The intuition behind this counterintuitive result is that an increase in the overall traffic input may not necessarily increase latency on a specific edge. Indeed, there are instances in which the use of an edge is reduced when traffic input increases, as for edge (b, c) in Figure 1. General results for arbitrary vectors of traffic inputs' increments are thus not available, but some advancements have been made in the literature. An increase in the traffic input on an O-D, holding other inputs constant, is proven to increase the latency of that specific origin-destination[16], although positive changes in demand in all O-Ds do not guarantee higher congestion. The reason why proportional increases in demand are sure to induce monotonicity is based on the fact that, in general, one can only prove that the average latency of an origin-destination pair weighted by its change in demand is increasing (Dafermos and Nagurney, 1984). Proportional increases guarantee that the average O-D latency weighted by the level of demand behaves in the

[16] Hall (1978) proves that, under some symmetry assumptions, each source to target latency ℓ_i is a monotone nondecreasing function of r_i, for $r_i > 0$, when all other traffic rates r_j, $j = 1, 2, ..., k$, $j \neq i$ are held constant, using sensitivity analysis for convex programs. This result is extended to asymmetric cases in Dafermos and Nagurney (1984). Lin, Roughgarden, Tardos, and Walkover (2009) prove a monotonicity result in a single commodity network ($k = 1$) using a direct combinatorial proof.

same way[17]. Proposition 3 should be viewed as providing a restrictive, but tractable case for useful comparative statics[18].

B. *Optimal Network Design*

For given triple $(C, \overrightarrow{r}, l)$, a reasonable representation of a city planner's problem is unlikely to be well described by (1), due to the non-trivial costs of assigning agents to specific routes in a directed fashion[19] or to the cost of optimally taxing each specific edge in the graph to internalize marginal social costs[20]. In practice, city planners aim at finding restrictions on accessible vertices that allow reductions in latency costs. More formally, the problem is to find the subnetwork M of C, over which agents selfishly route themselves, displaying the lowest latency at Nash equilibrium. For given triple $(C, \overrightarrow{r}, l)$, the problem can be represented as the following network design program:

$$
\begin{aligned}
(2) \qquad & \underset{\tilde{M} \subseteq C}{Min} \{L(h)\} \\
& s.t.\ \Pi_i^{\tilde{M}} \neq \varnothing, \qquad \forall i \\
& \qquad h \text{ is a feasible Nash flow for } \left(\tilde{M}, \overrightarrow{r}, l\right)
\end{aligned}
$$

For every subgraph $\tilde{M}$ with rate $\overrightarrow{r}$ and latency functions l there is a unique Nash flow, hence (2) admits a solution. However, finding the minimum-latency subnetwork M in problem (2) is NP-hard even in single commodity settings (i.e. with a single source-target pair), as proven by Roughgarden (2006), the first to introduce this problem in theoretical Computer Science.[21] A solution M to (2) exists. In the remainder of this section and for our empirical

[17] The level of demand is what matters for total congestion comparison. It is precisely a little gain on a high-demand O-D that drives Fisk's paradox for instance.

[18] Other applications employing such type of proportional increases include Roughgarden (2002) and Roughgarden and Tardos (2002) in their bicriteria results.

[19] Randomizations based, for instance, on *licence* **license** plate numbers are not uncommon anti-congestion policies used to direct flows, but what we emphasize here is that the main instrument for congestion abatement available to city planners is based on restrictions on access to certain vertices of graph C.

[20] See Walters (1961).

[21] For linear latency functions Cole, Dodis, Roughgarden (2006) prove that taxes on arrows can also be used to improve total latency in the network and they are equally efficient as arrows' removal. For general latency

analysis we will assume that it is known to the planner.

As further discussed in the following section, our analysis will take city structure $(C, \overrightarrow{r}, l)$ as given and perform comparative statics on the optimal subnetwork map M with respect to N. In the empirical analysis we will discuss how the actual city structure can be represented by $(C, \overrightarrow{r}, l)$ and how the empirical transit restrictions approximate $(M, \overrightarrow{r}, l)$.

A comment is in order here. The idea of decreasing aggregate latency by reducing the number of available arrows within a network may appear prima facie counterintuitive. A more connected network should guarantee lower congestion. However, this is not always the case. Since Braess (1968) seminal work, a large literature has developed around this seemingly counterintuitive result. Large negative social externalities can arise from decentralized privately beneficial actions. Improving the connectedness of the network may further increase incentives towards privately beneficial but socially costly actions and further decrease aggregate efficiency. Figure 2(a) illustrates an example of a triple $(C, 1, l)$ displaying Braess' paradox. In Figure 2(b) by removing arrow (v_1, v_2) one obtains an optimal subgraph $(M, 1, l)$ with lower Nash equilibrium latency.[22]

We now discuss some results concerning total latency costs in optimal subnetworks. Let us first notice that, given city structure $(C, \overrightarrow{r}, l)$, increasing the traffic rate through the optimal subnetwork M to a higher level $\overrightarrow{r}' = (1+\gamma)\overrightarrow{r} > \overrightarrow{r}$ produces an increase in congestion, that is $L(h) \leq L(\bar{h})$, where h and $\bar{h}$ are feasible Nash flows for $(M, \overrightarrow{r}, l)$ and $(M, \overrightarrow{r}', l)$ respectively. This statement follows directly from Proposition 3. More importantly, even allowing reoptimization of the subnetwork for the new triple $(C, \overrightarrow{r}', l)$ is not sufficient to reduce total latency, as proven by the following proposition.

propositionor traffic rates $\overrightarrow{r}$ and $\overrightarrow{r}' = (1+\gamma)\overrightarrow{r}$ with $\gamma > 0$, let M and M' be the optimal subgraphs for $(C, \overrightarrow{r}, l)$ and $(C, \overrightarrow{r}', l)$ respectively. If h and h' are feasible Nash flows for $(M, \overrightarrow{r}, l)$ and $(M', \overrightarrow{r}', l)$ respectively and l satisfies strong monotonicity, then

functions, they prove that taxes are more powerful than arrows' removal. Empirically, arrows' removal and access restrictions are more common than taxes as mechanisms of congestion regulation in urban areas.

[22] See Roughgarden (2006).

$L(h) \leq L(h')$.

Proof. In Appendix.

This monotonicity result shows that if the traffic rate $\overrightarrow{r}$ going through the triple $(C, \overrightarrow{r}, l)$ increases, then total latency $L(h)$ in the optimal subnetwork has to increase. This theoretical result is novel, but much of its relevance comes from the empirical implications we derive in what follows.

Under general assumptions on graph and latency other properties of the optimal subgraph remain elusive. For instance, even for $k = 1$ and with strictly convex latency functions, $L(h)$'s convexity with respect to $\overrightarrow{r}$ is not generally given[23]. Notice further that there is no assurance of $M' \subseteq M$, i.e. optimal subgraphs will display a progressive shedding of edges as population grows. As a counterexample let us double traffic input to $r = 2$ in Figure 2(b). We can show that $M' = C$ is an optimal subgraph for $(C, 2, l)$, while for $r = 1$ the optimal subgraph M was obtained by removing edge (v_1, v_2). This implies $M \subset M'$.

III. Empirical Mapping

We now operationalize the model and show what are its useful insights in interpreting the mapping data.

A. *City Structure: Assumptions*

We begin by giving the empirical correspondences of the network $C = (V, A)$ and the optimal subnetwork M. Sites within a city can be represented by a set of points on a map. This corresponds to the vertex set V of C. Arrows A in the directed graph indicate the direction of flow allowed from vertex to vertex (or from site to site).

Directions of flow are endogenous within cities and we represent such restrictions through subgraph M, the solution to problem (2). With regard to M, a planner is allowed to restrict

[23] Hall (1978) p.214 reports a counter-example for strictly convex latency functions.

access to a subset of the vertex set V in order to minimize total latency. This matches quite accurately the design of traffic plans for urban automotive transit. In the rest of our analysis we will focus on automotive traffic for this reason and because a large share of movement within a city is by ground transportation.

Our main operational assumption will be that planners intend to minimize latency as a primary objective. Empirically, congestion mitigation commissions routinely implement traffic control plans restricting access to certain routes and allow traffic flows to self-route through the remaining accessible edges of the graph to guarantee mobility over a given set of routes $\{s_i, t_i\}$ $i = 1, ..., k$. We recognize, however, that city planners may also have other goals in designing M: minimizing accidents, favoring some routes over others in the name of city development or as the result of pork-barrel/NIMBY politics, etc. As long as these secondary concerns do not completely trump traffic control, our model should still display a reasonable fit of the data.

City structure is described jointly by C and by the latency functions l.[24] An arrow's latency is generally the result of historical design (think of the thick Medieval walls restricting entry points to an Italian city center), but also of progressive reoptimization by transportation authorities. Certain roads are designed to carry a specific amount of flow. In our analysis we will consider latency functions l as given and control for historical and physical constraints that may induce higher latency on specific edges. Notice, however, that while we take latency functions l as exogenous, the flows argument of the function (and hence the latency) are endogenous in our setting.

Concerning travel demand elasticity, for every i, we are allowing perfect elasticity of travel demand along each particular path π_i in the network, combined with an inelastic travel demand over the $s_i - t_i$ commodity. This assumption aligns with the FHWA statement that "*The magnitude of demand elasticity depends heavily on the scope and time frame over which*

[24] We take the network structure as predetermined and operate by finding optimal subnetworks. The endogenization of network structures is a topic addressed by a growing literature within Economics. See for instance Acemoglu, Bimpikis and Ozdaglar (2009) and Jackson (2005) for a survey. An empirical example of the analysis of endogenous city structure is Burchfield et al. (2006).

travel demand is being measured. For example, a demand elasticity measured with respect to a single facility includes trips diverted from other routes or time periods and would be much higher than demand elasticities measured over a corridor or region."[25] It is important, however, to assess how sensitive our theoretical results are to the relaxation of the commodity i perfectly inelastic travel demand assumption. Appendix B presents a generalization of the model to the case of elastic travel demand and discusses in detail how our main theoretical implications are robust along this dimension.

Finally, a core innovation in our empirical mapping is to avoid the necessity of extensive traffic data information. Such micro information, albeit available for very limited areas, is not systematically available for the very large cross section of cities we employ.[26]

B. *Empirical Implementation*

Traffic flows along the different arrows within a city can be mapped into equilibrium Nash flows. If a flow f is feasible and Nash in $(C, \overrightarrow{r}, l)$, then any two $s_i - t_i$ paths $\pi^C_{i,1}, \pi^C_{i,2} \in \Pi^C_i$ which carry flow must have the same latency ℓ^C_i. The same has to be true for latency ℓ^M_i in the optimal subgraph M with respect to the feasible Nash flow h in $(M, \overrightarrow{r}, l)$. Total latencies must further satisfy $L(h) = \sum_i \ell^M_i r_i \leq \sum_i \ell^C_i r_i = L(f)$.

For $j = M, C$ and for each $s_i - t_i$ commute, we can decompose latency $\ell^j_i = \delta^j_i \sigma^j_i$ by multiplying physical distance δ^j_i and equilibrium unit travel time (inverse speed) $\sigma^j_i = 1/v^j_i$ for a path in Π^j_i. Ideally, one would like to measure aggregate latency costs $L(h) = \sum_i \ell^M_i p_i n_i$ and all the components of this sum (distance, speed, share of flow from each residency location, location population at source). We show in the data section that some of this information is available for a sample of US cities (but not all of it) and employ it to test Proposition 4.

Without loss of generality let us order the paths in Π^C_i and Π^M_i in terms of increasing

[25] See: http://www.fhwa.dot.gov/planning/itfaq.htm

[26] A necessary drawback of this approach is that, without micro flow data, empirical testing whether traffic is actually at a Nash equilibrium is not feasible. We will specify explicitly our operational assumptions in the subsequent empirical analysis.

distance, so $\Pi_i^j = \left\{\pi_{i,1}^j, \pi_{i,2}^j, ...\right\}$ where $\delta_{i,1}^j \leq \delta_{i,2}^j \leq ...$ and so on. We do not require $\pi_{i,z}^j$ to carry flow in equilibrium for $z = 1, 2, ...$, but if it does, then we write $\ell_i^j = \delta_{i,z}^j \sigma_{i,z}^j$.

Let us define Δ_i the linear distance between s_i and t_i. Since M is a subgraph of C, by the triangle inequality it must be that $\Delta_i \leq \delta_{i,1}^C \leq \delta_{i,1}^M$ for any i. This is because Δ_i is the euclidean distance between source and target, while $\delta_{i,1}^C$ is the shortest-distance path through the network when all edges are accessible, and finally $\delta_{i,1}^M$ is the shortest-distance path through the subnetwork M.

Finally, let us notice that $\delta_{i,1}^M$ may correspond to a path $\pi_{i,1}^M$ carrying no flow in equilibrium (for instance, because $\pi_{i,1}^M$ latency even at zero flow is extremely high). However, at least one minimum-distance flow-carrying path $\pi_{i,X}^M \in \Pi_i^M$ exists, given that $r_i > 0$ and $\Pi_i^M \neq \varnothing$. By definition $\delta_{i,1}^M \leq \delta_{i,X}^M$, so for $\pi_{i,X}^M$ it must be that for any i:

$$\Delta_i \leq \delta_{i,1}^C \leq \delta_{i,X}^M. \tag{3}$$

Let $\bar{v}_i$ be the reference speed at zero flow on $\pi_{i,X}^M$, we also have $\frac{1}{\bar{v}_i}\delta_{i,X}^M \leq \sigma_{i,X}^M \delta_{i,X}^M$ for each i.[27]

The interesting point here is that Δ_i, $\delta_{i,1}^C$, $\delta_{i,X}^M$ have a direct empirical correspondence. Given a $\{s_i, t_i\}$ address pair, automated mapping applications accessible over the Internet (such as *Google Maps*, which we employ) output directions including Δ_i as the *euclidean distance*, $\delta_{i,1}^C$ as the *walking distance*, and $\delta_{i,X}^M$ as the *driving distance*. Google Maps further supplies a *reference travel time* $\frac{1}{\bar{v}_i}\delta_{i,X}^M$ based on speed limits. These are close empirical matches of our theoretical quantities.[28] An example for Midtown Manhattan is presented in

[27] $\bar{v}_i$ is sometimes referred to as free-flow speed or zero-flow speed in the literature.

[28] We choose maps.google.com out of available web mapping interfaces (such as MapQuest.com or maps.yahoo.com) for its response speed. Concerning directions, the Google maps algorithm's details are proprietary. However, the algorithm is a variant of Dijkstra's (1959) algorithm, the fastest algorithm to solve weighted-graphs problems with non-negative edges (precisely the type of problems allowed within our graph $C = (V, A)$). The algorithm starts from a source and computes the latency to all directly connected vertices, where latencies are approximated by edge length (if the shortest route is considered) or by edge length weighted by speed limit (if the fastest route is considered). The algorithm then moves onto the following set of connected verteces, and so on and so forth, until the shortest path to the target is found. Standard paths are stored in the Google Maps' cache to speed up computation.

Two caveats are in order here. First, while directions by car are operative in Google maps, walking directions are beta (in the empirical section we have to discard some bad output due to this issue). Second, Google maps currently does not allow for a shortest/fastest direction choice, but the output likely balances

Figure 3.

It is feasible to sample a large set of possible commutes from city address books and collect this information. Incidentally, as theoretically predicted, the data will confirm inequality (3), with only a very small number of bad exceptions due to heuristic approximations within Google.[29]

We can now present our main empirical remarks.

Remark 1. *An explicit test of Proposition 4 can be formulated as:*

If M and M' are optimal subgraphs of problem (2) for triple $(C, \overrightarrow{r}, l)$ and $(C, \overrightarrow{r}', l)$ respectively, where $r = \sum_{i=1}^{k} p_i w_i N$ and $r' = \sum_{i=1}^{k} p_i w_i N'$, and if $N < N'$, then

$$\sum_{i=1}^{k} \left(\delta_{i,X}^{M} \sigma_{i,X}^{M}\right) p_i w_i N \leq \sum_{i=1}^{k} \left(\delta_{i,X'}^{M'} \sigma_{i,X'}^{M'}\right) p_i w_i N'$$

(i.e. total latency increases).

Remark 2. *Given the following four claims:*

1. *If $N < N'$, per traveler latency increases*

$$\sum_{i=1}^{k} \left(\delta_{i,X}^{M} \sigma_{i,X}^{M}\right) p_i w_i \leq \sum_{i=1}^{k} \left(\delta_{i,X'}^{M'} \sigma_{i,X'}^{M'}\right) p_i w_i,$$

2. *If $N < N'$, total delay increases*

$$\sum_{i=1}^{k} \delta_{i,X}^{M} \left(\sigma_{i,X}^{M} - \frac{1}{\bar{v}_i}\right) p_i w_i N \leq \sum_{i=1}^{k} \delta_{i,X'}^{M'} \left(\sigma_{i,X'}^{M'} - \frac{1}{\bar{v}_i'}\right) p_i w_i N',$$

the two. We include more details on the interface in the Data Section. For our analysis we assume that if Google Maps gives a path as the driving directions output, there is flow on that path and it is the shortest flow-carrying path, $\pi_{i,X}^{M}$. This matches the fact that Google does not output necessarily the shortest driving path $\pi_{i,1}^{M}$ but puts some weight on speed as well.

[29] See Data Section for details.

3. If $N < N'$, per traveler delay increases

$$\sum_{i=1}^{k} \delta_{i,X}^{M} \left(\sigma_{i,X}^{M} - \frac{1}{\bar{v}_i} \right) p_i w_i \leq \sum_{i=1}^{k} \delta_{i,X'}^{M'} \left(\sigma_{i,X'}^{M'} - \frac{1}{\bar{v}_i'} \right) p_i w_i,$$

4. If $N < N'$, average reference latency increases

$$\sum_{i=1}^{k} \left(\delta_{i,X}^{M} \frac{1}{\bar{v}_i} \right) p_i w_i \leq \sum_{i=1}^{k} \left(\delta_{i,X'}^{M'} \frac{1}{\bar{v}_i'} \right) p_i w_i,$$

two sufficient conditions for Remark 1 and Proposition 4 to hold are: (a) Claim 2 and claim 4 jointly hold; (b) Claim 3 and claim 4 jointly hold.

In Section 5 we validate empirically claims 2-4. If claims 2 and 4 hold jointly, Remark 1 must hold, hence verifying Proposition 4.[30] If claims 3 and 4 hold jointly, claim 1 must also hold[31] and Proposition 4 will hold a fortiori.[32]

Testing Remark 1 offers evidence in support of Proposition 4, our main theoretical implication, but does not offer evidence on whether increases in total latency originate exclusively from slowing down of traffic or from city plans M allowing only longer paths through their edges (hence increasing average distance traveled) or both. As shown in the previous section there is no assurance $M' \subseteq M$ in general, so an open empirical question is whether:

Remark 3. *City networks satisfy the following two claims:*

1. If $N < N'$, average distance traveled increases

$$\sum_{i=1}^{k} \left(\delta_{i,X}^{M} \right) p_i w_i \leq \sum_{i=1}^{k} \left(\delta_{i,X'}^{M'} \right) p_i w_i,$$

[30] If claim 2 holds then $\sum_{i=1}^{k} \delta_{i,X}^{M} \left(\sigma_{i,X}^{M} - \frac{1}{\bar{v}_i} \right) p_i w_i N \leq \sum_{i=1}^{k} \delta_{i,X}^{M'} \left(\sigma_{i,X}^{M'} - \frac{1}{\bar{v}_i'} \right) p_i w_i N'$. It follows from claim 4 that:
$\sum_{i=1}^{k} \delta_{i,X}^{M} \left(\sigma_{i,X}^{M} - \frac{1}{\bar{v}_i} \right) p_i w_i N \leq \sum_{i=1}^{k} \delta_{i,X}^{M'} \left(\sigma_{i,X}^{M'} - \frac{1}{\bar{v}_i'} \right) p_i w_i N' \leq \sum_{i=1}^{k} \delta_{i,X}^{M'} \left(\sigma_{i,X}^{M'} \right) p_i w_i N' - \sum_{i=1}^{k} \delta_{i,X}^{M} \frac{1}{\bar{v}_i} p_i w_i N$. Adding $\sum_{i=1}^{k} \delta_{i,X}^{M} \frac{1}{\bar{v}_i} p_i w_i N$ to all elements of the inequality implies Remark 1.

[31] By adding claim 3 and claim 4.

[32] Notice however that Proposition 4 does not imply claims in Remark 2, nor under general assumptions on the directed graph C and latency functions l such claims will necessarily hold.

2. If $N < N'$, average speed of transit decreases

$$\sum_{i=1}^{k} \left(v_{i,X}^{M}\right) p_i w_i \geq \sum_{i=1}^{k} \left(v_{i,X'}^{M'}\right) p_i w_i,$$

where $v_{i,X}^{M} = 1/\sigma_{i,X}^{M}$.

A final step in the empirical mapping is required. Our comparative statics with respect to population N holds graph C constant. Empirically, C evolves over time within a city and differs across cities. Our approach is to hold specific first order statics of the directed graph C constant, in particular characteristics affecting the location of sites within the city (e.g. water basins and streams, wetlands, steep-sloped terrain, coastal position, latitude, longitude, elevation and variation in elevation over the city surface, etc.) and graph tortuosity.[33] This approach is admittedly rough, but to the best of our knowledge the only one implementable for general comparison across city networks.

Let us define the average path *tortuosity* for graph C and latency functions l:

$$t(C, l) = \sum_{i=1}^{k} \left(\delta_{i,1}^{C} - \Delta_i\right) w_i. \tag{4}$$

A low-t city is one where moving from origin to destination is on direct lines (think of walking through downtown Chicago), as opposed to a high-t city where moving on a O-D path is on winding lines (think of walking through downtown Siena). Tortuosity is occasionally mathematically defined through an arc-chord ratio, which in our case would deliver the measure $T(C, l) = \sum_{i=1}^{k} \left(\delta_{i,1}^{C}/\Delta_i\right) w_i$. Tortuosity parsimoniously helps capturing limitations in the availability of links within the network, and historical features of the city structure.

[33] As explicitly stated in Section 5.A, in order to saturate for potential nonlinearities a fourth order polynomial in tortuosity, as defined in (4), a fourth order polynomial in population density, and state fixed effects are also employed to condition on C.

Analogously to (4), we can define the average per path *distortion* an individual incurs due to congestion for given triple $(C, \overrightarrow{r}, l)$ and for M solving (2):

$$d\left(C, \overrightarrow{r}, l\right) = \sum_{i=1}^{k} \left(\delta_{i,X}^{M} - \delta_{i,1}^{C}\right) w_i. \tag{5}$$

When positive, (5) emphasizes how latency minimization induces excess circuitousness into the subnetwork. It provides an intuitive measure of the degree to which paths have to be more roundabout in presence of congestion relative to a theoretical benchmark where an individual is able to freely travel through the network alone. Since such distortions precisely obey the logic of the Braess' paradox, we refer to them as Braess' distortions. An analogous mean ratio for transit distortions is $D\left(C, \overrightarrow{r}, l\right) = \sum_{i=1}^{k} \left(\delta_{i,X}^{M} / \delta_{i,1}^{C}\right) w_i$.

Claim 1 of Remark 3, if verified, implies:

Remark 4. *If M and M' are optimal subgraphs of problem (2) for triple $(C, \overrightarrow{r}, l)$ and $(C, \overrightarrow{r}', l)$ respectively, where $r = \sum_{i=1}^{k} p_i w_i N$ and $r' = \sum_{i=1}^{k} p_i w_i N'$ with $N < N'$, then $\sum_{i=1}^{k} \left(\delta_{i,X}^{M} - \delta_{i,1}^{C}\right) w_i \leq \sum_{i=1}^{k} \left(\delta_{i,X'}^{M'} - \delta_{i,1}^{C}\right) w_i$ (i.e. average travel distortion increases).*

A useful result in assessing the economic importance of Braess distortions is that excess circuitousness in the subnetwork bounds from below the gains in terms of speed of transit that must accrue due to the restrictions implemented by the planner, or more formally:

corollaryiven triple $(C, \overrightarrow{r}, l)$ with optimal subnetwork M, for each $s_i - t_i$ the ratio between driving and walking distance $\delta_{i,X}^{M} / \delta_{i,1}^{C}$ is less or equal than $v_{i,X}^{M} / v_{i,1}^{C}$, the ratio between the speed on the shortest driving path in M and the speed at which it would be possible to drive on the shortest path in the unrestricted network C.

Proof. In Appendix.

If the average excess distortion measured in a city is $\sum_{i=1}^{k} \left(\delta_{i,X}^{M} / \delta_{i,1}^{C}\right) w_i = 1.1$, this corollary implies that traffic moves on average at least 10 percent faster in the restricted

network relative to what would emerge in equilibrium in the (counterfactual) unrestricted network.

IV. Data

A. *Latency Costs*

From the 2009 Annual Urban Mobility Report published by the Texas Transportation Institute (TTI), part of the Texas A&M University System, we collect data on city-level total delay $\sum_{i=1}^{k} \delta_{i,X}^{M} \left(\sigma_{i,X}^{M} - \frac{1}{\bar{v}_i} \right) p_i w_i N$, average delay per traveler $\sum_{i=1}^{k} \delta_{i,X}^{M} \left(\sigma_{i,X}^{M} - \frac{1}{\bar{v}_i} \right) p_i w_i$, and average traffic speed $\sum_{i=1}^{k} \left(v_{i,X}^{M} \right) p_i w_i$ measures. Unfortunately, total latency is not reported by TTI (2009a). The Annual Urban Mobility Report assesses time delays and roadway congestion costs data for 90 large urban areas in the United States.[34] Measures of urban roadway congestion are generated from raw data on urban arterial and freeway traffic supplied through the Highway Performance Monitoring System from the Federal Highway Administration. To the best of our knowledge, TTI's Annual Urban Mobility Report is among the most comprehensive and complete sources of congestion costs estimates available for any country and the most comprehensive in terms of road miles covered for US cities.

For cross-validation purposes and robustness to employing traffic data obtained from alternative methodologies, we also collect congestion rankings from the INRIX National Traffic Scoreboard. INRIX, one of the country's leading traffic and navigation services company, gathers and analyses raw data originated from its proprietary Smart Driver Network, including 1.6 million GPS-equipped probe vehicles (including taxis, shuttle vans, service delivery trucks, and consumer vehicles) generating speed, direction, and location information for the largest 100 metropolitan areas in the United States.[35]

[34] See http://mobility.tamu.edu/ums/congestion_data/. The Annual Mobility Report includes estimates of congestion costs for an aggregate of 439 urban areas. TTI however collects detailed data only on the original 90 areas.

The remaining 349 areas are rolled into 1 aggregate, in essence treating them as a single large urban area for analysis purposes and no disaggregated data exist. We thank David Schrank at TTI for clarifying this issue.

[35] See http://www.inrix.com/pressrelease.asp?ID=94

Table 1 reports the top ten metropolitan areas for congestion costs in both samples. The TTI and INRIX samples refer to traffic conditions in the most recent period available (2007 and 2009, respectively), for maximal comparability between congestion costs and the corresponding traffic path distortions generated by the (continuously updated) Internet mapping applications described below.

B. *Braess' Distortions and Tortuosity*

The process of measuring traffic path distortions within the city network C is a crucial methodological step in our analysis. Given the relative novelty of our approach we feel that extra scrutiny is necessary. We do not limit ourselves to the analysis of US urban areas for which TTI measures are available, but we extend the sample both cross-sectionally within the United States and cross-country, focusing on Italian cities, an arguably different sample in terms of urban structure and local administration organization.

Concerning US cities, we consider the top 500 most populous census designated places (CDP). These data are available from the US Census Bureau[36], which describes CDPs as "delineated to provide data for settled concentrations of population that are identifiable by name but are not legally incorporated under the laws of the state in which they are located."[37]

For the Italian city sampling we employ the cities considered in Guiso, Sapienza and Zingales (2008), a large cross-sectional study of Italian municipalities. The data includes

INRIX coverage is more limited relative to TTI, as only mainline lanes of limited access highways are included in the INRIX samples.

[36] The census places come from the 2000 vintage census data. Places in Puerto Rico are not considered in constructing this list. U.S. Census Bureau. Places, cartographic boundary files, descriptions and metadata are available at http://www.census.gov/geo/www/cob/pl metadata.html#cdp 2005

[37] Census designated places mostly conform to commonly recognized political boundaries, but there are several results of this worth mentioning. Minneapolis and St. Paul, Minnesota *for example,* are considered as two distinct CDP's, not a single metro area. In some cases (e.g. Sarasota and Bradenton, Florida) cities which in tandem constitute a large metro-area, do not individually fall in the top 500 CDP's and so are not considered in this analysis. Some CDP's do not exist in any commonly understood sense, like Paradise, NV, which includes much of downtown Las Vegas, and exists for practical purposes, only as a CDP. These limitations and additional heuristics in Google (described below) limit the full U.S. sample to 457 municipalities.

692 Italian municipalities and all Italian main urban areas (*Capoluoghi di Provincia*).

a. Path Sampling Population is located at C's sources s_i and travels towards C's sinks t_i. In order to generate a sizeable set of random commutes within a city we generate $\{s_i, t_i\}$ $i = 1, ..., \hat{k}$ address pairs using a parallel algorithm for the US and Italian samples.

According to our model, source addresses can be randomly sampled, but sampling should be weighted by source population. Consider collecting the addresses of all people called Smith living in Chicago (or Rossi, living in Florence). Let us assume that people called Smith locate at the sources $s_1, ..., s_k$ with probability equal to the sources' subpopulation shares $w_1, ..., w_k$, where $w_i = \frac{n_i}{N}$ and $\sum_i w_i = 1$. That is, we assume that a more populous suburb s_i has more people called Smith than a less populous suburb $s_{i'}$ with $w_{i'} < w_i$ and that the share of people named Smith is the same in both s_i and $s_{i'}$. Then, the distribution of people named Smith is going to mimic the empirical distribution of source population $s_1, ..., s_k$ and the list of their addresses will provide a representative subset of source locations. Considering a comprehensive list of last names further eliminates name-specific idiosyncrasies. We follow this approach.[38]

The top 100 most common surnames in the US are derived using the available Census (2009) totals for name frequency. The analog for the Italian version is based on data found on the website cognomix.it. For each place, and each surname calculated above, addresses are collected from phone directories.[39] These addresses are then randomly paired in order to create the endpoints for random routes through the city.

Notice that our path sampling approach focuses on residence-to-residence, not on residence-

[38] One caveat is in order. Our approach will not adequately capture city segregation by race in the case of highly race-specific last names. In particular, for racial minorities living in concentrated areas, names will not appear random across network sources s. This may be relevant for Asian or Hispanic or Middle-Eastern enclaves. Such distortions should be quantitatively small, however, and confined to a minority of the city population.

[39] We employed whitepages.com or paginebianche.it websites, for the U.S.or Italy respectively.

Census designated places that do not conform to normal political boundaries may not have entries in the white pages. In the top 500 most populous CDPs there are three that fail to return results from the white pages. They are: Arden-Arcade, CA; Paradise, NV and Sunrise Manor, NV.

In the Italian case it is possible to collect all data available, while from whitepages.com it is possible to collect at most 100 addresses per name-place pair.

to-business pairs, which are more representative of a typical work commute (but not necessarily of all trips within the city). Since residence-to-residence paths can cross business districts, these structural features will not be lost in our data. For instance, in a circular city model with residences uniformly located on a circumference of radius Φ and a central business district of radius $\Phi/2$ (hence, covering 1/4 of the city disk) the probability of any random commute passing through the business district is 1/3. [40]

For each city C, $\hat{k}$ is sufficiently large to guarantee asymptotic convergence of sample moments, such as $\hat{d}(C, \overrightarrow{r}, l) = \sum_{i=1}^{\hat{k}} \left(\delta_{i,X}^{M} - \delta_{i,1}^{C}\right) w_i$, to population moments, such as $d(C, \overrightarrow{r}, l)$.[41]

Finally, since estimates of p_i are not generally available, in computing per traveler city averages based on Google we will proceed under the assumption of a constant share of travelers out of each location (i.e. constant probabilities $p_i = p$). This will allow to focus on all the empirical implications of the model, albeit at a cost in terms of generality.

b. Google Maps Data Once random commutes $\{s_i, t_i\}$ are generated, the various travel distances between the points must be calculated. This is accomplished using the Google Maps API, through which we can programmatically query the Google Maps service and determine the travel distance by car, $\delta_{i,X}^{M}$, and by foot, $\delta_{i,1}^{C}$, for each commute. Figure 3 presents an example of a commute in Midtown Manhattan, where the driving distance clearly exceeds the walking distance due to the specific restrictions of flow along some of the

[40] The potential difference in the distribution of business and residence locations can also be addressed directly by picking target destinations from business directories. We experimented with spot sampling from White and Yellow Pages in a subset of U.S. cities without any evidence of systematic distortions in the data. We ultimately decided against the use of Yellow Pages because of the difficulty in systematically weighting "employment per business phone number", a necessary component in assessing the likelihood of a given business commute (e.g. a large employer will be the target of more commutes than a small one, but they may have the same number of phone numbers listed, as the large employer may only list its main number).

[41] In the Italian data the number of pairs generated is a truncated function of the population of the Italian census place. For census places with populations in the lowest or highest quartile, 0.5 percent of the relevant quartile boundary is used as the target number of pairs to generate. For census places with populations in the interquartile interval, the number of pairs generated is equal to 0.5 percent of the actual population. In the US sample, 0.5 percent of the population of the 75th percentile city is used for all cities where enough addresses are available. This results in $\hat{k} = 622$ pairs for each US city, unless the number of addresses collected is less than this total.

graph's arrows. We also collected the latitude and longitude of the points in question, which allows us to calculate the euclidean distance Δ_i, between the points as well.[42] This process relies on the ability of Google to translate the given addresses into latitude and longitude coordinates via a process called geocoding.[43]

In this project, each pair of points required two requests to the API, one to get the distance based on driving directions, and one to get the distance based on walking directions. From these requests, the travel distance, latitude and longitude, and the accuracy of the location for both points were parsed from the JSON response and stored in a new data set. Aggregate measures of these values, calculated for each census place, constitute the final output of the procedure. The following items are constructed for each census place: 1. The mean value of driving distance traveled; 2. The mean travel duration at reference speed; 3. The mean value of the difference between driving and walking distances $d\left(C, \overrightarrow{r}, l\right)$; 4. The mean value of the difference between walking and euclidean distances $t(C, l)$; 5. The mean value of the ratio of walking to euclidean distances $D\left(C, \overrightarrow{r}, l\right)$; and 6. The mean value of the ratio of driving to walking distances $T(C, l)$. For robustness checks we constructed all the corresponding medians as well.

Some summary information about the data collection per city and its outcomes are presented in Table 2. Much more data was collected for the US case than for the Italian case when employing googlemaps.com. Reasons for differences in the number of pairs attempted and the number realized are: 1. Failure to properly geocode the address with high enough accuracy. The Google Maps API returns a code indicating the accuracy of the mapping, we accept values whose accuracy is indicated to be street-level or higher. 2. The Google Maps API returns a code indicating failure to generate directions for the given points. Naturally, such points have nothing to contribute to our analysis. 3. Paired addresses map to the same

[42] The euclidean distance is calculated using the Haversine Formula, which is a formula for calculating shortest distances on a spherical approximation of the earth. It does not take into account changes in elevation.

[43] The Google Maps API is fully described in that service's documentation at google.com, and those interested in the details of using the service are advised to look there.

location, resulting in zero distance for all measures. Such results are dropped.

As Table 2 also shows, the rate of failures in the Italian data is about three times that of the US data. This is likely due to the relative difficulty of geocoding Italian addresses, to the small size of some of the Italian cities in our sample, and poorer mapping information for Italian cities.

Table 3 presents the summary statistics for both the TTI US sample and the full US city sample. Units for distances are meters. On average the typical sampled path is 14.5 km in driving distance in the TTI sample and 10.2 km in linear distance, reasonable commute lengths. Delays' units are measured in annual excess hours, while travel duration at reference speed (approximated speed limits) are reported in seconds. Traffic congestion on average produces 31 hours of delay per year per traveler (including cost of time and fuel the TTI economic losses are substantial, as evident in the congestion cost estimates reported in the Table). The average time duration of the typical commute at reference speed is about 16.5 minutes (990 seconds). This seems a reasonably representative figure if compared with the US Census 2009 American Community Survey average daily time to work of 25.1 minutes (computed at actual speed). Finally, notice that both driving/walking and walking/linear distance ratios are on average larger than 1, as required by (3).

V. Empirical Results

A. *Remark 1: The Evidence*

Lacking data on total latency, it is not feasible to test the main implication of the model (Proposition 4) directly. In this section we follow an indirect approach.

We begin by showing that larger city populations N are positively correlated with longer total delays (Remark 2, claim 2), longer delays per traveler (Remark 2, claim 3), higher average reference travel durations (Remark 2, claim 4), longer average distance traveled (Remark 3, claim 1), and lower average speed (Remark 3, claim 2). These five claims are directly testable using either TTI data or Google Maps data. We present both uncondi-

tional and conditional scatterplots in Figure 4 to Figure 8. The set of controls for the conditional regressions in the right panels includes: state fixed effects (hence elasticities are estimated employing within-state variation only), a fourth order polynomial in tortuosity $t(C, l)$, a fourth order polynomial in population density (a typical control for city structure), a coastal city dummy, latitude, longitude, km^2 of lakes/swamps/ocean, km of coast, km of rivers/streams, mean elevation, standard deviation of elevation, mean terrain slope, standard deviation of slope, percent area suitable for development[44].

In all reported regressions we cluster our standard errors at the state level. All variables are expressed in natural logarithms, to reduce the incidence of outliers.

Figures 4-8 make clear that claims 2-5 of Remark 2 and 1-2 of Remark 3 are all strongly supported by the data. All regressions, unconditional and conditional, are statistically significant at standard confidence levels. The unconditional elasticity of total delay to population is estimated at 1.10. The unconditional elasticity of delay per traveler to population is estimated at 0.34. The unconditional elasticity of travel duration at reference speed to population is estimated at 0.13. The unconditional elasticity of distance traveled to population is estimated at 0.22. The unconditional elasticity of travel speed to population is estimated at -0.08.

Figure 9 also reports scatterplots comparing TTI to the Couture, Duranton, and Turner (CDT, 2012) and INRIX congestion indexes and population, a reduced-form result without a direct link to our model, but reassuring of TTI data's reliability. In particular, the CDT inverse travel speed index is reassuring, as based on observed travel speeds by a sample of commercial and privately owned vehicles from the National Household Transportation Survey.

In order to test Remark 1 (i.e. Proposition 4) and claim 1 of Remark 2 we apply a conservative multiple hypothesis testing approach based on the Bonferroni inequality (Savin, 1984). This is necessary as we are jointly testing multiple one-sided claims. For each null

[44]See Appendix C for variable definition and data sources.

hypothesis (i.e. that population elasticities of total delay, per traveler delay, and reference travel duration are nonpositive), the Bonferroni correction requires the respective p-values to fall below α/q, where α is the confidence level and $q = 3$ is the number of multiple hypotheses jointly considered (Remark 2, claims 2-4). We report the relevant one-sided hypothesis χ^2 test statistics and their p-values (both unconditional and conditional) in Table 4. We reject each null hypothesis substantially below the .0033 (= .01/3) confidence threshold with the sole exception of the unconditional χ^2 statistic for reference travel duration, which has a p-value of 0.0077. According to the Bonferroni inequality we reject the null that Proposition 4 fails at 5 percent confidence level for the unconditional model and at 1 percent for the conditional model. Claim 1 of Remark 2 is supported at equal levels of statistical significance.

B. *Remark 4: Braess' Distortions*

Our model is predicated upon the assumption that city planners may restrict access to network edges with the objective of minimizing total latency at Nash equilibrium. Under general assumptions the model is silent on how increases in traffic rates affect the optimal M (a task requiring far more stringent assumptions on the triple $(C, \overrightarrow{r}, l)$). However, we are able to present direct empirical evidence on how population correlates with the average path distortion $d\,(C, \overrightarrow{r}, l)$ that an individual incurs due to congestion and show that Remark 4 holds in the data.

Figure 10 reports that larger city populations N are positively correlated with larger Braess' distortions in the TTI sample, both conditionally and unconditionally. The unconditional elasticity of travel distortion to population is estimated at 0.37. Relative to Figure 7, where the estimated elasticity was 0.22, this result is quantitatively larger, reflecting the improvement in holding city structure C constant across cities when focusing on deviations relative to the shortest path available $\delta^C_{i,1}$.

Figure 11 shows that the positive relationship between Braess' distortions and city size holds in the full US sample as well, and not just in the more limited TTI sample. Figure

12 presents kernel densities estimated for the overall sample of US cities. For completeness we report here both average excess driving to walking distance $d\left(C, \overrightarrow{r}, l\right)$ and average ratio $D\left(C, \overrightarrow{r}, l\right)$. For both measures Figure 12 highlights a relatively high dispersion in terms of transit distortions across cities and heavily right-skewed distributions.

Moving to the Italian sample, Figure 13 again reports a positive and robust correlation between path distortions and city population. Standard errors in all regressions are clustered at the provincial level for the Italian sample.

The unconditional population elasticities range between 0.63 for the full US sample and 0.39 for the Italian sample, quantitatively comparable estimates. The quantitative implication in the most comprehensive US sample is that an extra 10 percent population moving through the same city network is associated with an increase in the distance one has to travel by car of *an additional* 6.3 percent. Going from A to B, where A and B are located on average 6.7 km from one another, where the measure of distance is the euclidean distance and where initial travel distortion is 1.78 km, requires 112 ($= 1,780 * 0.063$) extra meters of travel when the city population moving through the network increases by 10 percent.

To the best of our knowledge this represents a novel fact in the analysis of urban structure and one that points to a pervasive presence of Braess' distortions in more populated networks. In addition, this fact has an interesting correspondence to a related literature on the theoretical pervasiveness of the Braess' paradox in networks (Frank, 1981; Fisk and Pallottino, 1981; Valiant and Roughgarden, 2010; Chung and Young, 2010).

C. *Binding the Gains from Braess' Distortions in Cities*

It is important to assess how much one can possibly gain from imposing Braess' distortions in cities. For a flow f feasible and Nash in $(C, \overrightarrow{r}, l)$, Roughgarden and Tardos (2002) define the Braess' ratio as:

$$\Theta\left(C, \overrightarrow{r}, l\right) = \max_{\tilde{M} \subseteq C} \left[\min_{i=1}^{k} \frac{\ell_i^C}{\ell_i^{\tilde{M}}} \right]. \tag{6}$$

Intuitively, the Braess' ratio indicates the largest congestion amelioration that can be obtained by excluding some edges of graph C. A sizeable literature has developed around placing theoretical bounds around Braess' ratios, including Lin, Roughgarden, Tardos, and Walkover (2009) among the others. In this section we take a complementary route and present *empirical* bounds for the value of Braess' ratio in cities.

It is unfeasible to focus on all possible traffic plans for all cities in our sample as required by (6), so we focus on the implemented M, modifying the definition of Braess ratio to:

$$\beta\left(C, \overrightarrow{r}, l\right) = \min_{i=1}^{k} \frac{\ell_i^C}{\ell_i^M}. \tag{7}$$

Notice that when $k = 1$ then $\beta\left(C, \overrightarrow{r}, l\right) = L(f)/L(h)$. Re-expressing total latencies in terms of average distance and speed, we can bound the Braess' ratio:

$$\beta\left(C, \overrightarrow{r}, l\right) = \frac{\delta_{i,X}^C \sigma_{i,X}^C}{\delta_{i,X}^M \sigma_{i,X}^M} \leq \frac{\delta_{i,X}^C \sigma_{i,X}^C}{\delta_{i,X}^M \frac{1}{\bar{v}_i}} \leq \frac{\delta_{i,1}^C \sigma_{i,1}^C}{\delta_{i,X}^M \frac{1}{\bar{v}_i}} \leq \frac{\delta_{i,1}^C \frac{1}{v_i^0}}{\delta_{i,X}^M \frac{1}{\bar{v}_i}},$$

where i is the path satisfying (7) and v_i^0 the slowest speed on the shortest path in Π_i^C.

Google Maps allows to operationalize the upper bound by providing measures of $\delta_{i,1}^C$ and $\delta_{i,X}^M$. TTI calibrates a set of plausible approximations for $\bar{v}/v^0$ in congested arterial and freeway traffic at heavy ($\bar{v}/v^0 = 1.21$), severe ($\bar{v}/v^0 = 1.33$), and extreme conditions ($\bar{v}/v^0 = 1.51$)[45]. These calibrations are necessary, as we do not observe the counterfactual speed in the unrestricted network C. Assuming $\bar{v}/v^0$ on all arrows, we could potentially focus on the sampled $\min_{i=1}^{\hat{k}} \left[\frac{\bar{v}}{v^0} \frac{\delta_{i,1}^C}{\delta_{i,X}^M}\right]$ for each city for consistency with Roughgarden and Tardos (2002). However, a statistical approach more robust to idiosyncratic bad searches is to focus on sample means of walk-drive ratios $\frac{\bar{v}}{v^0} \sum_{i=1}^{k} \left(\delta_{i,1}^C / \delta_{i,X}^M\right) w_i$.

For the US sample, $\sum_{i=1}^{k} \left(\delta_{i,1}^C / \delta_{i,X}^M\right) w_i = .92$ implies an upper bound on the Braess' ratio for that city of 1.11 in heavy conditions, or a gain of at most 11 percent from imposing

[45] Exhibit A-6. Daily Traffic Volume per Lane and Speed Estimating Used in Delay Calculation in TTI (2009b).

Braess' distortions. Similarly we assess a 22 percent gain in severe conditions, and 39 percent gain in extreme counterfactual conditions.

According to the Federal Highway Administration 2010 Traffic Volume Trends, in the US 3,000 billion miles were driven annually in 2009, about 60% of which were driven in urban areas. This implies 1800 billion urban miles. Assuming a (conservatively high) average speed of 40 miles per hour (coinciding with the TTI highway and freeway estimated average speed), a value of time of \$16.01 per person-hour and \$105.67 per truck-hour as per TTI (2009a), and assuming 95% of urban hours are driven by car, this implies a total urban latency of \$922 billion. It follows an upper bound on the economic value of Braess' distortions of \$103 billion assuming heavy conditions in the counterfactual network, of \$204 billion assuming severe conditions, and of \$360 billion in the extreme case.

While this exercise is not useful in assessing potentially informative lower bounds on the value of such distortions (impossible to obtain without further assumptions on the counterfactual equilibrium at C), it emphasizes under reasonable calibrations the potential of large gains stemming from Braess' distortions.

VI. Conclusions

This paper offers a contribution to the analysis of congestion costs in cities. We present a theoretical model of congestion costs and endogenous traffic distortions that allows a straightforward mapping into empirically measurable quantities for a large sample of United States and Italian cities.

We make use of widely available Internet mapping applications, verify the general predictions of our model in the data, and further estimate positive and large population elasticities of traffic distortions in cities. We show that our results hold tightly across samples of large and small cities within the United States and within a large sample of Italian cities. This ultimately provides a novel set of cross-city stylized facts useful in the analysis of congestion costs. We further interpret these findings as pointing in the direction of a large role for

Braess' paradox in more populated city networks. From a methodological standpoint, our empirical path distortion and path tortuosity measures for cities can be considered useful summary statistics of network structures, possibly of use to geography and transportation economists.

Finally, one can further envision settings where network structure affects the matching, search, and interaction of individuals within the urban community. This is a relatively difficult problem to tackle quantitatively in Economics and in Sociology, but one for which we offer a methodological addition. This suggests potential applications of our city network measures, ranging from the analysis of social and cultural capital accumulation within cities (e.g. Are cities with more difficult/distorted city networks characterized by lower or higher levels of social capital?) to the study of within-city location decisions of businesses.

REFERENCES

[1] **Acemoglu, Daron, Kostas Bimpikis, and Asu Ozdaglar,** 2009. "Price and Capacity Competition," *Games and Economic Behavior*, vol. 66, no. 1, 1-26.

[2] **Acemoglu, Daron and Asu Ozdaglar,** 2007. "Competition in Parallel-Serial Networks" *IEEE Journal of Special Areas of Communication, Special Issue on Non-Cooperative Behavior in Networking* 25, pp. 1180-1192.

[3] **Anas, Alex, Richard Arnott and Kenneth A. Small**, 1998. "Urban Spatial Structure" *Journal of Economic Literature*, 36, 1426-1464.

[4] **Arnott, Richard, Andre de Palma and Robin Lindsey,** 1993. "A Structural Model of Peak-Period Congestion: A Traffic Bottleneck with Elastic Demand", *American Economic Review*, 83, 161-179.

[5] **Arnott, Richard and Kenneth A. Small,** 1994. "The economics of traffic congestion." *American Scientist* 82: 446-455.

[6] **Arnott, Richard, Tilmann Rave and Ronnie Schöb,** 2005. *Alleviating Urban Traffic Congestion*. MIT Press.

[7] **Beckmann, Martin, C. Bartlett McGuire and Christopher B. Winsten**, 1956. *Studies in the Economics of Transportation*. Yale University Press.

[8] **Braess, Dietrich**, 1968. "Über ein Paradoxon aus de Verkehrsplanung" *Unternehmensforschung,* 12, 258-268.

[9] **Burchfield, Marcy, Henry Overman, Diego Puga and Matthew Turner,** 2006. "Causes of Sprawl: A Portrait from Space", *Quarterly Journal of Economics* 121(2), .

[10] **Census Bureau US,** 2009. *Genealogy data: Frequently occurring surnames from census 2000* - US Census Bureau, December 29, 2009.

[11] **Chen, Po-An and David Kempe**, 2008, "Altruism, Selfishness, and Spite in Traffic Routing." *Proceedings of ACM EC 2008*, Chicago, IL

[12] **Chung, Fan and Stephen J. Young**, 2010, "Braess's Paradox in Large Sparse Graphs" WINE.

[13] **Cole, Richard, Yevgeniy Dodis and Tim Roughgarden,** 2006. "How much can taxes help selfish routing?", *Journal of Computer and System Sciences*, 72(3), 444-467.

[14] **Couture, Victor, Gilles Duranton, and Matt Turner,** 2012 "Speed", *mimeo*, University of Toronto.

[15] **Dafermos, Stella and Anna Nagurney**, 1984. "Sensitivity analysis for the asymmetric network equilibrium problem". *Mathematical Programming*, 28:174–184.

[16] **Dijkstra, Edsger,** 1959. "A Note on Two Problems in Connection with Graphs" *Numerische Matematik* 1: 269:271.

[17] **Dewees, Donald N.**, 1979. "Estimating the time costs of highway congestion." *Econometrica* 47: 1499-1512.

[18] **Duranton, Gilles and Matt Turner,** 2011 "The Fundamental Law of Road Congestion: Evidence from US Cities", *American Economic Review*, 101(6), 2616-52.

[19] **Fisk Caroline,** 1979. "More paradoxes in the equilibrium assignment problem", *Transportation Research* 13B, 305-309.

[20] **Fisk Caroline and S. Pallottino**, 1981. Empirical evidence for equilibrium paradoxes with implications for optimal planning strategies. *Transportation Research*, A15(3):245-248.

[21] **Frank, Marguerite**, 1981. "The Braess Paradox" *Mathematical Programming* 20: 283-302.

[22] **Fujita, Masahisa, Paul R Krugman, and Anthony Venables**, 1999, *The Spatial Economy: Cities, Regions and International Trade*, MIT Press.

[23] **Fujita, Masahisa and Jacques-Francois Thisse**, 2002, *Economics of Agglomeration: Cities, Industrial Location and Regional Growth*, Cambridge University Press.

[24] **Gordon, Peter, Ajay Kumar, and Harry W. Richardson**, 1989. "The Influence of Metropolitan Spatial Structure on Commuting Time" *Journal of Urban Economics*, 26, 138-151.

[25] **Guiso, Luigi, Paola Sapienza, and Luigi Zingales**, 2008. "Long Term Persistence". *National Bureau of Economic Research Working Papers 14278.*

[26] **Hall, Michael A.,** 1978. "Properties of the Equilibrium State in Transportation Networks" *Transportation Science*, 12(3), 208-216.

[27] **Hartgen, D.T.** 2003 (a). *The Impact of Highways and Other Major Road Improvements on Urban Growth in Ohio*, The Buckeye Institute, Columbus, OH.

[28] **Hartgen, D.T.** 2003 (b). *Highways and Sprawl in North Carolina*, The John Locke Foundation, Raleigh, NC.

[29] **Jackson**, **Matthew O.** 2005. "A Survey of Models of Network Formation: Stability and Efficiency," in *Group Formation in Economics: Networks, Clubs, and Coalitions*, Gabrielle Demange and Myrna Wooders, (eds.) Cambridge University Press: Cambridge.

[30] **Lin, Henry, Tim Roughgarden, and Eva. Tardos,** 2004. "A Stronger Bound on Braess's Paradox", *Proceedings of the Fifteenth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2004, New Orleans, Louisiana, USA*, 340-341.

[31] **Lin, Henry, Tim Roughgarden, Eva Tardos, and Asher Walkover,** 2009. "Stronger Bounds on Braess's Paradox and the Maximum Latency of Selfish Routing", Stanford University *mimeo.*

[32] **Lucca, David O. and Francesco Trebbi,** 2009. "Measuring Central Bank Communication: An Automated Approach with Application to FOMC Statements" *National Bureau of Economic Research Working Papers 15367.*

[33] **Quinet, Emile and Roger Vickerman**, 2004. *Principles of Transport Economics*, Edward Elgar Publishing.

[34] **Roughgarden Tim,** 2002. "Selfish Routing", Ph.D. Thesis, Cornell University.

[35] **Roughgarden Tim,** 2006. "On the Severity of Braess's Paradox: Designing Networks for Selfish Users is Hard", *J. Comput. Syst. Sci.*, 72(5), 922-953.

[36] **Roughgarden Tim and Eva Tardos,** 2002. "How Bad is Selfish Routing?", *Journal of ACM,* 49(2), 236-259

[37] **Savin, N.E.**, 1984. "Multiple Hypothesis Testing," *Handbook of Econometrics*, Zvi Griliches and M. D. Intriligator (ed.), edition 1, vol. 2, ch. 14, 827-879 Elsevier.

[38] **Small, Kenneth A. and Mogens Fosgerau,** 2010. "Marginal congestion cost on a dynamic network with queue spillbacks", mimeo UC Irvine.

[39] **Small, Kenneth A. and Erik T. Verhoef**, 2007. *The Economics of Urban Transportation*, Routledge, New York.

[40] **Small, Kenneth A. and Xuehao Chu,** 2003. "Hypercongestion", *Journal of Transport Economics and Policy* 37: 319-352.

[41] **Tsai, Yu H.** 2005 "Quantifying Urban Form: Compactness Versus Sprawl" *Urban Studies* 42, 141-161.

[42] **Texas Transportation Institute**, 2009a. *Urban Mobility Report*, College Station, Texas. Available: http://mobility.tamu.edu/ums/

[43] **Texas Transportation Institute**, 2009b. *Urban Mobility Report Methodology*, College Station, Texas. Available: http://mobility.tamu.edu/ums/report/methodology.stm

[44] **Valiant, Gregory, Tim Roughgarden,** 2010. "Braess's paradox in large random graphs" *Random Structures & Algorithms* 37(4), pp. 495–515.

[45] **Vickrey, William,** 1969. "Congestion Theory and Transport Investment". *American Economic Review,* 59, 251–261.

[46] **Walters, Alan,** 1961. "Theory and Measurement of Private and Social Cost of Highway Congestion". *Econometrica* 29(4), 676–699.

[47] **Wardrop, John Glen.** 1952. "Some Theoretical Aspects of Road Traffic Research" *Proc. Inst. Civ. Eng. Part II*, 1, 325-378.

[48] **Youn, H., M. T. Gastner, and H. Jeong**, 2008. "Price of Anarchy in Transportation Networks: Efficiency and Optimality Control" *Phys. Rev. Lett.* 101, 128701.

[49] **U. S. Department of Transportation**, 2009. **"**Assessing the Full Costs of Congestion on Surface Transportation Systems and Reducing Them through Pricing", available at: ost-pxweb.dot.gov/policy/reports/Costs%20of%20Surface%20Transportation%20Congestion.pdf

[50] **Verhoef, Erik T.,** 2001. "An integrated dynamic model of road traffic congestion based on simple car-following theory: Exploring hypercongestion," *Journal of Urban Economics* 49: 505-542.

[51] **Xie, Feng and David Levinson,** 2006. "Measuring the Structure of Road Networks". *Geographical Analysis* (forthcoming).

VII. Online Appendix A: Proofs

Proof of Proposition 1. Suppose $f^{*\prime}$ is the feasible minimum-latency flow in the triple $(C,(1+\gamma)\overrightarrow{r},l)$. Consider the flow $f''=\frac{1}{1+\gamma}f^{*\prime}$ constructed so all flows f''_a on every arrow $a\in A$ are scaled down by a factor $\frac{1}{1+\gamma}$ relative to $f^{*\prime}_a$. Function $l_a(.)$ is non-decreasing, so $L(f'')=\sum_a l_a(f''_a)f''_a\leq\sum_a l_a(f^{*\prime}_a)f^{*\prime}_a=L(f^{*\prime})$.

Notice that flow f'' is feasible in the triple $(C,\overrightarrow{r},l)$. Hence, by definition of minimum latency it follows $L(f^*)\leq L(f'')$. Hence, $L(f^*)\leq L(f^{*\prime})$.

This proves the proposition.

Proof of Proposition 3. Our objective is to prove $L(f')\geq L(f)$ for any $\gamma>0$. Let us restate $L(f')=\sum_{i=1}^k\ell'_ir'_i=\sum_{i=1}^k\ell'_i(1+\gamma)r_i$ and $L(f)=\sum_{i=1}^k\ell_ir_i$. Consider that:

$$\begin{aligned}L(f')-L(f)&=\sum_{i=1}^k\ell'_i(1+\gamma)r_i-\sum_{i=1}^k\ell_ir_i\\&\geq\sum_{i=1}^k\left(\ell'_i-\ell_i\right)r_i.\end{aligned}\tag{8}$$

Notice that for any $\gamma>0$ the expression $\sum_{i=1}^k\left(\ell'_i-\ell_i\right)r_i$ is nonnegative if $\sum_{i=1}^k\left(\ell'_i-\ell_i\right)\gamma r_i$ is nonnegative.

By following the approach of Dafermos and Nagurney (1984, Theorem 4.2) we are now going to show that:

$$\begin{aligned}&\sum_{i=1}^k\left(\ell'_i-\ell_i\right)\gamma r_i\\=\;&\sum_{i=1}^k\sum_{\pi\in\Pi_i^C}\left(\ell'_i-\ell_i\right)(f'_\pi-f_\pi)\geq 0.\end{aligned}$$

For each origin-destination pair $i=1,...,k$ let us define the following partition of the path set $\Pi_i^C=\Pi_i^{C1}\cup\Pi_i^{C12}\cup\Pi_i^{C2}$ based on which paths are carrying flow in which equilibrium, so that:

$$\begin{aligned}\Pi_i^{C1}&=\left\{\pi\in\Pi_i^C|f'_\pi=0,f_\pi>0\right\};\\\Pi_i^{C12}&=\left\{\pi\in\Pi_i^C|f'_\pi>0,f_\pi>0\right\};\\\Pi_i^{C1}&=\left\{\pi\in\Pi_i^C|f'_\pi>0,f_\pi=0\right\}.\end{aligned}$$

Consider now that this allows us to write:

$$\begin{aligned}
&\sum_{i=1}^{k} \sum_{\pi \in \Pi_i^C} \left(l_\pi(f') - l_\pi(f)\right)\left(f'_\pi - f_\pi\right) \\
= &\sum_{i=1}^{k} \Bigg[\sum_{\pi \in \Pi_i^{C12}} \left(\ell'_i - \ell_i\right)\left(f'_\pi - f_\pi\right) \\
&+ \sum_{\pi \in \Pi_i^{C1}} \left(l_\pi(f') - \ell_i\right)\left(f'_\pi - f_\pi\right) \\
&+ \sum_{\pi \in \Pi_i^{C2}} \left(\ell'_i - l_\pi(f)\right)\left(f'_\pi - f_\pi\right) \Bigg]. \qquad (9)
\end{aligned}$$

At Nash equilibrium it must hold that:

$$\sum_{\pi \in \Pi_i^{C1}} \left(l_\pi(f') - \ell_i\right)\left(f'_\pi - f_\pi\right) \leq \sum_{\pi \in \Pi_i^{C1}} \left(\ell'_i - \ell_i\right)\left(f'_\pi - f_\pi\right), \qquad (10)$$

given that for every $\pi \in \Pi_i^{C1}$ $f'_\pi = 0$ and $l_\pi(f') \geq \ell'_i$.

Symmetrically, it must also hold:

$$\sum_{\pi \in \Pi_i^{C2}} \left(\ell'_i - l_\pi(f)\right)\left(f'_\pi - f_\pi\right) \leq \sum_{\pi \in \Pi_i^{C2}} \left(\ell'_i - \ell_i\right)\left(f'_\pi - f_\pi\right), \qquad (11)$$

given that for every $\pi \in \Pi_i^{C2}$ $f_\pi = 0$ and $l_\pi(f) \geq \ell_i$.

From (9), (10), and (11) it follows that:

$$\sum_{i=1}^{k} \sum_{\pi \in \Pi_i^{C}} \left(l_\pi(f') - l_\pi(f)\right)\left(f'_\pi - f_\pi\right) \leq \sum_{i=1}^{k} \sum_{\pi \in \Pi_i^{C}} \left(\ell'_i - \ell_i\right)\left(f'_\pi - f_\pi\right).$$

Note that $l_a(x)$ is a non-decreasing function by assumption, therefore $l_\pi(f) = \sum_{a \in \pi} l_a\left(f_a\right)$ is non-decreasing. In addition, $l_a(x)$ is non-negative, so $l_\pi(f)$ is non-negative. Hence, if $f'_\pi - f_\pi < 0$ it must be $l_\pi(f') - l_\pi(f) < 0$ and if $f'_\pi - f_\pi > 0$ it must be $l_\pi(f') - l_\pi(f) > 0$. This implies:

$$0 \leq \sum_{i=1}^{k} \sum_{\pi \in \Pi_i^{C}} \left(l_\pi(f') - l_\pi(f)\right)\left(f'_\pi - f_\pi\right) \leq \sum_{i=1}^{k} \sum_{\pi \in \Pi_i^{C}} \left(\ell'_i - \ell_i\right)\left(f'_\pi - f_\pi\right) = \sum_{i=1}^{k} \left(\ell'_i - \ell_i\right)\gamma r_i$$

By (8) and the fact that $\gamma > 0$, this proves the proposition.

Proof of Proposition 4. Define $\hat{h}$ the feasible Nash flow for $(M', \overrightarrow{r}, l)$. By optimality of M with respect to the triple $(C, \overrightarrow{r}, l)$, it follows that $L(h) \leq L(\hat{h})$.

By the monotonicity result in Proposition 3, it must follow that for the two triples $(M', \overrightarrow{r}, l)$ and $(M', \overrightarrow{r}', l)$ the inequality $L(\hat{h}) \leq L(h')$ holds.

It follows that:

$$L(h) \leq L(\hat{h}) \leq L(h').$$

This proves the proposition.

Proof of Corollary 1. By definition $\ell_i^M = \sigma_{i,X}^M \delta_{i,X}^M$. Define $\tilde{\sigma} = \ell_i^M / \delta_{i,1}^C$. Hence, $\tilde{\sigma}/\sigma_{i,X}^M = \delta_{i,X}^M / \delta_{i,1}^C$.

By optimality of M it follows $\tilde{\sigma} \leq \ell_i^C / \delta_{i,1}^C \leq \sigma_{i,1}^C$, where the last inequality holds strictly if $\pi_{i,1}^C$ does not carry flow in equilibrium.

Then, $\delta_{i,X}^M / \delta_{i,1}^C \leq \sigma_{i,1}^C / \sigma_{i,X}^M$ or, in terms of speed gains, $\delta_{i,X}^M / \delta_{i,1}^C \leq v_{i,X}^M / v_{i,1}^C$.

This proves the corollary.

VIII. Online Appendix B: Endogenous Travel Demand.

In this appendix we provide a generalization of our model to endogenous travel demand. Let us consider city C where traffic operates in two different time periods: peak and off-peak. Let us maintain our assumption on latency functions l and define the traffic rate on a source-target $s_i - t_i$ at peak time as $r_i = p_i n_i$ and the traffic rate on a source-target $s_i - t_i$ at off-peak time as $\tilde{r}_i = (1 - p_i) n_i$. We allow p_i, the per period share of n_i that undertakes trip $s_i - t_i$ at peak time, to be endogenously determined. That is, everybody living at s_i has to reach t_i, but can decide when. Following Arnott, de Palma, and Lindsey (1993) we assume individuals obtain a disutility $u_i \geq 0$ from taking the trip in off-peak time, with u_i constant within the source population (but not necessarily across sources).

Since our focus is on the elasticity of travel demand, it is sensible to focus on a case where individual decision making matters, that is the Nash flow of the network (as opposed to the optimal directed flow, for instance) for the triple $(C, \overrightarrow{n}, l)$, where we indicate $\overrightarrow{n}$ the vector $[n_1, n_2, ..., n_k]$. Our previous notation with inelastic travel demand was $(C, \overrightarrow{r}, l)$ and the different notation is driven by the fact that now input traffic rates are endogenous and not given.

For commodity $s_i - t_i$, define the Nash delay $\ell_i(r_i)$, for given peak input traffic rate r_i. A resident of s_i travels at peak time if $-\ell_i(r_i) \geq -u_i - \tilde{\ell}_i(n_i - r_i)$, and travels off-peak otherwise. Indifference between on-peak and off-peak latency $\tilde{\ell}_i$ implies:

$$-\ell_i(p_i n_i) = -\tilde{\ell}_i((1 - p_i) n_i) - u_i. \tag{12}$$

Hall (1978) shows any increment of input traffic rate r_i produces an increase in ℓ_i, so path latency is monotonic increasing in input traffic rates, and the function is continuous by assumption. The same has to hold for a Nash equilibrium off-peak latency $\tilde{\ell}_i$. Monotonic and increasing path latencies imply there is a unique intersection p_i^* that solves (12). Furthermore, it must also be true that $p_i^* \geq 1/2$, that is, there is more traffic at peak than off peak, as people will trade off increased congestion with higher utility of traveling at peak times.

(12) pins down the selection between peak and off-peak travel uniquely. Hence, input traffic rates can be taken as exogenous when solving for a Nash equilibrium within the network at peak and off peak. This implies that there exists a unique Nash equilibrium for $(C, \overrightarrow{n}, l)$, since, given input peak traffic rate vector $\overrightarrow{r}$, there exists a unique Nash equilibrium for the peak triple $(C, \overrightarrow{r}, l)$ (applying Proposition 2) and, given input off-peak traffic rate vector $\overrightarrow{\tilde{r}}$, there exists a unique Nash equilibrium for the off-peak triple $\left(C, \overrightarrow{\tilde{r}}, l\right)$ (again, applying Proposition 2).

Given the uniqueness of the Nash equilibrium, we can further explore some interesting comparative statics. In the case of an increment of the population at the source, n_i, peak

input traffic rate $r_i = p_i^*(n_i) * n_i$ will respond directly proportionally because:

$$\begin{aligned}
\frac{\partial r_i}{\partial n_i} &= p_i + \frac{\partial p_i(n_i)}{\partial n_i} n_i = \\
&= p_i - \frac{\frac{\partial\left(\ell_i(p_i n_i) - \tilde{\ell}_i((1-p_i)n_i)\right)}{\partial n_i}}{\frac{\partial\left(\ell_i(p_i n_i) - \tilde{\ell}_i((1-p_i)n_i)\right)}{\partial p_i}} n_i \\
&= p_i - \frac{p_i \ell_i'(p_i n_i) - (1-p_i)\,\tilde{\ell}_i'((1-p_i)n_i)}{\ell_i'(p_i n_i) + \tilde{\ell}_i'((1-p_i)n_i)} \\
&\geq p_i - (1-p_i)\frac{\ell_i'(p_i n_i) - \tilde{\ell}_i'((1-p_i)n_i)}{\ell_i'(p_i n_i) + \tilde{\ell}_i'((1-p_i)n_i)} \\
&\geq p_i - (1-p_i) \\
&\geq 0
\end{aligned}$$

where we make use of the Implicit Function Theorem and the last inequality follows from $p_i^* \geq 1/2$. So, when the population at the source increases, input traffic at peak will increase. This result further implies that off-peak input traffic $\tilde{r}_i$ will increase a fortiori, given that, not only population at the source increases, but also people will reallocate from peak to off-peak travel. This further implies that the intuition of total latency at Nash being increasing in city population (Proposition 3) will be satisfied in this setting, both in peak and off-peak times. One can verify that all assumptions of Proposition 3 hold within the network both at peak and off peak for given input traffic rates. Define total cost in $(C, \overrightarrow{n}, l)$ as $\Lambda(C, \overrightarrow{n}, l) = \sum_i \ell_i r_i + \sum_i \left(\tilde{\ell}_i + u_i\right)(n_i - r_i)$. The argument above proves:

propositionf population $N' > N$ such that $n_i' = (1+\gamma)\,n_i$ for any commodity i and $\gamma > 0$, then $\Lambda(C, \overrightarrow{n}, l) \leq \Lambda(C, \overrightarrow{n'}, l)$.

A final issue in this extension is the definition of the optimal network design problem. City planners aim at finding restrictions on accessible vertices that allow reductions in latency costs, but now will need to weigh off-peak versus peak delays. More formally, the problem is to find the subnetwork M of C, over which agents selfishly route themselves, displaying the lowest latency at Nash equilibrium both off-peak and at peak. Notice that in this case the city manager by picking a subnetwork M of C will be also implicitly picking an input traffic rate at peak and off peak.

Condition 1. *City planners impose a single traffic plan at peak and off-peak, or for given*

quadruple $(C, \overrightarrow{n}, l)$*, the problem can be represented as the following network design program:*

$$\begin{aligned} &\underset{\tilde{M} \subseteq C}{Min} \left\{ \sum_i \ell_i r_i + \sum_i \left(\tilde{\ell}_i + u_i \right) (n_i - r_i) \right\} \\ &s.t.\ \Pi_i^{\tilde{M}} \neq \varnothing, \qquad \forall i \\ &h \text{ is a feasible Nash flow for } \left(\tilde{M}, \overrightarrow{r}, l \right) \text{ so that for every } \pi \in \Pi_i^{\tilde{M}}\ l_\pi(h) = \ell_i \\ &\tilde{h} \text{ is a feasible Nash flow for } \left(\tilde{M}, \overrightarrow{\tilde{r}}, l \right) \text{ so that for every } \pi \in \Pi_i^{\tilde{M}}\ l_\pi(\tilde{h}) = \tilde{\ell}_i \end{aligned}$$

Using (12) the problem becomes:

$$\begin{aligned} &\underset{\tilde{M} \subseteq C}{Min} \left\{ \sum_i \left(\tilde{\ell}_i + u_i \right) n_i \right\} \\ &s.t.\ \Pi_i^{\tilde{M}} \neq \varnothing, \qquad \forall i \\ &h \text{ is a feasible Nash flow for } \left(\tilde{M}, \overrightarrow{r}, l \right) \text{ so that for every } \pi \in \Pi_i^{\tilde{M}}\ l_\pi(h) = \ell_i \\ &\tilde{h} \text{ is a feasible Nash flow for } \left(\tilde{M}, \overrightarrow{\tilde{r}}, l \right) \text{ so that for every } \pi \in \Pi_i^{\tilde{M}}\ l_\pi(\tilde{h}) = \tilde{\ell}_i \end{aligned} \tag{13}$$

The problem (13) for the city planner admits a solution. The city planner can pick any subnetwork $\tilde{M}$, find the unique p_i^* for any commodity i, compute the total latency of the Nash equilibrium of the off-peak network $\left(\tilde{M}, \overrightarrow{\tilde{r}}, l \right)$ where input traffic at each source s_i is $\tilde{r}_i = (1 - p_i^*)\, n_i$, then pick the minimum latency $\tilde{M}$. Let us indicate with M a solution to (13).

propositiononsider population $N' > N$ such that $n_i' = (1 + \gamma)\, n_i$ for any commodity i and $\gamma > 0$ and M and M' optimal subgraphs for $(C, \overrightarrow{n}, l)$ and $(C, \overrightarrow{n}', l)$ respectively. Then $\Lambda(M, \overrightarrow{n}, l) \leq \Lambda(M', \overrightarrow{n'}, l)$.

Proof. By optimality of M to problem (13) it follows that $\Lambda(M, \overrightarrow{n}, l) \leq \Lambda(M', \overrightarrow{n}, l)$. By Proposition 5, $\Lambda(M', \overrightarrow{n}, l) \leq \Lambda(M', \overrightarrow{n'}, l)$. Hence, $\Lambda(M, \overrightarrow{n}, l) \leq \Lambda(M', \overrightarrow{n'}, l)$.

This proves the proposition.

Proposition 6 mirrors closely Proposition 4 and it is helpful in underlying the robustness of our main comparative statics for the analysis in the main text to the specifics of source-target travel demand.

IX. Online Appendix C: Geographic Controls.

i. Latitude: Latitude measured in degrees. Source: United States Geological Survey (USGS) Earth Explorer

URL: http://edcsns17.cr.usgs.gov/NewEarthExplorer/

ii. Longitude: Longitude measured in degrees. Source: USGS Earth Explorer

The following variables were measured using Geographing Information Software tools (ARCGIS), combined with Geospatial Modeling environment tools (Hawth's tools).

iii. Lakes sqkm: Total lake area in sq. km. within 50km radius. Scale 1:2,000,000. Source: USG.S/ National Atlas of the US URL: http://www.nationalatlas.gov/mld/hydrogm.html

vi. Swamps sqkm: Total swamps or marshes area in sq. km. within 50km radius. Scale 1:2,000,000. Source: USG.S./ National Atlas of the US

v. Ocean sqkm: Total ocean area in sq. km. within 50km radius. Scale 1:2,000,000. Source: USG.S./ National Atlas of the US

vi. Coastal City: Dummy variable, takes value 1 if city is within 100km of the ocean or the great lakes (Definition taken from Saiz, Albert 2010 "The Geographic Determinants of Housing Supply," The Quarterly Journal of Economics, MIT Press, vol. 125(3), pages 1253-1296, August). Scale 1:2,000,000. Source: USG.S./ National Atlas of the US

vii. Coast km: Total km of ocean coastline within 50km radius, low resolution. Source: USG.S./ National Atlas of the US

viii. Riv Str km: Total km of rivers and streams within 50km radius. Only available for the contiguous USA. High Resolution. Source: NOHRSC

ix. Mean Elevation: Mean elevation in meters within 50km radius at 30m/ 1arc second resolution. Source: NASA's Shuttle Radar Topography Mission for Continental USA and Hawaii, Version 2, and Alaska's StateWide Mapping Initiative for Alaska.

URL: http://www2.jpl.nasa.gov/srtm/

http://www.alaskamapped.org/

x. SD Elevation: Standard deviation of elevation in meters within 50km radius at 30m/1 arc second resolution. Source: NASA's Shuttle Radar Topography Mission for Continental USA and Hawaii, Version 2, and Alaska's StateWide Mapping Initiative for Alaska.

xi. Mean Slope: Mean slope in percent rise within 50km radius at 30m/1arc second resolution. Source: NASA's Shuttle Radar Topography Mission for Continental USA and Hawaii, Version 2, and Alaska's StateWide Mapping Initiative for Alaska.

xii. SD Slope: Standard deviation slope in percent rise within 50km radius at 30m/1arc second resolution. Source: NASA's Shuttle Radar Topography Mission for Continental USA and Hawaii, Version 2, and Alaska's StateWide Mapping Initiative for Alaska.

xiii. Slope less than 15: Percent area within 50km radius with a slope of less than 15%, or area suitable for development. (Definition taken from Saiz 2010). Source: NASA's Shuttle Radar Topography Mission for Continental USA and Hawaii, Version 2, and Alaska's StateWide Mapping Initiative for Alaska.

xiv. Density: persons per sq. mile. US Census Bureau.

Figure 1

This example is originally due to Fisk (1979) and slightly modified here. The graph below depicts the entire directed network. Consider the following three source-target commodities: (a, b) indicated with i= 1, (a, c) with i= 2, and (b, c) with i= k= 3. Latency functions are described in the graph. Assume traffic inputs are: r_1= 1, r_2= 20, r_3= 100. For commodity i= 1, 3 there is only one available path each, while for commodity 2 $\Pi_2 = \{\pi_{ac}; \pi_{abc}\}$ where π_{ac} indicates path a→c and π_{abc} indicates a→b→c. The Nash equilibrium solution requires finding how commodity 2 traffic self-routes through this network. This implies finding flows f_{ac} and f_{abc} such that conditions $r_2=f_{ac}+f_{abc}$; $\ell_1= r_1+f_{abc}$; $\ell_2= f_{ac} + 90 = f_{abc}+ r_1 + f_{abc} + r_3$; $\ell_3= r_3+f_{abc}$ hold. This problem is solved for f_{ac}= 17 and f_{abc}= 3 and equilibrium latencies ℓ_1= 4, ℓ_2= 107, ℓ_3= 103. Notice that total latency $L(f)= \Sigma_i \ell_i r_i$= 12,444.

Assume now traffic inputs increase to: r_1= 4, r_2= 20.1, r_3= 100.1. The Nash equilibrium flows are now f_{ac}= 18 and f_{abc}= 2 and ℓ_1= 6, ℓ_2= 108.1, ℓ_3= 102.1. Notice that now total latency drops to $L(f)= \Sigma_i \ell_i r_i$= 12,417. The intuition of this counterintuitive result is that by increasing congestion on edge a→b Nash equilibrium redirects individuals travelling form a to c away from path π_{abc} and onto the alternative path π_{ac}. This ends up ameliorating congestion on the b→c edge where the benefit is particularly large, as it ameliorates conditions for the large group r_3~ 100 (b, c) commuters.

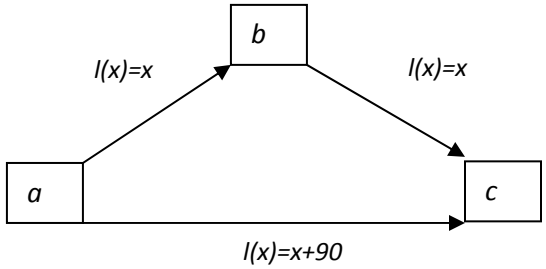
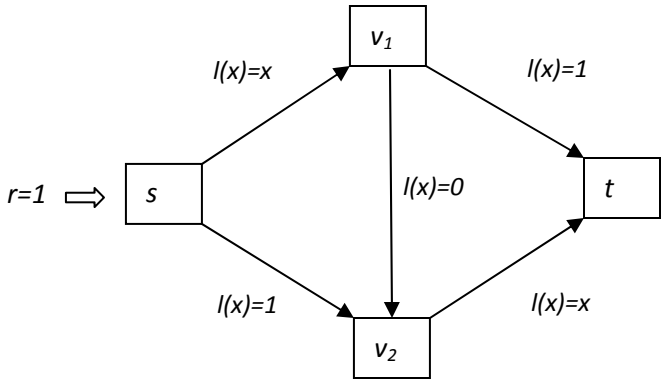
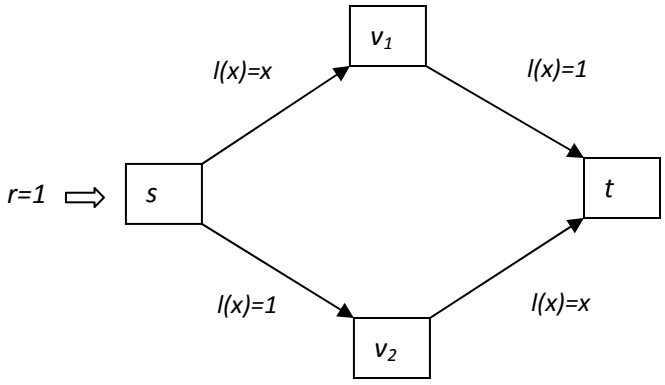


Figure 2

Graph (a) depicts the entire directed network. Traffic rate is 1. The minimum-latency flow directs 1/2 the traffic on path s→v_2→t and 1/2 on path s→v_1→t inducing a total latency of 3/2. The unique Nash flow routes through s→v_1→v_2→t inducing a total latency of 2. In (b) the optimal subgraph is depicted. By removing the v_1→v_2 arrow, total latency in the unique Nash flow is now 3/2.

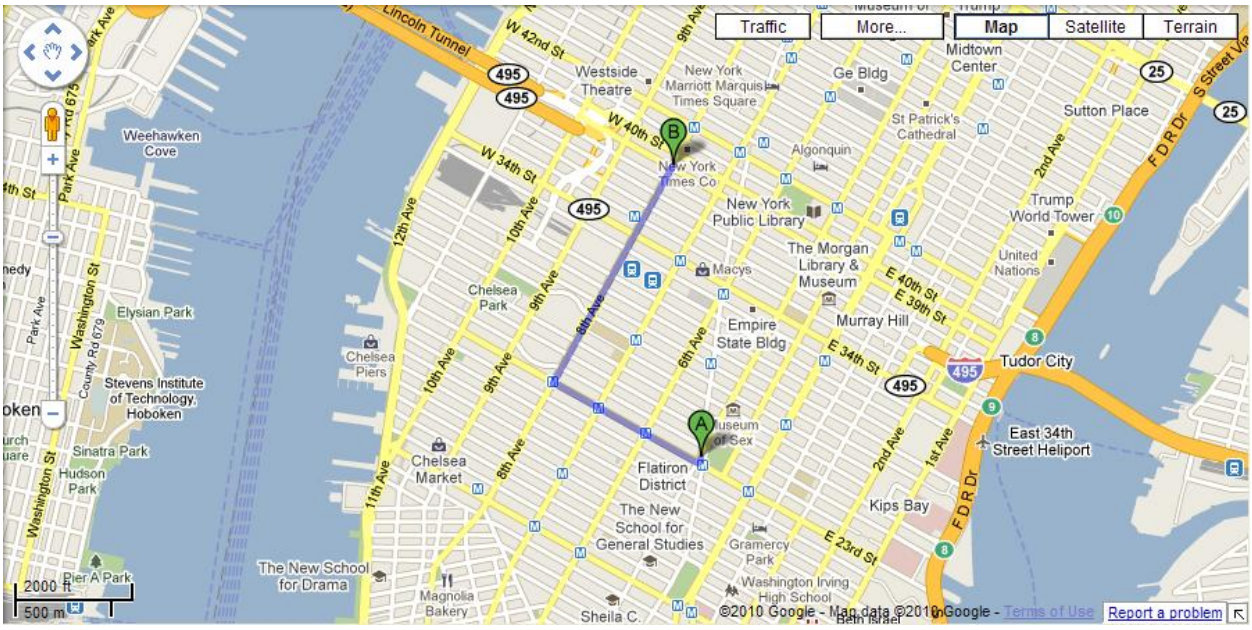


(a)



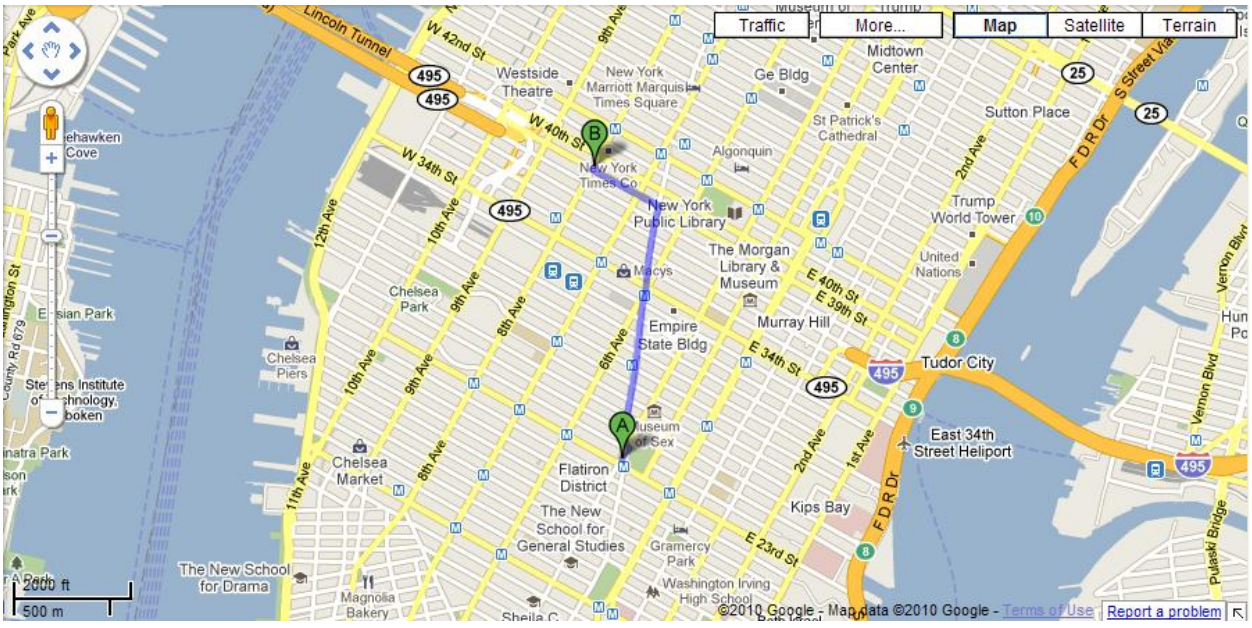
(b)

Figure 3



Example of driving directions: A→B, New York City, NY. Distance 2.2 km.

Source: http://maps.google.com/



Example of walking directions: A→B, New York City, NY. Distance 1.8 km.

Source: http://maps.google.com/

Table 1

Texas Transportation Institute 2009 Urban Mobility Report ***Congestion Costs Ranking 2007*** ***Top Ten Most Congested***	***INRIX Scorecard*** ***Congestion Rankings 2009*** ***Top Ten Most Congested***
Los Angeles-Long Beach-Santa Ana CA	Los Angeles-Long Beach-Santa Ana CA
New York-Newark NY-NJ-CT	New York-Northern New Jersey-Long Island NY-NJ-PA
Chicago IL-IN	Chicago-Joliet-Naperville IL-IN-WI
Atlanta GA	Washington-Arlington-Alexandria DC-VA-MD-WV
Miami FL	Dallas-Fort Worth-Arlington TX
Dallas-Fort Worth-Arlington TX	Houston-Sugar Land-Baytown TX
Washington DC-VA-MD	San Francisco-Oakland-Fremont CA
San Francisco-Oakland CA	Boston-Cambridge-Quincy MA-NH
Houston TX	Seattle-Tacoma-Bellevue WA
Detroit MI	Philadelphia-Camden-Wilmington PA-NJ-DE-MD

Table 2

Country	***Avg. # of Addresses Sampled (per city)***	***Avg. # of Pairs Attempted***	***Avg. # of Usable Results***
USA	5791	601	513
Italy	860	106	70

Table 3

Variable	*Obs.*	*Mean*	*Std. Dev.*
TTI U.S. Sample			
Population (1000)	98	569.38	968.58
Density (persons/sq mile)	98	2240.40	901.52
Total Delay (1000 hours/year)	98	47205.40	79213.57
Delay per Peak Traveler (hours/year)	98	31.21	15.13
Average Traffic Speed	98	39.79	4.07
Total Congestion Costs ($millions/year)	98	994.15	1690.53
Congestion Cost per Peak Traveler ($/year)	98	652.03	321.43
Average Traveled Distance (meters)	98	14557.06	7280.21
Average Latency at Reference Speed (seconds)	98	989.96	313.16
Average Travel Distortion (Drive-Walk)	98	1818.65	975.75
Average Path Tortuosity (Walk-Linear)	98	2442.15	1358.94
Average Travel Distortion (Drive/Walk)	98	1.14	0.04
Average Path Tortuosity (Walk/Linear)	98	1.27	0.07
Full U.S. Sample			
Population (1000)	457	213.52	486.90
Average Traveled Distance (meters)	457	9385.81	5568.38
Average Latency at Reference Speed (seconds)	457	749.35	275.60
Average Travel Distortion (Drive-Walk)	457	944.38	782.68
Average Path Tortuosity (Walk-Linear)	457	1780.73	982.04
Average Travel Distortion (Drive/Walk)	457	1.11	0.04
Average Path Tortuosity (Walk/Linear)	457	1.32	0.07

Table 4

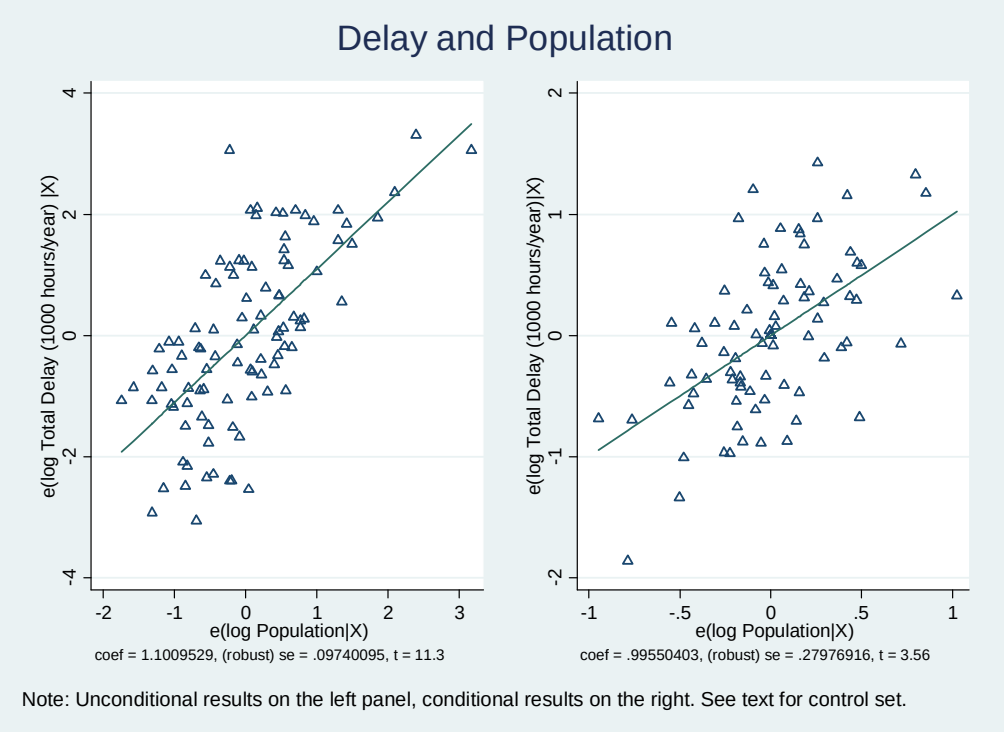
Multiple Hypothesis Model: Log Total Delay = $\alpha_1+\beta_1$*logPop+ϵ_1 Log Per Traveler Delay = $\alpha_2+\beta_2$*logPop+ϵ_2 Log Reference Duration = $\alpha_3+\beta_3$*logPop+ϵ_3		***Unconditional*** ***χ^2 Test P-values***	***Conditional on Controls*** ***χ^2 Test P-values***
Tests:	***Reference:***	(1)	(2)
H_0: $\beta_1 \leq 0$ H_1: $\beta_1 > 0$	*Remark 2, Claim 2*	**<0.001**	**<0.001**
H_0: $\beta_2 \leq 0$ H_1: $\beta_2 > 0$	*Remark 2, Claim 3*	**<0.001**	**<0.001**
H_0: $\beta_3 \leq 0$ H_1: $\beta_3 > 0$	*Remark 2, Claim 4*	**<0.008**	**<0.001**
H_0: $\beta_1 \leq 0$ Or $\beta_3 \leq 0$ H_1: $\beta_1 > 0$ And $\beta_3 > 0$	*Remark 1 [Jointly Remark 2, Claims 2&4]*	**<0.05**	**<0.01**
H_0: $\beta_2 \leq 0$ Or $\beta_3 \leq 0$ H_1: $\beta_2 > 0$ And $\beta_3 > 0$	*Remark 2, Claim 1 [Jointly Remark 2, Claims 3&4]*	**<0.05**	**<0.01**

Notes: Controls in column (2) include state fixed effects, a fourth order polynomial in tortuosity, a fourth order polynomial in population density, coastal city dummy, latitude, longitude, km² of lakes/swamps/ocean, km of coast, km of rivers/streams, mean elevation, standard deviation of elevation, mean terrain slope, standard deviation of slope, percent area suitable for development.

Variance covariance matrix clustered at state level.

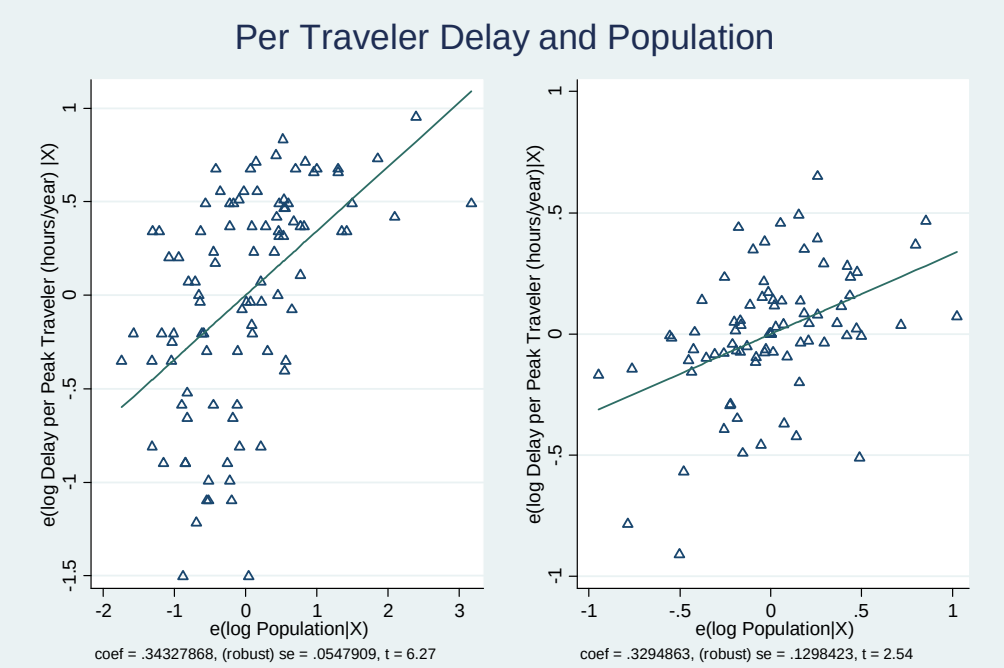
Chi-squared p-values are based on a Bonferroni $\alpha/3$ confidence level required for all tests.

Figure 4



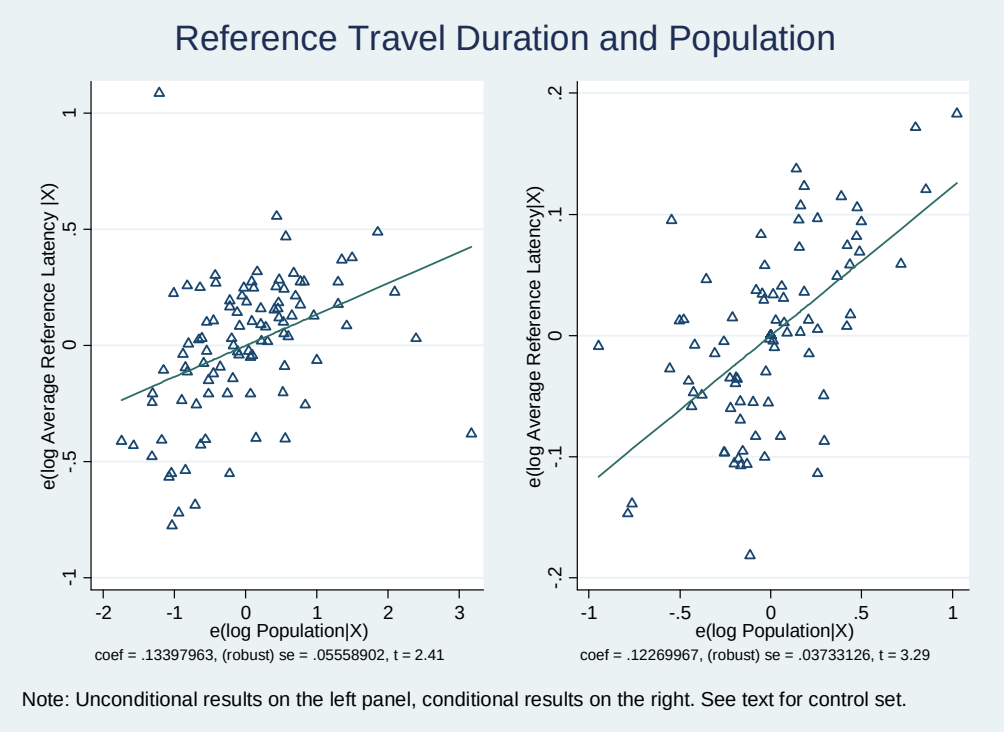
Note: Unconditional results on the left panel, conditional results on the right. See text for control set.

Figure 5



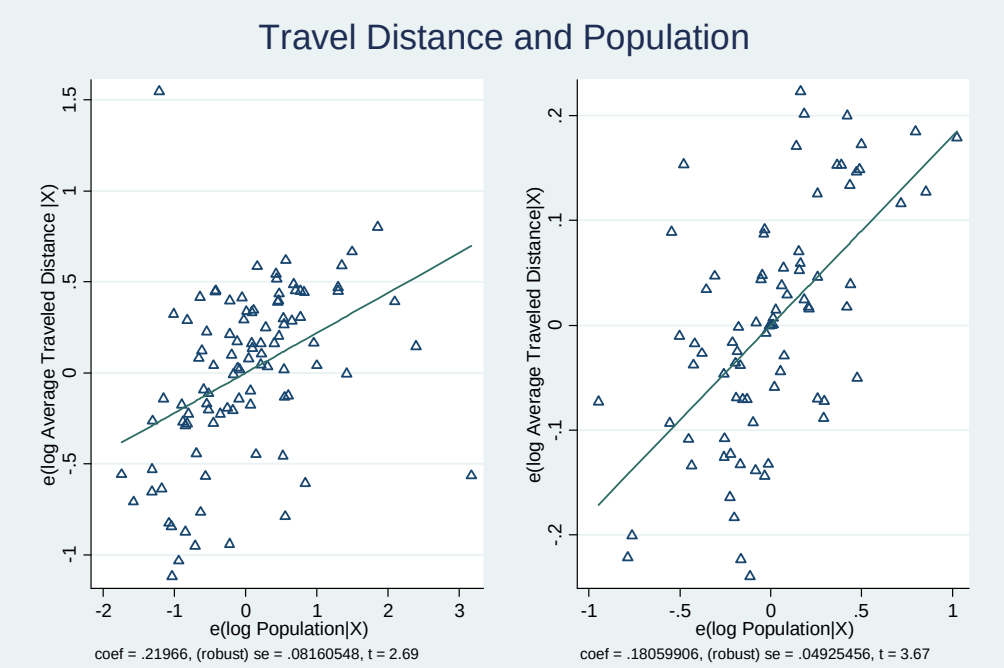
Note: Unconditional results on the left panel, conditional results on the right. See text for control set.

Figure 6



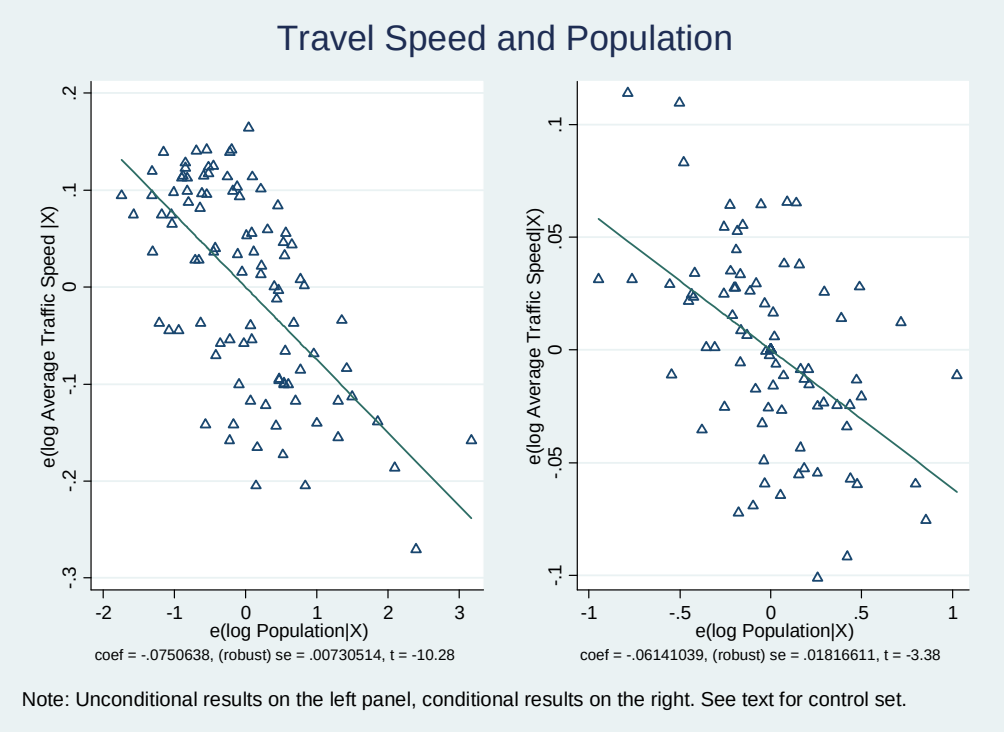
Note: Unconditional results on the left panel, conditional results on the right. See text for control set.

Figure 7



Note: Unconditional results on the left panel, conditional results on the right. See text for control set.

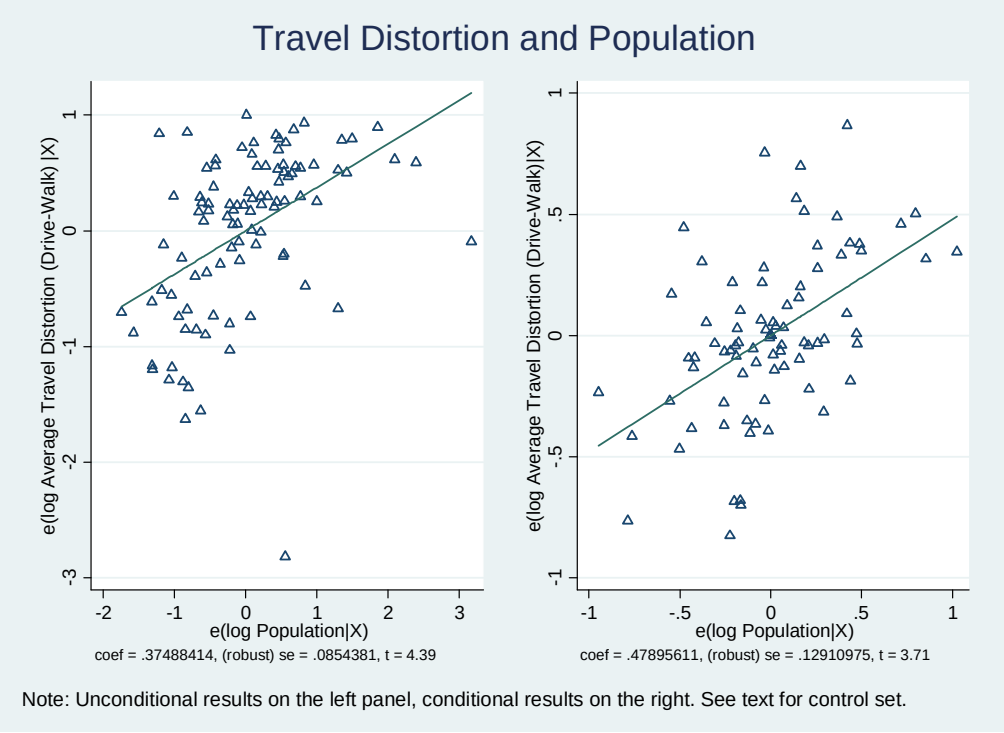
Figure 8



Note: Unconditional results on the left panel, conditional results on the right. See text for control set.

Figure 9

Figure 10



Note: Unconditional results on the left panel, conditional results on the right. See text for control set.

Figure 11

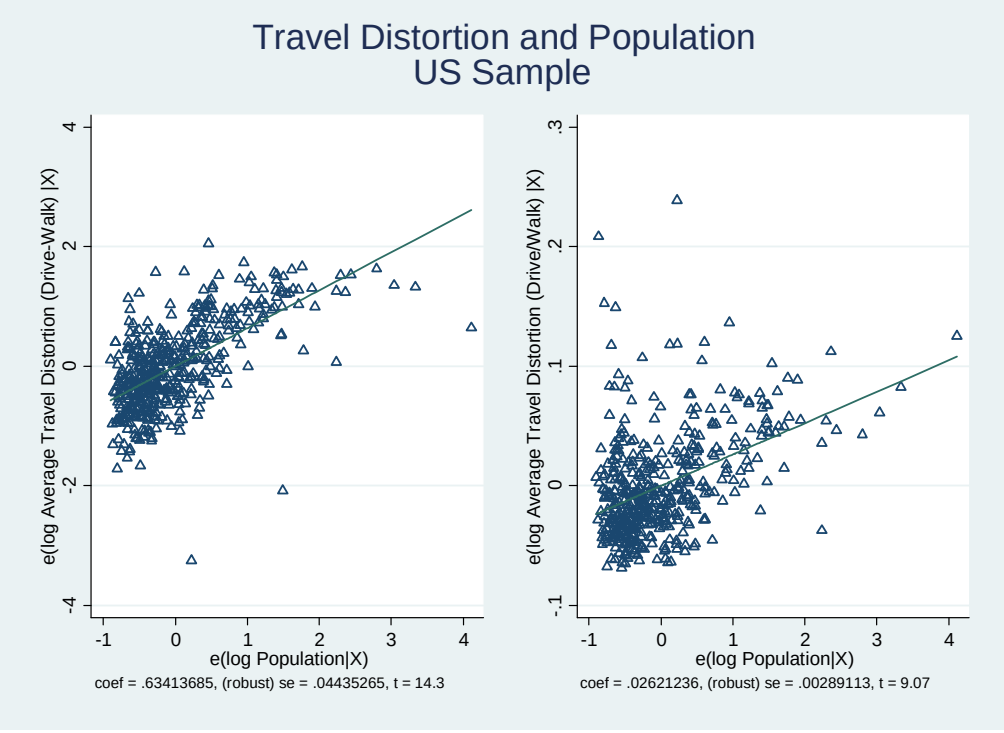
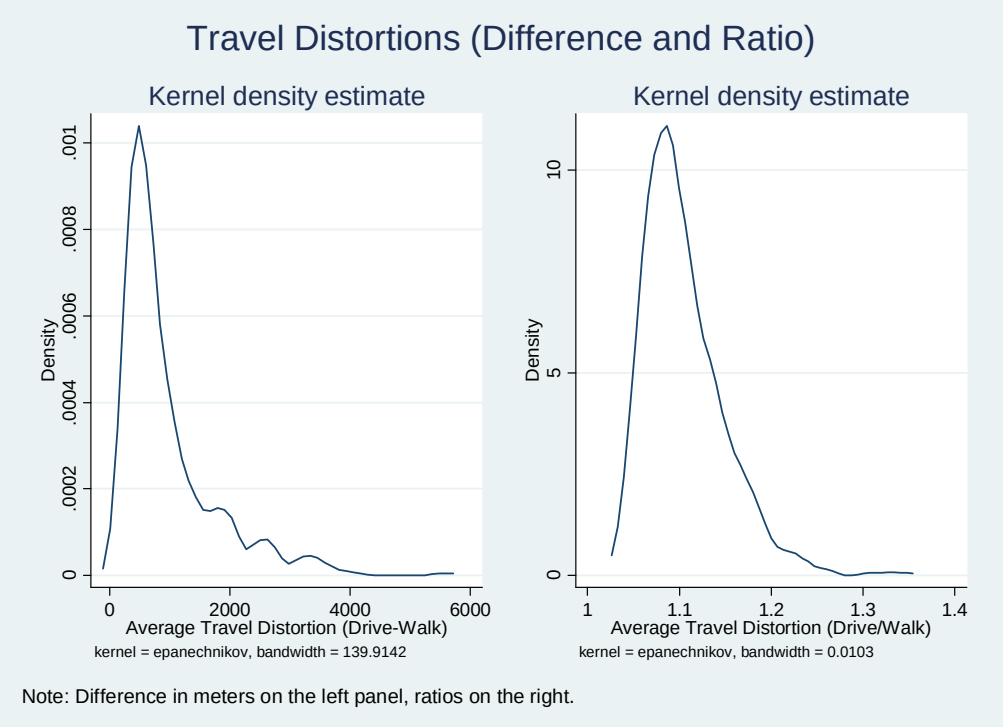


Figure 12



Note: Difference in meters on the left panel, ratios on the right.

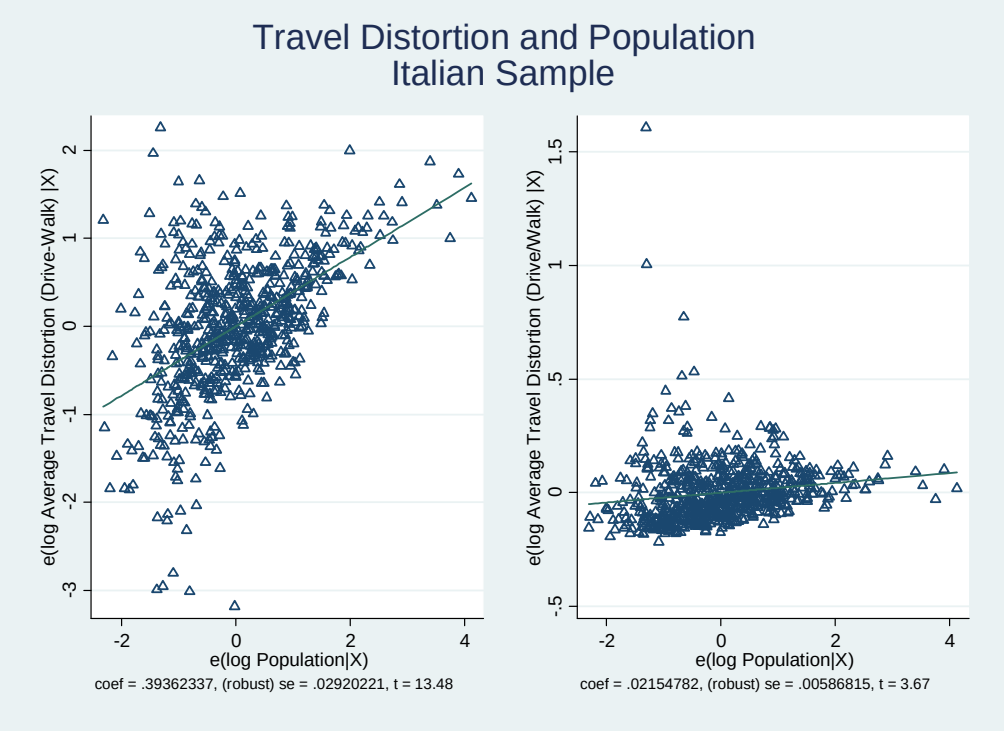



Figure 13