Does building highways reduce traffic congestion?*

Huibin Chang[†] Debarshi Indra[‡] Abhradeep Maiti[§]

August 10, 2020

Abstract

In a seminal study, Duranton and Turner (2011) finds evidence that points to the existence of the fundamental law of highway congestion in the US. They build a causal model using an instrumental variable (IV) approach that yields an estimate of 1.03 for the elasticity of vehicle miles traveled (VMT) to the stock of interstate highways in US metropolitan areas. The result means that government efforts to alleviate traffic congestion by expanding capacity are likely to fail — any increase in the stock of highways is accompanied by a commensurate increase in VMT, leaving travel times unaffected. In this article, we explore the impact of unobserved heterogeneity on the fundamental law. We begin by using a simple partial equilibrium model to demonstrate how metropolitan statistical areas (MSAs) that are identical in most respects but have different initial congestion levels respond differently to added capacity due to individual differences. These differences in MSAs gives rise to heterogeneity in the elasticity of VMT to capacity. We derive conditions under which the elasticity decreases with the initial congestion level. We then revisit the empirical analysis in Duranton and Turner (2011) using the instrumental variable quantile regression (IV-QR) model due to Chernozhukov and Hansen (2005, 2006, 2008). The IV-QR model allows us to incorporate variation in the elasticity due to the presence of unobserved differences across MSAs. Moreover, it allows us to evaluate the impact of changes in the stock of interstate highways on the entire conditional distribution of VMT, not just the impact on the conditional mean as in Duranton and Turner

*The authors thank Alex Anas and participants at the RSAI World Congress 2018, SMU Conference on Urban and Regional Economics 2019, seminar at Dongbei University of Finance and Economics for helpful comments. The opinions expressed in this article are the authors' own and do not reflect the view of Dongbei University of Finance and Economics, ABInBev, or IIM Kashipur.

[†]Corresponding author: Institute for Advanced Economic Research, Dongbei University of Finance and Economics, China. E-mail: huibinchang@icloud.com.

[‡]Growth Analytics Center, Anheuser-Busch InBev, Bangalore, India.

[§]Indian Institute of Management Kashipur, Kashipur, India.

(2011). The IV-QR estimates show that as predicted by the simple partial equilibrium model, the elasticity declines as one goes up the quantile ladder, being more than one at the lower quantiles and less than one at the higher quantiles. The median IV-QR estimate being close to one. The IV-QR model implies that among observationally identical cities, expanding road capacity can lower the number of cities experiencing severe congestion, although the mean or median congestion levels are likely to remain constant. We also estimate the impact of increased road capacity on the unconditional distribution of VMT using the generalized quantile regression (GQR) model due to Powell (Forthcoming). The GQR estimates mirror the IV-QR estimates, but their conclusions are starker at the upper quantiles: building highways have no statistically significant impact on VMT at the highest quantiles. The GQR results imply that building roads can lower the total number of cities having different observed characteristics experiencing severe congestion levels. We further explore the mechanisms that drive the empirical findings by running simulations using a spatial general equilibrium model with an extensive road network calibrated to the Greater Los Angeles (LA) Region. In the general equilibrium model, besides commute trips, consumers have to travel to different zones to acquire consumption goods. The model considers route choice in the network and mode choice. Building roads affect road traffic through three channels: the total amount of consumption, mode choice, and the substitutions among goods sold at different locations. We find that the elasticity of VMT to capacity in LA is 0.321, and the elasticity decreases consistently with the initial congestion level. We report the welfare effect and the changes in other travel-related variables. Our results have important policy implications in that they show that while building roads is unlikely to change the mean or median congestion level, it can reduce the number of cities experiencing high congestion levels.

1 Introduction

In an influential study, Duranton and Turner (2011) find evidence that confirms the existence of the fundamental law of highway congestion in the U.S. The law implies that any increase in road capacity will generate a proportionate increase in road traffic, leaving congestion levels unchanged (Downs, 1962). The law has important policy implications: it means that governments cannot build their way out of the congestion problem, leaving congestion pricing the only viable solution (Duranton and Turner, 2011).

The primary force at work behind the fundamental law is the phenomenon of induced travel demand. As travel times fall in a network due to additional capacity, travel demand increases due to two reasons. First, generative demand: new trips from latent travel demand caused by increased consumption and the required non-commute driving, or by serving previously less accessible locations, help create new business establishments and relocate existing jobs and residences. For example, Duranton and Turner (2012) found evidence that the construction of highways causes jobs to increase in the U.S. Employing a quite granular dataset, Gibbons et al. (2019) found similar evidence in the U.K. that improved accessibility brought about by new roads lead to increases in the number of firms, jobs, and worker productivity. Second, distributive demand: mode and route switches (Anas, 2015; Cervero, 2002).

Over the years, an extensive literature has found evidence of a positive traffic elasticity to changes in capacity in the range of 0.2 to 0.8 (Goodwin, 1996; Cervero, 2002). However, Duranton and Turner (2011) were the first to find an elasticity estimate close to one, 1.03 from their preferred regression model, thus confirming the fundamental law's existence. Using a similar methodology, Hsu and Zhang (2014) also find that the law holds in Japan, where the elasticity value ranges from 1.24 to 1.34.

Cities are complex systems. Their dynamics depend on historical, geographical, political, and human behavioral factors. While researchers observe many of these features, given the complexity of cities, it is likely that researchers do not account for many important city characteristics in their data. The presence of unobserved characteristics means that even among observationally identical cities, the demand response to an increase in road capacity can be different. We provide a partial equilibrium analysis that demonstrates the heterogeneity in demand response. As road capacity increases, the cost of travel falls, increasing traffic along the demand curve. Meanwhile, the travel demand curve also shifts to a higher level due to the presence of unobserved characteristics such as those related to historical, planning, and cultural features, and, as we shall focus on in this paper, those related to the *volume* and the *pattern* of non-commute trips. The heterogeneity in demand shifts potentially generates heterogeneity in the elasticity of traffic to road capacity across metropolitan areas, and therefore, the existence of the fundamental law may not be a universal phenomenon across cities. We explore this idea in this paper.

Our simple partial equilibrium model echos with Heckman's assertion in his Nobel lecture

(Heckman, 2001): in econometric models, heterogeneity cannot be restricted to model intercepts but must be included in model slopes as well. Accordingly, to explore the implications of unobserved heterogeneity on the existence of the fundamental law, we modify the econometric model in Duranton and Turner (2011), to a model with a nonseparable or nonadditive error (See Matzkin, 2007). This change makes the elasticity of traffic to road capacity now depend on unobserved city characteristics. The model in Duranton and Turner (2011) contains selection effect: the assignment of additional road capacity to cities is not random. We, therefore, choose the particular nonseparable model developed by Chernozhukov and Hansen (2005, 2006, 2008), which incorporates both nonadditive errors and the selection effect.

The Chernozhukov and Hansen (2005, 2006, 2008) model uses instruments to estimate conditional quantile functions in the presence of selection effect and is, therefore, called the instrumental variable quantile regression (IV-QR) model. IV-QR allows us to study the impact of increased road capacity on *conditional quantiles* of the traffic distribution, not just the conditional mean. Besides, we also estimate the impact of increased road capacity on the unconditional quantiles of the traffic distribution using the generalized quantile regression (GQR) model due to Powell (Forthcoming). The GQR model nests the IV-QR model as a special case and allows conditioning on covariates even when interest lies in the unconditional distribution of a variable.

Using the IV-QR model, Duranton and Turner's (2011) preferred specification yields elasticity estimates of 1.45 at the 10^{th} percentile, 1.19 at the 25^{th} percentile, 0.92 at the median, 0.82 at the 75^{th} percentile, and 0.80 at the 90^{th} percentile respectively. The elasticity estimates monotonically decline as we go up the quantile ladder, and the fundamental law holds only below the median. The GQR estimates concur with the IV-QR results: elasticity estimates go down as we go to higher quantiles, and the estimates are more than one below the median but less than one above it. Specifically, the elasticity is 1.27 at the 10^{th} percentile, 1.22 at the 25^{th} percentile, 1.02 at the median, 0.86 at the 75^{th} percentile, and -0.74 at the 90^{th} percentile, although the last estimate is not statistically significant. The IV-QR and GQR results show that highway construction can reduce the mass at the congestion distribution's upper tail.

Lastly, we further explore the processes behind the empirical findings by employing a spatially detailed general equilibrium model with a road network consisting of 210 nodes and 696 arcs to simulate the effect of building roads on traffic. The model is calibrated carefully to the

Greater LA Region using multiple geocoded data sources to capture the spatial allocations of jobs, population, income, consumption, and commute and non-work trips. Both arc choice in the network and mode choice are modeled explicitly as well as the utility maximization that determines location patterns of the non-work travel. By design, building new roads will impact traffic as measured by vehicle miles traveled (VMT) via three channels. First, the number of total trips by all modes which includes both commutes and non-work trips. A crucial assumption we made in this part is that for a particular origin-destination (OD) pair, the quantity of non-work trips between them is proportionate to the dollar amount of the purchases spent by the residents of the two locations at the other location. For example, the number of non-work trips between two locations A and B is a linear function of the purchases made in A by residents of B *and* the purchases made in B by residents of A. Any exogenous shock that lowers the gross price (inclusive of travel costs) or increases the disposable income (net of commute costs) would lead to more trips hence more VMT. Second, given total travel demand, VMT is also determined by the share of car mode. Building new roads is likely to lower the relative cost of car and induce consumers to switch to driving. Third, since the modeled region contain 97 distinct zones and consumers travel to different zones to acquire consumption goods, the changes in relative gross prices would alter the consumption bundles hence the location patterns of the non-work trips. We find that in the baseline simulation, the VMT elasticity to roads is 0.321 in the Greater LA Region. We also report the welfare changes brought about by a one-percent capacity increase along with model predictions of other travel-related variables. We then carry out a sequence of simulations that differ only in the number of non-work trips and VMTs in the starting points and are otherwise identical. We find that the VMT elasticity to capacity decreases consistently (from 0.356 to 0.291) in initial VMT, and this qualitative outcome is robust against different values of the mode choice elasticity and the elasticity of substitution among shopping locations.

2 The source of heterogeneity: a partial equilibrium analysis

Travel demand. Suppose that $N + x$ is the aggregate demand for VMT. Total VMT can be thought of as consists of two parts: inelastic travel demand N and discretionary travel demand x. Specifically, N is determined by factors such as the size of the workforce in the metropolitan area, urban sprawl, and relevant topological and planning features. For example, one source contributing to urban areas having different levels of N is population. Empirical evidence suggests that travel demand derived from commutes is quite inelastic. For instance, using data from Germany, Gutierrez–i–Puigarnau and van Ommeren (2010) estimated that commute distance, hence travel time, has a positive but numerically very weak effect on the number of days workers choose to work. Evidence such as this suggests that a MSA with a larger workforce tend to have a larger N that can be treated as exogenous and independent of congestion level. Other important sources that contribute to MSAs having different levels of N are those related to density, historical land prices, geographic characteristics, climate, and culture. x, on the other hand, represents the portion of VMT that is elastic and varies with congestion level. VMT derived from shopping trips is such an example. Because driving is generally perceived as a disutility, when facing severe traffic, consumers may cancel a shopping trip or reduce the number of such trips by accomplishing more tasks in each trip, thereby reducing VMT.

Another fact that motivated the inverse demand function in this paper is that the consumer's willingness to pay, even at the same level of VMT, is unlikely to remain constant across different metropolitan areas. If we plotted the inverse demand curves of different urban areas in the same graph, it is likely that these curves will not overlap. A general example that causes different locations to have different demand curves is the amenities attached to various locations, denoted by A. We therefore write the inverse demand for VMT as $P(N + x, A)$. To recognize the fact that the inverse demand can be shifted by amenities which include infrastructures is crucial in that the source of heterogeneity investigated here is indeed the heterogeneous responses in demand to changing levels in A. We assume that:

$$P_n = \frac{\partial P(N + x, A)}{\partial (N + x)} < 0 \tag{1}$$

$$P_A > 0 \tag{2}$$

Assumption (1) simply states the well-established fact that the willingness to pay decreases with VMT. Assumption (2) says that the willingness to pay is higher in places with higher amenity levels. Moreover, we further assume that amenity level is an increasing function of road capacity k and city size N[1] (positive agglomeration externality):

$$A_k > 0, \quad A_N > 0 \tag{3}$$

Travel cost. The private average cost of driving, $C(N + x, k)$, takes the usual form that is increasing and convex in VMT, and it is assumed that the cost goes down as capacity increases. That is:

$$C_n = \frac{\partial C\,(N + x, k)}{\partial\,(N + x)} > 0, \quad C_{nn} > 0, \quad C_k < 0 \tag{4}$$

FIGURE 1 HERE

Equilibrium. The equilibrium VMT is determined by the intersect of the downward sloping inverse demand and the upward sloping private average cost curve in the $-VMT space, as shown in Figure 1. In previous analyses of infrastructure improvements such as Duranton and Turner (2011), building new roads is captured by an outward shift and the flattening of the cost curve, which would move the equilibrium VMT along the demand curve. We argue, however, that this is only part of picture. Building new roads and improving upon existing ones will most likely create more accessible locations and make driving relatively more enjoyable, which, in turn, results in a higher demand given the level of willingness to pay. In other words, road improvements not only shifts and flattens the cost curve, it also shifts the inverse demand curve. To see how improving road capacity would affect the equilibrium VMT, note that the equilibrium condition is given by:

$$P\,[N + x(N, k), A] = C\,[N + x(N, k), k] \tag{5}$$

where $x(N, k)$ is the equilibrium level of discretionary travel demand.

Simple comparative statics yields:

$$x_k = \frac{P_A A_k - C_k}{C_n - P_n} > 0 \tag{6}$$

The inequality in the above follows as long as assumptions (1)-(4) are satisfied. Note that (6) says that discretionary travel demand x as well as total travel demand $N + x$ increase with road

[1] City size here should be interpreted broadly: it reflects the size of the workforce, the geographic expanse, and other location-specific characteristics that affect the inelastic portion of the travel demand.

capacity, and that the magnitude of this increase is greater if the willingness to pay is more sensitive to amenities and road capacities, which is to say, that if $P_A A_k > 0$, which is captured by an capacity-induced outward shift of the inverse demand curve, is greater. The magnitude is also greater if travel cost reduction, $C_k < 0$, is large (more negative) given a certain amount of capacity expansion, which is shown as the outward shift of the cost curve in Figure 1. The adjustment in equilibrium VMT is also shown in Figure 1. Note that $(C_n - P_n) > 0$ measures the gap between the marginal cost and the marginal willingness to pay in the extensive margin (changes along the curves). The condition described by the equation $(C_n - P_n)\, x_k = P_A A_k - C_k$ says that when there is a capacity shock to the equilibrium, the total change in the difference between the cost and the willingness to pay in the extensive margin equals the total change in the difference between the cost and the willingness to pay in the intensive margin ($P_A A_k - C_k > 0$).

Next we turn to the effect of N on equilibrium VMT. As noted before, N can be interpreted as the size of the workforce in the metropolitan area, it can also represents the base VMT level that is inelastic due to economic and planning characteristics. Differentiate (5) with respect to N gives us the following:

$$1 + x_N = \frac{P_A A_N}{C_n - P_n} > 0. \tag{7}$$

Equation (7) says that total equilibrium travel demand $N + x(N, k)$ increases with N. This outcome is once again guaranteed as long as assumptions (1)-(4) are met. Note that we could write (7) as:

$$\frac{d\left[N + x\left(N, k\right)\right]}{dN} = 1 + x_N = \frac{P_A A_N}{C_n - P_n} > 0. \tag{8}$$

While it is possible that $x_N < 0$, which says that *discretionary* travel demand goes down as the inelastic portion of the travel demand increase, the net effect on total travel demand as defined by the sum of inelastic demand *and* elastic demand will always be positive and $x_N > -1$ will always hold. The reason for this is that the demand curve will always be shifted rightward due to the amenity effect caused by a larger N, as can be seen in Figure 2. The equation $(C_n - P_n)(1 + x_N) = P_A A_N$ states that as N changes, the change in the difference between the cost and the willingness to pay in the extensive margin (the left-hand side) is equal to the change in the willingness to pay induced by adjustments in amenities (the right-hand side).

FIGURE 2 HERE

We are interested in the elasticity of aggregate VMT with respect to capacity and how this elasticity reacts to variations in N, or the initial level of VMT. First note that the elasticity of VMT with respect to capacity is given by:

$$\varepsilon_{\text{VMT},\,k} = \frac{d\left[N + x(N,k)\right]}{dk}\frac{k}{N + x(N,k)}$$

$$= \frac{k}{N + x(N,k)} \cdot x_k > 0 \tag{9}$$

Take the derivative of $\varepsilon_{\text{VMT},\,k}$ with respect to N:

$$\frac{d\varepsilon_{\text{VMT},\,k}}{dN} = x_{kN} \cdot \frac{k}{N + x\left(N,k\right)} - x_k x_N \cdot \frac{k}{\left[N + x\left(N,k\right)\right]^2} \tag{10}$$

$$= \left[x_{kN} - \frac{x_k x_N}{N + x(N,k)}\right]\frac{k}{N + x\left(N,k\right)} \tag{11}$$

FIGURE 3 HERE

FIGURE 4 HERE

First notice that the term multiplying the bracket in (11) is always positive. The sign of $d\varepsilon_{\text{VMT},\,k}/dN$ is therefore determined by the relative magnitudes of the two terms inside the bracket. In Appendix 8.1, we discuss the sufficient conditions for $x_{kN} < 0$ and it can be seen that these conditions are met under not at all restrictive assumptions. When x_{kN} is negative, and if the second term in the bracket is not a large positive number, the elasticity of VMT with respect to k decreases in N. This is likely to be true in big metropolitan areas. One could think of a case where city size or inelastic VMT demand N is very large, so that $C_{nn} \gg 0$, which implies that x_{kN} will be sufficiently negative (see Appendix 8.1). Meanwhile, a very large N also makes the second term in the bracket in equation (11) goes to zero. Therefore, a sufficiently large N would guarantee that $\frac{d\varepsilon_{\text{VMT},\,k}}{dN} < 0$.

Capacity elasticity of VMT decreasing in N has significant implications. If instead we take this elasticity as a constant, the conclusion one naturally arrives at is that one could *indiscriminatingly* adopt building roads as an effective means to alleviate congestion if the elasticity is small, or forgo such a costly strategy as a remedy for congestion. However, as shown in this section, the capacity elasticity of VMT in general varies with metropolitan area sizes and, specifically, with the portion of VMT demand that is inelastic. Such variations are substantiated and quantified in the IV-QR estimations and simulated in a more comprehensive and realistic general equilibrium model in the next sections. The variations in the elasticity stem from the fact that the demand response to a capacity change — a shift of the demand for VMT — grew weaker as initial VMT increases. The intuition is that as metropolitan areas become more congested, newly built roads are met with muted demand increases because congestion is already at a high level. On the other hand, if new roads/lanes are added in a less congested location, the generative demand is most likely to be greater than that in a previously congested location. We depict such a situation in Figure 3 and 4. In Figure 3, the increase in road capacity induces a relatively large shift in demand and, while the cost reduction is relatively small as initial congestion is mild, the demand effect and the cost effect together significantly increase the equilibrium VMT. In Figure 4, since initial VMT and congestion in the large city is high, although the cost reduction due to building new roads is stronger than that in the small city, the demand response is weaker. The *percent increase* in equilibrium VMT, as a result, is also smaller than that in the small city given the same *percent increase* in capacity.

3 The IV-QR and GQR methods

3.1 IV: Estimation of the conditional mean function under selection bias

In their article, Duranton and Turner (2011) (DT from now on) are interested in estimating the causal impact of "an increase in the stock of roads on driving in cities" (Duranton and Turner, 2011, Page 2619). In Equation (12), we reproduce their regression model in the potential outcome framework[2], where Y denotes Vehicle Miles Traveled (VMT), d denotes stock of interstate highways lane miles, x denotes a vector of controls[3], U captures unobserved MSA

[2]See Heckman and Robb (1986); Imbens and Angrist (1994); Chernozhukov and Hansen (2005).

[3]The controls include observable MSA characteristics such as present and past populations, geographical features (elevation range, ruggedness, heating degree days, cooling degree days, sprawl), socioeconomic features (share of population with at least some college education, log mean income, share poor, share of manufacturing employment, and an index of segregation), census division dummies, and year dummies. See the appendix in Duranton and Turner (2011) for more details.

characteristics, i index MSAs and t index years[4]. The goal of DT was to consistently estimate α: the elasticity of MSA VMT to interstate highway lane miles in MSA.

$$ln(Y_{it}^d) = \alpha ln(d_{it}) + \beta x_{it} + U_{it}^d \quad (12)$$

In the potential outcome framework, Equation (12) refers to the structural model which relates the potential outcome Y^d to a particular realization of the treatment variable $D = d$ and controls $X = x$, plus an unobservable U^d. Equation (12) describes a mechanism that produces a distribution of the outcome variable: VMT, among observationally identical cities. In the potential outcome framework, $\alpha = \frac{\partial E[ln(Y_{it}^d)]}{\partial ln(d_{it})}$ is interpreted as the structural treatment effect as both y and d are real valued variables, where α measures the impact of the treatment on the conditional mean of the outcome variable.

Another feature of the potential outcome framework is that it separates the definition of the causal effect we are trying to measure: α, from the treatment assignment mechanism. As DT point out, the assignment of roads to cities is not random but it in fact depends on contemporaneous travel demand in these cities. For example, Equation (13) might represent the supply function for roads in cities, where z represents a vector of observed city characteristics which determine the stock of highways but not VMT, and W represents unobserved city characteristic impacting the stock of highways.

$$ln(D_{it}) = \zeta ln(y_{it}) + \eta x_{it} + \theta z_{it} + W_{it} \quad (13)$$

In equilibrium, for every observational unit, the observed VMT Y_{it} and observed highway capacity D_{it} must satisfy both Equations (12) and (13), giving rise to endogeneity or selection effect in the structural model as represented by Equation (14).

$$ln(D_{it}) = \frac{\zeta\beta + \eta}{1 - \alpha\zeta} X_{it} + \frac{\theta}{1 - \alpha\zeta} Z_{it} + \frac{\zeta}{1 - \alpha\zeta} U_{it} + \frac{1}{1 - \alpha\zeta} W_{it} \quad (14)$$

Given a random sample $\{Y_{it}, D_{it}, X_{it}, Z_{it}\}_{i=1,...,N;t=1,...T}$ the observed model can therefore be written as a system of equations.

$$ln(Y_{it}) = \alpha ln(D_{it}) + \beta X_{it} + U_{it} \quad (15)$$

[4]Capital letters denote random variables and lower case letters denote realizations of those random variables.

$$ln(D_{it}) = \mu X_{it} + \sigma Z_{it} + V_{it} \tag{16}$$

As Equation (14) shows, D_{it} is correlated with U_{it}, therefore OLS estimates of α based on Equation (15) is not consistent since the orthogonality condition: $E\left(U_{it}|D_{it}, X_{it}\right) = 0$ is violated. In order to get a consistent estimate of α in the presence of selection effect, DT rely on an instrumental variable (IV) estimator. Their instruments include: the routes of major expeditions of exploration between 1835 and 1850, major rail routes in 1898, and the proposed routes of interstate highways in a preliminary plan of the network (Baum-Snow, 2007). For these variables to be valid instruments for the stock of interstate highways, they must satisfy three conditions: 1) They must not appear in Equation (12), 2) $Cov(Z_{it}, U_{it}) = 0$, 3) The coefficients in front of Z_{it} in Equation (15) cannot be all zero (See Wooldridge, 2010).

Before we move on it is important to mention one more point. DT observe the same metropolitan area over three decades (1983, 1993, and 2003), which allowed them to also use certain panel data regression models (fixed effects and first-difference) to estimate α. However, as DT mention, while these panel data estimators can remove biases due to the presence of time invariant components in U, they do not remove biases emanating from the presence of selection effect, which appear as time variant components in U. Given our goal in this paper to incorporate MSA level heterogeneity in the estimation of α, we chose to use methods which do not exploit the panel nature of data to remove selection biases, but instead rely on instrumental variable techniques. However, we must mention that we could have used the panel structure of the data to reach our goal, for example, by using the non-linear panel data model with interactive fixed effects described in Freyberger (2018). This model would have allowed us to incorporate both temporal and spatial heterogeneity in α and address selection bias more satisfactorily than traditional panel data models. The interactive fixed effect models allow for time variant components in U that can be correlated with the treatment variable (Bai, 2009) and their presence have been shown remove a substantial amount of estimation bias (Moon et al., 2018).

3.2 IV-QR: Estimation of the conditional quantile function under selection bias

In Equation (12), U_{it}^d is a scalar random variable that captures all unobserved MSA characteristics, and appears as an additive term in the structural equation. It therefore has no impact on the structural treatment effect α. To allow for heterogeneity in the treatment effect due to unobserved city characteristics we borrow from the literature on econometric models with nonseparable errors (See Matzkin, 2003, 2007)[5], and rewrite Equation (12) as Equation (17), where the model parameters are now random variables being functions of U_{it}^d.

$$ln(Y_{it}^d) = \alpha \left(U_{it}^d\right) ln(d_{it}) + \beta \left(U_{it}^d\right) x_{it} \tag{17}$$

Given the presence of selection bias in our model, we follow the assumptions made in Chernozhukov and Hansen (2005) (CH from now on) regarding the data generating process, in which case the nonseparable error model in Equation (17) can be interpreted as a conditional quantile function. As CH mentions, in the absence of selection bias: if the assignment of roads to cities were random, Equation (17) is the standard quantile regression model and estimation can proceed as laid out in Koenker and Bassett (1978). However, in the presence of selection bias the appropriate model is the instrumental variable quantile regression model (IV-QR) developed by Chernozhukov and Hansen in a series of papers (Chernozhukov and Hansen, 2005, 2006, 2008).[6]

The IV-QR model comprises primarily of five assumptions:

1. Potential Outcomes: Conditional on $X = x$, for each d, $Y_d = q\left(d, x, U_d\right)$, where $U_d \sim U(0, 1)$ and $q\left(d, x, \tau\right)$ is strictly increasing in τ.

2. Independence: Given $X = x$, U_d is independent of Z.

3. Selection: $D \equiv \delta\left(Z, X, V\right)$ for some unknown function δ and random vector V.

4. Rank Invariance or Rank Similarity: Given $X = x$, $Z = z$, for some d and d'; (a) $U_d = U_{d'}$, (b) $U_d \sim U_{d'}$, are identically distributed.

[5] For an alternative but similar approach, called the linear correlated random coefficients model (see Masten and Torgovitsky, 2016; Hoderlein et al., 2017), where the coefficients in the linear model are also assumed to be random and are allowed to be correlated with some of the observable factors affecting the outcome variable.

[6] See Autor et al. (2017) for applications of the IV-QR model.

5. Observed Variables: $Y \equiv q(D, X, U_D), D \equiv \delta(Z, X, V), X, Z$.

Below we briefly explain these assumptions since they are important in understanding the regression estimates we present later.

The assumptions 1-5 puts certain restrictions on the data generating process behind the IV-QR model. Assumption 1, comprises of a monotonicity and scalar heterogeneity assumption.[7] It implies that a scalar random variable U captures all of the unobserved characteristics of the city, however, this random variable does not enter the model additively, unlike in Equation 1. Besides, U is uniformly distributed and takes values between 0 and 1: U is a rank variable, it ranks observationally identical cities in terms of VMT. The strict monotonicity assumption rules out VMT functions for cities with the same observed covariates and treatment levels but different rank variables from crossing one another.

Assumption 2 rules out any correlation between the rank variable U and the instruments Z. It is similar to the conditional independence assumption made in additive linear IV models. Assumption 3 is just a restatement of Equation (5) in the IV-QR setting. It allows the IV-QR model to incorporate selection bias: the treatment effect is allowed to be determined by both observed and unobserved city characteristics.

As Chernozhukov and Hansen (2005) stress, Assumption 4 is the most important in the list of assumptions mentioned above. Assumption 4(a), the rank invariance assumption implies that U is constant and does not change with different treatment levels. According to Chernozhukov and Hansen (2005), this assumption is more likely to hold if one has a rich set of controls. As Chernozhukov and Hansen (2005) argue, since U is assumed to capture multi-dimensional unobserved characteristics, the rank invariance assumption might be too strong. This is certainly true in our case where we observe the same city in different decades, most likely unobserved characteristics change of time. The rank similarity assumption: 4(b), assumes U to be IID random variable and allows it to deviate across treatment levels in a non-systematic way, from some typical U.[8]

Assumption 4 also rules out serial correlation in our model. It also has some important implications for the interpretation of the treatment effect $\alpha(\tau)$. If rank invariance holds then

[7] Hoderlein and Mammen (2007) presents a model with nonseparable error without the assumption of scalar heterogeneity and monotonicity and discusses the conditions under which the marginal effect is identified. However, that model does not allow for a selection effect on the observed variable whose marginal effect we are interested to estimate.

[8] Dong and Shen (2018) propose a test for rank invariance or similarity.

increasing the stock of interstate highways by one percent will increase VMT by $\alpha(\tau)$ percent for cities at the τ^{th} conditional quantile. However, if rank similarity holds, then $\alpha(\tau)$ is the treatment effect at the τ^{th} conditional quantile: we can infer the effect of treatment on the conditional distribution of VMT but we cannot exactly pinpoint the treatment effect on a particular city.

The five assumptions mentioned above yields Equation (18) (See Theorem 1 in Chernozhukov and Hansen, 2005), which can be interpreted as saying that 0 is the $\tau'th$ quantile of $Y - q(D, X, \tau)$. The estimation problem is then to find a function such that 0 is the solution to the quantile regression of $Y - q(D, X, \tau)$ on X, Z as shown in Equation (19). In the appendix we detail the IV-QR estimation steps which can be seen as a 2SLS version for instrumental variable quantile regression.

$$P[Y < q(D, X, \tau)|X, Z] = \tau \tag{18}$$

$$0 \;\; \epsilon \;\; arg \;\; \underset{f \epsilon \mathscr{F}}{min} \;\; E\rho_{\tau}[Y - q(D, X, \tau) - f(X, Z)] \tag{19}$$

GQR: Estimation of the unconditional quantile function under selection bias

It is important to understand that the IV-QR model estimates the conditional treatment effect: $\alpha\left(\tau\right)$ shows how the distribution of VMT changes at different percentiles in response to changes in road capacity, conditional on observed city characteristics. However, for policy makers it might be more interesting to know the unconditional treatment effect. As Powell (Forthcoming) notes, unlike the unconditional mean function, the unconditional quantile function cannot be derived from the conditional quantile function. For example, since higher population is related to higher VMT, the 10^{th} percentile of the distribution of VMT conditional on city population, say at 1 million, is likely to be much higher than the 10^{th} percentile of the unconditional distribution of VMT . This means that the conditional treatment effect $\alpha\left(\tau\right)$ does not provide any information on the unconditional treatment effect.

To estimate the unconditional treatment effect we use the Generalized Quantile Regression (GQR) model due to Powell (Forthcoming), which still allows the researcher to condition on covariates and has the advantage that it nests the IV-QR model as a special case. In the GQR model, the covariates do not necessarily need to appear as additive terms unlike in the IV-QR model. Instead, the covariates can now influence the distribution of the rank variable

$$U_{it}^{d*} = f\left(U_{it}^{d}, X_{it}\right)$$

where U_{it}^{d*} is still an uniform random variable taking values strictly between zero and one. However, now the distribution of the rank variable depends on the observed covariates of the cities. The modified regression model is shown in Equation (20).

$$ln(Y_{it}^{d}) = \alpha_0 + \alpha_1\left(U_{it}^{d*}\right) ln(d_{it}) \tag{20}$$

Just like IV-QR, the GQR model is built on five assumptions:

1. Potential Outcomes: Y_d is the outcome given policy variables given d; $q\left(d, \tau\right)$ represents the τ^{th} quantile of Y_d.
2. Conditional Independence: $Y_d|X, Z \sim Y_d|X$ for all d.
3. Selection: $D = \delta(Z, X, V)$ for some unknown function δ and random vector V.
4. Rank Similarity: $P\left(Y_d < q\left(d, \tau\right)|X, Z, V\right) = P\left(Y_{d'} < q\left(d', \tau\right)|X, Z, V\right)$ for all d, d'.
5. Observed random vectors consists of $Y := Y_D, D, X, Z$.

These assumptions yield a set of two conditions shown in Equations (21) and (22) which can be used to estimate the GQR model parameters. The details of the estimation procedure are given in the appendix.

$$P(Y \leq q(D, \tau)|X, Z) = P(Y \leq q(D, \tau)|X) \tag{21}$$

$$P(Y \leq q(D, \tau)) = \tau \tag{22}$$

Equation (21) states that the instruments do not provide any additional information on the probability that the outcome variable is less than or equal to the quantile function. Equation (22) is the unconditional counterpart to the Equation (18) in the IV-QR setting. Note, that if the researcher does not have any control variables and all the variables are included in the model as treatment variables, Equation (21) and Equation (22) reduces to Equation (23) which is exactly condition (18). Therefore, the GQR model nests the IV-QR model as a special case.

$$P(Y \leq q(D,\tau)|Z) = P(Y \leq q(D,\tau)) = \tau \tag{23}$$

4 Estimation results

In this section, we report elasticity estimates of MSA VMT to IH lane miles from quantile regression (QR), instrumental variables quantile regression, and generalized quantile regression models. We focus only on the effect of IH lane miles since it is the only measure of road capacity that is comparable across decades in the DT data set. All the regression estimates presented in this section have counterparts in DT, the only difference being that we use the quantile regression framework while DT work under the classical regression setting. As we proceed, we always compare our results with the corresponding results from DT.

Before we proceed it will be useful to mention what trends are we are looking for in the elasticity estimates. In DT, the authors were looking to find if the mean elasticity estimate is greater than, equal to or less than one. In addition, they were checking if the elasticity estimates were changing over the three decades that comprised there data set. In our case we are interested in some additional trends. First, as per the simple partial equilibrium model, we check if the elasticity estimates decline as we move to higher quantiles. Second, we check if the median elasticity is greater than, equal to or less than one, and compare it against the mean estimates of DT. Points one and two taken together imply that we are looking to see if the elasticity estimates are greater than one below the median and lower than one above the median. This is in addition to checking if the elasticity estimates are changing over time.

TABLE 1 HERE

We begin with decade wise elasticity estimates based on univariate quantile regressions (Table 1). For comparison purposes, we also present mean estimates from OLS regressions as in DT. There are a few trends that are clear in Table 1. First, at every quantile, the elasticity estimate is greater than one, meaning that congestion increases in response to highway construction at every point in the conditional distribution function of MSA VMT. Second, the mean and median estimates are reasonably close in every decade, implying that the conditional distribution is symmetric. Third, in every decade we mostly observe elasticity falling as one moves to higher quantiles: while congestion increases in response to highway construction at every point in the conditional distribution of VMT, it rises less at the higher quantiles as compared to the lower quantiles. Finally, the elasticity estimates in the 25^{th}-75^{th} percentile range decline as we go forward in time.

TABLE 2 HERE

As a next step, following DT, we estimate several decade wise multivariate regressions whose results are shown in Table 2. Adding controls lowers the elasticity estimates across the board to less than one. Again, the mean and median elasticity are reasonably close–being around 0.92 in 1983, and around 0.76 in 1993 and 2003 respectively. In Table 2, the monotonic decline in elasticity is visible prominently in Columns (2) and (8), corresponding to DT preferred specification[9]. The heterogeneity in elasticity when it declines monotonically, and as measured by the difference in elasticity between the 10^{th} and 90^{th} percentile, is around 0.32 in Column (2) and around 0.14 in Column (8) respectively. In Table 2, elasticity declines between 1983 and 1993 across model specifications, but is stable between 1993 and 2003.

TABLE 3 HERE

In Table 3, we present elasticity estimates from different model specifications based on pooling the cross-sectional units across time. One side effect of this approach is that we lose the time variation in elasticity as seen in Tables 1 and 2. In Column 1 in Table 3, a model with only time dummies as controls, we see that the elasticity estimates are always greater than one but they decline monotonically from 1.40 at the 10^{th} percentile to 1.19 at the 90^{th} percentile. The median elasticity in Column 1 is equal to 1.25, again close to the mean elasticity estimate from a pooled OLS regression. In Columns 2-4, with additional controls, the elasticity estimates are always less than one, being around 0.80 at the median, but declining as we go up the quantiles. The largest variation in elasticity is seen in DT's preferred specification in Column 3 where the estimates go down from 0.99 at the 10^{th} percentile to 0.72 at the 90^{th} percentile. The picture changes completely when we add MSA fixed effects in Columns 5-9.

[9] The controls in this specification are unambiguously exogenous unlike socioeconomic characteristics which may be endogenous themselves

The elasticity estimates no longer show any variation around the median and are close to one at every quantile, being slightly lower than the respective mean estimates. However, as argued by Koenker (2004), in a quantile regression setting including fixed effects may not make sense especially when the number of time units per cross-sectional unit is small as is the case in DT's analysis. One must keep in mind that no within transformation is available in the quantile regression setting unlike in linear panel data models and one must estimate the individual fixed effects as parameters.

TABLE 4 HERE

TABLE 5 HERE

TABLE 6 HERE

TABLE 7 HERE

The presence of fixed effects in the models in Table 3 controls for additive time-invariant unobserved MSA characteristic, and accounts for its correlation with the treatment variable, but it does not resolve the bias due to the presence of selection effect: the correlation of the unobserved MSA characteristic with the treatment variable. To address this problem, DT use a pooled IV estimator. In Table 4, we present DT's 2SLS estimates along with our corresponding pooled IV-QR[10] results. In Tables 5-7 we replicate the regressions mentioned in Table 4, but use only one instrument in each model, namely, 1947 planned interstates in Table 5, 1898 railroads in Table 6, and 1835 exploration routes in Table 7 respectively. Across all the models in Table 4, elasticity declines monotonically as we go up the quantiles. In addition, as can be seen from the values of the Kolmogorov-Smirnov statistic, this variation in elasticity is statistically significant, except for the model in Column (2). Further, we see that except Column 1, where the elasticity estimates are greater than one at all the quantiles, the pattern that emerges is one where elasticity is greater than one at the 10^{th} and 25^{th} percentile, elasticity is close to one at the median, and less than one at the 75^{th} and 90^{th} percentiles. In particular, the model in Column 3, which is DT's preferred model, the elasticity is 1.45 at the 10^{th} percentile, 1.19 at the 25^{th} percentile, 0.92 at the median, 0.82 at the 75^{th} percentile, and 0.80 at the 90^{th} percentile. Figure 5 shows the monotonically declining pattern of elasticity corresponding to

[10] See Chen (2019) for details on the R package used to generate the results.

Column (3) in Tables 4-10 for percentiles between 0.02 and 0.98 at intervals of 0.01 with the associated 95% confidence interval.

TABLE 8 HERE

TABLE 9 HERE

TABLE 10 HERE

FIGURE 5 HERE

In Table 5 and 6, DT's preferred specification yields similar results. In Table 7 however, the IV-QR model with only 1835 exploration routes as instrument, the elasticity from DT's preferred model (Column 3) is much lower at the 75^{th} and 90^{th} percentile as compared to Tables 4, 5, and 6 — being 0.67 at the 75^{th} percentile and 0.63 at the 90^{th} percentile respectively. This is generally true for all the models when only 1835 exploration routes is used as instrument. Across Tables 5-7, the heterogeneity in elasticity is also statistically significant in Column (3).

In Tables 8-10, following DT, we report results from decade wise IV-QR models, with ln 1898 railroads, and ln 1947 planned interstates as instruments. In 1983 and 1993, DT's preferred specification: Column (3), does not yield a clear monotonic decline in elasticity and cluster near one, being always greater than one in 1983. The 2003 elasticity estimates corresponding to Column (3), however, shows a clear monotonic decline in elasticity with the median elasticity at 0.81 and the elasticity at 75^{th} and 90^{th} percentile being 0.72 and 0.59 respectively.

TABLE 11 HERE

In Table 11, we present elasticity estimates from the GQR model using all three instruments. First, across all models in Table 11, the elasticity generally declines as we move up the quantiles. Second, the GQR results show that the elasticity is almost always greater than one at the 10^{th}, 25^{th}, and 50^{th} percentiles. Interestingly, except in Column (1), the elasticity at the median is very close to the mean effect found in DT: 1.03. However, except Column (1), for all other models elasticity declines below one at the 75^{th} and 90^{th} percentiles. In fact, in Columns (3), (4), and (5), the elasticity at the 90^{th} percentile is negative but not statistically significant from zero. For comparison purposes, in Figure 6 we plot the elasticity at different quantiles corresponding to Column (3) or DT's preferred specification.

FIGURE 6 HERE

5 The simulation model

5.1 Spatial details of the model: the Greater LA Region

In the next two sections, we turn to a general equilibrium model with a road network, hereafter the LA TRAN model, to explore the effect of increased capacity and how this capacity effect is related to the initial VMT, or more generally, to city size *before* the infrastructure improvement. Since LA TRAN is a spatial general equilibrium model, we shall first briefly describe the Greater LA Region to which the model is calibrated. Then sections 5.2 - 5.5 discuss the model in detail. Section 5.6 explains calibration. We report simulation results in Section 6.

FIGURE 7 HERE

FIGURE 8 HERE

In our model, the Greater LA Region is divided into 97 zones[11] which span six counties[12] as shown in Figure 7. Connecting these 97 zones, is the LA TRAN road network which include 210 nodes (Figure 9) that represent the zones and waypoints and 696 arcs (Figure 8) that are aggregations of real-world roads. We therefore have a 97-by-97 origin-destination (OD) matrix.

5.2 The model structure I: A short description

The consumers in the model consist of both employed and unemployed consumers. The residential locations and employment locations are assumed to be fixed. In essence, this assumption exclude the possibility of relocations so that any changes in VMT could only stem from adjustments in the demand for consumption bundles, mode choices, and arc choices, i.e. consumption and driving behaviors. In doing so, we can focus on a well-defined short-run capacity elasticity of VMT, whereas in the long run such an elasticity is also a function of many confounding factors that are of income, technological, demographic, regulatory nature which also drive relocations. For example, Baum-Snow (2010) documented that roads connecting city centers and suburban areas are responsible for the suburbanization pattern in the U.S. where both jobs and residences are relocated from central cities to suburbs. Using a spatial general equilibrium model RELU-TRAN, Anas (2015) illustrated that as employment and residential real estate

[11] See Li et al. (2014).

[12] These six counties are Imperial, Los Angeles, Orange, Riverside, San Bernardino, Ventura.

developments are suburbanized as a result of new roads and cheaper land prices, jobs and residences chase each other in space in a cyclical fashion which keeps the travel times at a stable and tolerable level so that the suburbanization process becomes sustainable. However, these long-run effects are beyond the scope of the present study.

In our model, employed consumers commute daily on workdays. Besides commute trips, both employed and unemployed consumers make discretionary non-work trips between their residential location and various shopping locations. Non-work trips are required for acquiring consumption goods. We assume that the number of non-work trips between a residential location i and a shopping location j is linearly related to the dollar amount of expenditure spent in j by consumers from i. Specifically, the consumer maximizes utility over consumption of housing and goods while taking into account transportation costs incurred by consumption and commuting. Goods sold at different locations (zones) are considered imperfect substitutes with constant elasticity of substitution and the utility-maximizing consumer chooses the optimal bundle of goods sold at different locations, hence the shopping trip patterns.

Given travel demands generated by commute trips and shopping trips, the probability of choosing a particular mode of travel between each origin-destination (OD) pair is determined by the generalized costs of different modes. It is assumed that the costs of travel are exogenous for modes other than driving. In general, if travel demand is increased between a particular OD pair, worsening congestion will cause driving to become more expensive, and the change in relative costs would lead to some drivers switching to other modes of travel.

With the OD matrix generated by the utility-maximizing consumers and the mode choice probabilities determined by the relative costs of different modes of travel, it remains to determine the vehicle traffic flow equilibrium on the LA TRAN network. There are 210 nodes in the road network, 97 of which represent the zones, and the rest are waypoints that represent intersections of roads. There are 696 arcs connecting the 210 nodes so that for every OD pair, there are multiple paths that each consists of a series of arcs. For a trip originates from i and terminates in j, a driver faces multiple choices of arc combinations that could take him from the origin to the destination. Given travel demand (OD matrix), the network equilibrium is a probabilistic traffic flow assignment on all the arcs such that the expected disutility of travel is the lowest for all OD pairs.

5.3 Solving the model

As shown in Figure 10, the LA TRAN model can be viewed as consisting of several interrelated parts that are solved sequentially and iteratively. To solve the model, for example, the sequence could start with the utility maximization problem (UMP) which we will describe in detail below. Solving the UMP gives us the non-work trips matrix. Because each consumer's residential and employment locations are assumed to be fixed, simply adding the exogenous commute trips and the endogenous shopping trips gives us the OD matrix.

FIGURE 10 HERE

Suppose that mode choice probabilities are known, the OD matrix of *car trips* can be obtained from multiplying the all-modes OD matrix by mode choice probabilities. The car trips matrix is then loaded onto the road network to solve for equilibrium flows, travel times, and monetary costs on all arcs. We assume that the time and monetary costs of other (non-driving) modes remain unchanged.

TABLE 12 HERE

TABLE 13 HERE

Now that the generalized travel costs of all modes are known, mode choice probabilities and the across-modes average travel times and costs between all location pairs can be calculated. Next, the UMP needs to be solved again using updated travel costs, which, in turn, will generate an updated shopping trip matrix that will be used to update the all-modes OD matrix and the car trip matrix. The updated traffic flow equilibrium would give updated mode choice probabilities. This iterative process continues until all model variables converge to the equilibrium.[13] The rest of this section will explain in detail the three parts of the model and the welfare analysis.

5.4 The model structure II: A detailed description

5.4.1 The consumer

In the rest of sections 5 and 6, many symbols will be used and all of them are listed in Table 12 and Table 13 for reference. The UMP of a consumer whose residential and job locations are

[13]The model is considered converged once the maximum of the errors of all model variables between two consecutive iterations is below a tight tolerance level.

in model zone i and j, respectively, and whose income belongs to type-f is given by:

$$\max_{Z_{z|ijf}, h_{ijf}} U_{ijf} = \alpha_{if} \ln \left[\left(\sum_{z} \iota_{z|ijf} \cdot \left(Z_{z|ijf} \right)^{\sigma_f} \right)^{\frac{1}{\sigma_f}} \right] + \beta_{if} \ln \left(h_{ijf} \right) + \gamma_{ijf} G_{ijf} \tag{24}$$

The budget constraint:

$$\Delta_j \cdot \left[days \cdot hours \cdot w_{jf} \cdot \left(1 - t^i_{jf} \right) - days \cdot g_{ijf} \right] + m_{if} \cdot \left(1 - t^i_{jf} \right) \geq \sum_{z} P_{z|if} Z_{z|ijf} + R_i h_{ijf} \tag{25}$$

where i and j represent residential and employment locations with $j = 0$ indicating a consumer being unemployed. f denotes the income group to which a consumer belongs. $Z_{z|ijf}$ is the consumption of goods purchased in zone z by a consumer of type (i, j, f). $1/(1 - \sigma)$ is the elasticity of substitution among goods sold at different locations with $\sigma < 1$. $\iota_{z|ijf}$ is a calibrated parameter that captures the attractiveness of location z to consumer-(i, j, f). h_{ijf} is the consumption of housing goods of a consumer who lives in location i, works in location j, and is of type f. G_{ijf} is the two-way across-modes weighted average travel time between location i and j for consumer f. α_{if} and β_{if} are expenditure share parameters. γ_{ijf} determines both the magnitude of the disutility from commute times and value of time (VOT), i.e. the marginal rate of substitution between average travel time and disposable income: $\frac{\partial U / \partial G_{ijf}}{\partial U / \partial M_{ijf}}$, where

$$M_{ijf} = \Delta_j \cdot \left[days \cdot hours \cdot w_{jf} \cdot \left(1 - t^i_{jf} \right) - days \cdot g_{ijf} \right] + m_{if} \cdot \left(1 - t^i_{jf} \right) \tag{26}$$

is the disposable income.

In the budget constraint of the consumer, $\Delta_{j=0} = 0$ and $\Delta_{j=1,\ldots,97} = 1$ because unemployed consumers do not earn wage income, neither do they incur commute cost. w_{jf} is the hourly wage payment offered at location j for worker type f, t^i_{jf} the associated income tax rate, g_{ijf} the two-way monetary travel cost between i and j for road user f. *days* and *hours* are the number of work days in a year and the first bracket in the budget constraint is the wage income after commuting cost. m_{if} is the non-wage income, R_i the rent.

For each unit of the consumption good, consumers pay a price inclusive of transportation cost, or delivered price $P_{z|if}$:

$$P_{z|if} = p_z (1 + t^s_z)(1 + s_{izf} g_{izf}) \tag{27}$$

where p_z is mill price of goods sold in zone z, t_z^s the sales tax rate at location z. The parameter s_{izf} represents the number of trips required for *each dollar* spent on goods at z. We calibrate this parameter in the initial equilibrium so that the model-generated non-work trips computed from observed consumer expenditure data matches observed non-work trips. Solving the utility maximization problem yields:

$$\begin{aligned}\tilde{U}_{ijf} &= \alpha_{if} \ln \alpha_{if} + \beta_{if} \ln \beta_{if} + \ln M_{ijf} - \beta_{if} \ln R_i \\ &+ \frac{\alpha_{if}(1-\sigma_f)}{\sigma_f} \ln \left(\sum_z \iota_{z|if}^{\frac{1}{1-\sigma_f}} P_{z|if}^{\frac{\sigma_f}{\sigma_f - 1}} \right) \\ &+ \gamma_{ijf} G_{ijf}\end{aligned} \tag{28}$$

We can then compute the daily travel demand from location i to z:

$$TRIP_{izf} = N_{izf}^e + \frac{s_{izf}}{days} \sum_{j=0}^{97} N_{ijf} P_{z|if} Z_{z|ijf} \tag{29}$$

where N_{izf}^e is the number of daily *two-way* commute trips between i and z. Note that N_{izf}^e also represents the number of employed consumers of type (i, z, f). The Walrasian demand for goods is solved from the UMP:

$$Z_{z|ijf} = \frac{\iota_{z|if}^{\frac{1}{1-\sigma_f}} P_{z|if}^{\frac{1}{\sigma_f - 1}}}{\sum_{z'}^{97} \iota_{z'|if}^{\frac{1}{1-\sigma_f}} P_{z'|if}^{\frac{\sigma_f}{\sigma_f - 1}}} \alpha_{if} M_{ijf} \tag{30}$$

The summed term in Equation (29) is the total dollar value of consumption spent in zone z by consumers who live in zone i and of income group f. Hence the entire second term in (29) is the number of daily two-way shopping trips between location i and z by f type consumers. Also note that in our model daily *commute trips* are fixed for all OD pairs, whereas shopping trips are endogenous and are functions of travel costs.

It is useful to point out that in LA TRAN, besides a convex congestion function, which will be explained later in Section 5.4.3, travel demand in terms of both the number of trips and the location pattern (OD-matrix) will also change as a result of infrastructure improvements. Similar to the partial equilibrium example given in Section 2, an increase in density or capacity of roads, *ceteris paribus*, will not only flatten the average cost curve, the demand curve will also be shifted to the right due to the lowering of transportation costs.[14] Given the same change in

[14]For example, see Small and Verhoef (2007).

the average cost function, if the *demand response* differs by MSAs, then the same increase in road capacities at different locations will result in different responses in VMT. The heterogeneity in the demand responses across geographic units is what is driving the heterogeneity in the elasticity of VMT with respect to infrastructure provisions.

5.4.2 Mode choice

There are four modes of travel in the LA TRAN model. Besides driving ($m = 1$), $m = 2, 3, 4$ represent bus, rail, and other, respectively. Once travel demands are generated from the UMP, the endogenous probabilities of choosing driving for all OD pairs are given by:

$$PROB_{m=1|ijf} = \frac{\exp\left\{\Theta\left(\wp_{ji|f} + \wp_{ij|f}\right) + \mathrm{K}_{1|ijf}\right\}}{\exp\left\{\Theta\left(\wp_{ji|f} + \wp_{ij|f}\right) + \mathrm{K}_{1|ijf}\right\} + \sum_{m'=2}^{4} \exp\left\{\Theta\left(GC_{m'|ji} + GC_{m'|ij}\right) + \mathrm{K}_{m'|ijf}\right\}} \tag{31}$$

where, in (31), Θ is the dispersion parameter in the multinomial logit model, $\wp_{ji|f}$ is the generalized one-way travel cost of driving that includes both time and pecuniary costs between j and i (more on this later, in Equation 40). $\mathrm{K_{m|jif}}$ is the exogenous disutility associated with each choice situation. For $m = 2, 3, 4$, $GC_{m|ji}$ represent the exogenous generalized travel costs of non-driving modes which are the sum of the monetized time cost of $TIME_{m|ji}$ and the exogenous pecuniary cost $MCOST_{m|ji}$:

$$GC_{m|ji} = vot_f \cdot \frac{TIME_{m|ji}}{60} + MCOST_{m|ji}; \quad m = 2, 3, 4. \tag{32}$$

Similarly, the probabilities of choosing rail, bus, and other ($m = 2, 3, 4$) are:

$$PROB_{m\neq1|ijf} = \frac{\exp\left\{\Theta\left(GC_{m|ji} + GC_{m|ij}\right) + \mathrm{K}_{m|ijf}\right\}}{\exp\left\{\Theta\left(\wp_{ji|f} + \wp_{ij|f}\right) + \mathrm{K}_{1|ijf}\right\} + \sum_{m'=2}^{4} \exp\left\{\Theta\left(GC_{m'|ji} + GC_{m'|ij}\right) + \mathrm{K}_{m'|ijf}\right\}} \tag{33}$$

Next, one-way vehicle trips, which will be loaded onto the road network, can be calculated using the all-modes OD matrix $TRIP_{izf}$ given above and mode choice probabilities:

$$AUTOTRIP_{izf} = \frac{TRIP_{izf} \times PROB_{1|izf} + TRIP_{zif} \times PROB_{1|zif}}{passengers\ per\ vehicle_f} \tag{34}$$

5.4.3 Network equilibrium

Now that Equation (29) gives us the trip matrix $Trip_{ijf}$ generated by the *endogenous* consumption patterns and the exogenous residential and job locations, and that Equation (31) and (33) describe how mode choice is determined for each OD pair (i, j), the endogenous travel demand by car can be loaded onto the road network of 210 nodes and 696 arcs to compute the traffic assignment equilibrium. In this network equilibrium, each OD pair chooses a series of arcs, i.e. a route, in the network that give the lowest expected disutility of travel. This procedure is an extension of the Markovian dynamic programming algorithm based on Baillon and Cominetti (2008) and later adapted in the applications of the spatial GE model RELU-TRAN such as Anas (2020).

The congested travel time $time_a$ (measured in minutes) on a particular arc, a, is given by the Bureau of Public Roads (BPR) power function:[15]

$$time_a = time_a^{free-flow}\left[1 + b_a\left(\frac{\sum_f flow_{af}}{capacity_a}\right)^C\right] \tag{35}$$

where $time_a^{free-flow}$ is the free-flowing travel time on arc a; b_a and C are the congestion parameters, $capacity_a$ is arc (road) capacity and $flow_{af}$ the traffic flow of drivers from income group f on arc a. Note that $capacity_a$ is an abstract value calibrated to match model generated congested travel times with observed driving times in the initial equilibrium. The monetary costs *per passenger* on arc a is given by:

$$mcost_{af} = \frac{[price_{fuel} \times F(speed_a)] \times length_a}{passengers\ per\ vehicle_f} \tag{36}$$

where $speed_a = \frac{length_a}{time_a/60}$ is the *average* speed on arc a measured in mile-per-hour as a function of the length of the arc and the congested travel time. $price_{fuel}$ is gasoline price, $F(speed_a)$ is technological fuel intensity (TFI), or gasoline usage *per mile* as a function of speed:[16]

$$F(speed_a) = \sum_{n=0}^{6}(-1)^n c_n speed_a^n = \frac{1}{FE} \tag{37}$$

where FE stands for fuel efficiency as measured by $mile/gallon$. In the above fitted polynomial

[15] Bureau of Public Roads (1964).

[16] $c_0 = 0.122619$, $c_1 = 0.0117211$, $c_2 = 0.0006413$, $c_3 = 0.000018732$, $c_4 = 0.0000003$, $c_5 = 0.0000000024718$, $c_6 = 0.000000000008233$. See Davis et al. (2009).

function, if we plotted speed on the horizontal axis and TFI on the vertical axis, the TFI takes a U-shaped form with its minimum at around 45 $mile/hour$. Below this speed, a slower speed caused by congestion means a higher TFI (or lower fuel efficiency); above this speed, a *higher* speed which may be induced by improvements in infrastructure or mitigated congestion means a *higher* TFI, i.e. less congestion is causing more gasoline use.

The non-linear relationship between vehicle speed and TFI described by Equation (37) is the source of a special kind of externality: the externality in gasoline use. To our knowledge, we are the first to consider this type of externality in economic modelings of congestion. For example, in our calibrated model, the average speed over major roads and local roads in the initial equilibrium are around 43 and 30 $mile/hour$[17], respectively. Since both speeds are slower than 45 $mile/hour$, the optimal (minimum) TFI speed, an overall relief of congestion that speeds up traffic would not only lower each driver's use of gasoline — a reduction in the *private average cost*; it would also cause the extra gasoline use of all drivers imposed by each *individual* driver to go down — a reduction in the *social marginal cost*. This gasoline use externality caused by the non-linearity of the TFI function is calculated and presented alongside with the familiar time delay externality in our simulations. Its magnitude, as it turns out, is numerically inconsequential given the parameters in our model.

The generalized vehicle cost $gcost_a$ on arc a is defined as the sum of monetary and time costs:

$$gcost_{af} = vot_f \times \frac{time_a}{60} + mcost_{af} \tag{38}$$

For each trip set out from its origin to destination, it sequentially chooses which road (arc) to take at every intersection (node) that they reach during the journey. Each trip chooses a route that consists of roads that give the lowest expected disutility for the trip. The process is iterated updating travel times and monetary costs until an equilibrium is reached. Baillon and Cominetti (2008) showed that a unique equilibrium exists for the network model. The multinomial logit probability of choosing arc a at node o, given destination node d and driver income type f is:

[17] These speeds are not unrealistically slow, especially in the Greater LA Region, because they are the average speed that has taken into account not only congestion, but also road characteristics such as stops, bottlenecks, and other planning and topological features.

$$Pr_{a|df} = \frac{\exp\left[-\Omega_{df}\left(gcost_{af} + \wp_{\pi(a)|df}\right)\right]}{\sum_{a' \in A_o^+} \exp\left[-\Omega_{df}\left(gcost_{a'f} + \wp_{\pi(a')|df}\right)\right]}, \quad \sum_{a \in A_o +} \mathrm{P}_{a|d} = 1, \quad \forall d; \quad a \in A_o \tag{39}$$

For a trip leaving node o for destination d, $Pr_{a|df}$ is the probability of choosing road a, $\pi(a)$ is the end node of the road a, A_o^+ the set of all roads that are outgoing from node o. Ω_{df} is the dispersion parameter of the disutility shock, $gcost_{af}$ the generalized vehicle cost on road a given by Equation (38) and $\wp_{o|df}$ the expected disutility of driving between nodes o and d for a traveler of type f is given by:

$$\wp_{o|df} = -\frac{1}{\Omega_{df}}\left[\sum_{a' \in A_o^+} \exp\left[-\Omega_{df}\left(gcost_{a'f} + \wp_{\pi(a')|df}\right)\right]\right] \tag{40}$$

Note the iterative nature of Equation (40): in $\left(gcost_{a'f} + \wp_{\pi(a')|df}\right)$, $gcost_{a'f}$ is the generalized vehicle cost *given that* road a' is chosen, and $\wp_{\pi(a')|df}$ is the expected disutility from the end node of a' to the destination node d, and A_o^+ is the set of all a' that are available for the driver at the intersection o. Notice that, in equilibrium, from a node i to a destination node d, the total number of car trips, $x_{i|df}$, equals the sum of car trips that *originated from* i and other trips *started elsewhere* whose path crosses i and terminates at d:

$$x_{i|df} = TRIP_{i|df} + \sum_{a \in A_o^-} v_{a|df} \tag{41}$$

where, in Equation (41), the first term on the right is the total daily trips from i to d by f consumers that was given by Equation (29). Similar to A_o^+ which stands for the set of all *outgoing* arcs from node o, A_o^- is the set of all *incoming* arcs to node o. The second term, therefore, represents the traffic that is passing through (instead of originated from) node i and will terminate at node d. It follows that traffic volume on arc a with destination d, $v_{a|df}$, is given as follows:

$$v_{a|df} = x_{i(a)|df} Pr_{a|df} \tag{42}$$

where the subscript $i(a)$ means that i is the starting node of arc a. Equation (42) states that total traffic volume on arc a equals total vehicle trips from i to d times the probability of choosing arc a given d and f. Finally, we sum over all destination nodes to get the daily flow on arc a by type f consumers:

$$flow_{af} = \sum_{d} v_{a|df} \tag{43}$$

The network model solves for equilibrium flows, $flow_{af}$, on all the arcs in the network such that, for each OD pair, the series of arcs chosen would give the lowest disutility of travel. Using equilibrium $flow_{af}$, expected network equilibrium travel times τ_{id} and expected monetary cost between i and d μ_{idf} can be computed easily.

Moreover, since the model also treats zonal congestion in each model zone,[18] we assume there is one intrazonal arc in each model zone and the flow on each zonal arc, $Zflow_i$, is given by:

$$Zflow_i = AUTOTRIP_{ii} + ACCESS\_AUTOTRIP_i + EGRESS\_AUTOTRIP_i \tag{44}$$

In Equation (44), $ACCESS\_AUTOTRIP_i$ is a fraction of all vehicle trips leaving i for all destinations and, similarly, $EGRESS\_AUTOTRIP_i$ a fraction of all vehicle trips arriving i from all origins.

Given the length, free-flowing time, and capacity of local arcs, local congested travel times can be calculated in a similar fashion as we did for network equilibrium (Equation 35). Then, to calculate the equilibrium *driving* time inclusive of intrazonal driving between i and d, $\mathbb{T}_{id}$, we add together network equilibrium vehicle travel time (not including intrazonal travel) τ_{id}, intrazonal vehicle travel time in origin zone $Ztime_{ii}$, and intrazonal vehicle travel time in destination zone $Ztime_{dd}$:

$$\mathbb{T}_{id} = weight^{access} \times Ztime_{ii} + \tau_{id} + weight^{egress} \times Ztime_{dd} \tag{45}$$

The equilibrium monetary *vehicle* cost between i and d for a consumer from income group f, $\mathbb{M}_{idf}$, which includes both network and local monetary costs, is calculated in the same manner:

$$\mathbb{M}_{idf} = weight^{access} \times Zmcost_{iif} + \mu_{idf} + weight^{egress} \times Zmcost_{ddf} \tag{46}$$

in which $Zmcost_{iif}$ is the intrazonal monetary cost in zone i. Lastly, the two-way *across-modes average* travel time G_{izf} and *across-modes average* monetary travel cost g_{izf} are given by the

[18] This can be thought of as congestion that takes place within each model zone i.

following equations:

$$G_{izf} = PROB_{1|izf} \times (\mathbb{T}_{iz} + \mathbb{T}_{zi}) + \sum_{m>1} PROB_{m|izf} \times \left(TIME_{m|iz} + TIME_{m|zi}\right) \tag{47}$$

$$g_{izf} = PROB_{1|izf} \times (\mathbb{M}_{izf} + \mathbb{M}_{zif}) + \sum_{m>1} PROB_{m|izf} \times \left(MCOST_{m|iz} + MCOST_{m|zi}\right) \tag{48}$$

5.5 Welfare analysis

We measure welfare changes in three parts. The first part is the change in utility levels measured by equivalent variations (EVs). The second and third parts are dollar values of the time delay and gasoline use externalities. We give a breakdown of the derivations in the welfare analysis in the rest of this section.

5.5.1 Equivalent variations (EVs).

Recall the indirect utility function (28), which is repeated here for convenience:

$$\begin{aligned}\tilde{U}_{ijf} &= \alpha_{if} \ln \alpha_{if} + \beta_{if} \ln \beta_{if} - \beta_{if} \ln R_i \\ &+ \ln M_{ijf} \\ &+ \frac{\alpha_{if}\left(1-\sigma_f\right)}{\sigma_f} \ln \left(\sum_z \iota_{z|if}^{\frac{1}{1-\sigma_f}} P_{z|if}^{\frac{\sigma_f}{\sigma_f - 1}} \right) \\ &+ \gamma_{ijf} G_{ijf}\end{aligned} \tag{49}$$

First note that rent and *mill* prices of goods are exogenous in the model. When there is a shock such as an increase in road capacity, as the network model re-equilibrates, the new equilibrium traffic flows on arcs yield new equilibrium arc choice probabilities (Equation 39), mode choice probabilities (Equation 31, 33), congested travel times (Equation 45, 47), monetary travel costs (Equation 36, 48).

In turn, any adjustments in travel times and costs will affect utility via three channels: First, disposable income after commute cost (M_{ijf}, given by Equation 26), which appears in the second line of Equation (49); secondly, prices of goods inclusive of transportation costs

($P_{z|if}$, given by Equation 27), which appears in the third line of (49); and, thirdly, average across-modes travel time (G_{ijf}, given by Equation 47), that appears in the fourth line of Equation (49).

In the general equilibrium context considered here, it is appropriate to use the EV or compensating variation (CV)[19] to approximate the dollar value of changes in utility levels. An important assumption we maintain is that both mode and arc choices are outcomes of the network equilibrium, in the sense that the impact of one consumer's mode and arc choices on the across-modes average travel times G_{izf} are negligible. In other words, G_{izf} are the *market equilibrium* average travel times that are external to individual consumers. With this assumption, G_{izf} is treated as an exogenous parameter in the UMP (Equation 24), and the utility levels are *deterministic* in nature as opposed to be stochastic.

Let the initial indirect utility level be $\tilde{U}^{base}$ and suppose that the indirect utility level reaches $\tilde{U}^{policy}$ after the policy shock, which, in this case, is an expansion in road capacities. Then, using Equation (49), EV_{ijf} can be calculated as follows:

$$\begin{aligned} EV_{ijf} = \exp \Big[& \tilde{U}_{ijf}^{policy} - \alpha_{if} \ln \alpha_{if} - \beta_{if} \ln \beta_{if} + \beta_{if} \ln R_i \\ & - \frac{\alpha_{if}\left(1-\sigma_f\right)}{\sigma_f} \ln \left(\sum \iota_{z|ijf}^{\frac{1}{1-\sigma_f}} \underbrace{\left[p_z\left(1+t_z^s\right)\left(1+s_{izf}\cdot g_{izf}^{base}\right)\right]}_{P_{z|if}}^{\frac{\sigma_f}{\sigma_f-1}} \right) \\ & - \gamma_f G_{ijf}^{base} \Big] - M_{ijf}^{base} \end{aligned} \tag{50}$$

5.5.2 Externalities

Time delay externality. The time delay externality imposed by *one extra driver* on arc a is the difference between the social marginal cost and the driver's private average cost. It is straightforward to compute the externality measured in minutes from Equation (35). To include this externality as part of the welfare change, we have to convert time delay into dollar value using flow-weighted VOTs. Although the marginal rate of substitution between disposable income and across-modes travel time varies as the model transitions to a new equilibrium after a policy shock, we keep VOTs as constants throughout the simulations, which are set to 50% of hourly wages.[20] We verified that assuming away the endogeneity of VOTs — that is, treating

[19] we compute both the EVs and the CVs, which give highly similar numerical results. We present only EVs here.

[20] See Small (2012) for a comprehensive review of the valuation of travel time.

VOTs as constants as opposed to using the endogenous marginal rate of substitution between income and travel time as VOT — has a negligible effect on our numerical results. Using Equation (35), it is easy to show that the time delay externality imposed by one driver on arc a is:

$$\begin{aligned} externality_a^{time} &= SocialMarginalCost_a^{time} - PrivateAverageCost_a^{time} \\ &= c \cdot b_a \cdot time_a^{free-flow} \cdot \left(\frac{\sum_f flow_{af}}{capacity_a}\right)^c \cdot \frac{\sum_f flow_{af} \cdot vot_f}{\sum_f flow_{af}} \end{aligned} \tag{51}$$

Gasoline use externality. As explained before, because the TFI (gasoline use per mile) is a U-shaped non-linear function of vehicle speed which reaches its minimum around 45 $mile/hour$, each extra driver on arc a imposes an externality in gasoline consumption because the driver would change the average flow speed on the arc. This externality can be derived from the TFI function given by(37):

$$\begin{aligned} &externality_a^{gasoline} \\ &= SocialMarginalCost_a^{gasoline} - PrivateAverageCost_a^{gasoline} \\ &= pfuel \cdot \left(\sum_{n=0}^{6} (-1)^{n+1} n \cdot c_n \cdot speed_a^{n+1}\right) \cdot 60 \cdot b_a \cdot c \cdot t_a^{free-flow} \cdot \left(\frac{\sum_f flow_{af}}{capacity_a}\right)^c \end{aligned} \tag{52}$$

5.6 Calibration

5.6.1 Calibration approach

All variables and parameters in the simulation model are either observed or calibrated to represent as realistic as possible a starting equilibrium point. Whenever possible, we set variables to observed values. Some parameters are set in a way such that the model-generated variables (such as non-work trips) match exactly the target values. Other parameters are set to match target elasticities. We draw on existing econometric work to set the target elasticities within well-accepted ranges. In the case that a parameter is neither observable nor does a functional relation exist between its value and a target value, we make reasonable and necessary assumptions.

5.6.2 Data

The 2000 Census Transportation Planning Package (CTPP) and Southern California Association of Governments (SCAG) data were used to estimate the model's commuting flows between zones. The US Census of Population and Housing was used to calibrate expenditure shares (between housing and goods), non-work income by residence location, and population at the census tract level. The census tract level data are then aggregated to the model zone level. Using the observed location choice patterns, we derive the location distributions (among 97 zones) of the population (11.8 million) and of jobs (6.6 million). As noted before, the residence-job location patterns are assumed to be fixed throughout the simulations.

5.6.3 The value and disutility of travel times

The value of time vot_f is an important parameter as it determines the magnitude of the *pecuniary* benefit of time savings. It is set exogenously to half of the hourly wage (see, for example, Abrantes and Wardman, 2011; Small, 2012) at the outside of the simulations. Another related parameter we calibrate is the coefficient of commute time disutility. Given the value of $\gamma_{ijf}(< 0)$ in the utility function (Equation 28), the consumer's marginal rate of substitution between average travel time and disposable income:

$$-\frac{\partial M_{ijf}}{\partial G_{ijf}} = \frac{60}{days \cdot hours} \frac{\sum\limits_{ij} N_{ijf} M_{ijf} \gamma_{ijf}}{\sum\limits_{ij} N_{ijf}} = vot_f \tag{53}$$

We calibrate the values for γ_{ijf} so that the weighted average marginal rate of substitution between travel time and income for each income group equals vot_f. [21]

5.6.4 Elasticity of substitution and the demand elasticity

The elasticity of substitution among goods sold at different zones is $\frac{1}{1-\sigma_f}$. The elasticity of demand for goods of course also depends on σ_f:

$$\frac{\partial \ln Z_{z|ijf}}{\partial \ln P_{z|if}} = -\frac{1}{1-\sigma_f} \left(1 - \sigma_f \frac{\iota_{z|if}^{\frac{1}{1-\sigma_f}} P_{z|if}^{\frac{\sigma_f}{\sigma_f - 1}}}{\sum_{z'}^{97} \iota_{z'|if}^{\frac{1}{1-\sigma_f}} P_{z'|if}^{\frac{\sigma_f}{\sigma_f - 1}}} \right) \tag{54}$$

[21] Since both M_{ijf} and G_{ijf} are endogenous variables in the model, we also experimented replacing the exogenous vot_f with the endogenous value of $\partial M_{ijf}/\partial G_{ijf}$, but the difference in the numerical results is negligible.

Note that the second term in parentheses approaches to zero when the number of zones is large. The demand elasticity with respect to own price is therefore approximately $-\frac{1}{1-\sigma_f}$. The two elasticities determined by σ_f are consequential since, firstly, the demand elasticity determines how sensitive consumption, hence non-work trips (and total trips) are to gross price changes, and, secondly, the elasticity of substitution affects how consumers adjust the shopping destination combinations as relative prices change. The demand elasticity has an impact on the extensive margin of VMT while the elasticity of substitution changes VMT on the intensive margin. Unfortunately, we do not know what is the value of σ_f but we experimented with different elasticities and as it turns out, the simulation results are robust with regard to σ_f due to the effects on the extensive and intensive margins cancelling out each other. We set $\sigma_f = 0.5$ so that the elasticity of substitution is 2 and the demand elasticity is approximately -2.

5.6.5 Mode choice elasticity

Besides the total number of trips and the shopping location patterns, another important margin of adjustment in VMT following a capacity shock is mode choice. The probability of choosing car was given by Equation (31). We can use it to derive the choice elasticity with respect to expected disutility of driving:[22]

$$\frac{\partial \ln PROB_{m=1|ijf}}{\partial \ln(\wp_{ij} + \wp_{ji})} = \Theta(\wp_{ij} + \wp_{ji})(1 - PROB_{m=1|ijf}) \tag{55}$$

Indra (2014) found that the left-hand side of (55) to be -0.1 in LA. Hence we calibrate the value of Θ so that the trip-weighted average elasticity is equal to -0.1.

5.6.6 Congestion parameters and arc capacities

Recall that the congested time on a segment of road is calculated from the BPR function (35). We observe the value of $time_a^{free-flow}$ and set b_a to the standard 0.15. More crucial is the value of parameter C which captures how rapidly traffic slows down with increasing flows. While it is common to set $C = 4$ in engineering studies, any value of C that is greater than one is consistent with the convexity of the private average cost of driving. We settled on $C = 1.2$ in order to match the model-generated average congestion index $\frac{time_a}{time_a^{free-flow}}$ with the Texas A&M Transportation Institute (TTI) congestion index for the LA region in 2000. The TTI index for

[22] See Train (2009).

LA during *peak hours* is 1.37. With $C = 1.2$, the model-generated *all-day* congestion index is 1.26.

Moreover, given the parameter values in the BPR function (35), arc capacities are calibrated as follows: Since we observe from CTPP and SCAG datasets zone-to-zone daily flows and travel times, we choose a vector of arc capacities to minimize the Weighted (by flow) Average Percentage Error between model-generated times and observed times. In simulations, building roads is realized by universally increasing the exogenous capacities of all arcs by a given percentage.

6 Simulation results

We present the simulation results in two parts. The first part explains the baseline simulation in which we simply increase the capacity for all arcs in the network by one percent. We examine the changes in the new equilibrium brought about by the capacity expansion as well as how and why welfare components change. The second part presents results of multiple simulations — runs with different initial VMT levels and are otherwise the same — with a one-percent capacity increase. These simulations are designed to simulate different MSAs with everything similar except initial non-work trips. One of the critical findings in the second part of this section is that the increase in VMT in response to road capacity expansion *decreases* with initial VMT. Put it differently, given the same percent increase in capacity, the increase in driving distance is smaller if the pre-expansion congestion level is higher. This qualitative conclusion is robust within reasonably wide ranges of several key elasticities.

6.1 Baseline simulation: the effect of capacity increase

We perturb the initial equilibrium by a one-percent road capacity increase. Table 14 reports key findings. The first thing we should notice is that aggregate VMT increases by 0.321 percent following the one-percent capacity expansion. Two points should be made clear about the elasticity before we go further. Firstly, 0.321 may seem to be on the lower side vis-ï¿œ-vis our regression estimates. This is because unlike in the IV-QR where there might be several unobserved characteristics that are correlated with lane miles, the simulation presented here allows for *only* road capacity to change while keeping all the other environment parameters constant. Also, the capacity increase is modeled as a widening of the existing roads, while ruling out the construction of new roads which would have made previously unreachable locations accessible to businesses and consumers alike. An important difference between improving existing roads and building new roads is that in the long run, new infrastructure induces firms and consumers to relocate to the newly accessible locations, and the associated production and consumption activities entail yet more driving. In other words, the shock in our simulations is a narrowly-defined capacity increase. Secondly, as we shall soon explore in more detail, the capacity increase gives rise to behavior adjustments in three margins that may pull VMT in different directions. The 0.321 percent increase in aggregate VMT is the *net* elasticity that takes into account, for example, the fact that per-trip distance shortened by -0.4 percent after the capacity shock.

TABLE 14 HERE

As road capacity expands, congestion improves, and travel times shorten. This expected outcome is also implied by Equation (35). The shortened travel times indicate a faster speed. The monetary travel costs go down marginally as gasoline consumption is reduced due to a slightly higher speed.[23] In the model, demand for VMT changes due to changes in three margins. The first is the extensive margin, or total trips by all modes. Other things (mode choice and route choice) being equal, more total trips would result in more miles traveled. This is the combined price and income effect caused by less expensive driving. The second is the intensive margin where the car share of trips changes. Given fixed total demand, consumers switching to cars would also cause VMT to rise. This is the mode choice effect. The third margin is embedded in the CES preference structure which results in consumers traveling to and shopping at different locations (Equation 24). Because the gross price (Equation 27) changes with the monetary travel costs, consumers adjust the quantity combination of goods sold at different locations, thereby altering the OD-patterns and VMT. This is the substitution effect.

From Table 14, we can see in detail how VMT changes due to adjustments in those three margins. A). Total trips by all modes. Because travel costs are lowered by improved infrastructure, both gross price (inclusive of travel costs) and net income (after commute costs) move in the direction that induces more consumption, which, in turn, generates more non-work trips.

[23] Although it is true in our calibrated model that a faster speed reduces gasoline use per mile, this is not always the case. Because of the U-shaped relationship between the TFI and speed, increased speed could also lead to more gasoline use per mile under a faster average speed.

However, since the across-modes average monetary cost changes only slightly (-0.1%), its effects on gross price and net income are small, and the number of total trips increases by merely 0.02 percent. B). Mode choice. Because the generalized travel cost becomes lower for driving, the share of car goes up 0.66 percent point at the cost of other modes. C). The substitution effect. Notice that while congestion is mitigated by increased capacity, both the average speed on local roads, which are dominantly used for short-distance within-zone trips, and the average per-trip distance decrease. This is due to consumers adjusting their consumption bundles in favor of more frequent trips to local merchants. Above, the VMT change described in A) represents the generative demand response while B) and C) are the distributive demand effect. It is important to note that even without the substitution effect, the qualitative relation between VMT and capacity presented in Table 14 would still prevail because both the price and income effects and the mode choice effect induce more trips and, given fixed per-trip distance, more VMT.

Next we turn to the welfare implications of increased road capacity. As explained before, the overall welfare change is calculated as the aggregate of three measures: the EV, the time delay externality, and the gasoline use externality. As it turns out, the magnitude of the gasoline externality is negligible, but it does not mean this is always the case in other settings. A different MSA, for instance, may have a lower average speed due to severe congestion, and the TFI function is more steeply and negatively sloped at a low speed. Alternatively, technological changes may even make gasoline use more sensitive to speed, which would increase the magnitude of the gasoline externality. Yet another possibility is that when gasoline price is high relative to income and VOT, the gain (loss) from reducing (increasing) gasoline use externality would carry a higher weight against changes in the utility level.

The EV, as shown in Equation (49), is also determined by three factors. The first two factors are already explained before, which are the price effect and the income effect. As transportation costs become lower, gross prices, too, become lower. For the same reason, the disposable income increases. The third factor affecting the EV is that the across-modes average travel time enters the utility function directly, therefore a shortened travel time is associated with a higher utility level. The EV is +$23.16, the time delay externality increases by 29 cents, and the gasoline externality decreases by less than 1 cent. The overall welfare change, which is the sum of the three, is +$22.87 per consumer per year. It is not surprising that an increase in capacity, which aims at mitigating congestion, would cause time delay externality to go up, although slightly. The reason is that even though per-trip time shortens, both car trips and VMT increase, dominating the intensive margin (shorter times) to exacerbate time delay externality. Lastly, a word of caution is in order. While it is clear how different components of the welfare calculation affect the net outcome, the magnitude of the benefit depends crucially on VOT, which we take as a parameter (50% of hourly wage) in the model. If the real VOT is higher (lower) than 50% of hourly wage rate, the benefit measured in dollar would also be greater (smaller). An reliable estimation of VOT would make the welfare calculation more dependable, and it is especially important when the cost of a proposed infrastructure project is known and a cost-benefit analysis is carried out.

6.2 Simulating VMT demand in MSAs with different initial non-work trips

So far in this section we explored the effect of capacity increase in a *single* MSA benchmark case. We described how and why VMT changes the way it does following a capacity shock. In the rest of the section we show how the adjustment in VMT caused by a capacity increase differs if initial VMT were to differ from the benchmark. We find that, all else being equal, the elasticity of VMT with respect to capacity *decreases* with initial VMT.

In calibrating the model, one of the conditions is that parameters are set so that the model-generated number of commute trips and non-work trips match the observed (from SCAG) values. To create different scenarios with different initial VMTs, we perturb the initial *non-work trips* up and down while keeping everything else unchanged in the calibration. By keeping the number and pattern of *commute trips* unchanged, we assume that the spatial pattern of residences and jobs remain unchanged. Essentially, we create different fixed points to serve as starting points with different levels of traffic due only to non-work trips.

TABLE 15 HERE

Table 15 shows key travel-related variables under initial equilibrium points with different levels of non-work trips, hence different initial VMTs. The initial number of non-work trips are perturbed by 20 percent increment (except that instead of reducing non-work trips to 0, we set it to 10% of benchmark to represent the minimum level). As the target non-work trips increase from the minimum level — 10 percent of benchmark — to 200 percent of benchmark, VMT expectedly increases. Because location patterns, incomes, VOT, gasoline price, and road characteristics are all kept unchanged in these scenarios, increasing VMT means heavier traffic load, and initial equilibrium travel times increase with initial VMTs. Time delay externality also increases considerably with congestion. Prolonged travel times then indicate a lower average speed, which causes fuel economy to drop and gasoline use to rise.

TABLE 16 HERE

Table 16 shows the effect of a one-percent capacity increase under different initial VMTs. The second column shows that the elasticity of VMT with respect capacity steadily decreases from 0.356 to 0.291 as initial VMT increases. That is, when we are keeping income, population, and location patterns constant, VMT becomes less sensitive to added roads as the urban area becomes more congested. Why is this the case?

Recall from the benchmark simulation that equilibrium VMT changes due to three reasons: changes in the total number of trips by all modes (generative), changes in mode choice probabilities, and changes due to shopping location substitutions (distributive).[24] In the columns labeled "trips" we can see that the elasticity of trips by all modes increases (from 0.004 to 0.019) with initial VMT. Because the share of car among all modes also goes up as a result of the newly added capacity, the number of car trips increase both because of the intensive margin (mode share) and the extensive margin (trips by all modes).

What may seem puzzling is that, on the one hand, the number of total trips and car trips becomes more elastic with respect to capacity as initial VMT goes up while VMT becomes less elastic on the other. The reason why this happens, as shown in the VMT/trip column in Table 16, is that as initial VMT increases, so does the initial congestion level. In a congested MSA, the capacity-induced changes in relative gross price (inclusive of travel costs) between local and further-away shopping locations is greater than that in a less congested MSA, leading to a more pronounced substitution of local trips for long-distance trips, shortening the average per-trip distance. In other words, when capacity is increased, as initial VMT becomes greater, the effect of consumers taking shorter trips dominates the effect of the fact that there are more car trips, resulting in the capacity elasticity of VMT decreases with initial VMTs. Put

[24] In reality, another important factor that determines VMT is the size and the spatial pattern of local labor markets. However, as we have done in the estimations where we control for population and the size of labor supply, we are keeping in the simulations here the size and the spatial pattern of the labor market undisturbed to single out the *short run* effect of varying initial traffic volume levels, which echos the unobserved heterogeneity in our regression analysis.

it differently, compared to locations where consumers initially drive fewer miles, consumers in more congested MSAs react to new roads by adding fewer *extra* VMTs than those who live in less congested MSAs. This is true even when residents in congested MSAs respond to new roads by making a greater number of trips than those residents in less congested MSAs.

To what extent does the above conclusion depend on the elasticity of substitution among goods sold at different locations? Indeed, we do not know what is the true value of σ_f (Equation 24), assuming the CES structure is a good approximation to the consumer's preference over goods/locations. We use $\sigma = 0.5$ in the simulations, which makes the elasticity of substitution among locations 2. One might ask that if σ takes a smaller or negative value — suggesting inelastic substitution among locations, would the VMT elasticity with respect to capacity still decline with initial VMT?

We did a sensitivity test in which we carried out the same simulations presented in this section, but under different values of σ. We find that the qualitative conclusion, i.e., capacity elasticity of VMT decreases with initial VMT, holds under a wide range of σ. As seen in the baseline simulation, building roads encourages more trips (by all modes) and a higher car share that has a postive effect on VMT. At the mean time, uniformly increasing road capacity also prompts substitutions which shortens average per-trip distance that has a negative effect on VMT. Therefore, whether or not the VMT response grows more elastic with respect to capacity as initial VMT increases depends on the relative magnitudes of the two opposing forces. When the elasticity of substitution is small, so is the price elasticity of demand for goods and travel. Consequently, As σ becomes smaller or more negative, it would simultaneously dampen the substitution effect that shortens average trip length and weakens the demand effect that induces more trips. The net effect, as shown by our simulations, is that the substitutions effect always numerically dominates the demand effect and that capacity elasticity of VMT decreasing with initial VMT would prevail under both high and low values of σ.

7 Conclusion

We investigate the implications of unobserved MSA heterogeneity on the existence of the fundamental law of highway congestion. Using a quantile regression framework, we show that due to unobserved heterogeneity MSA VMT reacts differently to an increase in highway capacity. We show that while building highways is unlikely to change the conditional and unconditional mean and median congestion levels in the U.S., it can lower the number of MSAs inflicted by extreme congestion. We propose a simple theory that shows that when facing a capacity shock, not only does the private average cost of driving falls, but different urban areas tend to have different demand responses. Such variation in demand responses causes the VMT elasticity to road capacity to be different across different MSAs. The theory also suggests that under mild conditions, MSAs with high initial congestion levels will have smaller VMT elasticity to road capacity than MSAs with low initial congestion. We further illustrate using a spatial general equilibrium model how heterogeneity in initial congestion levels drives variation in VMT elasticity to capacity, consistent with our empirical findings.

8 Appendix

8.1 The derivation of x_{Nk}

Cross effect. We derive the comparative statics for $x_{Nk} = x_{kN}$. One way to do it is to differentiate equation (7) with respect to k.

$$x_{Nk} = \frac{(C_n - P_n) \cdot \frac{\partial P_A A_N}{\partial k} - P_A A_k \cdot \frac{\partial (C_n - P_n)}{\partial k}}{(C_n - P_n)^2} \Rightarrow$$

$$x_{Nk} = \frac{(C_n - P_n) \cdot [A_n (P_{An} x_k + P_{AA} A_k) + P_A A_{nk}] - P_A A_n [(C_{nn} x_k + C_{nk}) - (P_{nn} x_k + P_{nA} A_k)]}{(C_n - P_n)^2} \tag{56}$$

The denominator of equation (56) is always positive in equilibrium. We will examine the terms in the numerator:

$$\begin{aligned} &\overbrace{(C_n - P_n)}^{+} \cdot \left[\overbrace{A_n}^{+} \left(\overbrace{P_{An}}^{-} \overbrace{x_k}^{+} + \overbrace{P_{AA}}^{-} \overbrace{A_k}^{+} \right) + \overbrace{P_A}^{+} \overbrace{A_{nk}}^{?} \right] \\ &- \overbrace{P_A A_n}^{+} \cdot \left[\left(\overbrace{C_{nn} x_k}^{+} + \overbrace{C_{nk}}^{-} \right) - \left(\overbrace{P_{nn}}^{+/0} \overbrace{x_k}^{+} + \overbrace{P_{nA}}^{-} \overbrace{A_k}^{+} \right) \right] \end{aligned} \tag{57}$$

First note that the first bracket in (57) captures how the magnitude of the shift of the demand curve caused by a change in N would respond to a change in k. The second bracket in equation (57) captures how the gap between the marginal cost (with respect to n) and the marginal willingness to pay would respond to a change in k.

Notice that if A_{nk} is not very big, which requires that the *marginal* amenity benefit of agglomeration is not very sensitive to road capacity changes, then the first bracket in equation (57) is negative. Moreover, if the marginal cost C_n is not very sensitive to capacity changes — so that C_{nk} is not very negative, or that the exponent in the congestion function is relatively large so that C_{nn} is large, then the bracket in the second term in (57) is positive and the numerator of equation (56) will be negative.

For example, one sufficient condition for equation (56) to be negative is that C_{nn} is sufficiently large, which is to say that congestion exacerbates rapidly as road use increases. Another sufficient condition is that agglomeration is important to amenities, so that A_n is sufficiently large. Alternatively, if road capacity is important to amenities, so that A_k is large. Yet another example of equation (56) taking on negative values is that the demand is linear ($P_{nn} = 0$), and $A_{nk} = 0 = C_{nk}$.

8.2 Instrumental Variable Quantile Regression Estimation

As Chernozhukov and Hansen (2006) mention, the estimation method for the IV-QR model can be viewed as the quantile regression analog of two stage least squares. The estimation of the IV-QR model involves solving the following optimization problem

$$\hat{\alpha}(\tau) = \ arg \ \underset{\alpha \epsilon \mathscr{A}}{inf} \ ||\hat{\gamma}(\alpha, \tau)||_{A(\tau)} \tag{58}$$

where

$$\left(\hat{\beta}(\alpha,\tau), \hat{\gamma}(\alpha,\tau)\right) = \ arg \ \underset{(\beta,\gamma)\epsilon\mathscr{B}\times\mathscr{G}}{inf} Q_n(\tau,\alpha,\beta,\gamma) \tag{59}$$

and

$$Q_n(\tau,\alpha,\beta,\gamma) \equiv \frac{1}{n}\sum_{i=1}^{n} \rho_\tau \left(Y_i - D_i'\alpha - X_i'\beta - \hat{\Phi}_i(\tau)'\gamma\right) \cdot \hat{V}_i(\tau) \tag{60}$$

Above, $||x||_A = \sqrt{x'Ax}$, where $A(\tau)$ is any uniformly positive definite matrix, $\hat{\Phi}_i(\tau) \equiv \hat{\Phi}(\tau, X_i, Z_i)$

is a $\alpha \times 1$ vector of instruments, and $\hat{V}_i(\tau) \equiv \hat{V}(\tau, X_i, Z_i)$ is a positive weight function. The estimation problem is solved in two-steps.

1. For a given value of τ which lies in the interval $(0, 1)$, define a grid of values for α, say, $\{\alpha_j, j = 1, ..., J\}$. For every value of α in the grid run the τ-quantile regression of $Y_i - D_i'\alpha_j$ on the controls$X_i$and the instruments $\hat{\Phi}_i(\tau)$ to get estimates of $\hat{\beta}(\alpha_j, \tau)$ and $\hat{\gamma}(\alpha_j, \tau)$.

2. Pick the value of α from the grid, call $\hat{\alpha}(\tau)$, that yields the smallest value of $||\hat{\gamma}(\alpha_j, \tau)||$. The value of $\hat{\beta}(\tau)$ is then obtained from $\hat{\beta}(\hat{\alpha}(\tau), \tau)$.

8.3 Generalized Quantile Regression Estimation

In this section we describe the steps involved in estimating the GQR model parameters (see Powell, Forthcoming). As the GQR model nests the IV-QR model, the estimation strategy in this section also applies for estimating the conditional treatment effect in Equation (15). The GQR model estimation strategy is to use a GMM approach–the assumptions of the GQR model imply a set of two moment conditions shown in Equation (16), while the assumptions of the IV-QR model imply a single moment condition shown in Equation (17). While all the model parameters in the GQR model can be estimated in a single step, Powell (Forthcoming) proceeds with a three step approach for computational ease.

$$\begin{aligned} E\left\{Z_i\left[1\left(Y_i \leq \gamma + \alpha(\tau) D_i\right) - F\left(X_i'\beta(\tau)\right)\right]\right\} = 0 \quad (GQR) \quad (a) \\ E\left[1\left(Y_i \leq \gamma + \alpha(\tau) D_i\right) - \tau\right] = 0 \qquad\qquad (GQR) \quad (b) \end{aligned} \tag{61}$$

$$E\left\{Z_i\left[1\left(Y_i \leq \gamma + \alpha(\tau) D_i + X_i'\beta(\tau)\right) - \tau\right]\right\} = 0 \quad (IV - QR) \tag{62}$$

Step I: The set Θ defines the combinations of α and γ such that the second GQR moment condition holds for given τ. Powell (Forthcoming) notes that for any given values of α and τ one can find a value $\hat{\gamma}(\alpha, \tau)$ to satisfy Equation (17).

$$\Theta \equiv \left\{(\alpha, \gamma) \,|\tau - \frac{1}{N} < \frac{1}{N}\sum_{i=1}^{N} 1\left(Y_i \leq \gamma + \alpha D_i\right) \leq \tau\right\} \tag{63}$$

$$Set \quad \hat{\gamma}(\alpha, \tau) \quad such \quad that \quad \tau - \tfrac{1}{N} < \tfrac{1}{N}\textstyle\sum_{i=1}^{N} 1\left(Y_i - \alpha D_i \leq \gamma\right) \leq \tau \tag{64}$$

Step II: Given α, γ, τ such that $(\alpha, \gamma) \in \Theta$, the parameter vector β is estimated using maximum likelihood where the function F is assumed to be either a logit or probit.

$$\hat{\beta}(\alpha, \gamma, \tau) = \underset{\beta}{max} \sum_{i=1}^{N} 1 (Y_i \leq \gamma + \alpha D_i) \, lnF \left(X_i' \beta\right) + 1 (Y_i > \gamma + \alpha D_i) \, ln \left(1 - F \left(X_i' \beta\right)\right) \tag{65}$$

Step III: In the final step given τ the parameters α and γ are estimated by minimizing the weighted distance of the vector of moments $\widehat{g}$ so that Equation (16) holds as closely as possible.

$$g_i (\alpha, \gamma, \beta (\alpha, \gamma, \tau)) = Z_i \left[1 (Y_i \leq \gamma + \alpha D_i) - F \left(X_i' \beta (\alpha, \gamma, \tau)\right)\right] \tag{66}$$

$$\widehat{g} (\alpha, \gamma, \beta (\alpha, \gamma, \tau)) = \frac{1}{N} \sum_{i}^{N} g_i (\alpha, \gamma, \beta (\alpha, \gamma, \tau)) \tag{67}$$

$$\hat{\alpha}(\tau), \hat{\gamma(\tau)} \quad = \quad \underset{\alpha,\gamma}{min} \widehat{g} (\alpha, \gamma, \beta (\alpha, \gamma, \tau))' \hat{A} \widehat{g} (\alpha, \gamma, \beta (\alpha, \gamma, \tau)) \tag{68}$$

References

Abrantes, Pedro A.L. and Mark R. Wardman, "Meta-analysis of UK values of travel time: An update," *Transportation Research Part A: Policy and Practice*, 2011, *45* (1), 1 – 17.

Anas, Alex, "Why are urban travel times so stable?," *Journal of Regional Science*, 2015, *55* (2), 230–261.

_ , "The cost of congestion and the benefits of congestion pricing: A general equilibrium analysis," *Transportation Research Part B: Methodological*, 2020, *136*, 110 – 137.

Autor, David H., Susan N. Houseman, and Sari Pekkala Kerr, "The Effect of Work First Job Placements on the Distribution of Earnings: An Instrumental Variable Quantile Regression Approach," *Journal of Labor Economics*, 2017, *35* (1), 149–190.

Bai, Jushan, "Panel data models with interactive fixed effects," *Econometrica*, July 2009, *77* (4), 1229–1279.

Baillon, J-B and Roberto Cominetti, "Markovian traffic equilibrium," *Mathematical Programming*, 2008, *111* (1-2), 33–56.

Baum-Snow, Nathaniel, "Did Highways Cause Suburbanization?," *The Quarterly Journal of Economics*, May 2007, *122* (2), 775 – 805.

_ , "Changes in Transportation Infrastructure and Commuting Patterns in US Metropolitan Areas, 1960-2000," *American Economic Review*, May 2010, *100* (2), 378–82.

Bureau of Public Roads, *Traffic Assignment Manual*, U.S. Department of Commerce, Urban Planning Division, Washington D.C., 1964.

Cervero, Robert, "Induced travel demand: research design, empirical evidence, and normative policies," *Journal of Planning Literature*, 2002, *17*, 2–20.

Chen, Yu-Chang, "Instrumental variable quantile regression in R," Technical Report 2019.

Chernozhukov, Victor and Christian Hansen, "An IV model of quantile treatment effects," *Econometrica*, January 2005, *73* (1), 245–261.

— **and** — , "Instrumental quantile regression inference for structural and treatment effect models," *Journal of Econometrics*, 2006, *132* (2), 491 – 525.

— **and** — , "Instrumental variable quantile regression: A robust inference approach," *Journal of Econometrics*, 2008, *142* (1), 379–398.

Davis, Stacy C., SW Diegel, and RG Boundy, *Transportation Energy Data Book, Oak Ridge National Laboratory*, Oak Ridge, Tennessee (ORNL-6984), 2009.

Dong, Yingying and Shu Shen, "Testing for rank invariance or similarity in program evaluation," *Review of Economics and Statistics*, 2018, *100* (1), 78–85.

Downs, Anthony, "The law of peakhour expressway congestion," *Traffic Quarterly*, 1962, *16*, 393 – 409.

Duranton, Gilles and Matthew A. Turner, "The fundamental law of road congestion: Evidence from US cities," *American Economic Review*, 2011, *101* (6), 2616–52.

— **and** — , "Urban Growth and Transportation," *The Review of Economic Studies*, 03 2012, *79* (4), 1407–1440.

Freyberger, Joachim, "Non-parametric panel data models with interactive fixed effects," *Review of Economic Studies*, 2018, *85*, 1824–1851.

Gibbons, Stephen, Teemu Lyytikäinen, Henry G. Overman, and Rosa Sanchis-Guarner, "New road infrastructure: The effects on firms," *Journal of Urban Economics*, 2019, *110*, 35 – 50.

Goodwin, Phil B., "Empirical evidence on induced traffic," *Transportation*, 1996, *23* (35-54).

Gutierrez–i–Puigarnau, Eva and Jos N. van Ommeren, "Labour supply and commuting," *Journal of Urban Economics*, 2010, *68* (1), 82–89.

Heckman, James J., "Micro data, heterogeneity, and the evaluation of public policy: Nobel lecture," *Journal of political Economy*, 2001, *109* (4), 673–748.

— **and Richard Robb**, "Alternative methods for solving the problem of selection bias in evaluating the impact of treatments on outcomes," in Howard Wainer, ed., *Drawing Inference from Self-Selected Samples*, Springer, New York 1986, pp. 63–107.

Hoderlein, Stefan and Enno Mammen, "Identification of marginal effects in nonseparable models without monotonicity," *Econometrica*, 2007, *75* (5), 1513 – 1518.

_ **, Hajo Holzmann, and Alexander Meister**, "The triangular model with random coefficients," *Journal of Econometrics*, 2017, *201*, 144–169.

Hsu, Wen-Tai and Hongliang Zhang, "The fundamental law of highway congestion revisited: Evidence from national expressways in Japan," *Journal of Urban Economics*, 2014, *81*, 65–76.

Imbens, Guido W. and Joshua Angrist, "Identification and estimation of local average treatment effects," *Econometrica*, 1994, *62* (2), 467–475.

Indra, Debarshi, *Essays in applied econometrics*, State University of New York at Buffalo, 2014.

Koenker, Roger, "Quantile regression for longitudinal data," *Journal of Multivariate Analysis*, 2004, *91*, 74–89.

_ **and Gilbert Bassett**, "Regression Quantiles," *Econometrica*, January 1978, *46* (1), 33–50.

Li, Wenwen, Richard L. Church, and Michael F. Goodchild, "The p-compact-regions problem," *Geographical Analysis*, 2014, *46* (3), 250–273.

Masten, Matthew A. and Alexander Torgovitsky, "Identification of instrumental variable correlated random coefficients model," *The Review of Economics and Statistics*, 2016, *98* (5), 1001–1005.

Matzkin, Rosa L., "Nonparametric estimation of nonadditive random functions," *Econometrica*, 2003, *71* (5), 1339–1375.

_ , "Nonparametric Identification," in James J. Heckman and Edward E. Leamer, eds., *Handbook of Econometrics*, Vol. 6, Elsevier Science, 2007, pp. 5307–5368.

Moon, Hyungsik Roger, Matthew Shum, and Martin Weidner, "Estimation of random coefficients logit demand models with interactive fixed effects," *Journal of Econometrics*, October 2018, *206* (2), 613–644.

Powell, David, "Quantile treatment effects in the presence of covariates," *The Review of Economics and Statistics*, Forthcoming.

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

_ **and Erik T. Verhoef**, *The economics of urban transportation*, New York: Routledge, 2007.

Train, Kenneth E., *Discrete choice methods with simulation*, Cambridge university press, 2009.

Wooldridge, Jeffrey M., *Econometric analysis of cross section and panel data*, second ed., MIT Press, 2010.

Figures and tables

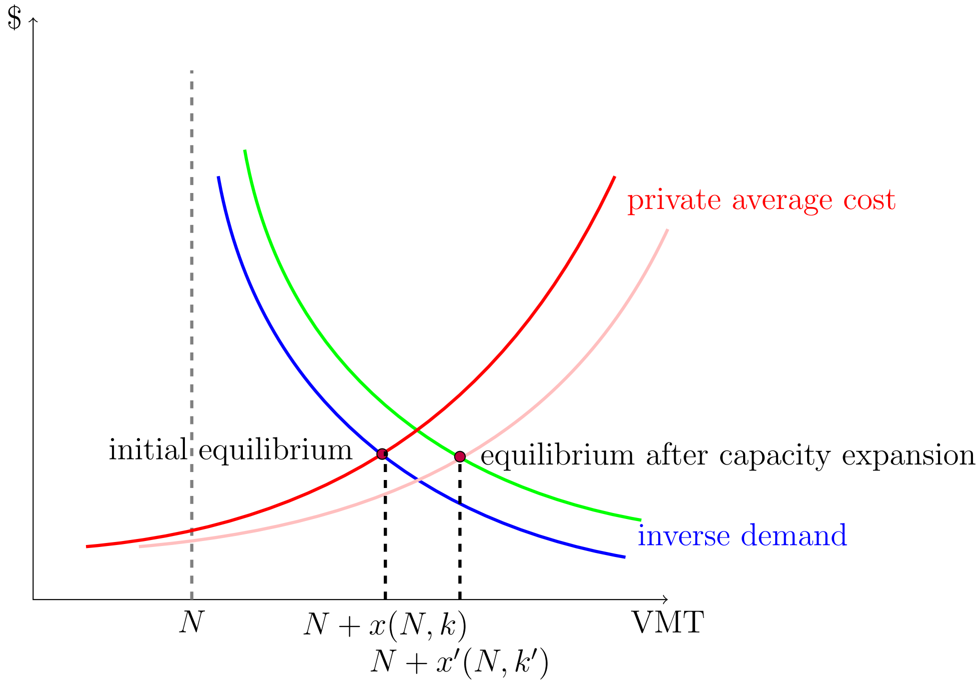

Figure 1: Equilibrium VMT under k and k'

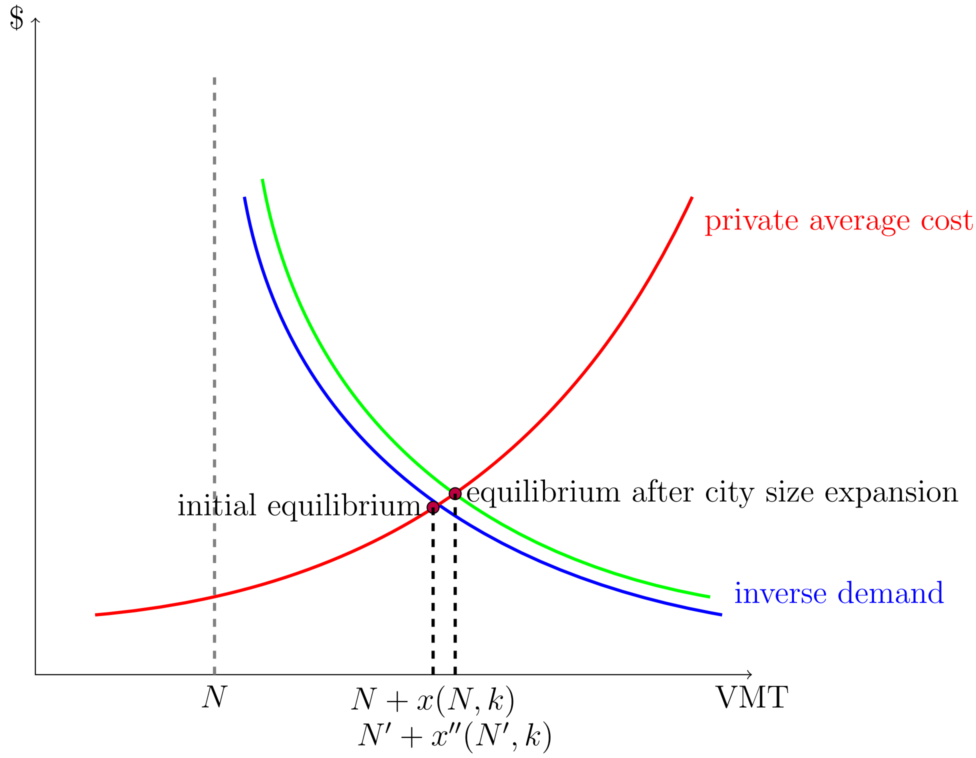

Figure 2: Equilibrium VMTs under N and N'

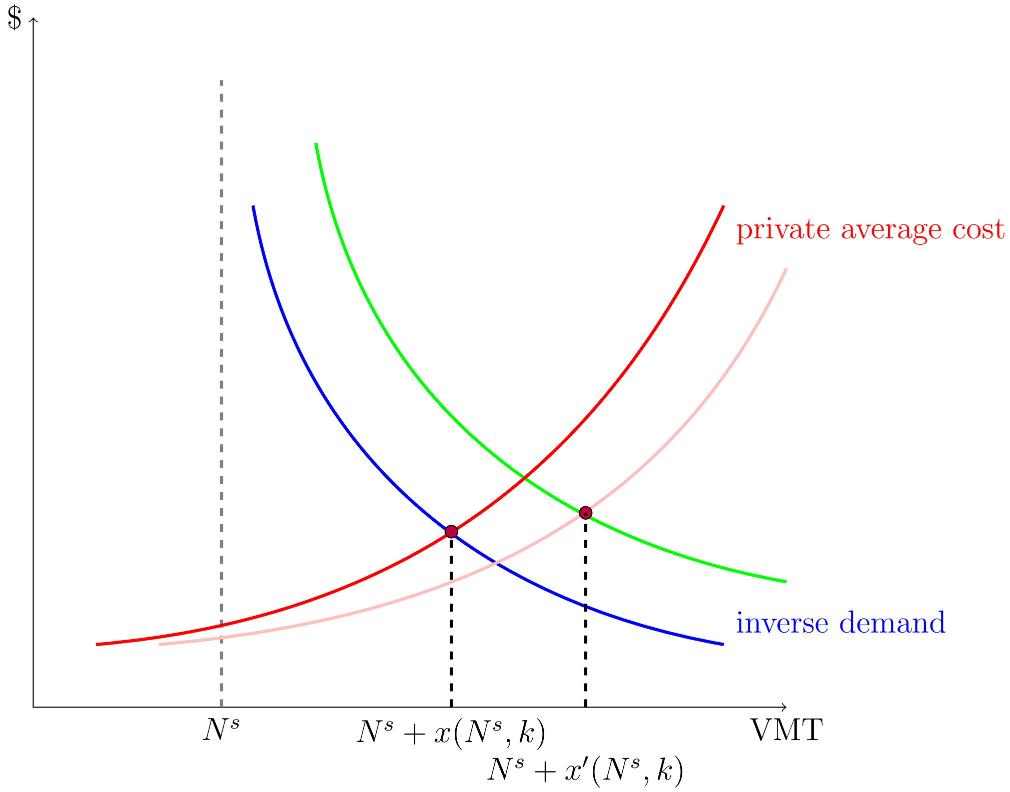

Figure 3: capacity expansion in a small city

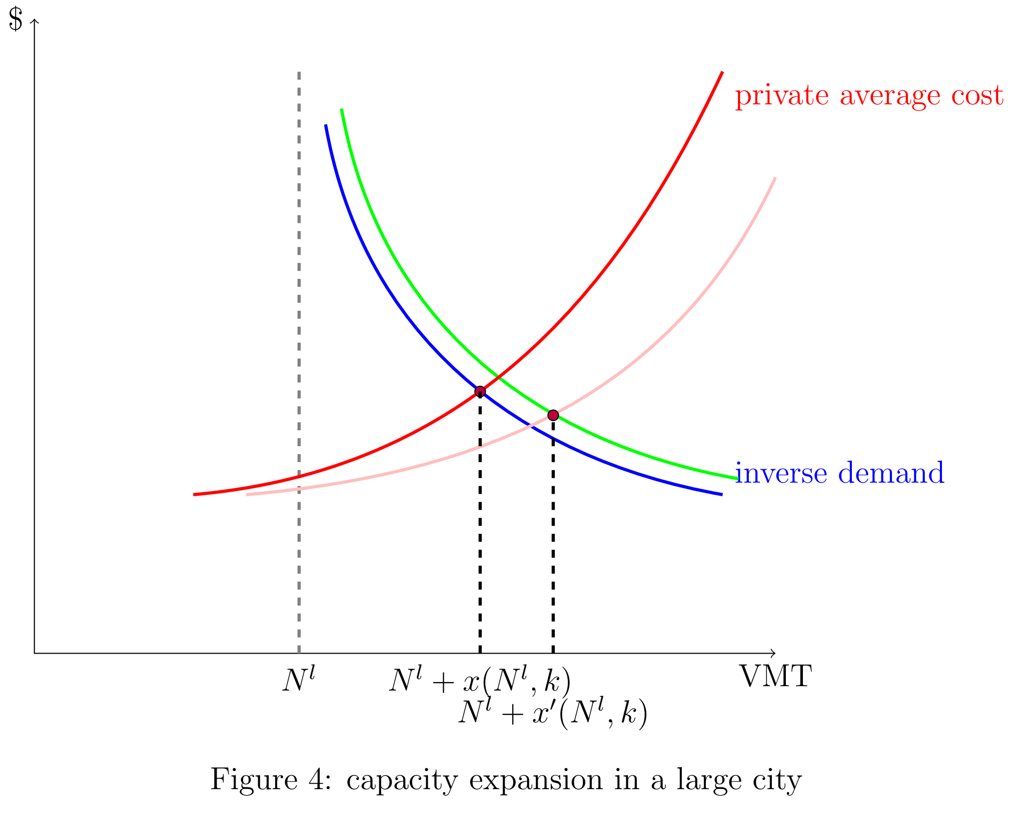

Figure 4: capacity expansion in a large city

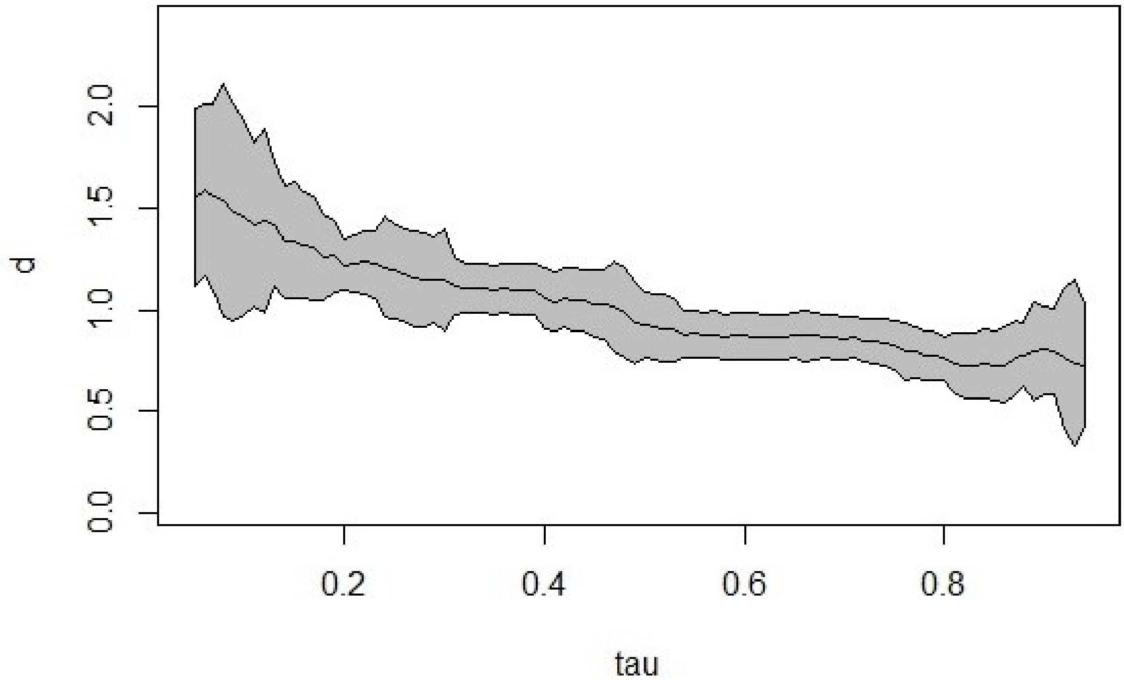

Figure 5: IV-QR estimates of elasticity of VMT to stock of MSA interstate highways stock at different quantiles

Note: Alpha is the elasticity estimate of VMT to IH lane miles in MSA based on instrumental variables quantile regression (IV-QR). Tau indexes the quantile. Dependent variable is ln(VMT) for interstate highways, entire MSAs. Treatment variable is ln(IH lane mile). Controls include ln(population), year dummy, geographical features, census division dummies. Instruments include ln(1835 exploration routes), ln(1898 railroads), and ln(1947 planned interstates). Number of observations = 684 covering 1983, 1993, and 2003. The shaded area indicates the 95% confidence interval. The horizontal dashed line indicates the 2SLS estimate equal to 1.03.

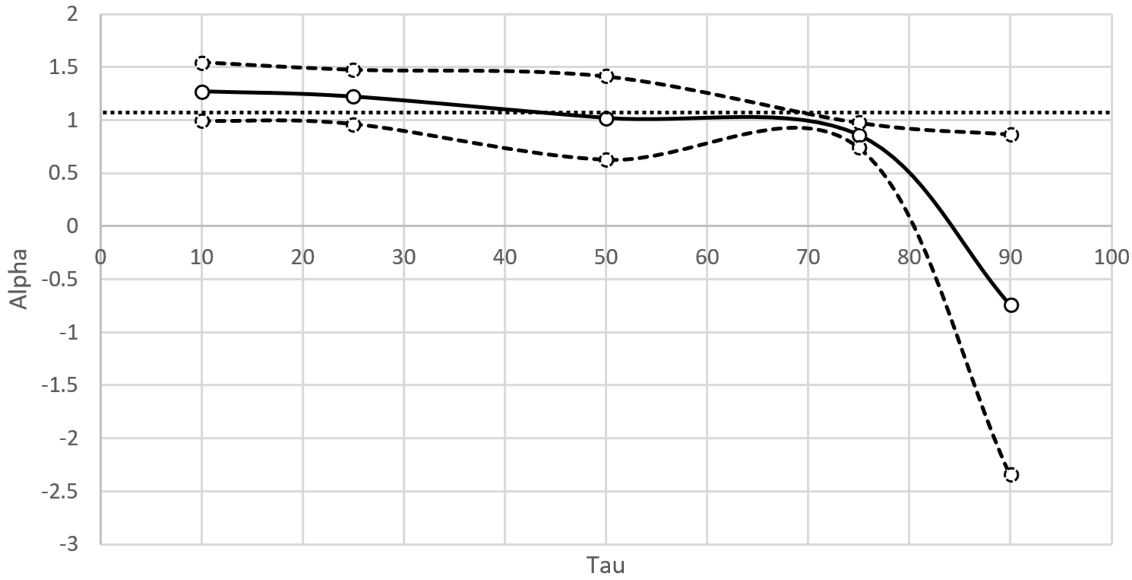

Figure 6: GQR estimates of elasticity of VMT to stock of MSA interstate highways at different quantiles

Note: Alpha is the elasticity estimate of VMT to IH lane miles in MSA based on generalized quantile regression (GQR). Tau indexes the quantile. Dependent variable is ln(VMT) for interstate highways, entire MSAs. Treatment variable is ln(IH lane mile). Controls include ln(population), year dummy, geographical features, census division dummies. Instruments include ln 1835 exploration routes, ln 1898 railroads, and ln 1947 planned interstates. Number of observations = 684 covering 1983, 1993, and 2003. The dashed lines indicate the 95% confidence interval around the point estimate given by the solid black line. The horizontal dotted line indicates the 2SLS estimate equal to 1.03.

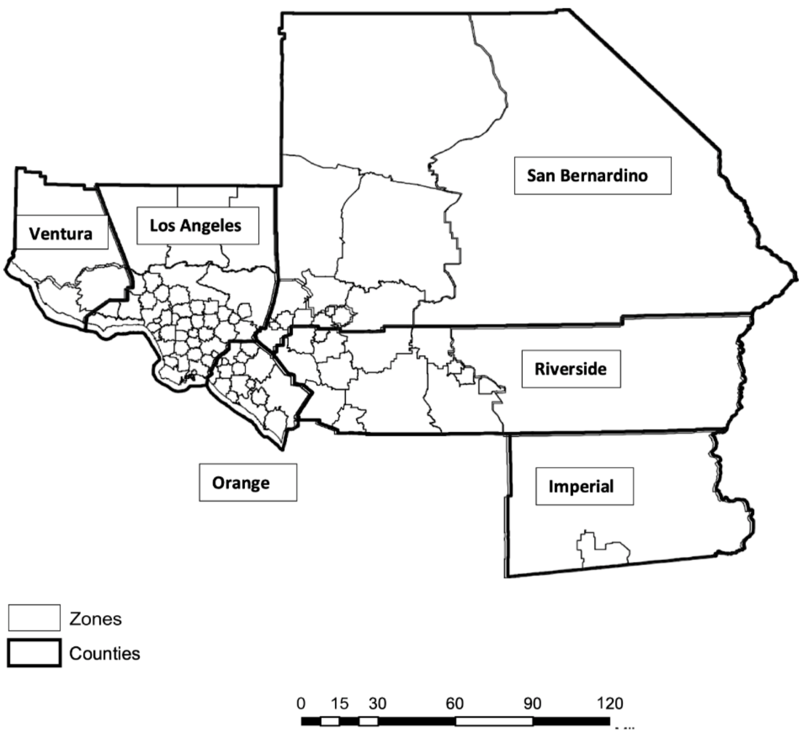

Figure 7: Counties and model zones in LA TRAN

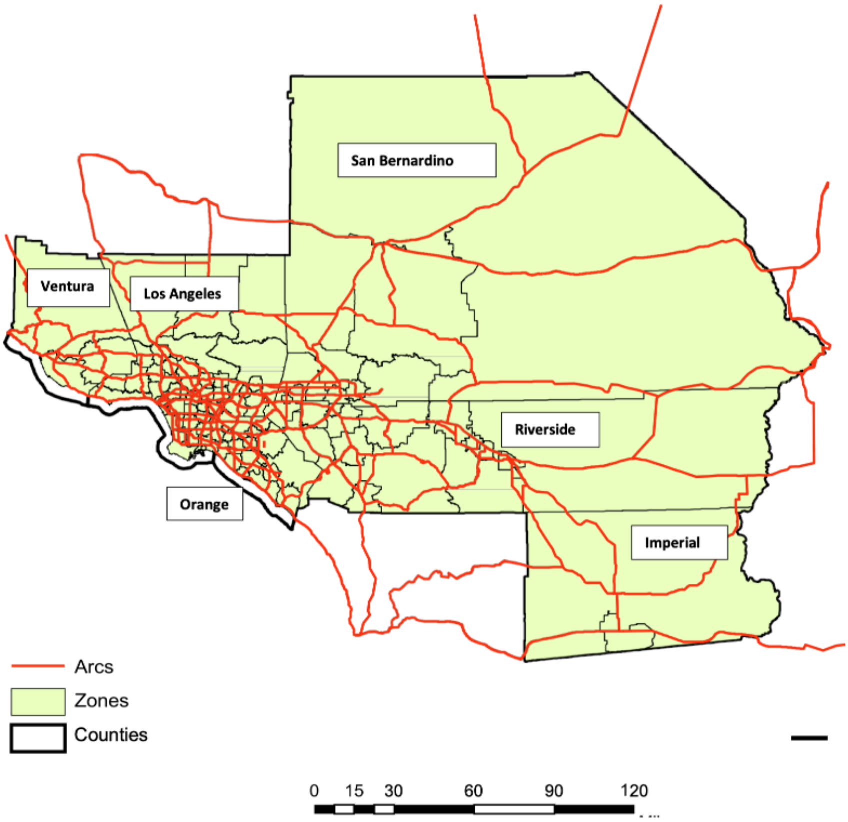

Figure 8: Arcs in the network

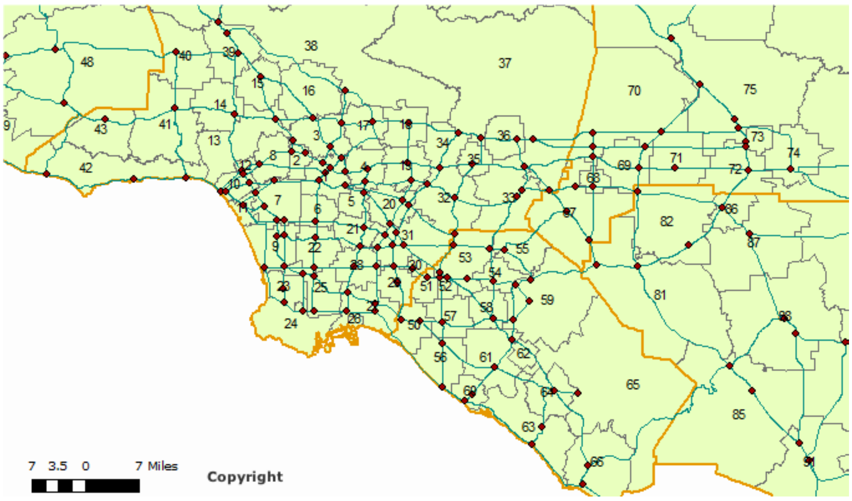

Figure 9: Nodes and arcs in the population centers

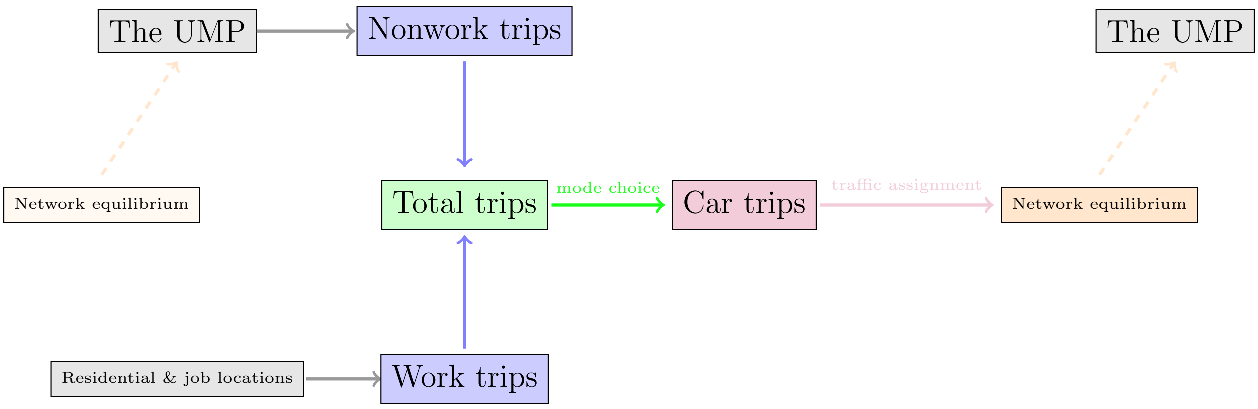

Figure 10: LA TRAN algorithm

Panel A: Univariate Quantile Regression by Decade; Panel B: Univariate OLS by Decade				
Dependent variable: ln(VMT) for interstate highways, entire MSAs				
Year:		1983	1993	2003
		(1)	(2)	(3)
Panel A	Quantiles			
ln(IH Lane mile)	10%	1.25***	1.37***	1.42***
		(0.14)	(0.12)	(0.06)
	25%	1.39***	1.34***	1.28***
		(0.04)	(0.04)	(0.03)
	50%	1.29***	1.25***	1.22***
		(0.03)	(0.02)	(0.02)
	75%	1.24***	1.22***	1.20***
		(0.03)	(0.02)	(0.02)
	90%	1.16***	1.22***	1.18***
		(0.05)	(0.04)	(0.04)
Panel B				
ln(IH Lane mile)		1.24***	1.25***	1.23***
		(0.04)	(0.02)	(0.02)
N			228	

Table 1: Effect of lane miles on VMT

Notes: Robust standard errors in parentheses.
*,**,*** Significant at 10% level, 5% level, and 1% level respectively.
Quantile regression results obtained using "qreg" command in Stata 15.

Panel A: Quantile Regression by Decade; Panel B: OLS by Decade										
Dependent variable: ln(VMT) for interstate highways, entire MSAs										
		1983	1983	1983	1993	1993	1993	2003	2003	2003
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A	Quantiles									
ln(IH Lane mile)	10%	0.96***	1.08***	1.00***	0.78***	0.82***	0.89***	0.75***	0.87***	0.81***
		(0.05)	(0.06)	(0.09)	(0.10)	(0.06)	(0.77)	(0.10)	(0.06)	(0.07)
	25%	0.90***	1.03***	1.04***	0.71***	0.88***	0.88***	0.68***	0.84***	0.82***
		(0.06)	(0.05)	(0.04)	(0.06)	(0.06)	(0.07)	(0.07)	(0.06)	(0.05)
	50%	0.87***	0.92***	0.90***	0.72***	0.76***	0.76***	0.71***	0.78***	0.73***
		(0.04)	(0.05)	(0.06)	(0.05)	(0.04)	(0.03)	(0.05)	(0.04)	(0.05)
	75%	0.87***	0.78***	0.80***	0.77***	0.79***	0.81***	0.80***	0.75***	0.74***
		(0.05)	(0.06)	(0.06)	(0.05)	(0.04)	(0.04)	(0.04)	(0.03)	(0.04)
	90%	0.74***	0.76***	0.72***	0.75***	0.72***	0.68***	0.76***	0.73***	0.71***
		(0.07)	(0.01)	(0.04)	(0.05)	(0.85)	(0.04)	(0.03)	(0.04)	(0.08)
ln(population)		Y	Y	Y	Y	Y	Y	Y	Y	Y
Geography			Y	Y		Y	Y		Y	Y
Census Divisions			Y	Y		Y	Y		Y	Y
Past populations				Y			Y			Y
Socioeconomic characteristics				Y			Y			Y
Panel B										
ln(IH Lane mile)		0.92***	0.94***	0.92***	0.73***	0.76***	0.77***	0.71***	0.75***	0.76***
		(0.06)	(0.06)	(0.05)	(0.05)	(0.04)	(0.04)	(0.05)	(0.04)	(0.04)
N						228				

Table 2: Effect of lane miles on VMT

Notes: Robust standard errors in parentheses.
*,**,*** Significant at 10% level, 5% level, and 1% level respectively.
Quantile regression results obtained using "qreg" command in Stata 15.

Panel A: Pooled Quantile Regression; Panel B: Pooled OLS										
Dependent Variable: ln(VMT) for interstate highways, entire MSAs										
MSA Sample		All	All	All	All	All	All	All	Big	Small
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A	Quantiles									
ln(IH lane miles)	10%	1.40***	0.84***	0.99***	0.96***	1.02***	1.01***	1.01***	0.99***	1.03***
		(0.06)	(0.06)	(0.02)	(0.03)	(0.04)	(0.05)	(0.06)	(0.01)	(0.05)
	25%	1.34***	0.81***	0.91***	0.88***	1.02***	1.01***	1.01***	0.99***	1.03***
		(0.02)	(0.03)	(0.03)	(0.03)	(0.07)	(0.11)	(0.09)	(0.01)	(0.08)
	50%	1.25***	0.79***	0.82***	0.82***	0.99***	0.98***	0.99***	1.01***	1.01***
		(0.02)	(0.03)	(0.03)	(0.03)	(0.01)	(0.01)	(0.01)	(0.01)	(0.11)
	75%	1.21***	0.79***	0.79***	0.80***	1.01***	1.01***	0.98***	0.99***	0.98***
		(0.01)	(0.02)	(0.03)	(0.03)	(0.01)	(0.01)	(0.11)	(0.07)	(0.11)
	90%	1.19***	0.74***	0.72***	0.72***	1.01***	1.01***	0.98***	0.99***	0.98***
		(0.02)	(0.03)	(0.01)	(0.03)	0.00	0.00	(0.04)	(0.10)	(0.04)
ln(population)			Y	Y	Y		Y	Y	Y	Y
Year dummy		Y	Y	Y	Y	Y	Y	Y	Y	Y
Geography				Y	Y					
Census divisions				Y	Y					
Socioeconomic characteristics					Y			Y		
Past populations					Y					
MSA fixed effects						Y	Y	Y	Y	Y
Panel B										
ln(IH lane miles)		1.24***	0.82***	0.86***	0.85***	1.05***	1.06***	1.05***	1.05***	1.12***
		(0.02)	(0.05)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.08)
N		684	684	684	684	684	684	684	342	342
Number of MSA		228	228	228	228	228	228	228	114	114

Table 3: Effect of lane miles on VMT

Notes: Heteroskedasticity robust standard errors are given in all the columns. Standard errors have also been clustered around MSAs in columns 1,2,3, and 4. In columns 8 and 9, sub-samples are taken out of the entire sample above and below the median population size in 190 respectively.
*,**,*** Significant at 10% level, 5% level, and 1% level respectively.
Quantile regression results in columns 1, 2, 3, and 4 are obtained using "qreg" command in Stata 15.
Quantile regression results in columns 5, 6, 7, 8, and 9 are obtained using "qreg2" command in Stata 15.

Panel A: Instrumental Variables Pooled Quantile Regression; Panel B: Instrumental Variables (TSLS) Pooled Regression						
Dependent variable: ln(VMT) for interstate highways, entire MSAs						
Instruments: ln 1835 exploration routes, ln 1898 railroads, and ln 1947 planned interstates						
		(1)	(2)	(3)	(4)	(5)
Panel A	Quantiles					
	10%	1.48***	1.04***	1.45***	1.44***	1.35***
		(0.03)	(0.07)	(0.24)	(0.16)	(0.15)
	25%	1.39***	1.01***	1.19***	1.21***	1.19***
		(0.02)	(0.07)	(0.12)	(0.09)	(0.08)
ln(IH Lane mile)	50%	1.27***	0.86***	0.92***	0.91***	1.02***
		(0.02)	(0.08)	(0.08)	(0.08)	(0.09)
	75%	1.23***	0.77***	0.82***	0.83***	0.83***
		(0.02)	(0.07)	(0.06)	(0.06)	(0.08)
	90%	1.14***	0.67***	0.80***	0.72***	0.65***
		(0.02)	(0.08)	(0.11)	(0.15)	(0.24)
ln(population)			Y	Y	Y	Y
Year dummy		Y	Y	Y	Y	Y
Geography				Y	Y	Y
Census Divisions				Y	Y	Y
Socioeconomic characteristics					Y	Y
Past populations						Y
		Kolmogorov-Smirnov statistic				
Null hypothesis:						
No Effect	$\alpha(\tau) = 0$	72.74^a	15.39^a	18.28^a	57.05^a	15.85^a
Constant effect	$\alpha(\tau) = \alpha$	3.62^a	2.49^c	4.06^a	6.07^a	3.72^a
Exogeneity	$\alpha(\tau) = \alpha_{QR}$	6.47^a	3.20^b	3.70^b	6.27^a	4.31^a
Panel B						
ln(IH Lane mile)		1.32***	0.92***	1.03***	1.01***	1.04***
		(0.03)	(0.06)	(0.07)	(0.08)	(0.08)
N				684		

Table 4: Effect of lane miles on VMT

Notes: Robust standard errors in parentheses. ln(IH Lane mile) is the endogenous variable and is being instrumented.
*,**,*** Significant at 10% level, 5% level, and 1% level respectively.
Quantile regression results are obtained using the R library IV-QR.
a) Kolmogorov-Smirnov statistic > 95% critical value.
b) 90% critical value < Kolmogorov-Smirnov statistic < 95% critical value.
c) Kolmogorov-Smirnov statistic < 90% critical value.

Panel A: Instrumental Variables Pooled Quantile Regression; Panel B: Instrumental Variables (TSLS) Pooled Regression)						
Dependent variable: ln(VMT) for interstate highways, entire MSAs						
Instruments: ln 1947 planned interstates						
		(1)	(2)	(3)	(4)	(5)
Panel A	Quantiles					
ln(IH Lane mile)	10%	1.47***	1.04***	1.43***	1.49***	1.43***
		(0.04)	(0.08)	(0.29)	(0.24)	(0.15)
	25%	1.41***	1.05***	1.23***	1.21***	1.19***
		(0.02)	(0.08)	(0.06)	(0.12)	(0.09)
	50%	1.29***	1.07***	0.96***	0.95***	1.03***
		(0.02)	(0.08)	(0.09)	(0.07)	(0.08)
	75%	1.26***	0.86***	0.87***	0.85***	0.88***
		(0.03)	(0.09)	(0.08)	(0.06)	(0.11)
	90%	1.24***	0.85***	0.89***	0.91***	0.89***
		(0.04)	(0.13)	(0.21)	(0.18)	(0.22)
ln(population)			Y	Y	Y	Y
Year Dummy		Y	Y	Y	Y	Y
Geography				Y	Y	Y
Census Divisions				Y	Y	Y
Socioeconomic characteristics					Y	Y
Past populations						Y
		Kolmogorov-Smirnov statistic				
Null hypothesis:						
No Effect	$\alpha(\tau) = 0$	69.11^a	14.38^a	18.64^a	22.52^a	18.30^a
Constant effect	$\alpha(\tau) = \alpha$	4.83^a	3.14^b	4.25^a	7.53^a	4.07^a
Exogeneity	$\alpha(\tau) = \alpha_{QR}$	3.10^b	3.67^a	4.39^a	6.95^a	5.21^a
Panel B						
ln(IH Lane mile)		1.33***	1.00***	1.10***	1.08***	1.12***
		(0.05)	(0.11)	(0.13)	(0.13)	(0.15)
N				684		

Table 5: Effect of lane miles on VMT

Notes: Robust standard errors in parentheses. ln(IH Lane mile) is the endogenous variable and is being instrumented.
*,**,*** Significant at 10% level, 5% level, and 1% level respectively.
Quantile regression results are obtained using the R library IV-QR.
a) Kolmogorov-Smirnov statistic > 95% critical value.
b) 90% critical value < Kolmogorov-Smirnov statistic < 95% critical value.
c) Kolmogorov-Smirnov statistic < 90% critical value.

Panel A: Instrumental Variables Pooled Quantile Regression; Panel B: Instrumental Variables (TSLS) Pooled Regression						
Dependent variable: ln(VMT) for interstate highways, entire MSAs						
Instruments: ln 1898 railroads						
		(1)	(2)	(3)	(4)	(5)
Panel A	Quantiles					
	10%	1.49***	0.99***	1.36***	1.37***	1.34***
		(0.03)	(0.07)	(0.16)	(0.18)	(0.23)
	25%	1.39***	0.89***	1.30***	1.28***	1.18***
		(0.02)	(0.10)	(0.16)	(0.13)	(0.12)
ln(IH Lane mile)	50%	1.27***	0.77***	0.85***	0.83***	0.86***
		(0.02)	(0.09)	(0.07)	(0.13)	(0.14)
	75%	1.19***	0.57***	0.86***	0.85***	0.83***
		(0.02)	(0.07)	(0.08)	(0.09)	(0.12)
	90%	1.13***	0.59***	0.80***	0.74***	0.56*
		(0.02)	(0.06)	(0.14)	(0.14)	(0.28)
ln(population)			Y	Y	Y	Y
Year Dummy		Y	Y	Y	Y	Y
Geography				Y	Y	Y
Census Divisions				Y	Y	Y
Socioeconomic characteristics					Y	Y
Past populations						Y
		Kolmogorov-Smirnov statistic				
Null hypothesis:						
No Effect	$\alpha(\tau) = 0$	58.24^a	18.86^a	13.06^a	36.35^a	36.06^a
Constant effect	$\alpha(\tau) = \alpha$	11.29^a	3.29^b	4.71^a	4.47^a	3.76^a
Exogeneity	$\alpha(\tau) = \alpha_{QR}$	14.56^a	3.45^a	3.24^a	5.59^a	9.50^a
Panel B						
ln(IH Lane mile)		1.31***	0.83***	1.03***	1.00***	1.02***
		-0.06	-0.15	-0.18	-0.18	-0.22
N				684		

Table 6: Effect of lane miles on VMT

Notes: Robust standard errors in parentheses. ln(IH Lane mile) is the endogenous variable and is being instrumented.
*,**,*** Significant at 10% level, 5% level, and 1% level respectively.
Quantile regression results are obtained using the R library IV-QR.
a) Kolmogorov-Smirnov statistic > 95% critical value.
b) 90% critical value < Kolmogorov-Smirnov statistic < 95% critical value.
c) Kolmogorov-Smirnov statistic < 90% critical value.

Panel A: Instrumental Variables Pooled Quantile Regression; Panel B: Instrumental Variables (TSLS) Pooled Regression						
Dependent variable: ln(VMT) for interstate highways, entire MSAs						
Instruments: ln 1835 exploration routes						
		(1)	(2)	(3)	(4)	(5)
Panel A	Quantiles					
	10%	1.51***	1.05***	1.57***	1.73***	1.42***
		(0.06)	(0.21)	(0.46)	(0.64)	(0.28)
	25%	1.35***	0.84***	0.83***	1.04***	1.30***
		(0.03)	(0.10)	(0.15)	(0.22)	(0.19)
ln(IH Lane mile)	50%	1.26***	0.52***	0.83***	0.64***	1.12***
		(0.03)	(0.09)	(0.24)	(0.14)	(0.21)
	75%	1.18***	0.34***	0.67***	0.63***	0.53***
		(0.02)	(0.12)	(0.12)	(0.22)	(0.23)
	90%	1.07***	0.34	0.63***	0.54***	0.43
		(0.03)	(0.24)	(0.09)	(0.12)	(0.33)
ln(population)			Y	Y	Y	Y
Year Dummy		Y	Y	Y	Y	Y
Geography				Y	Y	Y
Census Divisions				Y	Y	Y
Socioeconomic characteristics					Y	Y
Past populations						Y
		Kolmogorov-Smirnov statistic				
Null hypothesis:						
No Effect	$\alpha(\tau) = 0$	48.10^a	10.13^a	8.14^a	19.71^a	8.44^a
Constant effect	$\alpha(\tau) = \alpha$	6.51^a	4.43^a	2.51^c	6.33^a	3.55^a
Exogeneity	$\alpha(\tau) = \alpha_{QR}$	8.60^a	4.02^a	2.47^c	5.95^a	2.96^b
Panel B						
ln(IH Lane mile)		1.25***	0.63***	0.75***	0.68***	0.72***
		(0.08)	(0.17)	(0.18)	(0.21)	(0.22)
N				684		

Table 7: Effect of lane miles on VMT

Notes:
Robust standard errors in parentheses. ln(IH Lane mile) is the endogenous variable and is being instrumented.
*,**,*** Significant at 10% level, 5% level, and 1% level respectively.
Quantile regression results are obtained using the R library IV-QR.
a) Kolmogorov-Smirnov statistic > 95% critical value.
b) 90% critical value < Kolmogorov-Smirnov statistic < 95% critical value.
c) Kolmogorov-Smirnov statistic < 90% critical value.

Panel A: Instrumental Variables Quantile Regression; Panel B: Instrumental Variables (LIML) Regression						
Dependent variable: ln(VMT) for interstate highways, entire MSAs						
Instruments: ln 1898 railroads, and ln 1947 planned interstates						
		(1)	(2)	(3)	(4)	(5)
Panel A	Quantiles					
	10%	1.45***	1.02***	1.29***	1.24***	1.23***
		(0.13)	(0.11)	(0.27)	(0.22)	(0.16)
	25%	1.44***	1.14***	1.30***	1.25***	1.32***
		(0.04)	(0.09)	(0.11)	(0.08)	(0.11)
ln(IH Lane mile)	50%	1.38***	1.15***	1.13***	1.02***	1.28***
		(0.03)	(0.13)	(0.26)	(0.09)	(0.12)
	75%	1.38***	1.01***	1.01***	0.88***	0.89**
		(0.06)	(0.25)	(0.19)	(0.13)	(0.28)
	90%	1.29***	0.98***	1.05***	0.87***	0.87***
		(0.13)	(0.40)	(0.40)	(0.13)	(0.21)
ln(Population)			Y	Y	Y	Y
Geography				Y	Y	Y
Census Divisions				Y	Y	Y
Socioeconomic characteristics					Y	Y
Past populations						Y
		Kolmogorov-Smirnov statistics				
Null hypothesis:						
No Effect	$\alpha(\tau) = 0$	44.49^a	20.03^a	22^a	70^a	28.33^a
Constant effect	$\alpha(\tau) = \alpha$	3.43^a	2.67^c	1.03^c	2.93^b	3.07^b
Exogeneity	$\alpha(\tau) = \alpha_{QR}$	NA	2.46^c	3.02^a	6.99^a	11.28^a
Panel B						
ln(IH Lane mile)		1.39***	1.09***	1.18***	1.15***	1.20***
		(0.04)	(0.10)	(0.11)	(0.13)	(0.16)
N				228		

Table 8: Effect of lane miles on VMT

Notes:
Robust standard errors in parentheses. ln(IH Lane mile) is the endogenous variable and is being instrumented.
*,**,*** Significant at 10% level, 5% level, and 1% level respectively.
Quantile regression results are obtained using the R library IV-QR.
a) Kolmogorov-Smirnov statistic > 95% critical value.
b) 90% critical value < Kolmogorov-Smirnov statistic < 95% critical value.
c) Kolmogorov-Smirnov statistic < 90% critical value.

Panel A: Instrumental Variables Quantile Regression; Panel B: Instrumental Variables (LIML) Regression						
Dependent variable: ln(VMT) for interstate highways, entire MSAs						
Instruments: ln 1898 railroads, and ln 1947 planned interstates						
		(1)	(2)	(3)	(4)	(5)
Panel A	Quantiles					
ln(IH Lane mile)	10%	1.53***	1.07***	1.62***	1.47***	1.63***
		(0.04)	(0.21)	(0.42)	(0.20)	(0.50)
	25%	1.39***	1.00***	1.22***	1.24***	1.33***
		(0.04)	(0.14)	(0.15)	(0.21)	(0.33)
	50%	1.27***	0.90***	0.90***	0.89***	0.91***
		(0.03)	(0.11)	(0.14)	(0.19)	(0.31)
	75%	1.23***	0.83***	0.92***	0.91***	1.06***
		(0.03)	(0.11)	(0.15)	(0.19)	(0.43)
	90%	1.14***	0.91***	1.02***	0.85***	0.79***
		(0.04)	(0.13)	(0.31)	(0.25)	(0.36)
ln(Population)			Y	Y	Y	Y
Geography				Y	Y	Y
Census Divisions				Y	Y	Y
Socioeconomic characteristics					Y	Y
Past populations						Y
		Kolmogorov-Smirnov statistic				
Null hypothesis:						
No Effect	$\alpha(\tau) = 0$	43.11^a	20.31^a	24.02^a	29.24^a	26.81^a
Constant effect	$\alpha(\tau) = \alpha$	5.85^a	1.78^c	5.72^a	3.36^a	2.02^c
Exogeneity	$\alpha(\tau) = \alpha_{QR}$	NA	3.82^a	6.47^a	10.18^a	4.10^a
Panel B						
ln(IH Lane mile)		1.33***	0.98***	1.13***	1.08***	1.13***
		(0.05)	(0.13)	(0.16)	(0.15)	(0.17)
N		228				

Table 9: Effect of lane miles on VMT

Notes:
Robust standard errors in parentheses. ln(IH Lane mile) is the endogenous variable and is being instrumented.
*,**,*** Significant at 10% level, 5% level, and 1% level respectively.
Quantile regression results are obtained using the R library IV-QR.
a) Kolmogorov-Smirnov statistic > 95% critical value.
b) 90% critical value < Kolmogorov-Smirnov statistic < 95% critical value.
c) Kolmogorov-Smirnov statistic < 90% critical value.

Panel A: Instrumental Variables Quantile Regression; Panel B: Instrumental Variables (LIML) Regression						
Dependent variable: ln(VMT) for interstate highways, entire MSAs						
Instruments: ln 1898 railroads, and ln 1947 planned interstates						
		(1)	(2)	(3)	(4)	(5)
Panel A	Quantiles					
ln(IH Lane mile)	10%	1.43***	0.95***	1.26***	1.07***	1.23***
		(0.05)	(0.10)	(0.23)	(0.18)	(0.20)
	25%	1.37***	0.95***	1.14***	1.13***	1.00***
		(0.04)	(0.12)	(0.26)	(0.36)	(0.16)
	50%	1.22***	0.72***	0.81***	0.84***	0.93***
		(0.03)	(0.11)	(0.10)	(0.14)	(0.15)
	75%	1.16***	0.68***	0.72***	0.63***	0.66***
		(0.03)	(0.15)	(0.11)	(0.12)	(0.21)
	90%	1.13***	0.70***	0.59***	0.74***	0.90
		(0.03)	(0.12)	(0.12)	(0.19)	(0.58)
ln(Population)			Y	Y	Y	Y
Geography				Y	Y	Y
Census Divisions				Y	Y	Y
Socioeconomic characteristics					Y	Y
Past populations						Y
			Kolmogorov-Smirnov statistic			
Null hypothesis:						
No Effect	$\alpha(\tau) = 0$	39.62^a	11.11^a	33.10^a	28.91^a	9.17^a
Location Shift	$\alpha(\tau) = \alpha$	4.86^a	2.52^c	2.74^c	2.55^c	2.19^c
Exogeneity	$\alpha(\tau) = \alpha_{QR}$	NA	2.66^c	2.59^c	5.81^a	3.04^a
Panel B						
ln(IH Lane mile)		1.26***	0.82***	0.93***	0.92***	0.97***
		(0.05)	(0.11)	(0.13)	(0.13)	(0.16)
N				228		

Table 10: Effect of lane miles on VMT

Notes:
Robust standard errors in parentheses. ln(IH Lane mile) is the endogenous variable and is being instrumented.
*,**,*** Significant at 10% level, 5% level, and 1% level respectively.
Quantile regression results are obtained using the R library IV-QR.
a) Kolmogorov-Smirnov statistic > 95% critical value.
b) 90% critical value < Kolmogorov-Smirnov statistic < 95% critical value.
c) Kolmogorov-Smirnov statistic < 90% critical value.

Panel A: Unconditional Instrumental Variables Pooled Quantile Regression; Panel B: Instrumental Variables (TSLS) Pooled Regression						
Dependent variable: ln(VMT) for interstate highways, entire MSAs Instruments: ln 1835 exploration routes, ln 1898 railroads, and ln 1947 planned interstates						
		(1)	(2)	(3)	(4)	(5)
Panel A	Quantiles					
	10%	1.40*** (0.31)	0.78*** (0.21)	1.27*** (0.14)	1.27*** (0.14)	1.21*** (0.12)
	25%	1.39*** (0.07)	1.08*** (0.38)	1.22*** (0.13)	1.22*** (0.13)	1.22*** (0.14)
ln(IH Lane miles)	50%	1.25*** (0.10)	1.03*** (0.19)	1.02*** (0.20)	1.02*** (0.20)	1.02*** (0.19)
	75%	1.24*** (0.10)	0.49*** (0.07)	0.86*** (0.06)	0.85*** (0.07)	0.85*** (0.06)
	90%	1.30*** (0.05)	0.51*** (0.17)	-0.74 (0.82)	-0.34 (0.65)	-0.48 (0.80)
ln(population)			Y	Y	Y	Y
Year dummy		Y	Y	Y	Y	Y
Geography				Y	Y	Y
Census Divisions				Y	Y	Y
Socioeconomic characteristics					Y	Y
Past populations						Y
Panel B						
ln(IH Lane miles)		1.32*** (0.03)	0.92*** (0.06)	1.03*** (0.07)	1.01*** (0.08)	1.04*** (0.08)
N				684		

Table 11: Effect of lane miles on VMT

Notes:
Robust standard errors in parentheses. ln(IH Lane mile) is the endogenous variable and is being instrumented.
*,**,*** Significant at 10% level, 5% level, and 1% level respectively.
Quantile regression results are obtained using the R library IV-QR.
a) Kolmogorov-Smirnov statistic > 95% critical value.
b) 90% critical value < Kolmogorov-Smirnov statistic < 95% critical value.
c) Kolmogorov-Smirnov statistic < 90% critical value.

Parameters & variables in the simulation model (Part 1)				
Symbol	**Description**	**Equation number**	**Value**	**Note**
i, j, z	Zone number	Subscript	0-97	
f	Income group	Subscript	1-4	
m	Mode of travel	Subscript	1-4	
a	Arc number	Subscript	1-696	
o, d, π	Node number	Subscript	1-210	Including i, j, z
$base$	Base equilibrium	Superscript		
$policy$	Policy simulation equilibrium	Superscript		
α, β	Preference parameters	24		Calibrated
ι	Preference parameter	24	1	Assumed
$1/(1-\sigma)$	Elasticity of substitution	24	2 ($\sigma = 0.5$)	Assumed
$1/(\sigma-1)$	Demand elasticity	54	-2	Assumed
γ	VOT parameter	24		Calibrated
vot	VOT	32	50% wage	Small (2012)
Z	Non-housing consumption	24		Calibrated
h	Housing consumption	24		Calibrated
G	Average travel time	24		Calibrated
g	Average travel cost	25		Calibrated
Δ	Employment status	25	1 or 0	
$days$	No. of work days in a year	25	250	Assumed
$hours$	No. of work hours in a day	25	8	Assumed
w	Hourly wage rate	25		implan.com
t^i	Income tax	25		Collected
t^s	Sale tax	27		Collected
m	Non-work income	25		Calibrated
P	Price inclusive of travel costs	25		Calibrated
p	Mill price	27	1	Normalized
R	Rent	25		Observed
s	No. of trips per dollar spent	27		Calibrated
N	No. of workers between an OD	29		Observed
$PROB$	Mode choice probabilities	31		Observed
$\Theta,\ \Omega$	Dispersion parameter	31	1	Assumed
K	Disutility constants	31		Calibrated
$\wp$	Expected disutility of driving	31		Calibrated
GC	Expected disutility of other modes	31		Calibrated
$TIME$	Non-driving travel time	32		Observed
$GCOST$	Non-driving travel cost	32		Observed
$AUTOTRIP$	Daily car (round) trips	34		Calibrated

Table 12: List of Symbols — part 1/2

Parameters & variables in the simulation model (Part 2)				
Symbol	**Description**	**Equation number**	**Value**	**Note**
$time$	Driving time on arc	35		Calibrated
b	BPR parameter	35	0.15	Assumed
c	BPR parameter	35	1.2	Calibrated
$flow$	Flow on arc	35		Observed
$capacity$	Capacity of arc	35		Calibrated
$price_{fuel}$	Gasoline price	36	\$1.6	Observed
$length$	Arc length	36		Observed
$speed$	Average arc speed	36		Calibrated
$F(speed)$	Gasoline use per mile	36		Calibrated
$mcost$	Monetary driving cost	38		Calibrated
$gcost$	Generalized driving cost	38		Calibrated
Pr	Arc choice probability	39		Calibrated
x	Flow between two nodes	41		Calibrated
v	Flow on an arc given an end node	42		Calibrated
$Zflow$	Intra-zonal flow	44		Calibrated
$ACCESS\_AUTOTRIP$	Auto trips leaving a zone	44		Calibrated
$EGRESS\_AUTOTRIP$	Auto trips arriving in a zone	44		Calibrated
$\mathbb{T}$	Time cost of driving between zones	45		Calibrated
$\mathbb{M}$	Money cost of driving between zones	46		Calibrated
$\tilde{U}$	Indirect utility	28		Calibrated
EV	Equivalent variation	50		Calibrated
$\partial \ln(PROB_{m\|ijf})/\partial \ln(\wp_{ij})$	Mode choice elasticity	55	0.1	Indra (2014)

Table 13: List of symbols - part 2/2

Baseline simulation: the effect of one percent increase of road capacity

	Initial level	% Δ after a 1% cap. Increase
Car vehicle-miles traveled (VMT)		
aggregate miles/year (ELASTICITY)	**77.8 bn**	**+0.321%**
miles/user/day	**26.2**	**+0.3%**
average miles/trip	**11.5**	**-0.4%**
Gasoline consumption		
aggregate gallons/year	**2.35 bn**	**+0.3%**
gallons/user/day	**0.791**	**+0.3%**
Average fuel economy (mi/gal)	**32.5**	**+0.01%**
Average vehicle speed (mi/hr)		
major roads and highways	**41.9**	**+0.06%**
local roads	**29.8**	**-0.02%**
Average per trip cost		
time (mins)		
car	**15**	**-0.1%**
all modes	**16**	**-0.6%**
Monetary (\$)		
car	**0.4**	**-0.02%**
all modes	**0.4**	**-0.1%**

	Initial level	Δ after a 1% cap. Increase
Welfare (\$/user/year)		**+22.87**
Equivalent variation (\$/user/year)	**-**	**+23.16**
Delay Externality		
hour/user/year	**36.9**	**+0.04**
\$/user/year	**331**	**+0.29**
Gasoline Use Externality		
gallon/user/year	**8.6**	**-0.002**
\$/user/year	**14**	**-0.003**
Avg. Wkr. Inc. AFTER travel cost (\$)	**50,930**	**+0.02%**
Price index inclusive of travel cost	**1.068**	**-0.001%**
Mode share, %		
Car	**80.6**	**+0.66**
Bus	**3.5**	**-0.13**
Rail	**0.2**	**-0.01**
Other	**15.7**	**-0.52**
Daily (round) trips, millions		
car trips	**12.9**	**+0.8%**
public transit trips	**0.6**	**-3.9%**
total trips by all modes	**15.7**	**+0.02%**

Table 14: Baseline simulation

Number of non-work trips as a percentage of baseline level	Vehicle miles traveled (mi)			Gasoline use (gal)		fuel economy	Vehicle speed (mi/hr)		Avg per trip time (mins)		Externality	
	/year	/user/day	/trip	/year	/user/day	mi/gal	major roads and highways	local roads	car	all modes	hr/user/year	gal/user/year
10	45.3 bn	15.3	16.0	1.35 bn	0.46	33.2	43.4	35.4	13.9	15	15.6	3.3
20	48.9 bn	16.5	15.1	1.46 bn	0.49	33.2	43.2	34.8	14.1	15	17.3	3.6
40	56.1 bn	18.9	13.8	1.68 bn	0.57	33.0	42.9	33.6	14.3	15	21.1	4.5
60	63.4 bn	21.3	12.8	1.90 bn	0.64	32.9	42.5	32.3	14.6	15	25.6	5.6
80	70.6 bn	23.8	12.1	2.12 bn	0.72	32.7	42.2	31.0	14.9	16	30.9	7.0
100 (baseline)	**77.8 bn**	**26.2**	**11.5**	**2.35 bn**	**0.79**	**32.5**	**41.9**	**29.8**	**15.2**	**16**	**36.9**	**8.6**
120	85.0 bn	28.7	11.1	2.57 bn	0.87	32.3	41.6	28.7	15.5	16	43.7	10.6
140	92.3 bn	31.1	10.7	2.80 bn	0.94	32.0	41.3	27.6	15.8	16	51.4	12.8
160	99.5 bn	33.5	10.4	3.03 bn	1.02	31.8	41.0	26.5	16.2	17	59.9	15.4
180	106.7 bn	36.0	10.1	3.26 bn	1.10	31.5	40.7	25.6	16.5	17	69.2	18.2
200	114.0 bn	38.4	9.9	3.50 bn	1.18	31.2	40.4	24.7	16.8	17	79.4	21.3

Table 15: Key initial equilibrium variables under different levels of non-work trips

Number of non-work trips as a percentage of baseline level	The effect of a one-percent road capacity increase under different initial VMTs											
	percent change							percent point chnge				
	Vehicle miles traveled (mi)		Gasoline use (gal)	fuel economy	Speed	Trips		Share of car trips	Welfare			
	aggr.	/trip	/year	mi/gal	local roads	car	all modes		Overall	EV	time delay externality	gasoline use externality
10	**+0.356**	-0.24	+0.35	+0.014	+0.03	+0.60	+0.004	+0.676	+23.41	+23.18	-0.22	-0.02
20	**+0.352**	-0.27	+0.34	+0.014	+0.02	+0.63	+0.007	+0.674	+23.45	+23.21	-0.22	-0.02
40	**+0.343**	-0.32	+0.33	+0.014	+0.01	+0.69	+0.011	+0.670	+23.45	+23.24	-0.19	-0.02
60	**+0.335**	-0.35	+0.33	+0.014	+0.001	+0.73	+0.015	+0.666	+23.35	+23.24	-0.10	-0.02
80	**+0.328**	-0.38	+0.32	+0.013	-0.01	+0.75	+0.017	+0.663	+23.16	+23.21	+0.06	-0.01
100 (baseline)	**+0.321**	**-0.40**	**+0.32**	**+0.011**	**-0.02**	**+0.77**	**+0.019**	**+0.659**	**+22.87**	**+23.16**	**+0.29**	**-0.003**
120	**+0.314**	-0.42	+0.31	+0.009	-0.03	+0.79	+0.020	+0.657	+22.48	+23.09	+0.59	+0.01
140	**+0.308**	-0.43	+0.31	+0.006	-0.04	+0.80	+0.021	+0.654	+22.02	+23.00	+0.96	+0.03
160	**+0.302**	-0.44	+0.30	+0.003	-0.05	+0.81	+0.021	+0.652	+21.49	+22.92	+1.38	+0.05
180	**+0.297**	-0.45	+0.30	-0.001	-0.05	+0.82	+0.020	+0.650	+20.85	+22.82	+1.89	+0.08
200	**+0.291**	-0.46	+0.29	-0.004	-0.06	+0.82	+0.019	+0.648	+20.20	+22.75	+2.43	+0.11

Table 16: The effect of capacity increase under different levels of initial non-work trips