Estimating the Social Cost of Congestion Using the Bottleneck Model

Final Report

METRANS Project 17-14

December 2018

Principal Investigator: Jinwon Kim

Department of Economics, CSU Long Beach

Disclaimer

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. This document is disseminated under the sponsorship of the Department of Transportation, University Transportation Centers Program, and California Department of Transportation in the interest of information exchange. The U.S. Government and California Department of Transportation assume no liability for the contents or use thereof. The contents do not necessarily reflect the official views or policies of the State of California or the Department of Transportation. This report does not constitute a standard, specification, or regulation.

Abstract

This paper uses the Vickrey (1969) bottleneck model to empirically measure the social cost of traffic congestion in the US. We estimate extra travel time over and above *hypothetical* free-flow travel time, which we call "queuing time", for each average commute trip. Our estimate implies that the annual social cost of congestion borne by all US commuters is 24 billion dollars. A higher level of congestion in a city may be attributed to a smaller per capita road stock in the city. This paper also empirically quantifies the optimal toll depending both on the commuter's arrival time and residential location.

Disclosure

The project was funded in entirety under U.S. Department of Transportation. The total funding amount was $52,724. The Principle Investigator of the project was Professor Jinwon Kim in the Department of Economics at California State University, Long Beach, and was carried out during the period: 08/14/2017 - 09/30/2018.

Acknowledgement

I would like to thank METRANS for funding this research. I thank Stephen Gibbons, Jonathan Hall, Jos van Ommeren, and conference and seminar audiences at the ITEA Meetings in Barcelona, the UEA Meetings in Vancouver, CSU Long Beach, UC Irvine, and especially Jan Brueckner for helpful comments.

Contents

List of Figures

List of Tables

1 Introduction

Economists have long understood that traffic congestion is a source of market failure and therefore a social problem. The traditional framework for analyzing the inefficiency cost of traffic congestion is based on Pigou (1920).[1] Despite the success of the conceptual Pigouvian framework, empirical estimates of its inefficiency cost are relatively rare. An empirical estimation of the Pigouvian inefficiency cost is challenging, since it requires identification of both the demand and the cost (supply) functions of travel that are hard to empirically conceptualize.

The Pigouvian congestion model itself has also a theoretical drawback, in that it does not treat commuters' trip timing decisions endogenously. A result of this omission is the *static* nature of traffic congestion, which is an oversimplification of the real-world's *dynamic* traffic congestion. The alternative bottleneck model of Vickrey (1969), later formalized in Arnott et al. (1990, 1993), treats the trip timing decision endogenously and has now become the workhorse model for analysis of traffic congestion involving dynamics. The main goal of this paper is to empirically measure the inefficiency cost from traffic congestion in the US using the bottleneck model framework.

The main insight from the bottleneck model is that for any commute trip, extra travel time over and above *hypothetical* free-flow time, which the model calls "queuing time" (or congestion delay), is purely social loss and thus constitutes the society's welfare loss. Using a detailed trip-level dataset of commuters in the US, this paper therefore estimates the queuing time for each average commute trip, and we then aggregate them to compute the social cost of congestion from all commute trips in the US.

Note that the bottleneck model and the Pigouvian model could be analyzed in a unified framework, as attempted in Arnott et al. (1993), but it would be fair to say that these models are quite different framework for the same phenomenon of traffic congestion. While the Pigouvian framework is based on individuals' route choice (or travel mode choice) behavior,

[1] See Brueckner (2011) for a modern exposition of the Pigouvian congestion model.

the bottleneck model is focusing on the trip timing choice of individuals. The key concept in the Pigou model is externality. While the bottleneck model does not ignore the fact that individual drivers generate congestion externalities, the individuals' failure in the bottleneck model is a coordination problem. The bottleneck model is more explicitly based on individual's micro-economic behavior. However, it is not the focus of the paper to analyze their difference and similarities of these two models. As far as we know, there is no earlier studies that estimated the cost of road congestion using the bottleneck model, so it is worthwhile to use this popular model to apply the data and estimate the social cost of congestion.

An identification issue arises because each commute trip's *hypothetical* free-flow travel time is not directly observable. For this, we exploit the model's other key result: that the other commuter on the same commute route who alternatively arrived earliest in the morning meets no queue, so her travel time is the free-flow time corresponding to the route. Our econometric model is therefore specified to estimate the expected difference in travel time by binary arrival timing choice, i.e., arrival at a peak time with queuing vs. arrival at an earlier timing with no queuing, conditional on that the commuters under comparison effectively traveled an identical route.

Our key identifying assumption is that travel time of an individual who arrives at a particular timing is determined solely by the characteristics of the route traveled, especially its physical distance, which allows us to focus on conditioning on the route traveled in our estimation. Our basic identifying strategy is to include extensive route characteristics as well as person characteristics related to the choice of commute route in our empirical specification. We then address the potential omitted variable bias concern by employing the instrumental variable estimation and the household fixed model estimation strategies.

Another empirical issue of the bottleneck model is that it considers only commuters whose travel demand is perfectly inelastic. To set up a consistent empirical framework, we also use only the sample of commute trips that took place in a finite time interval in a day (morning or evening congestion, respectively), which implies that the social cost estimated in this paper includes only the costs from commute trips. An advantage of using only the commute trips is that the number of all commutes in the US population is relatively well identified (compared to the other kinds of trips), so our inference about the commuter population would be accurate for this particular group of travelers.

The empirical literature on the cost of congestion usually estimates the relationship between the inverse of travel speed and the traffic density, which in the Pigouvian framework is the cost (supply) function of travel, using traffic data on particular road segments (e.g., Walters, 1961; Dewees, 1979; Fosgerau and Small, 2012). The main estimand in these studies is marginal external congestion cost of added car to the road segment that is used to gauge the Pigouvian congestion toll. A more recent paper by Li et al. (2018) estimates the cost function of travel for an entire city, Beijing, using a larger number of road segments in the city. Since traffic density is endogenous, they use a driving restriction policy as an exogenous variation (instrument) for traffic density. A limitation of these studies is that since the demand function of travel is not explicitly estimated, either computation of the Pigouvian deadweight loss is missing or computation of it relies on their assumption on the demand function. Moreover, traffic data used in these studies do not fully capture the traveler's congestion delays that occur over her whole travel route.

Other studies use trip-level datasets for estimation of the congestion cost. For example, Couture et al. (2018) estimate the speed-distance relationship that holds at the trip level and then construct each city's speed index, which allow them to identify the city-level supply function of travel. However, their Pigouvian deadweight loss is computed still without an explicitly estimated demand function. Akbar and Duranton (2017) is the first study that explicitly estimates the demand function of travel as well as the supply function of travel. They use a counterfactual travel data generated from Google Maps as well as a survey data on actual trips, which together allow them to identify the demand function of travel in a single city Bogota, Columbia.

Rather than using the Pigouvian framework, which is common in the earlier studies, this paper uses arguably a more detailed microeconomic model of Vickrey (1969) to empirically measure the social cost of congestion. We do not need to estimate the demand function of travel that is empirically challenging, nor the marginal external congestion cost, although our estimation still allows us to indirectly quantify marginal external congestion cost. Our approach is to directly estimate the inefficiency cost for each commute trip, which is more intuitive and practically more feasible. Our method can be applied to estimation of the inefficiency cost of congestion measured at the route, city, and country levels.

Our paper is also related to a few earlier empirical papers on the bottleneck model. In particular, Small (1982) uses a discrete choice econometric framework, in which trip timing choices are explicit, to estimate the parameters in the commuters' scheduling preferences. Hall (2018) uses a structural econometric framework to estimate the distribution of the scheduling preferences that are heterogeneous across individuals. Their work, however, is not directed to measuring the social cost of traffic congestion.

Our main estimation result indicates that morning commuters who arrived at work at a peak time (between 6:15 and 11:00) traveled on average for about 2 minutes longer than the other commuters on the same route who alternatively arrived earlier than 6:15. As expected, the queuing time tends to be longer (shorter) for the morning commutes with an arrival time closer to (farther from) the peak time, say 9:00. From the sample of evening commutes, we find that the expected queuing time for each evening commute is about 1.8 minutes, implying that the average daily queuing time for each worker is about 3.8 minutes, which is about 8% of the sample mean of daily commute time. While this individual inefficiency cost may sound small, aggregation over all workers yields a huge social cost. According to our computation, the annual social cost of congestion borne by all US workers is about 24 billion dollars.

We also examine the variation of congestion by city. Specifically, we construct the congestion ranking of large metropolitan statistical areas (MSA) based on our city-specific estimates for the mean queuing time. We then explore the determinants of traffic congestion in the city. We find that residents in a city with a smaller road stock per capita do experience a longer queuing time on average than those in a city with a larger road stock per capita, and that the level of congestion is higher in the city where the demand for vehicle travels is overall higher. We also examine within-city determinants of congestion and find some anticipated results.

The optimal policy in the bottleneck model is imposition of a time-varying congestion toll. The model predicts that the optimal toll would eliminate all the queues by inducing reschedule of departures, which would then recover all the welfare loss estimated in this paper. This optimal toll is exactly same as the queuing time cost as a function of arrival time, which could be imposed separately at each commute route. This paper makes a meaningful starting point toward empirical quantification of a detailed optimal toll by estimating the expected queuing time as a function of arrival time by trip distance for residents in particular cities.

The rest of the paper is organized as follows. Section 2 presents our conceptual and the empirical framework. Section 3 explains our dataset. Section 4 reports our estimation results. In Section 5, we investigate the variation in congestion over residential location and explore the determinants of congestion. In Section 6, we quantify the optimal congestion toll. Finally, Section 7 concludes.

2 Conceptual and empirical framework

2.1 The social cost of congestion in the bottleneck model

In this section, we briefly review the basic bottleneck model (Vickrey, 1969; Arnott et al., 1990) and explicitly define the social cost of congestion to be empirically measured. We review only essential elements of the model.

Consider a commute route with distance x. With the speed under the congestion-free condition normalized at 1, the free-flow travel time from home to work is then simply x. There is a congestion bottleneck at the entrance to the work. Or, the whole commute route itself may be regarded as a congestion bottleneck. The bottleneck allows at most ψ cars to pass per each time unit, which means that the parameter ψ is the capacity of the bottleneck.

There are a fixed N number of car commuters traveling this route, each of who has an identical scheduling preference.[2] Specifically, all the commuters want to arrive at work at the specific time t^* if there were no queue, which however is physically impossible because the bottleneck has a limited capacity. The resulting key feature of the Nash equilibrium is the trade-off between travel time and arrival at a preferable time closer to t^*. To formally illustrate this trade-off, let $Y(t)$ denote the individual's travel time as a function of arrival time $t \in [t_0, t_1]$, where t_0 and t_1 are the first and the last commuters' arrival times, respectively. It is shown that in equilibrium, $Y(t)$ increases with t from t_0 to t^* and decreases afterward until t_1. See Figure 1 for illustration of the $Y(t)$ function.

Figure 1 about here.

Excessive travel time relative to free-flow time for the commuter who arrives at $t \in (t_0, t_1)$ is the "queuing time", which equals $Y(t) - x$. is Since the first traveler has zero queue, $Y(t_0) = x$ holds, so that the queuing time of a commuter who arrives between t_0 and t_1 is $Y(t) - Y(t_0)$.[3] The model's key insight is that each individual's queuing time is pure social loss and thus constitutes the society's welfare loss. Intuitively, if the individuals were able to coordinate their trip timing choices, then they would depart and arrive only at the rate of ψ without forming any queue and could still arrive at work at the same time as under laissez-faire.

[2]The particular form of individual commuter's cost assumed in the basic bottleneck model (Arnott et al., 1990) is

$$\alpha(t - d) + \beta \max(t^* - t, 0) + \gamma \max(t - t^*, 0),$$

where d is departure time from home and t is arrival time at work. The first term is the travel time cost, and the second and the third term are the schedule delay cost incurred due to an earlier or a later arrival than the most preferred t^*. The parameters α, β, and γ are the unit cost placed on each cost component.

[3]One could interpret that there is a positive level of congestion in the Pigouvian model, whereas the bottleneck model has no congestion delay at the social optimum. However, it is not an entirely correct interpretation, because congestion arises in the Pigou model when speed and *traffic density* are negatively related. In the meantime, congestion in the bottleneck model is defined as a drop in traffic flows, not a drop in traffic density. Since flow and density are different, it is possible that congestion does not arise at the bottlneeck model but does occur in the Pigou model for the same road situation.

The inefficiency cost of congestion borne by all the route users is therefore the opportunity cost of their total queuing times. The total queuing times are computed by integrating $Y(t) - Y(t_0)$ over $t \in [t_0, t_1]$, with each arrival timing weighted by the rate of arrivals that is ψ. By dividing the total queuing times by N, we get the average queuing time per route user written as

$$\frac{\psi}{N} \int_{t_0}^{t_1} [Y(t) - Y(t_0)] dt. \tag{1}$$

This is the measure of welfare loss for this particular commute route. Note that the cost of congestion we define here is expressed in a "time" unit rather than in a monetary unit. The variables used in our empirical analysis are also in the time unit, but we can always multiply our estimate by the monetary unit cost of travel time, which we call simply "value of time", to obtain the social cost expressed in a monetary unit.[4]

While the standard bottleneck model assumes a single route, the real-world commuters travel heterogeneous routes, and we now formalize this. Let each of the heterogeneous routes be indexed by j, with $j \in [0, J]$. The distribution of the number of travelers over the routes is described by the density function denoted by $f(j) \equiv N_j/\bar{N}$, where $\bar{N}$ is the number of all commuters in the population and N_j is the number of users of particular route j. The average queuing time in the population is the average queuing time for each route integrated over the routes weighted by the route density function $f(j)$, which we write as

$$\int_0^J \int_{t_{0j}}^{t_{1j}} \frac{\psi_j}{N_j} [Y_j(t) - Y_j(t_{0j})] f(j) \; dt dj, \tag{2}$$

where $Y_j(t)$ is travel time of commuter on route j who arrived at a time $t \in [t_{0j}, t_{1j}]$ and $Y_j(t_{0j})$ is the travel time of a different commuter on the same route, who alternatively arrived

[4] The monetary unit cost of travel time is α in the scheduling cost form shown given in footnote 2. Note also that in the basic bottleneck model (Arnott et al., 1993), marginal external cost of added car to the route is derived and given by the expression (1) multiplied by 2α. This implies that estimation of (2) below in this paper would give indirectly an estimate for marginal external cost of added car to each average route.

in the first place at t_{0j} (with no queuing). Note that $Y_j(t_{0j})$ is simply the physical distance of commute route j.

We have so far assumed that individual commuters have a homogeneous scheduling preference (see footnote 2 for the scheduling preference). However, commuter preferences in the real world are heterogeneous. Our conceptual framework may assume that commuters' scheduling preferences are heterogeneous, but we can still show that the queuing time of any individual on a given route is a social loss, regardless of the particular individual's preference type (see Arnott et al., 1994). This implies that (2) as the average queuing time over the individuals having different preferences is the inefficiency cost per each average commute trip even under the heterogeneity of scheduling preferences.

2.2 Empirical framework

Our goal in this section is to estimate (2). Empirically, (2) is the expected gap in travel time between a commuter who arrives in a peak time $t \in (t_{0j}, t_{1j})$ and another commuter on the same route who alternatively arrives earliest at t_{0j}. At the moment, the distribution of routes, captured in the mass function $f(j)$, is set aside in our estimation. Our initial belief is that our data is a random sample of all commute routes that exist in the US, under which we do not have to explicitly weight the route-specific average queuing time by $f(j)$. We address the potential non-random sampling issue by adopting the weighting estimation strategy below (see Section 4.4).

We estimate the average difference in travel times by commuters' binary arrival timing choices (arrival at a peak time with queuing vs. arrival at an early time with no queuing), conditional on that the commuters under comparison effectively traveled an identical route. Our key identifying assumption is that travel time of a commuter traveling at a particular timing is determined solely by the route traveled by the commuter. Each traveler on a congested commute route would be atomic, implying that the only way she could change her travel time on the given route would be choosing a different trip timing or she must travel a different route.[5]

Given that our trip-level sample does not report the commuter's exact origin and destination and there is only a single commute trip for each route, we are not able to condition on the route traveled directly. Our basic strategy is therefore to include the direct route characteristic, the physical distance of the route traveled, as well as proxies for the route characteristics, such as population density around the commuter's residence, in our estimation. We then employ the instrumental variable estimation and the household-fixed effect model estimation strategies to address the potential omitted variable bias concern.

There may also be person characteristics capturing the commuter's scheduling preference, which therefore would help predicting her trip timing choice. The bottleneck model shows that commuters having a particular common preference tend to sort into a particular interval of arrival times.[6] The sorting of arrival times by preference type *within* route, however, is not an immediate empirical problem, since the queuing time of an individual who arrives at $t \in (t_{0j}, t_{1j})$ on route j is always measured by $Y_j(t) - Y_j(t_{0j})$, even when this commuter has a different scheduling preference from the other who arrives at t_{0j}, as long as their trips occur on the same route j. However, there is the possibility that commuter preferences correlated with trip timing choices may also be related to the choice of commute route that would actually determine the travel time. We therefore include person characteristics in order to control for their correlation with the choice of commute route.

We now specify our empirical framework. The outcome variable is travel time of commuter i who travels commute route j, which we denote by Y_{ij}. Note that although we do not observe commute route directly, it is useful to explicitly have the route subscript j in

[5]Note that the estimated relationship between trip timing and travel time may have a "causal" interpretation once our estimation is conditioned on the route traveled. The causal effect is defined as the difference between the commuter's actual travel time and her counterfactual free-flow travel time that would have been spent if she had alternatively chosen to arrive at the earliest timing with no queuing staying on the same route (see Angrist and Pischke, 2009).

[6]For example, from the scheduling cost form given in footnote 2, consider a group of commuters with a higher ratio of the unit cost of travel time to the unit cost of schedule delay (i.e., α/β or α/γ) compared to the other commuters. Arnott et al. (1994) show that these travelers tend to want to avoid a long queue and arrive at a non-peak time .

our empirical model. Also note that since each commuter in our sample makes only one trip, the commuter id is equivalent to the trip id.

Next, the treatment variable is the commuter's binary trip timing choice, which we denote by T_{ij}. The value for T_{ij} is 0 if commuter i on route j arrived early enough to have zero queuing and 1 if she arrived in a peak time with queuing. In our conceptual framework, the time at which queuing initially arises differs by route (see t_{0j} in (2)), but we are not able to allow the cut-off time for T_{ij} to vary by route. We instead seek a time, before which commuters on any route do not queue at all, and use it as the cut-off time for T_{ij}. The underlying assumption is that there exists at least one commuter on each route who travels before this early timing.[7] We searched for the best cut-off time by each 30-minute interval, such that there is a statistically significant gap in travel time just before and after this arrival timing (see the specification in column (2) in Table 2). The resulting the cut-off time is set at 6:15, implying that the treated group in our specification is the commuters who arrived at a time between 6:15 and 11:00 and the control group is those who arrived before 6:15 (but after 5:00). Note that we are excluding the trips completed either before 5:00 or after 11:00 since we suspect that these trips significantly differ from the regular commute trips considered in our estimation.[8]

Our estimating equation is written as

$$Y_{ij} = \alpha + \beta T_{ij} + \sum_k \gamma_k m_j^k + X_{ij}'\delta + u_j + w_{ij} + \epsilon_{ij}, \tag{3}$$

where m_j is one-way commute distance in 10 miles for route j. With the k superscripts

[7]This assumption would make sense if commuter preferences were quite heterogeneous, especially in terms of the most preferred arrival timing, t^*.

[8]About 79% out of the home-to-work morning commutes in our dataset have an arrival time between 5:00 and 11:00. Note that the departure time variable as a trip timing choice is not used to define the treatment status. The reason is that trip distances are heterogeneous, so departure time actually provides little information on whether the commuter actually queued or not. In the bottleneck model, each commuter's choice set $Y(t)$ is a common knowledge, which means that the commuter choosing a departure time in effect is also choosing the corresponding arrival time. An unexpected congestion delay would result in an arrival that is later than the intended time. However, the resulting positive correlation between arrival time and travel time is also a social loss (regardless of it is intended or not), so we must not exclude these congestion delays in our estimation.

denoting exponents, m_j^k are the polynomials of m_j, which are included to capture the nonlinearity of the relationship between trip distance and travel time (γ_k are the coefficients).

X_{ij} is the vector of the proxies capturing the route characteristics. Specifically, it includes population density around the resident's neighborhood, percentage of rental housing units (relative to owned units) in the neighborhood, urban-rural category measuring "contextual density" in the neighborhood, population size categories in the city where the commuter resides in (for smaller cities) as well as the city fixed effects (for larger cities). It also includes person characteristics, such as family income, to control for the preference related to residential and work location choices and thus to commute route choices. It also includes other person characteristics, such as age and gender, which would control for the driving habit affecting the travel speed.

Our empirical specification is also explicit about unobservable route and person characteristics. First, u_j captures unobservable route characteristics, such as the road capacity, the speed limit, the number of traffic signs on the route, and the proximity to highways. We can easily think of the possibility that u_j is correlated with both trip timing choices and congestion (travel time). For example, commuters traveling a highly congested route (for some unobservable reason) may tend to choose to travel at an off-peak time to avoid otherwise a long congestion delay. Second, w_{ij} is unobservable person characteristics, which may predict the commuter's trip timing choice and the choice of commute route. We further discuss the concerns related to these omitted variables below. Finally, ϵ_{ij} is the error term satisfying the zero conditional mean.

3 Data

For estimation, we use the 2009 National Household Travel Survey (NHTS) dataset. We also use the Highway Performance Monitoring System (HPMS) dataset below, but we explain this dataset only in Section 5. The entire NHTS dataset is composed of household, person, vehicle, and trip level samples. We mainly use the trip level sample, but we merge the person- and the household-level samples into the main trip-level sample to obtain each commuter's personal and household characteristics. Our dataset provides the key information at the trip level, such as the trip's departure and arrival timings, trip duration (travel time), and physical distance between the origin and the destination.

To construct our estimation sample, we select only commute trips made by car (of any vehicle type) driven by commuter herself. We exclude other commute modes (such as public transportation) and trips made for other purposes (such as shopping). For our estimation of morning congestion, we use morning commutes completed between 5:00 and 11:00. We also estimate evening congestion below using the sample of evening commute trips that took place between 15:00 and 23:00. We exclude commutes that took place on weekends.

For the survey, each household is assigned a specific travel day and instructed to record all the family members' trips that occurred on that day. It is therefore expected that each worker in the household reports one morning commute and one evening commute, respectively. There are nevertheless a few commuters (about 3% in the morning sample) who reported more than one commute trip in each of the morning and the evening intervals. We suspect that these observations are irregular, so we exclude these observations. Note however that each household may be composed of multiple workers and thus multiple commute trips in each time interval, and these trips are included in our estimation sample.

Table 1 about here.

Table 1 reports the summary statistics for the key variables computed from our morning commute sample. It shows that about 91% of morning commutes are completed at a time between 6:15 and 11:00 while the other 9% were completed between 5:00 and 6:15. The average arrival time at work is 7:47.[9]. Using the mean trip duration (travel time) of 22.88

[9]The arrival time variable is originally in the military format in the NHTS dataset, e.g., 06:10, 19:10, etc. We convert this raw variable into the one that follows the decimal system with the minute unit and normalize the value for 8:00 at 0. We use this converted variable to construct the variables used in our estimation

minutes and the mean commute distance of 13 miles, we calculate the average travel speed that is 34.2 miles per hour. From our sample of evening commutes, we find that the average travel time is 24.6 minutes and the average travel speed is 31.4 miles per hour (see Section 4.5 for the evening sample construction).

4 Estimation results

4.1 OLS estimation results

In our empirical specification (3), the main parameter of interest is β. Table 2 reports the OLS estimation results from the sample of morning commutes. Column (1) indicates that the estimate for β is 1.992, implying that if the average commuter had alternatively arrived before 6:15 while staying on the same route, she could then have reduced her travel time by 1.992 minutes (8.7% of the mean travel time in the sample).[10] Using the sample mean for commute distance, which is about 13 miles, the average queuing time per mile traveled is about 0.15 minutes (9 seconds).

Table 2 about here.

Column (2) in Table 2 presents the other specification, in which multiple arrival time interval dummies (instead of the single dichotomous T_{ij} variable) are included. The coefficient on each dummy indicates the predicted gap in travel time between a commuter who arrived at a time indicated by the dummy and the other who arrived before 6:15. For example, all other things equal, the average commuter who arrived at a time between 8:45 and 9:15 is predicted to spend 3.117 minutes more in commuting than those who arrived before 6:15. As expected, the average queuing time tends to be longer as the commuter's arrival time is closer to the peak time, say 9:00, and get shorter afterward. Panel (a) in Figure 2 plots

[10] We also used the log of travel time as the dependent variable and obtained 0.066 as the estimated coefficient on T_{ij}, which implies about a 6.6% average difference in travel times by binary arrival timing choices.

the estimated coefficient for each arrival time dummy from column (2) along with its 95% confidence interval. The figure clearly illustrates the queue dynamics that is consistent with the bottleneck model's prediction (see Figure 1). The figure also shows that the standard error of the coefficient estimator tends to increase as the arrival time gets closer to the noon. Note also that the average of the coefficients on the arrival time dummies from column (2) is 2.223, which is close in magnitude to the β estimate of 1.992 from column (1), which is an anticipated result.[11]

Figure 2 about here.

Table 2 also contains the estimated coefficients on the control variables, which show the dependency of travel time on these variables. For example, the household income has a negative coefficient, which implies that commuters with a higher income tend to reside in an area with a lower level of traffic congestion (e.g., less congested suburban area). We also find that travel time is longer for the commuters residing in a higher density area than those in a lower density area. Another interesting observation is that older drivers' travel times are on average longer than younger drivers, which implies a significant difference in driving habit by age.

4.2 Instrumental variables estimation results

Our OLS estimator for β may be biased due to the omitted route and person characteristics (see (3)). We now address this concern by employing the instrumental variable (IV) estimation strategy.

We seek an instrument for trip timing choice (T_{ij}) that is uncorrelated with unobservable determinants of the travel time. Our key identifying assumption is that the travel time of a commuter traveling at a particular point in time is determined by the route traveled

[11]In the bottleneck model, the commuters arrive at work at a constant rate of the bottleneck capacity during the peak hours (between 6:15 and 11:00), which implies that these numbers are predicted to be the same from the model.

rather than by the commuter herself. Therefore, our instrument predicting the commuter's trip timing position relative to the other travelers on the same route should be uncorrelated with unobserved route characteristics or unobservable person characteristics related to the choice of route. Furthermore, since the choice of route is almost equivalent to the choice of residential and work location, our instrument should be uncorrelated with this choice.

We use particular person characteristic, the commuter's job category, as the main instrument. In our dataset, each worker's job is categorized into manufacturing, service, clerical, and professional. We find from the data that the worker's job category is a strong predictor of her trip timing choice. For example, manufacturing workers tend to arrive especially earlier compared to the other workers.[12]

Our belief is that observationally identical workers who are in different job categories would not choose commute route (residential and job location) differently from each other in order to have a different level of congestion, considering that the included income variable would control for the worker's subjective value of travel time. There may be particular routes dominated by workers with a common job, e.g., freeway near a big manufacturing facility. For our IV estimator to be biased in such case, however, the level of congestion for these routes must systematically be different from that for the other routes, which however seems implausible, since the included neighborhood and city characteristics would control for the cluster of particular workers.

Table 3 about here.

Table 3 presents our instrumental variable regression results. Our IV sets are not weak according to the conventional F tests, and the over-identification p-values are large enough, indicating that our IV sets are valid.

Column (1) presents the second-stage regression result from the model using the manufacturing and the professional worker dummies as the instruments. We find that it is useful

[12] From the scheduling cost in footnote 2, the preference of the manufacturing workers would be characterized by an earlier t^* or a greater α/β than that of the other workers.

to still include the service category in our regression, so we omit only the clerical category in our first-stage regression (see column (3)). From this specification, the IV estimate for β is 2.283, which is a bit larger than the OLS estimate of 1.992.

In column (2), we report the second-stage regression result from the model that also excludes the worker's education status that is therefore used as IV. This specification yields a bit higher over-identification p-value than the model presented in column (1). In fact, the worker's education may help explaining the nature of the job and thus predicting the worker's trip timing more accurately. The estimate for β from this specification is 2.402, which is a bit larger than both the OLS estimate and the other IV estimate in column (1).

In columns (3), we present the first-stage regression result. We find that the proportion of manufacturing workers who arrived earlier than 6:15 is much higher than for the other workers, and the professional and the service workers also tend to arrive earlier than the clerical workers. We also find that workers with a fewer years of education more tend to arrive earlier than 6:15. Another notable result is that commuters traveling a longer distance more tend to arrive earlier than 6:15 compared to those traveling a shorter distance (see the negative coefficient on m).[13]

The next important issue is whether the arrival time dummy T_{ij} may be treated as an exogenous variable in our empirical specification. For each of the IV regressions, we find that both the Durbin and the Wu-Hausman p values are significantly larger than any usual significance levels, which means that we do not reject the null hypothesis that T_{ij} is exogenous. We also carried out a Hausman test for comparison between 2SLS and OLS, which also leads us to conclude that the OLS and the IV estimates are statistically indistinguishable.[14] We conclude that the omitted variables in our specification are not causing a significant bias in our OLS estimator. Note that this result does not imply that trip timing choice is indeed exogenous, but it suggests the necessity of including a rich set of the route and the person

[13]See Fosgerau and Kim (2017), who offer a theoretical explanation on this empirical relationship between commute distance and trip timing.

[14]The Hausman p value for each IV set presented in Table 3 is effectively equal to 1, so we do not reject the null that OLS is a consistent estimator.

characteristics in our estimation.

4.3 Household fixed effects model estimation

We further address the potential omitted variable bias concern by employing the household fixed effects model, which exploits multiple observations on commute trips within household. Each commuter in our sample makes only one commute trip, but there are households, each of which is composed of multiple workers and thus of multiple commute trips in the same morning. Our sample therefore has variations in the key variables within household.

We write the household fixed effects model as

$$Y_{ij} = \alpha + \beta T_{ij} + \sum_k \gamma_k m_j^k + X_{ij}' \delta + u_h + \tilde{u_j} + \tilde{w_{ij}} + \epsilon_{ij}, \tag{4}$$

where u_h is unobservable household characteristics common to the workers in the household. This model allows us to eliminate the possibility that unobservable route characteristics around the worker's residence, which is captured in u_h, may cause a biased outcome. The model, however, does not control for unobservable characteristics of the travel route near each household member's job location, which we denote by $\tilde{u_j}$. Also, since each worker within the household may have unobservable characteristics related to her own route choice, we still include $\tilde{w_{ij}}$ in the model.

Table 4 about here.

Table 4 presents the deviation-from-means estimation result for our household fixed effects model.[15] Column (1) indicates that the average queuing time in each morning commute trip is about 1.798 minutes, which is a bit smaller than the OLS estimate of 1.992. Column (2) presents the model, in which multiple arrival time interval dummies are included. We

[15] The household-fixed effects model does not include the family characteristics and the travel day in week dummies since these variables have no variation within household. So, we do not lose missing observations on these variables. Also, we do not exclude the commuters who make multiple commute trips in the same morning. As a result, the household fixed effects model has been estimated using a larger sample than that used for the OLS and the IV estimations.

find that the estimated coefficients from this model are also overall smaller than the OLS estimates from the same model (see column (2) in Table 2).

The next question is whether the household fixed effects model estimation is preferred over the OLS estimation. Note that the OLS estimate of 1.992 is in between the IV estimates (2.283 or 2.402) and the household fixed effects model estimate (1.798). Since we did not find a significant bias in our OLS estimate from the comparison between the OLS and the IV estimates above, it is therefore more difficult to choose the fixed effects model estimate over the OLS estimate. In fact, the smaller β estimate in the household fixed effects model than the pooled OLS estimate is counter-intuitive, since it would then imply that commuters with a smaller u_h (corresponding to a lower level of congestion) more tend to arrive earlier than 6:15.[16]

4.4 Weighting estimation

Under the assumption that our data is a random sample of commute routes in the US, we do not need to weight the route-specific average queuing time by the route mass function $f(j)$ (see (2)). However, the NHTS data contains several "add-on" areas (particular states or cities), where survey participants are oversampled. If the add-on areas are systematically more or less congested than the other areas, then our estimate for the average queuing time in the whole population may be biased.

We consider the weighting estimation methodology for this non-random sampling issue. A drawback of using the weighting estimation in our study is that since it is not equivalent to the maximum likelihood estimation, the standard likelihood ratio tests for over-identifying restrictions for our IV regression cannot be used. We could alternatively adopt the boot-

[16] We also carried out a Hausman test to compare the fixed effects and the random effects models. In both models presented in columns (1) and (2), the Hausman p-value is approximately 0, allowing us to safely reject the null hypothesis that the random effects model is unbiased. However, it does not imply that the fixed effects model estimate is preferred over our OLS estimate, because the fixed effects model allows us to include only the control variables that vary within household whereas the OLS estimation allows us to include a more extensive set of the control variables. The other issue is that the household fixed effects model does not utilize the information from the single-worker households.

strapping methodology used in Brownstone and Golob (2009) and Kim and Brownstone (2013)[17], but that would be useful only when we are certain to mainly use the weighting estimation, which however does not seem to be reasonable in our study. Furthermore, even when we are sure to use the weighting estimation, it is not yet clear as to which weight variable to use among various alternates. As suggested in Solon et al. (2015), we obtain alternate weighted estimates and compare them to our unweighted estimate, which allows us at least to judge to what extent, if any, and to which direction our unweighted estimate may be biased.

We use 3 different weight variables. The first is the one provided by the original 2009 NHTS. The NHTS weight variable is calculated to compensate for the different probabilities of being selected as a sample. For example, if residents in California were oversampled by a ratio of 2 to 1, then the weight of 1/2 is given to each Californian observation.[18] However, since our estimation sample is not the same as the entire NHTS sample, we calculate our own weight variables based on the MSA and the state population sizes, respectively. Like the NHTS weight calculation, we compare the ratio of the number of residents in a particular region (MSA and state) in our sample to that in the US population to construct each of the MSA- and state-based weight variables.

The weighting estimation using the NHTS weight, the MSA weight, and the state weight variables yields the β estimate of 1.646, 2.332, and 1.976, respectively. So, our unweighted estimate of 1.992 does not seem to deviate unidirectionally from these weighted estimates. We also carried out the formal test of DuMouchel and Duncan (1983) to judge whether there is a statistically significant difference between the weighted and the unweighted estimates. The test begins by estimating the model that adds the interaction term of the weight variable with each of all the explanatory variables to the original model.[19] The null hypothesis for our

[17] They tested for over-identifying restrictions by bootstrapping the variance of the difference between the restricted and the unrestricted reduced forms.

[18] The NHTS calculates their weight variable using a more detailed formula and the information on the phone numbers and their associated zip codes. The NHTS then applies the base weight computed in this way to a more complicated formula that accounts for non-responses in their household interview attempts.

[19] The explanatory variables interacted include all the variables presented in Table 2 and the constant

F test is that all the coefficients on the interaction terms are jointly zero. Our estimation result from the morning commutes indicates that when the NHTS weight and the state weight variables are used, the test does not reject the null hypothesis, which implies that there is no significant evidence that the unweighted estimate is biased due to the non-random sampling. We conducted the same analysis using the sample of evening commutes and found similar results. We therefore decided to mainly present the unweighted estimates and use them for our computation of the social cost of congestion.

4.5 Congestion delay in the evening

In this section, we estimate the expected queuing time for each evening commute trip. For estimation, we use the sample of commutes trips from work to home with an arrival time between 15:00 and 23:00. We assume that the commuters who arrived between 21:00 and 23:00 have zero queuing time while those who arrived between 15:00 and 21:00 may have a positive queuing time. The threshold time (t_{1j} in (2)) is set at 21:00 for the same reason as in the morning congestion estimation. In our sample, 95.3% of evening commutes are completed between 15:00 and 21:00, so they are in the treatment group.

Table 5 about here.

Table 5 reports the results for evening congestion, estimated using the unweighted OLS.[20] The estimated coefficient on the treatment variable is 1.818, which means that the commuters who arrived at a time between 15:00 and 21:00 on average spent 1.818 minutes (about 7.4% of the mean travel time in the evening sample) more in commuting than the other commuters on the same route who arrived after 21:00. Based on the mean commute distance of 12.87 miles, the average evening commuter queues for about 0.14 minutes (8.5 seconds) to travel

term, except for the MSA dummies and the day of week dummies. There are 28 interaction terms included.

[20] We also used an IV estimation with the same instruments used in the morning congestion estimation. However, we find that our estimation may suffer from the weak IV problem, although the F statistics is still statistically non-zero. The IV regression adding the service worker dummy to our instrument set seems to be a bit better, but we nevertheless still prefer to present only the OLS estimates. The endogeneity test indicates that the OLS and the IV estimates do not differ significantly.

each mile. In column (2), we present the model including multiple arrival time interval dummies. We find that evening congestion is in its peak for the commuters who arrive around 18:00-19:00. Panel (b) in Figure 2 plots the point estimate for the coefficient on each arrival time dummy from column (2) along with its 95% confidence interval. The figure validates the inversed U-shaped queue dynamics predicted by the bottleneck model.

4.6 Computation of the social cost of congestion

By summing up our OLS estimate for the expected queuing time in the morning and that in the evening (1.992 and 1.818, respectively), we conclude that the average daily queuing time per worker is about 3.81 minutes (about 8% of the sample average of daily commute travel time). The opportunity cost of the queuing time expressed in the dollar unit is calculated by multiplying our estimate (3.81 minutes) by the unit cost of travel time (or simply the value of time). From the extensive literature review by Small and Verhoef (2007), we find that 50% of the wage rate is most appropriate as the value of time for typical workers. Since the Bureau of Labor Statistics reports that the average hourly wage for the US workers in 2010 is about \$24, we therefore use \$12 per hour as the value of time. We conclude that the inefficiency cost of congestion borne by each US worker in a workday is \$0.762 ($= (3.81/60) * 12$).[21]

Here, we briefly comment on the magnitude of congestion delays estimated. Certainly, the congestion delay of about 2 minutes per commute trip sounds small considering the trips that occur on highly congested freeways in big cities. Indeed, our estimation indicates that commuters in particular big cities (such as Los Angeles) traveling a long distance experience a much longer congestion delay than just 2 minutes (see Table 10). This implies that this gap may be explained by the fact that we are using a random sample of all commute

[21] Our monetary inefficiency cost estimate could be underestimated for the following reason. We find below that the queuing time tends to be longer for residents in a larger and richer city, which suggests that the worker's queuing time may be positively correlated to her value of time. However, estimation of this correlation is beyond the scope of this paper, so we leave incorporation of this relationship for future research.

routes including the ones that are never congested, since these routes would offset the huge congestion in the other areas. It may also be explained by the fact that workers in a highly congested area may sort into the locations nearer the job to have a shorter distance for their regular commutes, which would contribute to a small congestion estimated in the absolute scale. Indeed, we find below that per-distance queuing time is not quite small especially for short-distance travelers (see Section 5.3).

We now calculate the inefficiency cost of congestion borne by all US commuters throughout the year. The Bureau of Labor Statistics reports that the total employment in the US in 2009 is about 142 million. The percentage of workers who commuted by a private vehicle (either driving alone or carpooling) in the US in 2009 is 86.1%[22], implying that about 122 million workers used cars for their regular commutes in that year. We also assume that there are 260 workdays in a year. We therefore multiply our estimate for daily inefficiency cost per worker ($0.762) by 122 million (workers) and 260 (days) to conclude that the social cost of congestion borne by all US commuters in a year is about 24 billion dollars. Our estimate is very similar in magnitude to that of Couture et al. (2018), but ours includes only the costs from commute trips while Couture et al. (2018) use the trips of all purposes for their estimation of the Pigouvian deadweight loss. So, we can say that our social cost estimate is a bit larger than that of Couture et al. (2018).

5 Where are more congested and why?

In this section, we investigate the variation of congestion by city and residential location within city.

[22]Source: U.S. Census Bureau, American Community Survey, 2009.

5.1 Congestion ranking of cities

We first construct our congestion ranking of the cities. In our morning commute sample, 18,992 commuters (49%) live in one of the major 49 MSAs, each of which has a population size that is larger than 1 million, and the rest (51%) live in the other area (either an MSA whose population size is smaller than 1 million or in an area that is not an MSA). Using the sample of morning commuters living in each of the 49 MSAs, we estimate the city-specific β, i.e., the expected queuing time for each morning commuter in the city, to construct our congestion ranking of the cities.

Table 6 about here.

Table 6 presents the OLS estimate for β and its 95% confidence interval estimated using the sample of each MSA. Among the 49 MSAs included in the dataset, we present only 17 MSAs, each of which has a sample size of at least 300, since a smaller sample size exhibits a too high standard error and thus an unreliable estimate. Among the 17 cities, we find that Miami and Los Angeles are notably congested, and Washington, San Francisco, and Jacksonville are also highly congested. The proportion of the queuing time out of the travel time for an average commuter in Miami is about 23% while that for Tampa is only 3%. We also find that typical commuters in an MSA whose population size is larger than 1 million have about 3.457 minutes of queuing time while those in the other area have the average queuing time of only 0.679 minutes (see the bottom two rows in the table). Our congestion ranking is comparable to the speed index developed by Couture et al. (2018) and the county-level congestion index of the Texas Transportation Institute (see Schrank and Lomax, 2009). We find that our ranking is overall consistent with these earlier congestion indices.

5.2 Determinants of congestion

We now address the question of why particular cities are more (or less) congested than the other cities. To allow congestion to depend on the city characteristics, we estimate the following model:

$$Y_{ij} = \alpha + \beta_0 T_{ij} + \beta_1 T_{ij} R_c + \sum_k \gamma_k m_j^k + X_{ij}' \delta + u_j + w_{ij} + \epsilon_{ij}, \tag{5}$$

where R_c indicates the characteristic of city c in which commuter i resides in. The expected queuing time, measured by the gap in travel times involved with peak and pre-6:15 arrivals, now depends on the city characteristic and is $\beta_0 + \beta_1 R_c$.[23]

Following Couture et al. (2018), we consider two key city characteristics that potentially determine its congestion level. The first is the road provision in the city in which the commuter resides in as a supply-side congestion determinant. The second is the aggregate daily vehicle mileage traveled (VMT) by the residents in the city as a demand-side determinant. We also investigate within-city determinants of congestion below.

To obtain the information on road stocks in the cities, we use the Highway Performance Monitoring System (HPMS) dataset, which reports the total length of roadways in miles at the county level. The HPMS classifies city roads into 6 categories: (i) interstate highways, (ii) other freeways and expressways, (iii) principal arterial, (iv) minor arterial, (v) collector, and (vi) local roads. We combine (i) and (ii) to define total freeways, which we denote by FWY. We combine (iii)-(v) to make the category of major urban roads, which we denote by MRU. Following Duranton and Turner (2011), we exclude local roads in our computation of the city's road stock, since they are not systematically reported.

Tables 7 about here.

We match each county in the HPMS dataset to each of the 48 MSAs classified in our NHTS estimation sample.[24] Table 7 reports the summary statistics of the road stock variables

[23]The bottleneck model suggests that the average (expected) queuing time for commuters traveling route j is longer as the ratio of the number of route users to the bottleneck capacity (i.e., N_j/ψ_j) is greater (see (1)). Effectively, we are testing this hypothesis using the city's road stock per capita as ψ_j and the city's aggregate daily VMT as N_j, regarding each city as a commute route j.

[24]The HPMS data does not report the road stock information for West Palm Beach, so these observations are dropped from our estimation sample.

constructed at the MSA level. The mean of the total MSA road length is 10,357 miles, with New York endowed with the longest 43,696 miles of roads. The mean of the total MSA FWY length is 294 miles, which is only 3% of the mean of the total MSA road length. However, the share of MSA vehicle mileage traveled (VMT) on FWY out of the total MSA VMT is much larger and is on average about 39% (see the third row from the bottom). The HPMS also offers the total MSA "lane" miles of FWY, whose mean is 1,709, which implies that the average number of lanes of FWY is about 5.8 ($= 1,709/294$). The mean of the total MSA miles of major urban roads (MRU) is 2,602, and the share of VMT on MRU out of the total MSA VMT is on average 47%.

Table 8 about here.

In columns (1)-(3) in Table 8, we report the estimation results for the specification (5), with each MSA road stock variable used for R_c. In these columns (as well as column (4)), we omit the MSA fixed effects and the MSA size category dummies since R_c varies only at the MSA level. We instead control for the population size of the MSA in which the commuter resides in. In column (1), the interacted city characteristic (R_c) is the total FWY length per 1,000 capita of the MSA in which the commuter resides. The estimated coefficient on the interaction term is -14.694, which is significantly negative. We use this estimate to compare the expected queuing time of a commuter in the MSA whose FWY stock is the largest among all the included MSAs to that of a commuter in the MSA whose FWY stock is the smallest. Given that Kansas City has the largest stock of about 0.26 FWY miles per 1,000 capita and Los Angeles has the smallest FWY stock of about 0.06 miles per 1,000 capita, the expected gap in queuing time between the two cities is about 2.94 minutes ($= 14.694 * (0.26 - 0.06)$).

In column (2), we use the "lane" miles of FWY per 1,000 capita as the city characteristic R_c and find that the coefficient on the interaction term is -2.179. Based on this estimate, the expected gap in queuing time between a commuter in the city with the largest FWY lane miles per capita (Kansas City) and the other in the city with the smallest stock per capita (Chicago) is about 2.01 minutes. Finally, in column (3), we use the total MSA MRU length per 1,000 capita as R_c and find that the coefficient on the interaction term is -2.194. The expected gap in queuing time between a commuter in the city with the largest MRU stock per capita (Grand Rapids) and the other in the city with the smallest MRU stock per capita (Miami) is about 2.11 minutes.

These results suggest that the variation of congestion by city is largely explained by the city's road stock per capita. This conclusion is comparable to that of Duranton and Turner (2011) who, unlike us, suggested that construction of new roads would not relieve traffic congestion in the city significantly since it would then induce a proportionate increase in the city's total VMT. A difference is that while we are investigating the level of congestion for the residents in the city, the dependent variable in Duranton and Turner (2011) is the city's total VMT. Another difference is that Duranton and Turner (2011) adopts the instrumental variable strategy to endogeneize their road stock variable while we do not have much causal interpretation here. Most of all, we are estimating using a cross-sectional variation across cities to show that congestion is lower in city with a higher level of road stock. In the meantime, Duranton and Turner (2011) focus on overtime adjustment to highway supply expansion *for given city*. These are quite different identification strategy for estimation of the relationship between road and congestion.

We next consider the demand-side determinant of congestion. The HPMS offers the estimate for aggregate daily VMT for residents in each city. We divide the total MSA VMT by the total MSA road miles (including both FWY and MRU) to construct the total VMT per road mile, which we use as the city characteristic R_c in (5). According the estimation result reported in column (4), the estimated coefficient on the interaction term is significantly positive and is 0.17. Based on this estimate, the expected gap in queuing time between a commuters in the city with the highest overall vehicle travel demand (Miami) and the other in the city with the lowest demand (Milwaukee) is about 3.4 minutes.[25] Thus, our demand-side variable seems to explain a bit larger portion of congestion variation than any of the

[25] In Miami, the aggregate daily VMT per road mile is 32.29, and that in Milwaukee is 12.27, so 3.4 minutes is computed by $0.17 * (32.29 - 12.27)$.

road stock variables.

Finally, we explore within-city determinants of congestion by adding each within-city characteristic interacted with T_{ij} to our main empirical model. In column (5) in Table 8, we consider population density around the neighborhood in which the commuter resides in, measured by population per 1/10 square mile (tract level), as the interacted variable. As expected, the estimated coefficient on the interaction term is positive, implying that commuters residing in a higher density neighborhood tend to have a longer queuing time than those in a lower density neighborhood. In column (6), urban-rural dummies interacted with T_{ij} are included.[26] As expected, the coefficients on the interaction terms are all positive, implying that commuters residing in urban, suburban, and second-city categories have on average a longer queuing time than those residing in the countryside (the left-out category). Finally, in column (7), we include the family income interacted with T_{ij} in our model. We find that queuing times are overall longer for higher-income workers, but the coefficient on the family income itself is negative. Our interpretation is that higher income earners (with a higher value of time) tend to live in a more congested area while trying to reduce her travel time, e.g., by taking an express lane.

5.3 Route distance and queuing time

Our estimation of the expected queuing time is conditioned on the trip distance. However, we expect that commuters traveling a longer distance queue for a longer time than those traveling a shorter distance. We further suspect that the relationship between queuing time and trip distance is concave, which means that the relationship between travel speed and trip distance is positive. This relationship may be due to a fixed time cost of travel, such as the time spent for parking, which would not depend on the trip distance. Also, commuters traveling a longer distance are more likely to be suburban residents who would pass less

[26] The urban-rural dummies measure the resident's neighborhood density in a larger geographic scope. Kim and Brownstone (2013) find that this variable is very helpful in explaining the resident's vehicle travel demand.

congested suburban locations. It may also be due to the fact that commuters in a highly congested area may choose to live nearer the job to have a shorter trip.

To investigate the relationship between queuing time and commute distance, we estimate our main model using each subsample of morning commutes sorted by commute distance. We consider 5 distance groups including $m_j \leq 0.5$, $0.5 < m_j \leq 1$, $1 < m_j \leq 2$, $2 < m_j \leq 4$, and $m_j > 4$, where m_j is trip distance in 10 miles for route j. The estimation result for each subsample is reported in Table 9. We find that the expected queuing time is longer as the commuter travels a longer distance. In particular, the expected queuing time of a commuter whose trip distance is shorter than 5 miles (m_j is in 10 miles) is only 0.328 minutes while that for a commuter whose distance is longer than 40 miles is 5.432 minutes.

Table 9 about here.

To see whether the relationship between queuing time and trip distance is concave, we divide the point estimate for β (expected queuing time) by the mean trip distance in mile from each subsample. The calculation results are presented in the next row. We find that from the second group (with the trip distance between 5 and 10 miles) to the last group (with the distance above 40 miles), per-mile queuing time tends to decrease. This confirms that the relationship between travel time and trip distance is overall concave, or equivalently, that travel speed is an increasing function of the trip distance.[27]

6 Congestion toll

In this section, we try to empirically quantify the optimal congestion toll. It is shown in the bottleneck model that the optimal toll imposed based on the commuter's arrival time potentially fully recovers the inefficiency cost from congestion. The toll achieves this goal by

[27]This result also suggests that our expected queuing time estimate conditioned on the trip distance may be sensitive to non-random sampling of commute routes, especially in terms of distance. Our estimation relies on the belief that the distribution of trip distances in the NHTS sample does not deviate much from its distribution in the population. Our weighting estimation in Section 4.4 supports the hypothesis that non-random sampling does not cause a significant bias.

rescheduling commutes, so that the commuters now depart only at the rate of the bottleneck capacity and thus do not form any queue with no delay in their arrival times relative to under laissez-faire. Under the optimal toll, each commuter attains the same commuting cost as under laissez-faire, which implies that the toll revenue is exactly the same as the inefficiency cost under laissez-faire. It is also shown that the optimal toll amount is exactly same as the laissez-faire queuing time cost as a function of arrival time, i.e., $\alpha(Y(t) - Y(t_0))$ for $t \in [t_0, t_1]$, where α is the monetary unit cost of travel time (value of time).

This paper has provided the framework to estimate the queuing time $Y(t) - Y(t_0)$ as a function of arrival time (t) conditional on the commute route, which would therefore quantify the optimal toll for the commuters traveling the given route (with an appropriate use of α). However, since we do not observe commute routes directly, we are unable to perfectly identify $Y(t) - Y(t_0)$ for each separate route. In this paper, we sort the commuters by city and commute distance, viewing the commuters sorted in each group traveling an identical route. We then estimate the queuing time as a function of t, i.e., $Y(t) - Y(t_0)$, for the commuters in each group.

Table 10 about here.

Table 10 shows the estimation results from two representative MSAs, one that is highly congested Los Angeles and the other that is less congested Greensboro. Each model in the table includes multiple arrival time interval dummies and is estimated separately for the 4 commute-distance groups using the sample of morning commuters residing in each city. For example, according to the table, a commuter in Los Angeles whose trip distance is 10-20 miles has the expected queuing time of 10.122 minutes if she would arrive around 8:00. When \$12 per hour is used as the value of time, this commuter would be charged a toll of about \$2.02. As expected, the expected queuing time as a function of arrival time and thus the optimal toll has an inversed U shape. Also, the commuters traveling a longer distance queue for a longer time and thus are charged a higher toll. We also find that the estimated coefficients on the arrival time dummies are much smaller for the Greensboro residents than for the Los Angeles residents, which implies that the residents in Greensboro are charged a much smaller (or zero) congestion toll compared to the Los Angeles residents.

A city government may adopt our framework to quantify a practically implementable toll that would be a function of arrival time and trip distance. A more detailed information on work-home locations (commute routes) of a larger sample of commuters would improve the estimates from ours. If the employer's interest were to be consistent with the city government's, then the employer would identify each employee's hypothetical travel times as a function of arrival time using real-time traffic information (e.g., Google Maps) and the employee's home address, which would then give the employee-specific time-varying toll charge.

7 Conclusion

This paper is the first study that formally used the bottleneck model framework to empirically measure the social cost of congestion. We find that the inefficiency time cost for each average daily US commuter is about 3.8 minutes (about 8% of the sample mean of travel time). Although this individual cost sounds small, our calculation shows that the annual social cost of congestion from commute trips wasted in the US is about 24 billion dollars.

We have also investigated the variation in congestion over time in a day and by residential location (especially by city). An important finding is that a higher level of congestion in a city may be attributed to an insufficient road stock in the city. This finding, however, does not imply that an increase in road stock would have a welfare benefit that is comparable to imposition of the optimal congestion toll. This paper has provided a meaningful starting point toward empirical quantification of the optimal congestion toll.

The most interesting extension from our estimation would use "big data", such as traffic data from Google Maps, as well as the information on actual commute routes (home and work addresses) to estimate the queuing times and the congestion tolls in a greater detail. Another future work may incorporate trips of all purposes (not just commutes) to estimate the social cost of congestion from all trips. One could also add some causal interpretations to our estimated relationship between road stock (or vehicle travel demand) and the level of congestion.

References

Akbar, P. A. and Duranton, G. (2017). Measuring congestion in a highly congested city: Bogota. Unpublished paper.

Angrist, J. D. and Pischke, J.-S. (2009). *Mostly harmless econometrics: An empiricist's companion*. Princeton University Press.

Arnott, R., de Palma, A., and Lindsey, R. (1990). Economics of a bottleneck. *Journal of Urban Economics*, 27(1):111–130.

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

Arnott, R., de Palma, A., and Lindsey, R. (1994). The welfare effects of congestion tolls with heterogeneous commuters. *Journal of Transport Economics and Policy*, 28(2):139–161.

Brownstone, D. and Golob, T. F. (2009). The impact of residential density on vehicle usage and energy consumption. *Journal of Urban Economics*, 65(1):91–98.

Brueckner, J. K. (2011). *Lectures on urban economics*. MIT Press, Cambridge.

Couture, V., Duranton, G., and Turner, M. (2018). Speed. *Review of Economics and Statistics*, Forthcoming.

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

DuMouchel, W. H. and Duncan, G. J. (1983). Using sample survey weights in multiple regression analyses of stratified samples. *Journal of the American Statistical Association*, 78(383):535–543.

Duranton, G. and Turner, M. (2011). The fundamental law of road congestion: Evidence from us cities. *The American Economic Review*, 101(6):2616–2652.

Fosgerau, M. and Kim, J. (2017). Commuting in a city with bottlenecks: Theory and evidence. Unpublished paper.

Fosgerau, M. and Small, K. A. (2012). Marginal congestion cost on a dynamic expressway network. *Journal of Transport Economics and Policy*, 46(3):431–450.

Hall, J. D. (2018). Can tolling help everyone? Estimating the aggregate and distributional consequence of congestion pricing. Unpublished paper.

Kim, J. and Brownstone, D. (2013). The impact of residential density on vehicle usage and fuel consumption: Evidence from national samples. *Energy Economics*, 40:196–206.

Li, S., Purevjav, A., and Yang, J. (2018). The marginal cost of traffic congestion and road pricing: Evidence from a natural experiment in beijing. Unpublished paper.

Pigou, A. (1920). *The Economics of Welfare*. London: Macmillan and Co.

Schrank, D. and Lomax, T. (2009). Urban mobility report. Technical report, Texas Transportation Institute.

Small, K. A. (1982). The scheduling of consumer activities: Work trips. *The American Economic Review*, 72(3):467–479.

Small, K. A. and Verhoef, E. T. (2007). *The Economics of Urban Transportation*. Routledge, London.

Solon, G., Haider, S. J., and Wooldridge, J. M. (2015). What are we weighting for?. *Journal of Human Resources*, 50(2):301–316.

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

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

Figure 1: Travel time as a function of arrival time

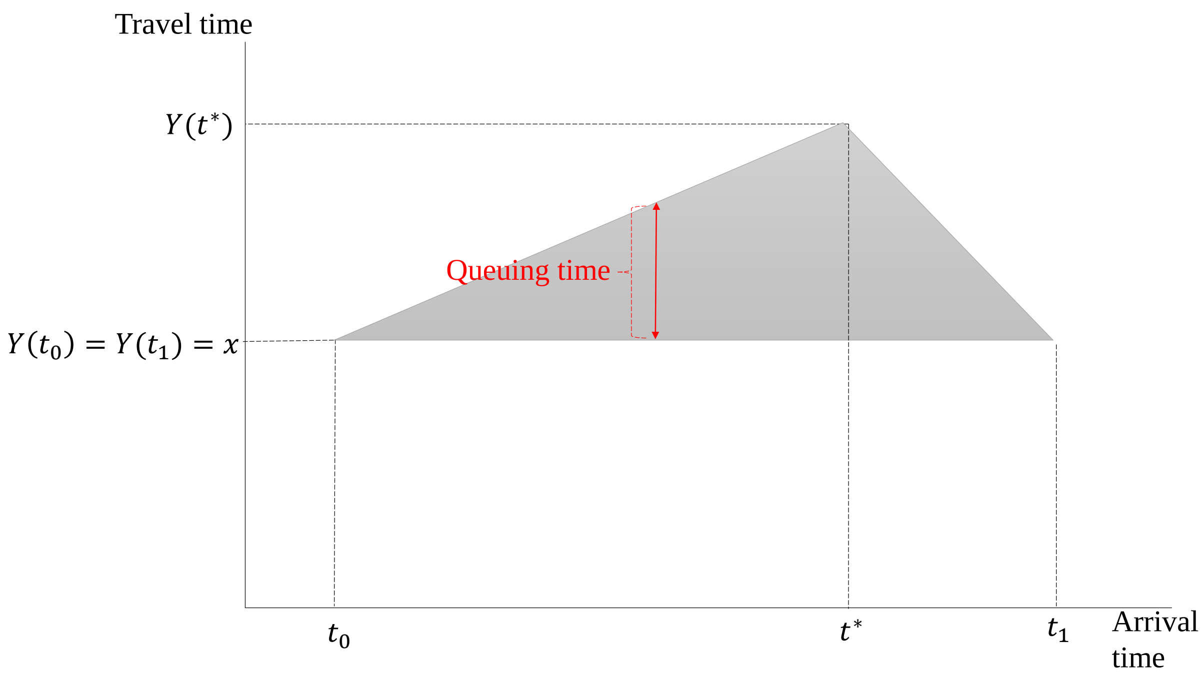
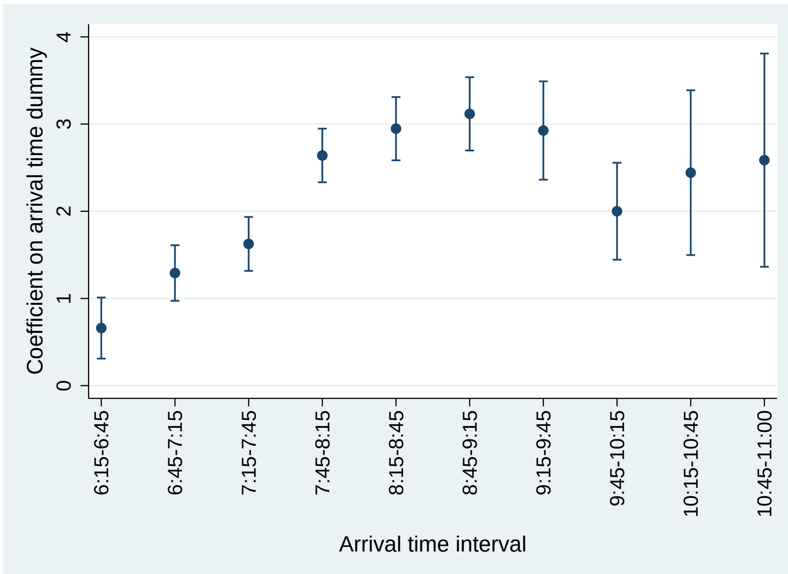
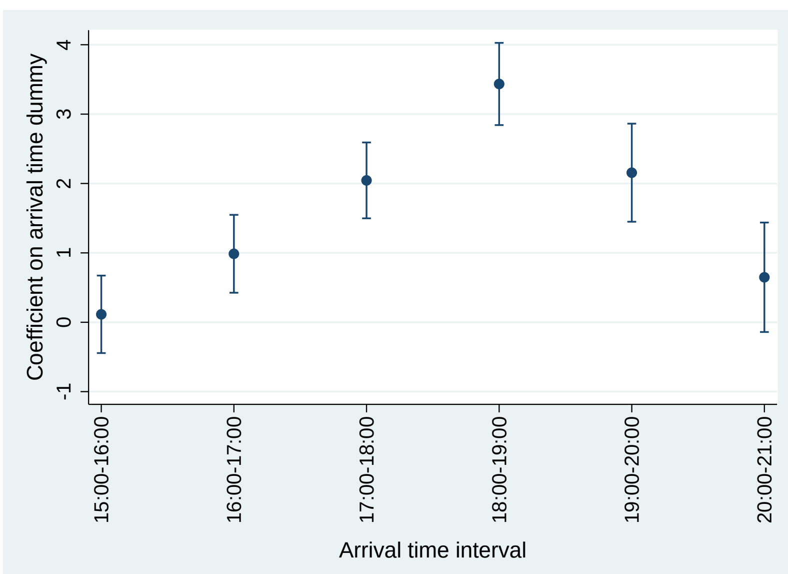


Figure 2: Queuing time estimate by arrival time

(a) Morning



(b) Evening



Notes: In panel (a), each dot represents the estimated coefficient on each arrival time interval dummy from the specification presented in column (2) in Table 2. Vertical spikes around each point estimate represent the 95% confidence interval constructed using robust standard errors. Panel (b) plots the estimated coefficients from the sample of evening commutes, which is presented in column (2) in Table 5, as well as their confidence intervals.

Table 1: Summary statistics

	Mean	Std Dev
Key variables		
Travel time (Y)	22.88	(17.35)
Arrived between 6:15 and 11:00 (T)	0.91	
Arrival time at work (in minute)	7:47 am	(69.02)
Commute distance in 10 miles (m)	1.30	(1.37)
Control variables [a]		
Population per 1/10 sq mile (population density) [b]	27.63	(39.24)
% of rental units relative to owned units [b]	25.11	(18.21)
Neighborhood category is urban [c]	0.08	
Neighborhood category is second city	0.17	
Neighborhood category is suburban	0.25	
Neighborhood category is town	0.50	
Family income in 1,000 dollars	78.29	(34.61)
Family size	2.88	(1.26)
Number of workers	1.78	(0.71)
Race is white	0.87	
Female	0.49	
Age	48.45	(11.90)
Bachelor degree	0.55	
Graduate degree	0.20	
Job category is service	0.21	
Job category is manufacturing	0.15	
Job category is professional	0.50	
Job category is clerical	0.15	

Notes: The summary statistics are calculated from the estimation sample of morning commutes. The sample size is 38,868. [a]The list of the control variables presented in this table is not exhaustive. [b]These variables are measured at the tract level. [c]The urban-rural category variable measures "contextual density", the density of a larger and more relevant geographic scope (see Kim and Brownstone (2013) for a more detailed explanation on this variable).

Table 2: OLS estimation results

Dependent variable: Travel time (Y)	(1)		(2)	
Trip timing choice variables				
Arrived between 6:15 and 11:00 (T)	1.992***	(0.139)		
Arrived between 6:15 and 6:45			0.661***	(0.179)
Arrived between 6:45 and 7:15			1.292***	(0.163)
Arrived between 7:15 and 7:45			1.625***	(0.158)
Arrived between 7:45 and 8:15			2.640***	(0.157)
Arrived between 8:15 and 8:45			2.947***	(0.185)
Arrived between 8:45 and 9:15			3.117***	(0.215)
Arrived between 9:15 and 9:45			2.926***	(0.288)
Arrived between 9:45 and 10:15			2.000***	(0.284)
Arrived between 10:15 and 10:45			2.443***	(0.482)
Arrived between 10:45 and 11:00			2.586***	(0.624)
Distance and its polynomials				
m	14.042***	(0.206)	14.137***	(0.206)
m^2	-0.674***	(0.066)	-0.689***	(0.066)
m^3	0.033***	(0.005)	0.033***	(0.005)
m^4	-0.0005***	(0.0001)	-0.0005***	(0.0001)
Other control variables				
Population density	0.019***	(0.002)	0.019***	(0.002)
% of rental units relative to owned units	-0.020***	(0.003)	-0.021***	(0.003)
Neighborhood category is urban [a]	1.138***	(0.262)	1.088***	(0.260)
Neighborhood category is second city	0.476***	(0.141)	0.459***	(0.141)
Neighborhood category is suburban	0.931***	(0.126)	0.884***	(0.125)
MSA pop size is 250k-500k [b]	0.092	(0.160)	0.104	(0.160)
MSA pop size is 500k-1,000k	0.439***	(0.151)	0.414***	(0.151)
Family income in 1,000 dollars	-0.0044***	(0.0015)	-0.0054***	(0.0015)
Family size	0.040	(0.058)	0.035	(0.057)
Number of workers	-0.173**	(0.072)	-0.170**	(0.071)
Number of vehicles owned	-0.082*	(0.042)	-0.061	(0.042)
Carpool	0.094	(0.196)	0.159	(0.195)
Owner-occupier	-0.007	(0.163)	-0.032	(0.162)
Has kids	-0.344**	(0.137)	-0.404***	(0.137)
Race is white [c]	-0.565***	(0.196)	-0.524***	(0.196)
Race is black	0.482	(0.332)	0.568*	(0.331)
Race is hispanic	0.463*	(0.239)	0.510**	(0.238)
Female	-0.091	(0.083)	-0.280***	(0.084)
Age	0.036***	(0.004)	0.037***	(0.004)
Bachelor degree [d]	-0.045	(0.103)	-0.175*	(0.103)
Graduate degree	0.103	(0.132)	-0.146	(0.132)
Job category is service	-0.283***	(0.104)	-0.443***	(0.105)
n	38,868		38,868	
R^2	0.788		0.789	

Notes: Robust standard errors are in parentheses. All models include a constant, MSA fixed effects, and day in week dummies. In each model, the left-out group is the commuters who arrived before or at 6:15. [a]The left-out group is those living in town or country. [b]The left-out group is those living in an MSA with a population size less than 250k. Residents in an MSA with a population size over 1,000k are controlled by the MSA fixed effects. [c]The left-out group is Asians and other minorities. [d]The left-out group is those with a high school diploma or a lower education level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 3: IV estimation results

Dependent variable:	Travel time (Y)		Travel time (Y)		T	
	(1)		(2)		(3)	
Trip timing choice variable						
Arrived between 6:15 and 11:00 (T)	2.283**	(1.029)	2.402***	(0.783)		
Distance and its polynomials						
m	14.049***	(0.210)	14.049***	(0.208)	-0.0236***	(0.0032)
m^2	-0.675***	(0.066)	-0.674***	(0.066)	0.0032***	(0.0008)
m^3	0.033***	(0.005)	0.033***	(0.005)	-0.00011**	(0.00004)
m^4	-0.0005***	(0.0001)	-0.0005***	(0.0001)	0.0000011*	(0.0000006)
Other control variables						
Population density	0.019***	(0.002)	0.019***	(0.002)	0.00005	(0.0006)
% of rental units relative to owned units	-0.020***	(0.003)	-0.020***	(0.003)	-0.00009	(0.00009)
Neighborhood category is urban	1.140***	(0.261)	1.142***	(0.261)	-0.0096	(0.0082)
Neighborhood category is second city	0.477***	(0.141)	0.476***	(0.141)	-0.0042	(0.0050)
Neighborhood category is suburban	0.930***	(0.126)	0.931***	(0.126)	0.0019	(0.0041)
MSA pop size is 250k-500k	0.093	(0.160)	0.094	(0.160)	-0.0039	(0.0065)
MSA pop size is 500k-1,000k	0.435***	(0.152)	0.434***	(0.151)	0.0145**	(0.0059)
Family income in 1,000 dollars	-0.0045***	(0.0015)	-0.0044***	(0.0015)	0.0002***	(0.0001)
Family size	0.040	(0.058)	0.041	(0.058)	-0.0001	(0.0019)
Number of workers	-0.174**	(0.072)	-0.178**	(0.072)	0.0042*	(0.0025)
Number of vehicles owned	-0.081*	(0.042)	-0.082*	(0.042)	-0.0005	(0.0015)
Carpool	0.089	(0.195)	0.091	(0.196)	0.0099*	(0.0060)
Owner-occupier	-0.008	(0.163)	-0.011	(0.163)	0.0031	(0.0063)
Has kids	-0.347**	(0.137)	-0.348**	(0.137)	0.0108**	(0.0045)
Race is white	-0.569***	(0.196)	-0.579***	(0.196)	0.0119**	(0.0060)
Race is black	0.486	(0.332)	0.480	(0.332)	-0.0130	(0.0094)
Race is hispanic	0.468**	(0.239)	0.472**	(0.239)	-0.0114*	(0.0069)
Female	-0.113	(0.112)	-0.125	(0.100)	0.0416***	(0.0031)
Age	0.036***	(0.004)	0.037***	(0.004)	-0.0005***	(0.0001)
Bachelor degree	-0.065	(0.121)			0.0449***	(0.0043)
Graduate degree	0.069	(0.172)			0.0839***	(0.0048)
Job category is service	-0.290***	(0.105)	-0.302***	(0.103)	-0.0322***	(0.0043)
Job category is manufacturing					-0.1596***	(0.0069)
Job category is professional					-0.0304***	(0.0037)
n	38,868		38,868		38,868	
R^2	0.788		0.788		0.074	
First-stage F statistic	267.1		247.2			
Overidentification test p-value	0.260		0.398			
Endogeneity test p-value	0.775		0.610			

Notes: Robust standard errors are in parentheses. All models include a constant, MSA fixed effects, and day in week dummies. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 4: Household fixed-effects model estimation results

Dependent variable: Travel time (Y)		
	(1)	(2)
Arrived between 6:15 and 11:00 (T)	1.798***	
	(0.311)	
Arrived between 6:15 and 6:45		0.704*
		(0.378)
Arrived between 6:45 and 7:15		1.231***
		(0.359)
Arrived between 7:15 and 7:45		1.596***
		(0.366)
Arrived between 7:45 and 8:15		2.542***
		(0.362)
Arrived between 8:15 and 8:45		2.730***
		(0.408)
Arrived between 8:45 and 9:15		2.588***
		(0.459)
Arrived between 9:15 and 9:45		2.513***
		(0.517)
Arrived between 9:45 and 10:15		2.161***
		(0.542)
Arrived between 10:15 and 10:45		1.805***
		(0.614)
Arrived between 10:45 and 11:00		2.748***
		(0.945)
m	14.985***	15.065***
	(0.446)	(0.442)
m^2	-1.043***	-1.060***
	(0.167)	(0.165)
m^3	0.073***	0.074***
	(0.016)	(0.016)
m^4	-0.0015***	-0.0016***
	(0.0004)	(0.0004)
Carpool	0.433	0.462
	(0.419)	(0.414)
Female	-0.267**	-0.465***
	(0.128)	(0.132)
Age	0.018*	0.022**
	(0.009)	(0.009)
Bachelor degree	-0.002	-0.056
	(0.246)	(0.247)
Graduate degree	-0.578*	-0.615*
	(0.327)	(0.327)
Job category is service	-0.498**	-0.638***
	(0.222)	(0.224)
n	41,724	41,724
Number of groups	33,960	33,960
R^2	0.795	0.796

Notes: Robust standard errors are in parentheses. All models include a constant and MSA fixed effects. Only the variables that have variation within household are included in each model. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 5: OLS estimation results from the sample of evening commutes

Dependent variable: Travel time (Y)		
	(1)	(2)
Arrived between 15:00 and 21:00	1.818***	
	(0.271)	
Arrived between 15:00 and 16:00		0.114
		(0.285)
Arrived between 16:00 and 17:00		0.987***
		(0.286)
Arrived between 17:00 and 18:00		2.045***
		(0.279)
Arrived between 18:00 and 19:00		3.435***
		(0.302)
Arrived between 19:00 and 20:00		2.156***
		(0.361)
Arrived between 20:00 and 21:00		0.648
		(0.403)
n	33,994	33,994
R^2	0.734	0.737

Notes: Robust standard errors are in parentheses. All models include the same control variables as in Table 2, as well as a constant, MSA fixed effects, and day in week dummies. In each model, the left-out group is the commuters who arrived after 21:00. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 6: Queuing time estimate by city

Sample category	β estimate	95% confidence interval for β	β per mile [b]	n
Miami-Fort Lauderdale, FL	6.227	[3.390, 9.064]	0.506	742
Los Angeles-Riverside-Orange County, CA	5.743	[4.331, 7.156]	0.395	1,819
Washington-Baltimore, DC-MD-VA-WV	4.749	[2.448, 7.049]	0.249	759
San Francisco-Oakland-San Jose, CA	4.651	[2.729, 6.573]	0.336	1,005
Jacksonville, FL	4.508	[2.073, 6.943]	0.300	315
Houston-Galveston-Brazoria, TX	4.116	[2.635, 5.597]	0.265	1,190
Dallas-Fort Worth, TX	4.020	[2.424, 5.616]	0.273	1,840
Atlanta, GA	3.551	[-0.629, 7.731]	0.220	309
New York-Northern New Jersey-Long Island, NY-NJ-CT-PA	3.190	[1.597, 4.783]	0.207	1,560
San Antonio, TX	3.123	[1.248, 5.000]	0.233	571
San Diego, CA	2.543	[1.385, 3.701]	0.187	1,618
Sacramento-Yolo, CA	2.410	[0.349, 4.472]	0.186	322
Austin-San Marcos, TX	2.244	[-1.667, 6.156]	0.171	493
Orlando, FL	2.011	[-1.604, 5.627]	0.155	303
Greensboro-Winston-Salem-High Point, NC	1.594	[0.019, 3.169]	0.129	1,590
Norfolk-Virginia Beach-Newport News, VA-NC	1.586	[-0.380, 3.552]	0.130	995
Tampa-St. Petersburg-Clearwater, FL	0.715	[-5.272, 6.702]	0.060	475
MSAs whose population is larger than 1 million [a]	3.457	[3.024, 3.890]	0.248	18,992
MSAs whose population is smaller than 1 million and non-MSAs	0.679	[0.346, 1.013]	0.056	19,876

Notes: The listed MSAs have a population size that is larger than 1 million and a sample size that is greater than 300. The estimated model is the same as the basic model presented in the first column in Table 2, except that we do not include the MSA fixed effects and the MSA population categories. [a]For estimation using this sample, we do include the MSA fixed effects. [b]For each MSA, the β estimate is divided by the MSA's mean trip distance in mile.

Table 7: Summary statistics of the MSA-level road provision variables

	Mean	Std Dev	Min	Max	Obs
Miles of all roadways	10,357	8,196	3,203	43,696	48
Miles of freeways (FWY) [a]	294	217	67	1,214	48
Miles of FWY per 1,000 MSA population	0.12	0.05	0.06	0.26	48
Lane miles of FWY	1,709	1,423	375	7,225	48
Lane miles of FWY per 1,000 MSA population	0.68	0.21	0.33	1.25	48
Miles of major urban roads (MRU) [b]	2,602	2,298	695	11,327	48
Miles of MRU per 1,000 MSA population	0.99	0.25	0.55	1.51	48
Share of vehicle mileage traveled on FWY	0.39	0.08	0.22	0.55	48
Share of vehicle mileage traveled on MRU	0.47	0.07	0.32	0.63	48
MSA population in 1 million	2.98	3.41	0.71	18.70	48

Notes: The data source is the 2008 edition of the Highway Performance and Monitoring System (HPMS), which has the information on road provision for all large MSAs (over 1 million) in our NHTS sample, except for the West Palm Beach MSA. [a]FWY includes the HPMS categories of "interstate highways" and "other freeways and expressways". [b]MRU includes the HPMS categories of "principle arterial", "minor arterial", and "collector". We exclude the HPMS category of "local roads".

Table 8: Estimation results from the models including interaction term

Dependent variable: Travel time (Y)							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Interaction term							
$T \times$ Miles of FWY per 1,000 capita	-14.694***						
	(4.107)						
$T \times$ Lane miles of FWY per 1,000 capita		-2.179**					
		(0.912)					
$T \times$ Miles of MRU per 1,000 capita			-2.194***				
			(0.813)				
$T \times$ Total VMT (in 1000) per road mile				0.170***			
				(0.039)			
$T \times$ Population density					0.0195***		
					(0.0042)		
$T \times$ Urban						2.518***	
						(0.525)	
$T \times$ Second city						0.712*	
						(0.383)	
$T \times$ Suburban						1.829***	
						(0.343)	
$T \times$ Family income in 1,000 dollars							0.0270***
							(0.0041)
Other variables (list of the reported variables is not exhaustive) [a]							
Arrived between 6:15 and 11:00 (T)	5.177***	4.946***	5.393***	-0.428	1.462***	1.270***	0.111
	(0.539)	(0.672)	(0.767)	(0.907)	(0.167)	(0.179)	(0.322)
Miles of FWY per 1,000 capita	1.350						
	(4.124)						
Lane miles of FWY per 1,000 capita		-0.852					
		(0.909)					
Miles of MRU per 1,000 capita			1.540*				
			(0.788)				
Total VMT (in 1000) per road mile				-0.101***			
				(0.038)			
MSA population in 1 million [b]	0.019	0.015	0.075***	0.079***			
	(0.016)	(0.015)	(0.015)	(0.013)			
Population density	0.0183***	0.0186***	0.0187***	0.0184***	0.0014	0.0188***	0.0188***
	(0.0022)	(0.0022)	(0.0022)	(0.0023)	(0.0043)	(0.0022)	(0.0022)
Urban	1.491***	1.500***	1.897***	1.731***	1.133***	-1.145**	1.131***
	(0.283)	(0.282)	(0.279)	(0.282)	(0.261)	(0.532)	(0.262)
Second city	0.636***	0.671***	0.838***	0.745***	0.466***	-0.175	0.482***
	(0.219)	(0.219)	(0.218)	(0.219)	(0.141)	(0.373)	(0.141)
Suburban	1.202***	1.224***	1.452***	1.347***	0.923***	-0.746**	0.932***
	(0.184)	(0.183)	(0.181)	(0.184)	(0.126)	(0.336)	(0.126)
Family income in 1,000 dollars	-0.0027	-0.0024	-0.0019	-0.0023	-0.0043***	-0.0044***	-0.0289***
	(0.0023)	(0.0023)	(0.0023)	(0.0023)	(0.0015)	(0.0015)	(0.0039)
MSA fixed effects [c]	No	No	No	No	Yes	Yes	Yes
n	18,788	18,788	18,788	18,788	38,868	38,868	38,868
R^2	0.743	0.743	0.742	0.742	0.788	0.788	0.788

Notes: Robust standard errors are in parentheses. [a]Other than the control variables reported in this table, all models include the same control variables presented in Table 2 as well as a constant and day in week dummies. [b]For the MSA population information, we use the 2008 edition of the HPMS, which we are using to acquire the MSA road provision information. [c]The models in columns (1)-(4) do not include the MSA effects, because they include the road provision and the total VMT variables, which are at the MSA level. In these columns, we use the sample of commuters who live in the 48 large MSAs. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 9: Queuing time estimate by trip distance

Dependent variable: Travel time (Y)					
	$m \leq 0.5$	$0.5 < m \leq 1$	$1 < m \leq 2$	$2 < m \leq 4$	$m > 4$
Coefficient on arrival between	0.328*	1.706***	2.143***	3.696***	5.432***
6:15 and 11:00 (β)	(0.168)	(0.217)	(0.256)	(0.403)	(1.216)
β per mile	0.116	0.216	0.144	0.132	0.093
n	12,570	9,121	9,790	5,893	1,494
R^2	0.309	0.185	0.207	0.279	0.711

Notes: Robust standard errors are in parentheses. All models include the same control variables as in Table 2, as well as a constant, MSA fixed effects, and day in week dummies. OLS is used for estimation. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 10: Queuing time estimate by arrival time and trip distance (Los Angeles and Greensboro)

Dependent variable: Travel time (Y)								
	Los Angeles MSA				Greensboro MSA			
	$m \leq 0.5$	$0.5 < m \leq 1$	$1 < m \leq 2$	$m > 2$	$m \leq 0.5$	$0.5 < m \leq 1$	$1 < m \leq 2$	$m > 2$
Arrived between 6:15 and 6:45	3.764	0.979	1.512	5.635**	-1.824	0.987	0.540	3.594
	(2.617)	(1.402)	(1.645)	(2.834)	(2.567)	(1.041)	(1.706)	(2.634)
Arrived between 6:45 and 7:15	2.039*	3.580**	3.734*	3.136	-0.811	0.847	1.738	1.287
	(1.118)	(1.484)	(1.938)	(2.400)	(2.549)	(1.026)	(1.634)	(2.267)
Arrived between 7:15 and 7:45	1.848**	2.310*	3.569**	6.998***	-1.502	1.085	0.008	2.857
	(0.903)	(1.341)	(1.509)	(2.653)	(2.599)	(0.912)	(1.618)	(2.214)
Arrived between 7:45 and 8:15	2.421***	3.479**	10.122***	11.232***	-0.691	3.194***	1.421	3.770*
	(0.823)	(1.584)	(1.680)	(2.664)	(2.603)	(0.961)	(1.518)	(2.131)
Arrived between 8:15 and 8:45	3.311***	5.313***	7.822***	14.841***	-0.603	3.065***	1.977	2.776
	(0.974)	(1.600)	(2.090)	(2.999)	(2.594)	(1.092)	(1.617)	(2.309)
Arrived between 8:45 and 9:15	1.763*	8.382***	7.461***	15.240***	-1.750	2.290**	0.243	4.970*
	(0.932)	(2.362)	(2.165)	(3.176)	(2.607)	(1.057)	(2.058)	(2.733)
Arrived between 9:15 and 9:45	2.048*	6.616**	9.475***	19.207***	-0.892	3.000**	3.889	3.167
	(1.144)	(2.734)	(2.878)	(3.510)	(2.604)	(1.517)	(2.458)	(3.300)
Arrived between 9:45 and 10:15	5.790**	5.589***	2.748	10.476**	-0.725	0.907	0.182	5.331
	(2.425)	(1.873)	(2.076)	(4.655)	(2.587)	(1.607)	(1.731)	(5.410)
Arrived between 10:15 and 10:45	2.970*	4.934**	6.496**	6.035	-1.582	5.011	3.295	8.877
	(1.620)	(2.321)	(2.703)	(4.984)	(2.561)	(3.857)	(4.189)	(5.508)
Arrived between 10:45 and 11:00	3.655***	2.228	-7.348***	9.861	2.263	4.354***	5.470	2.443
	(1.210)	(2.567)	(1.981)	(6.788)	(3.302)	(1.655)	(3.321)	(6.626)
n	542	394	459	424	426	447	467	250
R^2	0.338	0.223	0.283	0.612	0.341	0.174	0.280	0.759

Notes: Robust standard errors are in parentheses. All models include the same control variables as in Table 2 except for the MSA fixed effects and the MSA size dummies. OLS is used for estimation. The Los Angeles MSA includes Los Angeles, Riverside, and Orange Counties. The Greensboro MSA includes Greensboro, Winston-Salem, and High Point counties. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.