SJSU SAN JOSÉ STATE UNIVERSITY

MTI MINETA TRANSPORTATION INSTITUTE

Congestion Costs and Scheduling Preferences of Car Commuters in California: Estimates Using Big Data

Jinwon Kim
Jucheol Moon

CSUTC California State University Transportation Consortium

CALIFORNIA STATE UNIVERSITY LONG BEACH

CSU TRANSPORTATION CONSORTIUM

transweb.sjsu.edu/csutc

Mineta Transportation Institute

Founded in 1991, the Mineta Transportation Institute (MTI), an organized research and training unit in partnership with the Lucas College and Graduate School of Business at San José State University (SJSU), increases mobility for all by improving the safety, efficiency, accessibility, and convenience of our nation's transportation system. Through research, education, workforce development, and technology transfer, we help create a connected world. MTI leads the Mineta Consortium for Transportation Mobility (MCTM) funded by the U.S. Department of Transportation and the California State University Transportation Consortium (CSUTC) funded by the State of California through Senate Bill 1. MTI focuses on three primary responsibilities:

Research

MTI conducts multi-disciplinary research focused on surface transportation that contributes to effective decision making. Research areas include: active transportation; planning and policy; security and counterterrorism; sustainable transportation and land use; transit and passenger rail; transportation engineering; transportation finance; transportation technology; and workforce and labor. MTI research publications undergo expert peer review to ensure the quality of the research.

Education and Workforce

To ensure the efficient movement of people and products, we must prepare a new cohort of transportation professionals who are ready to lead a more diverse, inclusive, and equitable transportation industry. To help achieve this, MTI sponsors a suite of workforce development and education opportunities. The Institute supports educational programs offered by the Lucas Graduate School of Business: a Master of Science in Transportation Management, plus graduate certificates that include High-Speed and Intercity Rail Management and Transportation Security Management. These flexible programs offer live online classes so that working transportation professionals can pursue an advanced degree regardless of their location.

Information and Technology Transfer

MTI utilizes a diverse array of dissemination methods and media to ensure research results reach those responsible for managing change. These methods include publication, seminars, workshops, websites, social media, webinars, and other technology transfer mechanisms. Additionally, MTI promotes the availability of completed research to professional organizations and works to integrate the research findings into the graduate education program. MTI's extensive collection of transportation-related publications is integrated into San José State University's world-class Martin Luther King, Jr. Library.

Report 21-34

Congestion Costs and Scheduling Preferences of Car Commuters in California: Estimates Using Big Data

Jinwon Kim
Jucheol Moon

February 2022

A publication of the
Mineta Transportation Institute
Created by Congress in 1991

College of Business
San José State University
San José, CA 95192-0219

TECHNICAL REPORT DOCUMENTATION PAGE

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.	
21-34			
4. Title and Subtitle Congestion Costs and Scheduling Preferences of Car Commuters in California: Estimates Using Big Data		**5. Report Date** February 2022 **6. Performing Organization Code**	
7. Authors Jinwon Kim: 0000-0001-8209-4319 Jucheol Moon: 0000-0002-2885-0627		**8. Performing Organization Report** CA-MTI-2031	
9. Performing Organization Name and Address Mineta Transportation Institute College of Business San José State University San José, CA 95192-0219		**10. Work Unit No.** **11. Contract or Grant No.** ZSB12017-SJAUX	
12. Sponsoring Agency Name and Address State of California SB1 2017/2018 Trustees of the California State University Sponsored Programs Administration 401 Golden Shore, 5th Floor Long Beach, CA 90802		**13. Type of Report and Period Covered** **14. Sponsoring Agency Code**	
15. Supplemental Notes 10.31979/mti.2022.2031			
16. Abstract On average, California car commuters waste 4–5 minutes per morning commute due to congestion. Multiplied across all California car commuters, those few minutes entail a yearly total of approximately 2.3 billion hours of time wasted, costing 6 billion dollars. The objective of this study is to quantify congestion costs and determine how commuters adapt to the level of congestion they face (i.e., commuters' scheduling utility functions). To that end, this research developed a model of trip scheduling under congestion to construct California commuters' travel-time profiles, i.e., the menu of travel times that each individual would likely face according to alternate trip timing choices. The results show that commuters facing higher levels of congestion tend to avoid delays by arriving at an inconvenient edge time rather than commuting during the peak. Further, commuters are willing to accept about 0.5 additional minutes of schedule delay to reduce travel time by 1 minute. We found that for most commuters in our data, the travel time profile is much flatter than the estimated schedule utility, which implies that commuters tend to arrive around their own ideal arrival times, although the estimated utility function exhibits a moderate schedule inflexibility. This finding ultimately calls into question the existing bottleneck model's quantification of the economic cost of congestion as well as the optimal toll to ameliorate congestion.			
17. Key Words Traffic congestion, bottlenecks, congestion pricing, data management, artificial intelligence	**18. Distribution Statement** No restrictions. This document is available to the public through The National Technical Information Service, Springfield, VA 22161.		
19. Security Classif. (of this report) Unclassified	**20. Security Classif. (of this page)** Unclassified	**21. No. of Pages** 60	**22. Price**

Form DOT F 1700.7 (8-72)

DOI: 10.31979/mti.2022.2031

Mineta Transportation Institute
College of Business
San José State University
San José, CA 95192-0219

Tel: (408) 924-7560
Email: mineta-institute@sjsu.edu

transweb.sjsu.edu/research/2031

ACKNOWLEDGEMENTS

Jinwon Kim* and Jucheol Moon† thank Andre de Palma, Ken Small, Erik Verhoef, and seminar participants at the 2021 ITEA and the KER International Conference for helpful comments.

* Department of Economics, Sogang University, Seoul, Korea.
Email: jinwonkim@sogang.ac.kr

† Department of Computer Engineering & Computer Science, California State University, Long Beach.
Email: jucheol.moon@csulb.edu

CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. Introduction

Since Pigou (1920), scholars have sought to understand the phenomenon of traffic congestion. The growing number of recent empirical studies, which provide more reliable estimates on the cost of congestion by taking advantage of modern empirical tools and new data, confirm that the cost of congestion is substantial (e.g., Couture et al., 2018; Yang et al., 2020; Bento et al., 2021; Russo et al., 2021). However, these works usually focus on just one element of congestion models, especially the slope of the road cost function, which is perhaps insufficient as a detailed policy guide toward congestion toll that would be imposed on real-world drivers who are heterogeneous in multiple dimensions.

The recent availability of real-time traffic information—namely "big data" like Google Maps and the development of prediction tools based on artificial intelligence, usually known as "machine learning"—is enabling researchers to explore the problem of congestion with a more structural framework. Among others, Akbar and Duranton (2017) and Kreindler (2020) have taken initial steps in this direction.[1] The current paper is in line with these recent papers, but ours is more closely tied to Vickrey (1969), which has become the workhorse model for analyzing scheduling behavior and congestion dynamics in economics (see also Arnott et al., 1990, 1993).

Specifically, this paper develops a simple model of trip scheduling under congestion to formally define the congestion delays the commuter faces and identify how commuters adapt to the level of congestion they face (e.g., by adjusting route or timing choices). We then then empirically quantify the model's key elements by utilizing real-time travel time information from Google Maps and machine learning as a prediction tool. Our theoretical model features a salient fact which has mostly been ignored in the classical bottleneck models—that individuals live and work at different locations and therefore each faces their own travel-time profile. Specifically, the commuter in our model chooses her optimal trip timing to minimize commuting costs subject to the travel-time profile faced on her commuter route as a constraint. In this situation, there may be a unique optimal trip timing chosen by the individual on her commute route, which, in turn, implies that the shapes of congestion dynamics faced by individuals matter in their scheduling choices. Specifically, commuters facing a higher level of congestion during the peak hour would choose an inconvenient early or late timing to avoid an otherwise longer congestion delay—a finding for which we find statistically significant evidence.

The two key elements of the model are the travel-time profile faced by an individual and the scheduling utility, which jointly reveal the commuter's optimal arrival time choice. We empirically quantify both. A travel-time profile is the menu of travel times that an individual commuter faces on her commute route by alternate trip timing choices. We construct the systematic travel time predictions by alternate arrival time intervals, i.e., travel-time profiles, for a large survey sample of commuters in California by querying travel times at Google Maps for each origin-destination pair with random timing in a day over a half-year. Our travel survey dataset consists of 9,127 different zip code pairs of origin and destination, which

[1] There are other studies of mobility and congestion using big data, especially in developing countries: e.g., Akbar et al. (2019); Kreindler and Miyauchi (2020). See Selod (2021) for a literature review on the uses of big data in transportation research more broadly. Also, for a survey of big data uses in urban economics research, see Glaeser et al. (2018).

are traveled by 14,544 commuters. This implies that a zip code pair and the corresponding travel-time profile is almost unique to an individual commuter.

An especially useful quantity directly drawn from each constructed travel-time profile is the commuter's congestion delay (or "queuing time," borrowing from the bottleneck model's terminology), which is defined by the gap between the commuter's actual travel time and the counterfactual congestion-free travel time that would have been realized if she had traveled during times in which there was no congestion. The classical bottleneck model (Vickrey, 1969; Arnott et al., 1990, 1993) suggests that this congestion delay is a pure social loss, and the aggregation of these losses equals the economic inefficiency costs from congestion (see also Kim, 2019). Our constructed travel-time profiles exhibit an average congestion delay of about 4.7 minutes per trip (about 18% of the sample mean of travel time), which implies inefficiency costs borne by morning commuters in California of about 6 billion dollars.

The scheduling utility estimated in this paper is composed of a travel time cost and schedule-delay costs which arise from arriving earlier or later than the commuter's ideal arrival time. The estimated parameters in the utility function reveal commuters' willingness to pay for a schedule delay reduction expressed in the unit of travel times. The first step toward estimating these parameters is to assign an ideal arrival time to each commuter, as it directly defines schedule delays as attributes in our random utility framework. For this, we apply the machine learning method to predict the counterfactual arrival time in the absence of queuing, learning from a group of commuters who are plausibly assumed to arrive at their own ideal arrival times. We find large prediction errors in the machine learning estimates, but these are smaller than the errors generated from a linear regression. We update our estimates using our theoretically grounded conjecture and the scheduling utility estimates in the previous studies.

Our scheduling utility estimates indicate that commuters are willing to accept about 0.5 additional minutes of schedule delay to reduce travel time by 1 minute. While the shape of the estimated scheduling utility is similar to those presented in previous studies (Small, 1982; Kreindler, 2020), we newly discover that it is much steeper than travel-time profiles faced by commuters. Because travel time falls too slowly to compensate for the corresponding increase in schedule delay costs, this situation implies the tendency of commuters to choose an arrival time that is close to their ideal arrival times. This result also implies that commuters' ideal arrival times are quite dispersed, which in turn explains the relatively flat travel-time profiles. Given the big difference between travel-time profiles and indifference curves found in this paper, the results indicate that the classical bottleneck analysis (regarding its quantification of the economic inefficiency of congestion and design of a congestion toll) cannot directly apply under our model, calling for a welfare analysis in a similar bottleneck framework incorporating the heterogeneity in travel-time profiles and ideal arrival times.

This paper adds to the literature estimating motorists' scheduling preferences and the distribution of preferences. In his seminal work, Small (1982) formulated the scheduling utility, the so-called $\alpha - \beta - \gamma$ preference still used widely in the literature, and estimated these parameters. Peer et al. (2015) decomposed morning scheduling decisions of car commuters into long-run choices of arrival routines and short-run choices of departure times subject to the routines chosen in the long run to estimate both

short-run and long-run preferences.[2] Kreindler (2020) estimated the scheduling utility for drivers in Bangalore, India, using, like us, Google Maps to construct the menu of travel times and adopting an experimental setting to explicitly estimate the parameters in terms of a monetary unit. Hjorth et al. (2015) estimated scheduling preferences using stated preference data with the incorporation of non-linearity and travel time variability. Small et al. (2005) estimated the values of travel time and reliability as well as the distribution of those values. Hall (2021a) estimated the joint distribution of scheduling preference parameters as well as ideal arrival times to evaluate the effects of optimal time-varying tolling policy. Other papers estimate the preferences of public transit users using revealed preference data (Peer et al., 2016; Hörcher et al., 2017). Our current paper uses a much larger sample of commuters to more accurately estimate the parameters with an application of a machine learning method to predict commuters' ideal arrival times in a scheduling utility form that is allowed to be non-linear.

Our paper is also closely related to the recent empirical papers estimating the costs of congestion. Most of these papers focus on estimating Pigouvian deadweight loss by estimating the cost function of trips at the road or city levels (Walters, 1961; Couture et al., 2018; Yang et al., 2020; Russo et al., 2021). Meanwhile, Akbar and Duranton (2017) use big data to identify the demand curve as well as the cost curve to more accurately estimate the deadweight loss from congestion. Recent papers such as Tang (2021) and Tarduno (2021) adopt a quasi-experimental design to estimate congestion costs. Our approach to the estimation of congestion costs is most similar to that of Kim (2019), who also defines the congestion delay (queuing time) as the difference between the observed travel time and the counterfactual free-flow travel time to measure the economic cost of congestion defined in the standard bottleneck model (Vickrey, 1969; Arnott et al., 1990, 1993). However, unlike Kim (2019), who used only travel survey data and thus suffered from the identification problem, this paper observes counterfactual times at all hypothetical arrival time choices using Google Maps, which allows use to directly identify the congestion delay for each traveler.

The rest of the paper is organized as follows. Section 2 presents a model of trip scheduling under individual-specific travel-time profiles and also develops an empirical framework for measuring the congestion delays of individual commuters. In Section 3, we explain the data. In Section 4, we examine overall shapes of travel time and congestion delay profiles and quantify congestion costs. In Section 5, we estimate the causal effect of congestion on trip scheduling. In Section 6, we estimate the structural scheduling preference parameters. In Section 7, we discuss some implications of our findings. Finally, Section 8 concludes.

[2] Verhoef (2020) investigates whether incorporating the empirically validated difference between long- and short-run preferences affects optimal pricing of congested roads.

2. The Conceptual Framework

2.1 A Model of Commute Scheduling

This section presents a simple model of trip scheduling for a representative commuter. Every commuter travels their own route to work, simply because commuters all live and work in different places. Each commuter therefore faces her own menu of travel times according to alternate trip timing choices, i.e., the individual-specific travel-time profile.

The commuter's goal is to minimize commuting cost by choosing an arrival time, subject to the travel-time profile as a constraint. Commuting cost—with an arrival time t, denoted by $C(t)$—is assumed to have the following form:

$$C(t) = \alpha T(t) + \beta SDE(t)^{\omega_1} + \gamma SDL(t)^{\omega_2}, \tag{1}$$

where $T(t)$ is travel time for the arrival time of t. With t_* indicating the commuter's ideal arrival time, $SDE \equiv \max(t_* - t, 0)$ is schedule delay for arrivals earlier than t_*. The third term is the schedule-delay cost for arrivals later than t_*, with $SDL \equiv \max(t - t_*, 0)$. The parameters α, β, and γ are the unit costs placed on each component. The commuter's departure time is simply $d \equiv t - T(t)$. Hence, under the assumption that the commuter is fully informed about his travel-time profile, the commuter effectively chooses an arrival time (t) by choosing the corresponding departure time (d).

Note that while the exponent of the travel time is normalized at unity, the exponents placed on the schedule delays satisfy $\omega_1 \geq 1$ and $\omega_2 \geq 1$. It is useful to draw an iso-cost curve on the plane of arrival time t against travel time $T(t)$ to appreciate the meaning of this assumption (see Figure 1). The slope of the iso-cost curve is $(\beta/\alpha)\,\omega_1\,(t_* - t)^{\omega_1 - 1}$ for $t < t_*$ and $-(\gamma/\alpha)\,\omega_2\,(t - t_*)^{\omega_2 - 1}$ for $t > t_*$, so the iso-cost curve gets steeper as the arrival time deviates more from the ideal arrival time t_* if $\omega_1 > 1$ and $\omega_2 > 1$. In words, the required travel time reduction to maintain the cost from an increase in schedule delay increases as the arrival time deviates more from t_*. When $\omega_1 = \omega_2 = 1$, our cost form is the same as the customary $\alpha - \beta - \gamma$, the preference of Small (1982) also used in Arnott et al. (1990, 1993), under which the marginal rate of substitution between schedule delays and travel time is constant at β/α before t_* and γ/α after t_* (see Figure 2).[3]

[3] Other studies use a more general formulation by imposing non-separability of arrival time and travel time. Under that condition, the optimal trip timing depends on the trip duration (see Fosgerau and de Palma, 2012; Fosgerau et al., 2018; Fosgerau and Kim, 2019).

Figure 1: Trip Scheduling of the Commuter

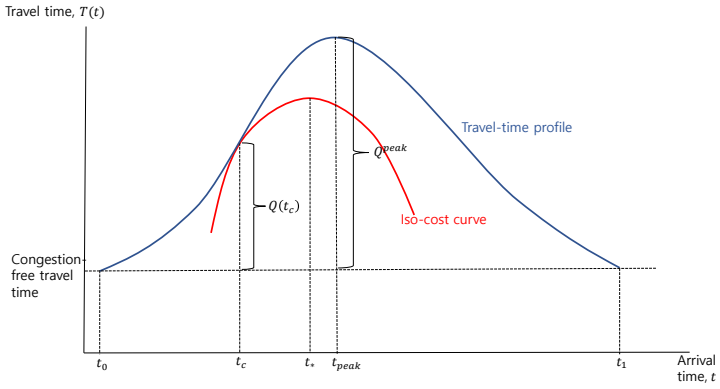
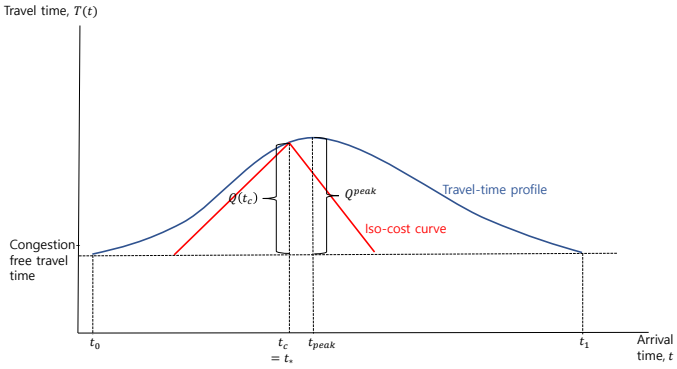
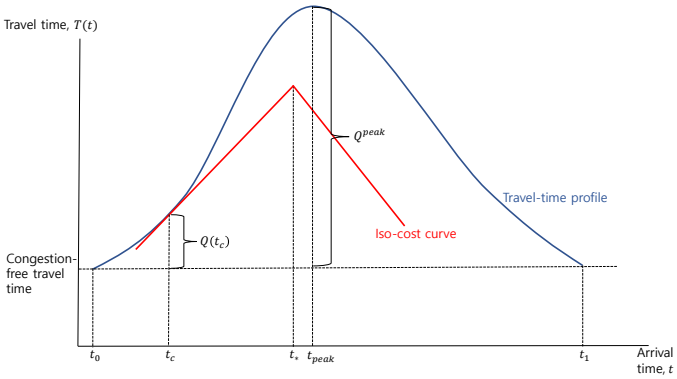


Figure 2: The Effect of Travel-Time Profile on Scheduling Choice



(a) Low level of congestion



(b) High level of congestion

Let $T^P(t)$ be the travel-time profile *faced* by the commuter. $T^P(t)$ is the expected travel time conditional on the commuter's arrival timing t. With t_0 being the time at which congestion sets in and t_1 being the time at which congestion ends, it is assumed that $T^P(t)$ is monotonically increasing from t_0 until t_{peak} and decreasing as t approaches t_1. The congestion-free travel of the commuter, $T^P(t_0) = T^P(t_1) \equiv T^{free}$, is the travel time if the commuter had traveled earliest or latest among her possible choices (an earlier arrival than t_0 or a later arrival than t_1 cannot be optimal if $t_0 < t_* < t_1$).

We can also define a congestion-delay profile, which represents the congestion dynamics faced by the commuter. We denote it by $Q(t) \equiv T^P(t) - T^{free}$. The congestion delay is zero for an arrival of either t_0 or t_1, i.e., $Q(t_0) = Q(t_1) = 0$. The maximum congestion delay on the commute route is $Q^{peak} \equiv Q(t_{peak}) = T^P(t_{peak}) - T^{free}$, where t_{peak} is the individual's arrival timing for which his travel time would be the longest.

To characterize the commuters' choice, consider, as an example, a bell-shaped travel-time profile, reflecting the real-world observations, which is illustrated in Figure 1. At any given arrival time choice t, any point below the travel-time profile is not feasible. So, at the cost-minimizing arrival time, the travel-time profile and an iso-cost curve meet while the iso-cost curve is placed at the lowest position possible. In the figure, t_c is the optimal arrival time chosen by the commuter. The following proposition generalizes the results on arrival time choice: for the proof, see Appendix B.

Proposition 1 *Assume* $t_0 < t_* < t_1$.

1. *If the iso-cost curves have a larger curvature than* $T^P(t)$ *globally over* $t \in [t_0, t_1]$, *then there exists a unique optimal arrival time,* t_c, *between* t_0 *and* t_1. *At* t_c, *the iso-cost curve and the travel-time profile are tangent.*

2. *If the iso-cost curves have a smaller curvature than* $T^P(t)$ *globally over* $t \in [t_0, t_1]$, *then there exists a unique optimal arrival time, which is* $t_c \in \{t_0, t_*, t_1\}$.

Proposition 1 suggests that each commuter has her own optimal trip timing most suitable to the travel-time profile she faces, which is a feature that departs from the classical bottleneck model, where commuters are indifferent with alternate trip timing choices in the equilibrium.

A significant implication of our model is that travel-time profiles seem to have an impact on scheduling choices. To explore this point, consider first the case where iso-cost curves have a larger curvature than the travel-time profile. In this case, note that the condition for optimal choice is that the slope of the travel-time profile equals the slope of the iso-cost curve. If the former is larger at one point, then a given delay (to an earlier time if $t < t_*$ or to a later time if $t > t_*$) would decrease the travel time more than the required fall in travel time for the constant cost, inducing a choice that is farther from t_* and closer to t_0 or t_1. This implies that commuters facing a steeper travel-time profile would tend to choose a time that is more distant from t_* and closer to t_0 or t_1, while those facing a flatter travel-time profile would tend to choose a time closer to t_*.

In Figure 2, we illustrate the other case where the iso-cost curves have a smaller curvature overall than the travel-time profile, specifically the linear iso-curves, by imposing $\omega_1 = \omega_2 = 1$. In panel (a), we assume that the commuter faces a lower level of congestion with a flat travel-time profile, according to which the commuter chooses to arrive at her ideal arrival time. In panel (b), the commuter is facing a higher level of congestion, and in this scenario the commuter chooses an inconvenient edge time that is largely different from t_*.

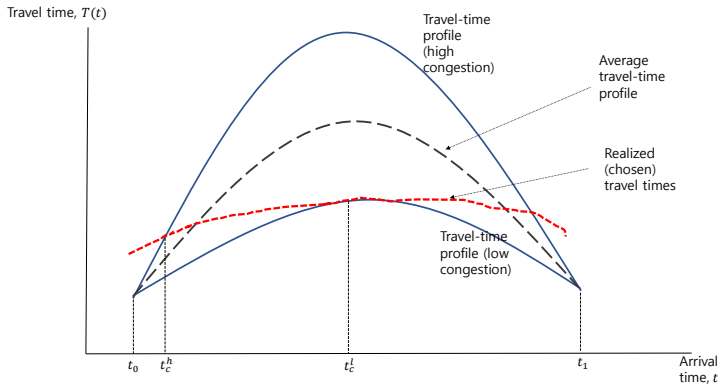
Based on these different choices, the effect of travel-time profiles on scheduling choices can be stated as follows:

Corollary 1.1 *An increase in the (absolute) slope of the travel-time profile induces an earlier arrival of the commuter if* $t_c < t_*$ *and a later arrival if* $t_c > t_*$. *Since* Q^{peak} *measures the overall slopes of the travel-time profile, all other things being equal, commuters facing a higher* Q^{peak} *tend to arrive at a time that is more different from their ideal arrival times, while those facing a lower* Q^{peak} *tend to arrive at a time that is closer to their ideal arrival times.*

This corollary states that commuters adapt to the congestion dynamics that they face, especially to the maximum congestion delay Q^{peak}, by choosing the trip timing most suitable to it. An interesting observation from this tendency is that a commuter facing a higher level of congestion may actually have a shorter "realized" congestion delay than when facing an otherwise a lower level of congestion, because the commuter could avoid the otherwise long queue by arriving at a non-peak time. For example, in Figure 2, the realized congestion delay, i.e., $Q(t_c) = T^P(t_c) - T^{free}$, is shorter in panel (b) than in panel (a), even though panel (b) illustrates a higher level of congestion with a higher Q^{peak}.

Figure 3 depicts another situation illustrating how two individuals with similar preferences but facing different levels of congestion make scheduling decisions. Based on Corollary 1.1, this scenario works from the assumption that the commuter facing a higher level of congestion chooses an edge time, denoted by t_c^h, while the one facing a lower congestion chooses a peak time, t_c^l. In the figure, the black dotted curve is the average travel-time profile, which connects the average travel time values of these two commuters by arrival time. Meanwhile, the red dotted curve in the figure connects the chosen combination of arrival time and travel time, exemplifying scatter plots on arrival times and travel times drawn from the data. The large difference between the two dotted curves is due to the commuters' adaptation behavior shown in in Corollary 1.1.

Figure 3: Average and Realized Travel-Time Profiles



The final comment is on the scope of this simple model. The model focuses on understanding scheduling choices of commuters when each of them faces their own congestion dynamics on their respective commute routes, without endogenizing the equilibrium congestion dynamics that are heterogeneous across

commuters. This paper decisively highlights the importance of incorporating the large heterogeneity in both commuter preferences and congestion dynamics faced by individuals in a congestion model used for welfare analysis and optimal policy design.

2.2 Measuring Commuters' Congestion Delays

Here, we provide the framework to measure individual commuters' queuing times empirically. With i being the commuter subscript, the individual-specific congestion dynamics faced by the commuter can be written as $Q_i(t) = T_i^P(t) - T_i^{free}$. With t_i indicating the chosen (observed) arrival time, the individual's "realized" queuing time is calculated by evaluating $Q_i(t)$ at t_i, i.e., $Q_i(t_i) \equiv T_i^P(t_i) - T_i^{free}$. Travel diary data mostly report the travel time outcome corresponding to $T_i^P(t_i)$, but not the congestion-free travel time T_i^{free}, which is the counterfactual travel time that would obtain if the traveler had alternatively chosen to travel earliest (or latest) in the morning and encountered no queues.[4] We overcome this identification challenge by constructing each commuter's queuing-time profile using the data on hypothetical trips queried via Google Maps.

We first construct the individual-specific travel-time profiles. Let $\widehat{T_i^P(m)}$ be the systematic prediction of travel times for the origin and destination pair, namely the "route," traveled by commuter i, where m is the set of alternate arrival time intervals, with $m \equiv \{1, 2, 3, \cdots, M\}$. The predicted travel time must be "systematic," in the sense that the prediction rules out any day-specific unanticipated shocks such as weather, traffic accidents, and road work that would affect travel time on the particular day, since commuters would make scheduling choices based on their systematic travel-time profiles.[5] To this end, we queried travel times for a long time horizon from Google Maps and then averaged out the travel time outcomes to construct the systematic travel-time predictions.

In the bottleneck model of Vickrey (1969), T_i^{free} is the travel time outcome if the traveler had arrived at t_{0i}, the time at which the road capacity is fully utilized and congestion sets in. We could use the travel-time prediction at $m = 1$, i.e., $\widehat{T_i^P(m=1)}$, for T_i^{free} from normal days (before COVID-19). Note however that the initial times t_{0i} vary significantly depending on the capacity and demand conditions of the routes. Thus, we may be overestimating T_i^{free} when considering some routes that are already congested from the first interval $m = 1$ (before 6:22 AM). We therefore use only travel time predictions conducted during the lockdown policy under COVID-19, especially counterfactual trips with pre-6:00 arrivals, to construct the individual-specific congestion-free travel time estimates, which we denote by $\widehat{T_i^{free}}$.

[4] It is useful to see how the counterfactual framework (see Angrist and Pischke, 2009) may be adopted to define the individual's congestion delay alternatively. Assume that there are only two alternate trip timing choices: a time when there is congestion and another time when there is no congestion. Let T_{1i} be the travel time outcome if the traveler chose to arrive in a time with congestion and T_{01} be the time if the traveler has chosen to travel at the time with no congestion (very early or very late in the morning). The congestion delay of the commuter is $T_{1i} - T_{01}$, which is the causal effect of trip timing on travel time (see Kim, 2019). The traveler cannot choose both, so researchers can only observe either of them and thus cannot identify the congestion delay directly using travel diary data.

[5] Indeed, Peer et al. (2015) showed that commuters would form "routine" in response to their systematic travel-time profiles and then make daily choices on trip timing.

The estimate for the individual-specific congestion delay profile is then written as $\widehat{Q_i(m)} \equiv \widehat{T_i^P(m)} - \widehat{T_i^{free}}$. The "realized" delay of commuter i is obtained by evaluating it at the chosen arrival time t_i, which is written as

$$\widehat{Q_i(t_i)} = \max\{\widehat{T_i^P(m)} D_m(t_i) - \widehat{T_i^{free}}, 0\}, \tag{2}$$

where

$$D_m(t_i) = \begin{cases} 0 & \text{if } t_i \notin m \\ 1 & \text{if } t_i \in m \end{cases}$$

is a binary variable which takes value 1 if the commuter's arrival time t_i belongs in the interval m and 0 otherwise.

Another quantity estimated below is the peak queuing time of individual i, written as the following:

$$\widehat{Q_i^{peak}} = \max\{\widehat{T_i^P(m_i^{peak})} - \widehat{T_i^{free}}, 0\}, \tag{3}$$

which evaluates the congestion dynamics at the arrival time interval m_i^{peak} at which the travel-time profile of the individual i reaches its peak. This expression gives the expected queuing time conditional on arrival in the peak congestion condition of the commuter's own route. Proposition 1 suggests that commuters with a higher $\widehat{Q_i^{peak}}$ tend to avoid the peak timing, which we test by regressing the binary choice variable for peak hour arrival on $\widehat{Q_i^{peak}}$ below.

In Section 4, we calculate congestion delays of a sample of commuters in California and use them to calculate congestion costs in California. We then test the "adaptation" hypothesis of Corollary 1.1 by examining how the maximum queuing time, Q_i^{peak}, affects scheduling choices. We also estimate the scheduling utility function to draw the iso-cost curves, which, together with the travel-time profiles, would reveal actual and ideal arrival times. As we will see, commuters in the real world have much steeper iso-cost curves than their travel-time profiles, as illustrated in panel (a) in Figure 1, which suggests that commuters' actual arrival times do not deviate largely from their ideal arrival times.

3. Data

The data on travelers' scheduling choices are obtained from the 2012 California Household Travel Survey (CHTS), which is based on a statewide survey conducted every ten years. From the entire CHTS sample, we select morning commutes made by passenger vehicles to construct the estimation sample. We further limit our attention to morning commutes whose arrival times are between 5:00 and 11:15. We exclude trips taken using public transit or ride sharing. There are several reasons for the selection of this sample. First, the choice set is relatively clear, so it is easier to estimate the scheduling preferences. Second, our model is closely tied to the Vickrey (1969) bottleneck model, and one of our goals is to test the power of this model. Third, time spent on commuting is the largest portion in total travel time spent in cars. In particular, while the typical American driver spent about 72 minutes driving per day in 2008 (Couture et al., 2018), commuters in our sample spent about 25 minutes for a morning commute only.

The CHTS dataset reports departure times from home and arrival times at work. We mainly use the arrival time variable, since departure time may suffer from a large heterogeneity in its values of commuters traveling difference distances even when they have a similar preference. We convert the original arrival time variable expressed in 24-hour military time into a continuous form, with midnight normalized at 0. For example, an arrival at 7:30 AM gives the arrival time value of 450 ($= 60 \times 7 + 30$).

The other important information we obtain from the CHTS sample is the zip codes of the home and work locations of each commuter. The geographic centers in each zip code (for home and work) serve as the origin and the destination when we query from Google Maps. In our final sample composed of 14,544 commuters, there are 9,127 different zip code pairs (with 1,462 different zip codes for home or work sites). So, although each zip code pair is not entirely unique to individuals (since commuters residing at different addresses may travel to or from the same zip codes), we can say that zip codes provide fairly detailed information on the origin and the destination of the commuters' trips.[6]

For each pair of zip codes ("route"), we construct the travel-time profile, i.e., the menu of travel times by commuter's alternate arrival time choices. In total, we queried 18,340,541 hypothetical trips from Google Maps in the period of January 6, 2020, until July 28, 2021 for this purpose. We classify the arrival times between 5:00 to 11:15 AM into 16 ($= M$) intervals, so that each interval spans 15 minutes, which is done to ensure that each interval contains modes of arrivals such as 7:30, 7:00 in the center. The first and the last intervals (with $m = 1$ or $m = 16$) are set to be much wider, so these intervals cover 5:00–6:22 and 9:52–11:15, respectively, which is done to reflect low arrival rates and slowly changing traffic conditions during these times.

For each route (each person), we calculate the mean of predicted travel times obtained from Google Maps from trips queried between January 6, 2021, and March 18, 2020 (the date just before the first lockdown policy enacted by the California government due to COVID-19 on March 19) to measure the systematic prediction of travel time by interval, i.e., $\widehat{T_i^P(m)}$ for $m = 1, 2, ..., 16$. We used travel time prediction from many days (more than 2.5 months) to rule out day-specific shocks such as accidents.

[6] For more information on how we identify routes for commuters traveling within the same zip code, see Appendix A.

Once we construct travel-time profiles for each of 9,127 different pairs of zip codes, we then match the profiles to individual commuters so that each has their own travel-time profile. Our final sample consists of 14,544 commute observations.

Finally, note that we must also assume that travel-time profiles constructed between January and March 2020 are similar to those of 2012 when the commuters in the CHTS sample made their scheduling choices. Because Google Maps queries are always real-time, we cannot go back and estimate the travel time that 2012 commuters would have experienced at that time. Since congestion is determined by long-term demand and supply factors, the cross-sectional variation in congestion across zip codes would remain stable. We provide further remarks on the data construction in Appendix A.

Table 1 shows descriptive statistics for the key variables, such as arrival time, travel time, and trip distance. As mentioned above, the arrival time variable (t) follows the decimal system, with 00:00 being normalized at 0. The sample mean for arrival time is 478, i.e., 7:58 AM ($478 = 7 \times 60 + 58$). The mean travel time is 25 minutes and the mean trip distance is 12 miles according to the CHTS sample. The Google Maps predicted travel times for the commuters—i.e., $\widehat{T_i^P(m)}D_i(m)$ from (2)—have a sample mean of 26. The mean distance from Google Maps is a bit longer that that from the CHTS sample, implying that people tend to travel distances shorter than the distance between the central locations of the zip codes reported in the CHTS. The average speed (distance divided by travel time) from Google Maps is around 33 miles per hour.

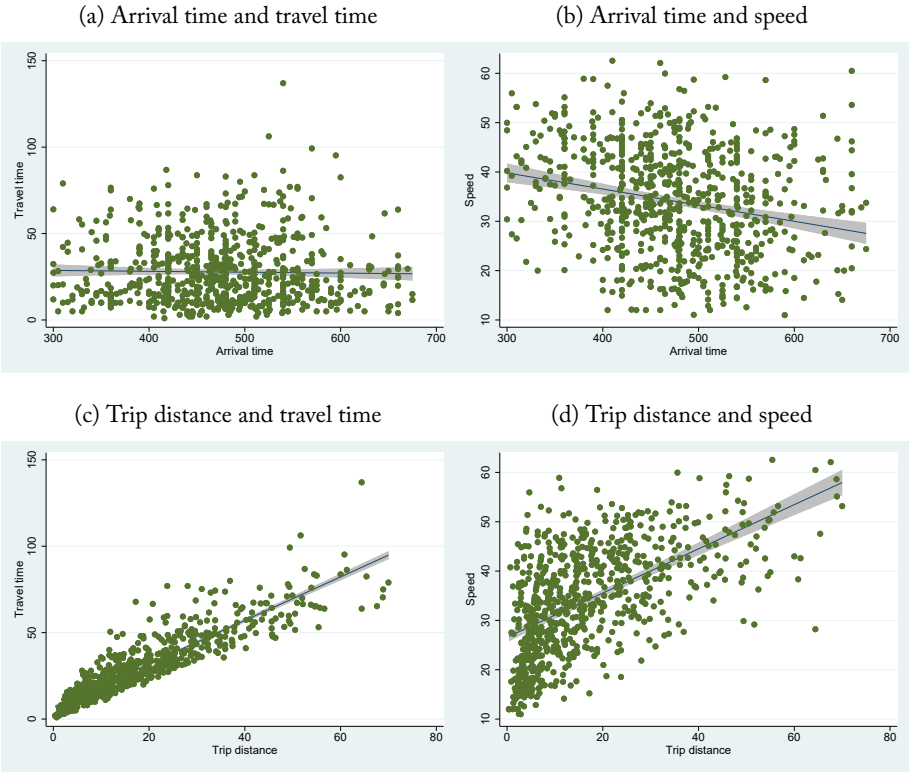
Figure 4 provides scatter diagrams to visualize relationships between variables. From panel (a), we find that travel times are overall constant over the entire morning time interval. However, panel (b) shows that the speed (ratio of distance to travel time) tends to decline as the arrival time becomes later. Panel (c) shows a monotonic relationship between trip distance and travel time. Panel (d) shows that the speed is higher for longer-distance commutes, reflecting a higher level of congestion for commuters residing closer to their jobs.[7]

Figure 5 shows scatter diagram showing the fitness of the Google Maps predicted values to the CHTS reported values. In each figure, the horizontal axis represents the CHTS reported values, and the vertical axis plots the predicted value matched to each CHTS observation. As shown in panel (a), Google Maps data match quite well with the survey respondents' reported travel times, but there are some errors—presumably because the survey respondents' perceptions of their travel times are not always accurate. Panel (b) shows that the distance values in the two datasets match better than the travel times. The correlation coefficient for the travel time variables in the two datasets is 0.62, and for the distance variables it is 0.78.

[7] This relationship is further explored in Couture et al. (2018) and Fosgerau and Kim (2019).

Finally, Figure 6 illustrates the long-run evolution of travel times faced by California commuters throughout the periods including the pre-COVID-19 pandemic period and subsequent periods. Each point on the curve indicates the monthly cross-sectional average of Google Maps predicted travel times on the condition that commuters arrive in the peak hour (7:52–8:52).[8] As shown in the figure, in January and February 2020, the average of expected travel times conditional on arrival in the peak hour is around 28–29 minutes, but it dropped significantly when the statewide lockdown policy was enacted on March 19, 2020. While there was some recovery in travel time during the summer seasons of 2020 and 2021, travel times remain significantly low relative to the pre-COVID-19 values.

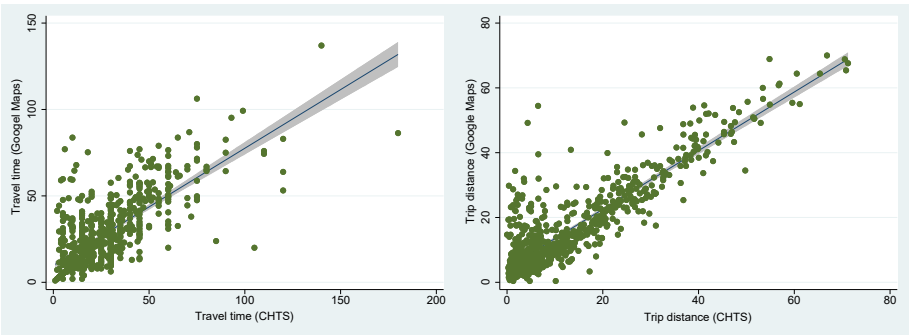
Figure 4: Scatter Diagrams Between Variables



Notes: The scatter diagrams are drawn to visualize the relationships between different variables. Only 5% of observations from the CHTS estimation sample (14,544) are used to draw the scatter diagrams. The fitted line as well as its 95% confidence band is drawn in each figure. Travel times and trip distances predicted at the commuters' chosen arrival time intervals are used to draw scatter diagrams. The speed variable plotted in panels (b) and (d) are calculated by dividing the trip distance (measured by Google Maps) by travel time (predicted by Google Maps) and multiplying it by 60 to normalize into a per-hour scale.

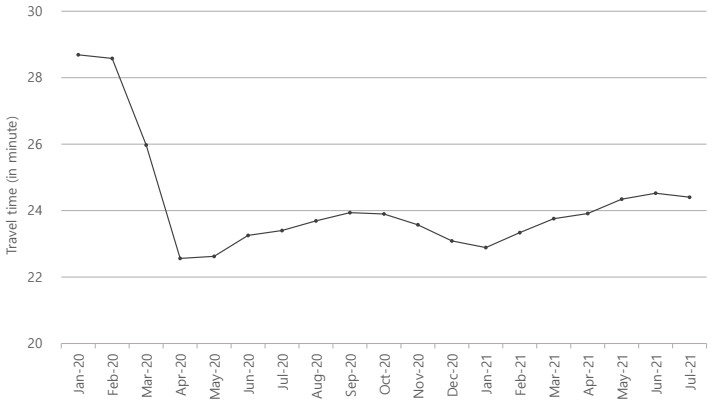
[8] Specifically, we first calculate each individual's expected travel time if she arrived in the peak hour (7:52–8:52) by taking the mean of travel times from Google Maps queried trips whose arrivals belong in the peak time interval over each month. These expected peak travel times are then averaged out from the commuters in the sample with the size of 14,544, which yields each point on the curve.

Figure 5: Degree of Fit of Google Maps to CHTS Data



Notes: The scatter diagrams are drawn to evaluate the degree of fit of the Google Maps predicted values to the CHTS reported values of the key variables. Only 5% of observations from the entire sample (14,544) are used to draw the scatter diagrams. The fitted line as well as its 95% confidence band is drawn in each figure.

Figure 6: Average of Expected Travel Time in the Peak Hour by Month



Notes: Each point plots the monthly average of expected travel times at the peak hour of the commuters in the sample. The expected travel time of the peak hour (7:52–8:52) of each commuter is first calculated by taking the mean of Google Maps travel times using trips whose arrival times are in the peak hour over the month, which is then averaged out across commuters to give each point.

Table 1: Summary Statistics

	Obs	Mean (SD)	Median	Min	Max
Key variables					
Arrival time in minutes (0:00 at zero)[a]	14,544	478.19	478	300	675
		(77.54)			
Travel time in minutes (CHTS)	14,544	25.07	20	1	295
		(18.50)			
Trip distance in miles (CHTS)	14,377	12.28	8.38	0	389
		(12.90)			
Travel time in minutes (Google Maps)[b]	14,544	26.08	21.5	1	137
		(17.35)			
Trip distance in miles (Google Maps)[c]	14,544	15.30	11.78	0.3	81
		(12.35)			
Speed (miles per hour, Google Maps)	14,544	33.42	33.38	6.4	67
		(10.45)			
Quantities measuring congestion					
Congestion-free travel time in minutes[d]	14,544	21.46	18	0.7	81
		(13.90)			
Congestion delay, realized at chosen arrival timing (in minutes)	14,544	4.65	2.08	0	71
		(6.75)			
Congestion delay per trip length (minute per mile)[c]	14,544	0.35	0.2	0	4.88
		(0.40)			
Congestion delay conditional on arrival is at the peak ($\widehat{Q^{peak}}$)	14,544	8.36	4.3	0	78
		(10.09)			
Selected control variables					
Household income[e]	13,460	106.1	87.5	5	300
		(68.74)			
Race is white	14,544	0.73		0	1
Gender is female	14,544	0.48		0	1

Notes: a. The arrival time variable follows the decimal system, with 00:00 being normalized at 0. The average arrival of 478 means 7:58 AM. b. Each commuter's Google Maps predicted travel time is the average of query outcomes between January 6 and March 18, 2020, for arrivals that fall in the interval chosen by the commuter. c. Note that trip distance may differ for given commute route (zip code pair), as Google Maps may suggest different routes. For the trip length, we use the distance of the route suggested by Google Maps at the commuter's chosen arrival time interval. d. Congestion-free travel time for each commute route (zip code pair) is estimated by taking the average of travel time outcomes for arrivals before 6:22 queried between March 19 and June 30. e. The original income variable reported in the CHTS is categorical, but we convert it to a continuous variable by choosing the mid-point of each interval.

4. Congestion Costs

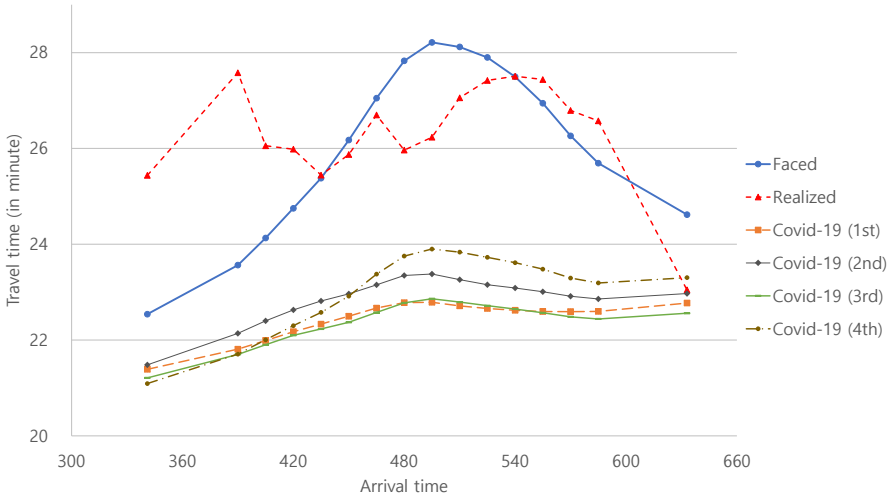
4.1 Travel Time and Congestion Delay Profiles

By taking the average of travel times of counterfactual trips whose arrival times fall in each interval m, which were queried between January 6 to March 18, 2020, we construct the individual-specific travel-time profile denoted by $\widehat{T_i^P(m)}$. This is the travel time that individual i would face if her arrival time belonged in the interval m. By arrival time interval m, we calculate the average of expected travel times, which is written as

$$\overline{\widehat{T(m)}} \equiv \frac{\sum_{i=1}^{N} \widehat{T_i^P(m)}}{N}, \tag{4}$$

where N is the size of the sample ($= 14,544$). This quantity is calculated for each $m \equiv \{1, 2, 3, \cdots, M\}$ and plotted in Figure 7. Since we use the same sample to calculate the average of expected travel time consistently for all intervals, we can interpret this curve as the average travel-time profile, or the travel-time profile of the commuter facing an average level of congestion. According to the drawn curve, a typical commuter would spend around 23 minutes if she chose to arrive before 7:00, and travel time is expected to be longer than 28 minutes if she can expect to arrive around the peak time, 8:15.

Figure 7: Travel-Time Profiles Averaged Out Across Commuters



Notes: The arrival time variable represented on the horizontal axis is in minutes, and time 00:00 is normalized at 0. The solid blue curve illustrates the average of travel-time profiles of the commuters in the sample. For this, we first construct each person's travel-time profile (systematic travel-time prediction conditional on the commuter's arrival time being in each interval) by taking the mean of travel times predicted by Google Maps using counterfactual trips whose arrival times fall in each interval over the data collection period. These are then averaged out across commuters by arrival time interval to give each point in the blue curve. The data collection period from Google Maps covers January 6 to March 18, 2020. Each point on the red dotted curve represents the average of the travel-time values calculated only using the commuters who actually chose that interval. The average travel-time profiles faced by travelers are also drawn in the figure by selecting different data collection periods to illustrate how the travel-time profiles have changed since the COVID-19 measures entered into force in California: COVID-19 1st period covers March 19, 2020, through July 31, 2020; COVID-19 2nd period covers August 1, 2020, through November 30, 2020; COVID-19 3rd period covers December 1, 2020, through March 31, 2021; and COVID-19 4th period covers April 1, 2021, through July 28, 2021.

The overall travel-time profiles that form as result of commuters' arrival time choices are much different from the profiles that are faced by commuters. To show this point, we calculate the mean predicted travel time for each arrival time interval using only commuters who actually chose to arrive in that interval, which is expressed as

$$\widetilde{\widehat{T(m)}} \equiv \frac{\sum_{i=1}^{N} \left(\widehat{T_i^P(m)} | t_i \in m \right)}{N_m}, \tag{5}$$

where N_m is the number of commuters arriving in m. This is plotted by m and illustrated by the red dotted curve in Figure 7. For example, 1,155 commuters arrived in the interval of 7:37–7:52 ($457 < t \leq 472$) and the mean of their expected travel time is 26.7, etc. As can be seen in the figure, the travel-time profile based only on the chosen arrivals is flatter than the profile faced by commuters, which is attributed to the fact that commuters tend to adjust their arrival time choices to the travel-time profiles that they face. Specifically, commuters traveling a longer distance and thus facing a long travel time at the peak would plan to arrive at a non-peak time. Since these travelers tend to arrive early in the morning, the mean travel time for early arrivals is quite high.

Figure 7 also shows how travel-time profiles changed from the travel demand shocks from COVID-19. In particular, we consider four sub-periods under the COVID-19 condition and draw mean travel-time profiles for each of them using Google Maps query outcomes drawn from each sub-period.[9] As anticipated, the travel-time profile is flattest for trips occurring just after the state's lockdown policy. We see some recoveries of the travel times for the next sub-periods, although they are still much flatter than under the pre-COVID-19 condition.

Now let us turn to queuing-time profiles. We can construct each individual's queuing-time profile using Google Maps by subtracting the individual-specific congestion-free travel time from the individual's travel-time profile, i.e., $\widehat{Q_i(m)} = \widehat{T_i^P(m)} - \widehat{T_i^{free}}$. To measure each individual's congestion-free travel time for each route, i.e., $\widehat{T_i^{free}}$, we use trips with pre-6:22 arrivals using queries conducted during the COVID-19 lockdown policy (between March 19, 2020, and June 30, 2020). We anticipate that the lockdown policy increases remote working and thus reduces physical car commuting, which would eliminate congestion for morning commute trips. As confirmed in Figure 6, expected travel times are significantly lower during these days than before COVID-19. Also, by focusing on pre-6:22 arrivals like in Kim (2019), we can safely choose only trips that have not encountered any queues. We taking the mean of travel time predictions for these trips to construct $\widehat{T_i^{free}}$ for each i. Note that $\widehat{Q_i(m)}$ is the queuing-time profile that is faced by the individual, which is distinguished from the realized queuing time (2) (the difference comes from whether the predicted travel time $T_i^P(m)$ is interacted with the commuter's choice to arrive in that interval).

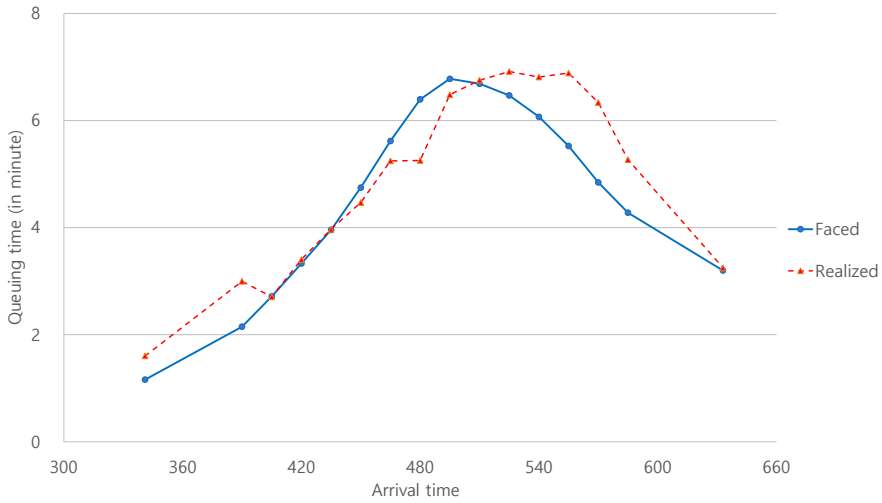
[9] The sub-periods include March 19 through July 31 (2020), August 1 through November 30 (2020), December 1 through March 31 (2021), and April 1 through July 28 (2021).

The mean of $\widehat{Q_i(m)}$ over the entire set of commuters in the sample for each arrival interval m is written as

$$\overline{\widehat{Q(m)}} \equiv \frac{\sum_{i=1}^{N} \widehat{Q_i(m)}}{N}, \tag{6}$$

which is plotted by the interval m and illustrated by the blue curve in Figure 8. This curve describes the average shape of the queue profiles faced by individuals. This curve has almost the same shape as the travel-time profile, since the queuing-time profile for an individual is obtained directly from her travel-time profile with only a level adjustment.

Figure 8: Average Queuing-Time Profiles Faced and Realized



Notes: The average queuing-time profile faced by commuters is illustrated by the solid blue curve. Each commuter's expected queuing-time profile is first constructed by subtracting the estimated congestion-free travel time from the Google Maps constructed travel-time prediction of each arrival time interval. These individuals' expected queuing times at respective arrival time intervals are then averaged out across commuters in the sample to construct the solid blue curve. Each point on the red dotted curve indicates the average of the expected queuing times calculated from only commuters who actually chose that arrival time interval.

Now let us define the queue time profile that results from commuters' scheduling choices. Specifically, we define the mean queuing time for commuters who actually arrived in the interval m as the following:

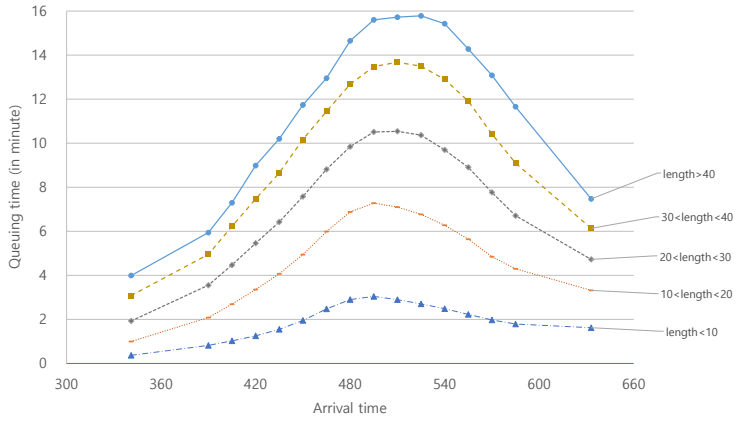
$$\overline{\widehat{Q(m)}} \equiv \frac{\sum_{i=1}^{N} \left(\widehat{Q_i(m)} | t_i \in m \right)}{N_m}, \tag{7}$$

where N_m is the number of commuters arriving in m. This is plotted by m and illustrated by the red curve in the same figure. Commuters arriving around the peak time on average experience a longer queue than those arriving at edge times, but the peak of the red curve is a bit later than the peak of the blue curve. Also, importantly, the red curve during the peak hours between 7:00 and 8:30 is placed below the blue curve during the same interval, meaning that the mean of queuing times for trips that actually chose this interval is smaller than the mean of all potential queuing times during this interval. This implies that commuters who would meet a relatively longer queue tend to avoid this interval, which possibly validates the hypothesis of adaptation behavior of commuters from Section 2.

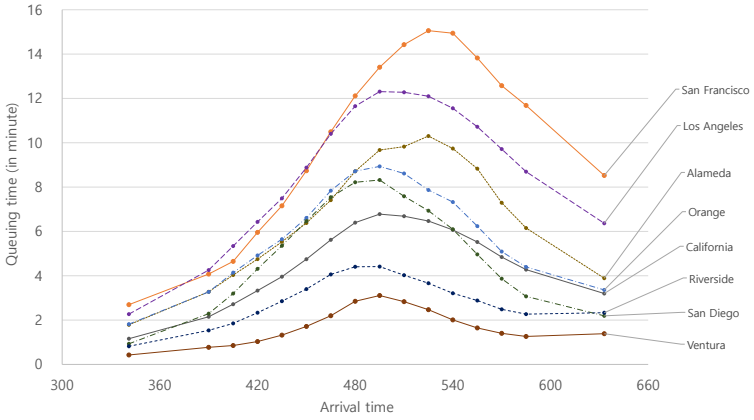
Figure 9 explores the heterogeneity of congestion profiles by different groups of commuters. In panel (a), we draw the queuing-time profiles faced by individuals by distance groups, which clearly shows that commuters traveling a longer distance experience a long queue, which is seen by the large gap in travel times between non-peak and peak time for the longer-distance groups. Panel (b) draws congestion profiles for commuters by county of employers. As anticipated, counties known to be highly congested, such as San Francisco and Los Angeles, have much steeper congestion profiles than counties such as San Diego and Ventura. To further explore the difference in congestion by county, we report the congestion ranking in Table A1, where the ranking is based on the mean of per-mile expected queuing time in the peak hour.

Figure 9: Variation of Congestion

(a) By trip length

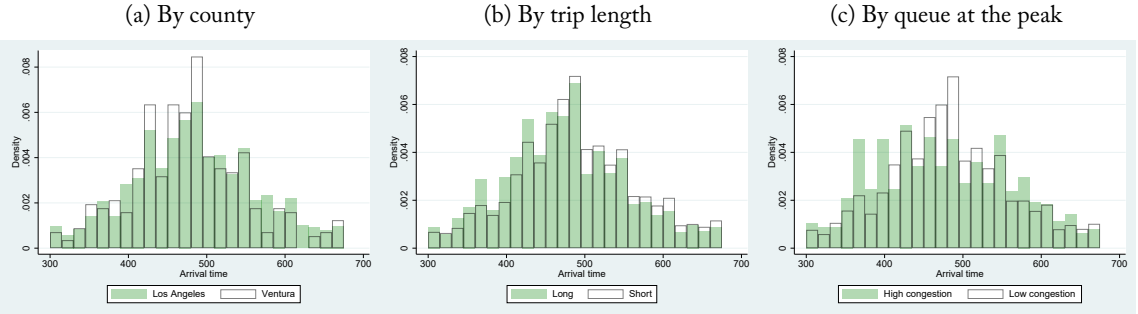


(b) By county



Notes: Panel (a) shows how queuing-time profiles (faced by individuals) differ by trip length. We classify commuters into five different groups sorted by trip length (the length of each pair of zip codes is the average of distances of different routes suggested by Google Maps by arrival time interval for the commuter) and draw the average queuing-time profiles of commuters in different groups. Panel (b) draws queuing-time profiles for different counties. Because congestion on commute route is best reflected in the county in which the worker's job is located, I classify according to county. Points on each curve are the average of expected queuing times at the arrival time interval calculated from commuters working in each county. Among 57 counties in California, we select several large counties to draw their average congestion dynamics. See Table A1 for the number of observations used to draw each profile.

Figure 10: Difference in the Distribution of Arrival Times



Finally, before empirically exploring scheduling choice patterns by congestion level (Section 5), it is useful to describe such patterns. In Figure 10, we draw histograms to illustrate how commuters' scheduling choices differ by congestion level. Panel (a) illustrates the distribution of arrival times for the relatively highly congested Los Angeles county and the less congested Ventura county. Panel (b) illustrates the distribution of arrival times for commuters whose trip lengths are longer than the median distance (11.78 miles) and those whose trip lengths are shorter. Panel (c) illustrates the distribution of arrival times for commuters whose maximum queue on their commute routes exceeds 30 minutes and that of the rest of the commuters. All the panels imply that commuters subject to a higher level of congestion tend to avoid the peak time.

4.2 Calculating Congestion Costs in California

Here, we aggregate congestion costs from our sample to estimate overall congestion costs in the California population. We have estimated the individual's queuing time that is "realized" as a result of commuters' scheduling choices, i.e., $\widehat{Q_i(t_i)}$ (see (2)), which is different from the queuing-time profile that is *faced* by the individual that was denoted by $\widehat{Q_i(m)}$ above (the difference comes from whether the predicted travel time $T_i^P(m)$ is interacted with the commuter's choice to arrive in that interval). The distribution of $\widehat{Q_i(t_i)}$ is illustrated in the second column in Table 2, and the distribution of peak queuing times that they would face in the peak point, i.e., $\widehat{Q_i^{peak}}$, is illustrated in the third column in the same table.

Table 2: Percentiles of Estimated Congestion Delays

	Delay, realized	Expected delay at the peak
5%	0.00	0.38
10%	0.17	0.82
25%	0.76	1.91
50%	2.08	4.30
75%	5.38	10.77
90%	12.81	22.42
95%	19.16	30.77
99%	32.21	46.38

Notes: The second column shows the distribution of the estimated individuals' realized delays (denoted by $\widehat{Q_i(t_i)}$). The third column shows the distribution of the maximum queuing times that each commuter would experience if choosing the peak time on her route (denoted by $\widehat{Q_i^{peak}}$).

The total of realized queuing times from our sample is expressed in the following:

$$TQ \equiv \sum_{i=1}^{N} \widehat{Q_i(t_i)} = \sum_{i=m}^{M} \sum_{i=1}^{N} \left(\widehat{Q_i(m)} | t_i \in m \right) \tag{8}$$

and the simple mean of average queuing times is

$$AQ \equiv \frac{TQ}{N}, \tag{9}$$

where N is the number of observations in our sample. We find that TQ for all 14,544 commuters in our sample is 67,669 minutes, with AQ being 4.65 minutes, which is about 20% of the sample mean of travel times. The median queuing time is 2.08. The median queuing time is much shorter than the mean queuing time because the queuing times are skewed toward longer queuing times.

We also calculate the weighted average of queuing times, using weights based on the proportion of the in-sample county population to the county population in California; this is to deal with the large heterogeneity in the congestion levels across counties and the different distribution of county observations from the population distribution. We calculate

$$AQ^w \equiv= \sum_{m=1}^{M} \sum_{i=1}^{N} w_i \left(\widehat{Q_i(m)} | t_i \in m \right), \qquad \sum_{i=1}^{N} w_i = 1. \tag{10}$$

The weight place for i is proportionate to the inverse of the proportion of the county observation in the sample to the county population; this is to compensate commuters from counties with fewer observations in our sample.[10] We find that the weighted mean of queuing times is about 5.14 minutes, which is a bit larger than the unweighted mean, implying an oversampling of commuters located in relatively less congested areas in the CHTS sample.

We use the weighted mean of queuing times to calculate the total congestion costs in California. We note that the total employment in the US in 2009 is about 142 million and that the percentage of workers who commuted by private vehicle (either driving alone or carpooling) in 2009 is about 86.1%, and we conclude that about 122 million workers used cars for their regular commutes in that year. We also assume that the number of work days for typical workers in the US in a year is 220 days, although this number should vary by worker. We therefore multiply our estimate for queuing time (5.14 minutes) by 18.9 million workers (i.e., the whole labor force) and 220 (days) to conclude that the total queuing times of all US car commuters in a year is about 2.3 billion hours. Following Kim (2019), who chose a value of time (VOT) estimate based on the extensive surveys conducted in Small and Verhoef (2007) and Wardman (2001), we use $16.8 as VOT, which then yields a total congestion cost of about 6 billion dollars from morning commutes.[11]

Now the important issue is whether the total congestion delay is equivalent to the economic inefficiency cost. The Vickrey (1969) bottleneck model suggests equivalence. But travel-time profile is not the same as the iso-cost curve, raising a question. Perhaps congestion delay is imperfect as a measure of economic inefficiency. We need to answer the question of what total travel times would be if all behaved collectively so as to minimize them. We discuss this issue further below.

[10] Specifically, we first calculate the ratio of county sample size in our CHTS sample to the county population size and assign a weight to each county by the inverse of this ratio. We then aggregate them all over individuals, which we based on to assign an individual's weight to make sure that the sum of weights is unity.

[11] One could interact the individual specific queuing time to apply different VOT to individuals with different incomes, which, however, must assume that we have information about the distribution of VOTs by income.

Finally, in Figure 11, we illustrate a cumulative function of total queuing time by arrival time, which is written as

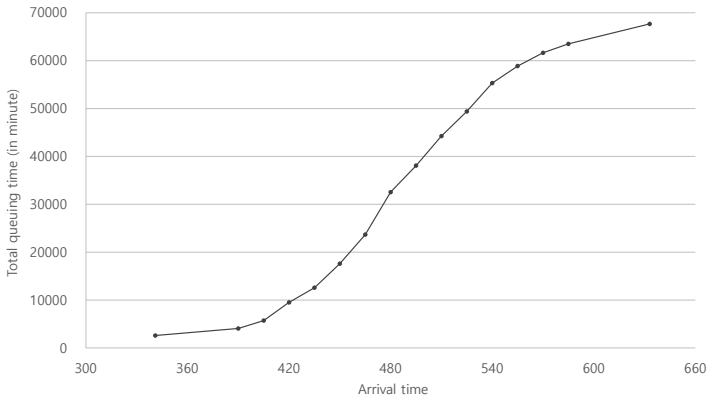
$$F(t) \equiv \int_{-\infty}^{t} f(m)dm, \tag{11}$$

where

$$f(m) = \frac{\sum_{i=1}^{N} \left(\widehat{Q_i(m)} | t_i \in m \right)}{l_m}, \tag{12}$$

where l_m is the width of the arrival time interval (minutes in the interval). Note that since the total queuing time for those arriving in m is divided by the interval width in minutes, $f(m)$ gives the rate of increase in total queuing time TQ by each minute. As expected, the $F(t)$ curve is increasing at the highest rate during the peak time, while the curve is flattened during non-peak times. This means that commuters who arrive in the peak times (7:30–9:00) contribute most to the aggregate congestion costs, which is due to both their long queue per person and the high rate of arrivals during these peak times.

Figure 11: Cumulative Function of Total Queuing Times



Notes: This graph shows how the total queuing time of the commuters in the sample, i.e., expression (8), increases over time in a given day. The vertical axis represents expression (11).

5. The Effect of Congestion on Scheduling Choices

The key theme of this paper is that people all have different routes (paths) to works. A natural question would then be, how would an individual commuter respond to the condition of her specific route? If there is certain adaptation pattern, how is it significant for the estimation of scheduling preferences? What is the implication of such a pattern on congestion costs and optimal tolling policy? We now turn to focus on these questions.

We use a regression model to test the adaptation hypothesis of whether commuters (randomly) allocated on a highly congested route tend to avoid the long queue at peak by choosing to arrive either earlier or later than the peak time. In effect, we aim at estimating the causal effect of congestion of the route on trip timing choice. The key explanatory variable should measure Q^{peak}, the longest queue if the commuter had arrived at the peak time t_{peak} (see Figure 1 visualizing Q^{peak}), which is measured by expression (3) defined above.

We estimate a simple model of binary choices between an arrival at the peak hour (7:52–8:52) and non-peak hours. In our sample, about 30% of commuters arrive during this peak hour. Our estimating equation is written as

$$arrpeak_i = \delta_0 + \delta_1 \widehat{Q_i^{peak}} + X_i\Gamma + \epsilon_i, \tag{13}$$

where $arrpeak_i$ is an indicator whose value is 1 if arrival is in the peak hour (between 7:52–8:52), X_i is the vector of observable worker and trip characteristics, and ϵ_i is the error term. The adaption hypothesis implies a lower chance of choosing the peak hour if the commuter were to travel a more congested route under which $\delta_1 < 0$. Importantly—unlike the realized queuing time (2)—travel time outcomes are not interacted with the commuter's trip timing choice in expression (3), which implies that the key explanatory variable is not a function of the arrival time itself, and there is no reverse causality issue, at least not by construction of the variables themselves.

Table 3 reports the estimation results. (Note that the CHTS classifies 21 industry categories and 25 occupation categories.) First, from the simple regression reported in column (1), we find a negative coefficient on the queue variable, confirming that commuters meeting a longer queue at the peak tend to arrive at a non-peak time. The simple correlation found in column (1) may be due to the tendency of longer-distance travelers to arrive at a non-peak time.

Column (2) controls for the trip distance, from which we find that we lose some significance, but we still find a negative estimated coefficient. In column (3), we include the home-county fixed effects as well as a set of personal characteristics as controls. In column (4), we include detailed information on the workers' occupation and the industry dummies. As we include more meaningful controls, we have a more significant adaptation pattern (a more negative estimate for δ_1), which implies that the failure to control for these variables would underestimate commuters' tendency to adapt.

We need to be careful about the possibility of sorting, under which scenario our estimate may be biased. Specifically, a higher level of congestion on a route at the peak hour may mean that the users of the

route tend to have a stronger preference toward the peak hour, so a commuter's arrival at the peak hour may explain the congestion on her route. While we try to control for factors determining commuters' scheduling preferences, the concern of omitted variables remains.

We use the instrumental-variation regression to address this concern. We use the mean congestion of workers who have the same workplace as the commuter. This variable reflects the level of congestion around the workplace of the commuter, not her particular scheduling preference. The IV estimates reported in columns (4)–(5) are a bit larger than the OLS estimates, implying that the adaptation pattern is a bit stronger than what is observed by the OLS estimates. This result is intuitive, since the unobservable characteristics related to the choice of arrival at the peak hour would be associated with a higher level of congestion along the route at the peak hour, which leads to an overestimation (less negativity) of the true effect of congestion on peak trip timing choice. Finally, column (7) reports the Probit estimation result, which exhibits a rather large effect of congestion on the peak-time arrival choice.

The adaptation tendency implied in our estimates is small. In particular, for an 8-minute increase (i.e., the sample mean for $\widehat{Q_i^{peak}}$) in the peak queuing time, there is only about a 2% reduction in the probability of choosing the peak hour. As we shall argue later, the small adaptation tendency does not imply that commuters are very inflexible. Our scheduling preference parameter estimates imply that commuters are somewhat inflexible, but not very inflexible. We will provide our explanation reconciling these findings below.

Table 3: Testing the Adaptation Hypothesis

Dependent variable: Dummy for arrival in 7:52–8:52

	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) IV	(6) IV	(7) Probit
Expected queuing time at peak	-0.00211***	-0.00114**	-0.00180***	-0.00296***	-0.00274***	-0.00322***	-0.00938***
	(-4.72)	(-2.31)	(-3.48)	(-4.74)	(-2.65)	(-3.03)	(-4.53)
Home/work distance (miles)		-0.00167***	-0.00112***	-0.000449	-0.000529	-0.000358	-0.00148
		(-4.70)	(-3.03)	(-1.05)	(-1.01)	(-0.68)	(-1.08)
Number of household members			-0.00363	0.00181	0.00180	0.00183	0.00555
			(-0.87)	(0.42)	(0.42)	(0.43)	(0.41)
Multi-worker household			0.000195	0.00131	0.00133	0.00128	0.00354
			(0.02)	(0.13)	(0.13)	(0.13)	(0.11)
Has students			0.0320***	0.0325***	0.0326***	0.0325***	0.0973***
			(2.84)	(2.83)	(2.84)	(2.84)	(2.78)
Homeowner			-0.00549	-0.00631	-0.00631	-0.00631	-0.0197
			(-0.48)	(-0.53)	(-0.53)	(-0.53)	(-0.53)
Annual family income in $1,000			0.000252***	0.0000779	0.0000776	0.0000782	0.000222
			(3.40)	(0.97)	(0.97)	(0.98)	(0.95)
Female			0.0669***	0.0400***	0.0400***	0.0399***	0.122***
			(7.55)	(4.05)	(4.08)	(4.06)	(4.11)
College degree or more			0.0898***	0.0741***	0.0740***	0.0741***	0.226***
			(9.51)	(6.91)	(6.94)	(6.95)	(6.96)
White			0.0154	0.0135	0.0137	0.0133	0.0431
			(1.57)	(1.32)	(1.34)	(1.31)	(1.36)
Has less than 5 full work days			-0.0424***	-0.0453***	-0.0454***	-0.0453***	-0.142***
			(-3.57)	(-3.70)	(-3.72)	(-3.72)	(-3.63)
Flexible worker			-0.0136	-0.0276**	-0.0276**	-0.0277**	-0.0809*
			(-1.01)	(-1.99)	(-2.00)	(-2.01)	(-1.90)
County fixed effects	No	No	No	Yes	Yes	Yes	Yes
Industry fixed effects[a]	No	No	No	Yes	Yes	Yes	Yes
Occupation fixed effects[a]	No	No	No	Yes	Yes	Yes	Yes
N	11,774	11,774	10,854	10,578	10,578	10,578	10,577
R-sq	0.002	0.003	0.025	0.051	0.051	0.051	

In Table 4, we estimate several alternative specifications to further test the adaptation hypothesis. First in column (1), we use only the sample of trips with arrivals earlier than 8:52 to observe the choice between an arrival at the peak (7:52–8:52) and an earlier arrival (than 7:52), and we find a slightly stronger adaptation tendency. In column (2), we use only the sample commutes arriving at the peak hour and later times (7:52–8:52 vs. later than 8:52) and find a similar adaptation tendency as those in Table 4.

In columns (3)–(4), we use the queuing time per mile as the explanatory variable. Regardless of inclusion of the distance variable as control, we again find a statistically significant effect of congestion on the tendency to choose the peak hour. However, the magnitude of this effect is still small. For a 10-minute increase in the queuing time per mile, there is only around a 3% lower chance of arriving at the peak hour.

In column (5), we modify the queuing time to the log scale and find that the elasticity of choice probability with respect to log of travel is actually small. For a 100% increase in the queuing time at the peak, we find that there is a 1.2% lesser chance of choosing the peak time as the arrival time. Finally, we used the log of queuing time per mile, which exhibits a similar estimate for the elasticity.

Overall, we find a statistically significant evidence of adaptation behavior, which is consistent with the hypothesis from our conceptual framework. Specifically, commuters (randomly) assigned on a route that is highly congested at the peak time tend to adapt to the condition by arriving at a non-peak time, i.e., congestion on a route causes commuters to adjust their trip schedule. This pattern is robust to the inclusion of many control variables, especially the commute distance, and commuters even adapt to the queuing time per mile traveled; addressing the omitted variable concern preserves our results by using IV. However, the adaptation tendency is quite weak. For a 100% increase in travel time at the peak hour, there is only a 3% lower chance of choosing the peak time. For the small adaptation tendency, we need an explanation, which we provide below after estimating the scheduling preferences in the next section.

Table 4: Testing the Adaptation Hypothesis, Alternative Specifications

	Dependent variable: Dummy for arrival in 7:52–8:52					
	(1)	(2)	(3)	(4)	(5)	(6)
	Early vs. Peak	Late vs. Peak				
Expected queuing time at peak	-0.00421***	-0.00296***				
	(-5.47)	(-2.88)				
Queuing time per mile			-0.0323***	-0.0245**		
			(-2.62)	(-2.00)		
Log of queuing time					-0.0123**	
					(-2.39)	
Log of queuing time per mile						-0.0126**
						(-2.35)
Home/work distance (miles)	-0.00113**	0.000573	-0.00166***		-0.00102**	-0.00170***
	(-2.23)	(0.83)	(-4.54)		(-2.37)	(-4.49)
Household and personal characteristics[a]	Yes	Yes	Yes	Yes	Yes	Yes
County fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Industry fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Occupation fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
N	8,082	5,574	10,578	10,578	10,317	10,317
R-sq	0.117	0.087	0.050	0.048	0.050	0.049

Notes: a. See Table 3 for the list of control variables. b. OLS are used in all specifications.

6. Estimation of Scheduling Utilities

In this section, we estimate the schedule preferences of California commuters. The knowledge on scheduling preferences is used to develop useful implications for tolling policy.

6.1 The Scheduling Utility Formulation and the Logit Model

An individual is assumed to choose an arrival time interval among 13 intervals, with each interval's width being 15 minutes from 6:22 to 9:52 AM. We exclude the trips ending outside of this range, since these travelers would have quite different preferences. The systematic utility that the commuter i obtains by choosing arrival time interval m is specified as

$$V_{mi} = \sum_{j=1}^{J} \left(-\alpha_j T_{mi}^j - \beta_j SDE_{mi}^j - \gamma_j SDL_{mi}^j \right), \tag{14}$$

where: T_{mi} is the travel time that i would face if choosing the interval m; $SDE_{mi} = \max(0, t_i^* - t_{mi})$, where t_i^* is i's ideal arrival time and t_{mi} is the mid-point time of the interval in which i's arrival time belongs; and $SDL_{mi} = \max(0, t_{mi} - t_i^*)$. Note that this specification allows ideal arrival time t_i^* to be individual-specific. Just below, we use machine learning to assign the individuals with their desired arrival times. We include the polynomials of the utility components to allow the non-linearity of the utility.

With ϵ_{im} being the individual-specific factors determining the utility, the utility is $V_{mi} + \epsilon_{mi}$. In the multinomial logit framework, in which the extreme value distribution of ϵ_{mi} is assumed, the probability of choosing alternative m from choice set M by decision maker i is given by:

$$P_{mi} = \frac{\exp V_{mi}}{\sum_{l=1}^{M} \exp(V_{li})}, \tag{15}$$

and the likelihood is written

$$LL \equiv \sum_{i=1}^{N} \log(\Pi_{mi}^{y_{mi}}), \tag{16}$$

where $y_{mi} = 1$ indicates that m is i's chosen alternative (and y_{mi} is otherwise 0). The maximum likelihood estimation is used to estimate the values of the preference parameters α_j, β_j, and γ_j.

6.2 Estimating Ideal Arrival Times

Conjecture

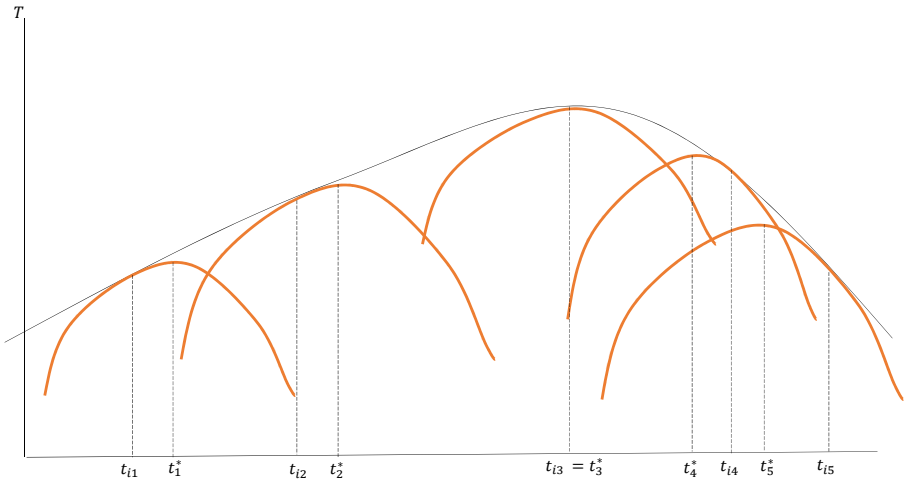
It is essential to first estimate individual-specific ideal arrival times, since they define the schedule delays as a key attribute considered in decision making. It is useful to first make a conjecture on them based on our framework. Our conjecture is mainly based on our finding that the adaptation tendency is quite weak. We suspect that the weak adaptation tendency is mainly due to rather flat travel-time profiles, even for

routes that are highly congested during rush hours. From the travel-time profiles drawn from Google Maps, the travel time on average increases only about 7 minutes during the first half of the morning period (from an early time, say 6:00 AM, to the peak time, 8:00 AM). These travel-time profiles should be much flatter than most commuters' iso-cost curves, according to earlier studies estimating scheduling preferences (Small, 1982; Kreindler, 2020), which report an estimated β to α ratio of 0.3–0.5, which implies a one-hour increase in travel time corresponding to two-hour changes in arrival times.

Because the condition for an arrival at a non-preferred time is a steeper travel-time profile that that of the iso-cost curve around the commuter's ideal arrival time (see Section 2), it is difficult for the flat travel-time profiles for congested routes to induce a rescheduling of arrivals. We can also say that commuters choosing non-peak times choose them mainly because these times are actually close to their preferred times and not because they adapt to potential long delays at the peak time.

This situation is described in Figure 12. In the graph, we hypothesize non-linear iso-cost curves, whose slopes are higher than that of the travel-time profile. Comparing each different individuals' ideal arrival times t_i^* with their actual arrival times t_i, we observe that commuters tend to arrive just near their ideal arrival times, so that one's ideal time is closer to one's actual arrival time than to others' arrival times.

Figure 12: Conjecture



Based on our framework, we therefore offer the conjecture that commuters' actual arrival times are overall quite close to their ideal arrival times. Since a commuter's actual arrival time tells us most of the information on her ideal arrival time according to our conjecture, we use the actual time to infer the ideal arrival time. We incorporate this conjecture in finally assigning ideal arrival times to individual commuters below.

Machine Learning Estimation and Update

We use our conjecture and the machine learning estimation to finally assign individuals with their own ideal arrival times. For the machine learning estimation, the first step is to define the "example group," i.e., the group of commuters who would reveal how agents would behave in a certain condition. Because our goal is to know commuters' ideal arrival times, we define the example group as commuters who actually arrive at their ideal arrival times. They arrive at their ideal times because they do not meet any congestion on their routes. In the variants of the bottleneck model in which residents are distributed over space and travel different distances, including Fosgerau and de Palma (2012), Fosgerau et al. (2018), and Fosgerau and Kim (2019), commuters traveling the shortest distance tend to arrive at their most preferred arrival times. Motivated by this observation, we define the example group of commuters as the commuters who have the same zip code for home and work, with a distance under 3 miles and a self-reported trip duration that is less than 10 minutes. There are 1,388 such commuters in the example group sample.

Figure 13 illustrates the distributions of the arrival times of the commuters in the example group and those of the commuters in the "learning group" (for whom the ideal times should be estimated). As expected, the arrival times of the commuters in the example group (ideal arrival times) are more concentrated around the peak than those of the learning group. However, the difference is not very large, which is consistent with our conjecture that actual arrivals are close to the ideal times.

The "machine" first learns how observed characteristics, such as incomes, occupation, and family size, affect the arrival times of commuters in the example group. We find that the information on workers' jobs (composed of 21 industry dummies and 24 occupation dummies) have an especially large contribution to the prediction of arrival times. To test the prediction power, we randomly select 30% of the 1,388 observations in the example group and predict their arrivals using the rest of the group (70%). We then plot predicted arrival times against the (known) actual arrival times in Figure 14. The figure indicates that machine learning tends to over-estimate the arrival times for earlier arrivals and under-estimate for later arrivals, so that the distribution of predicted ideal arrival times is systematically too concentrated around the center. This is probably because early or late arrivals are not quite predicted by any systematic factors, so they are regarded as being determined mainly by individual-specific factors. We try to correct these errors using our own conjecture below. While our machine learning estimates exhibit large systematic prediction errors, they still offer much better prediction power than out-of-sample linear regression prediction.[12]

Using the information learned from choices made by commuters in the example group, we finally predict counterfactual arrival times at work that would be chosen in the absence of queues, i.e., ideal arrival times, for the learning group. Since we assume that commuters traveling a short enough distance would arrive at their ideal times, we impose the average distance of the example group (1.4 miles) as the common trip distance for the learning group in predicting their counterfactual arrival times. Then, the machine

[12] From the predicted and actual values plotted in Figure 14 drawn from machine learning estimates, the correlation coefficient is 0.31, and the R-squared from their linear regression is about 0.1. When like plots are drawn from the predicted values made from out-of-sample linear regression prediction, the correlation between them is smaller.

learning effectively matches commuters with the same quality between the example and non-example group to make predictions on ideal arrival time conditional on the short trip distance.

Figure 13: Distribution of Arrivals for Example and Learning Groups

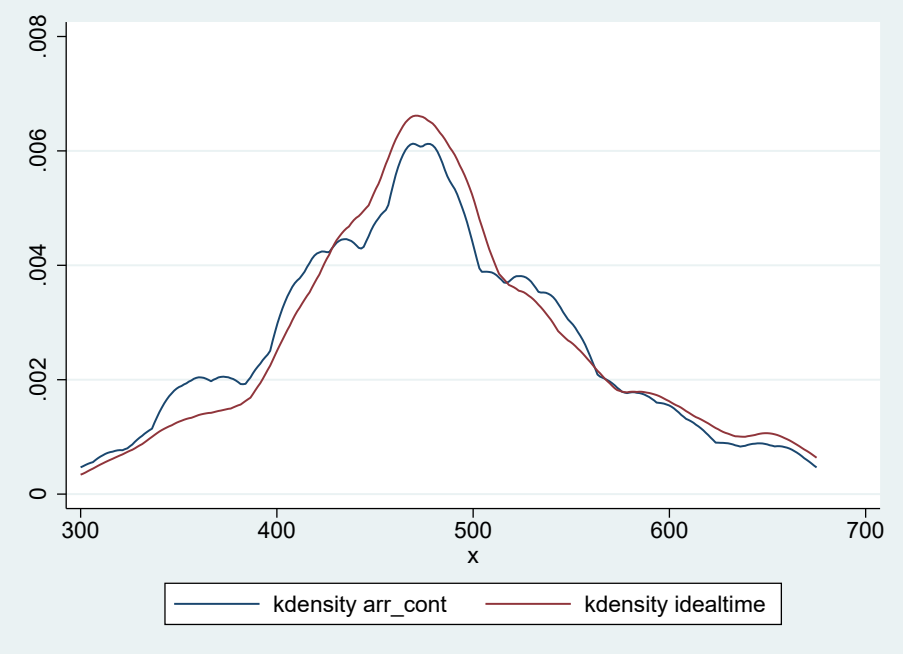
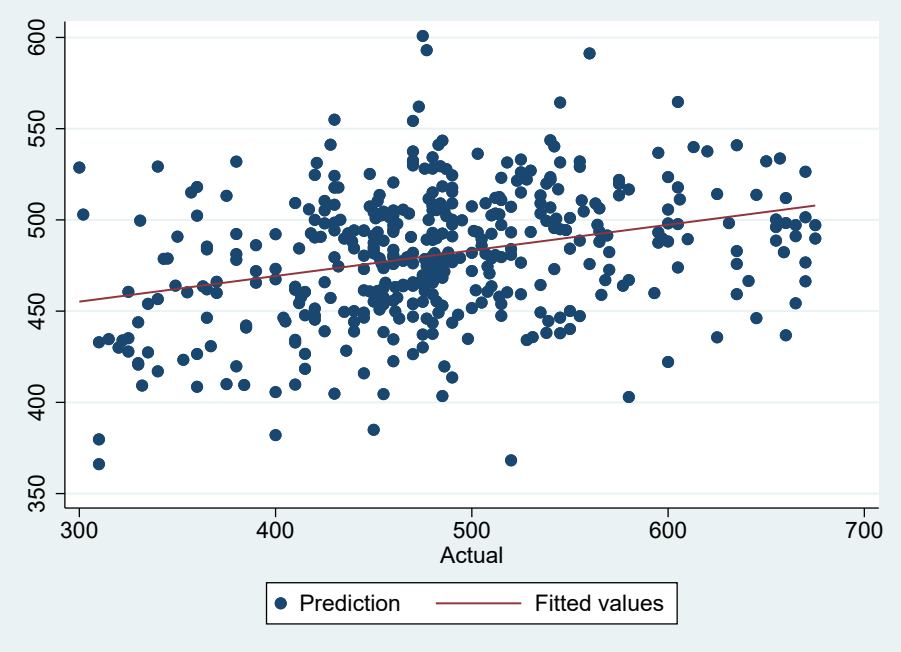


Figure 14: Scatter Plots for True and Predicted Arrival Times



Now, we try to correct the prediction errors generated from machine learning. We consider an error-correction mechanism which incorporates our conjecture that ideal arrival times are close to actual arrival times. Since the machine learning estimates do not utilize the information on commuters' actual arrival times directly, we directly incorporate them simply by weighting the two in the following form:

$$\widehat{t_i^*} \equiv \pi_0 + \pi_1 \widetilde{t_i^*} + (1 - \pi_1) t_i, \tag{17}$$

where $\widehat{t_i^*}$ is our final estimate for t_i^*, $\widetilde{t_i^*}$ is the machine learning estimate for ideal arrival time, t_i is i's actual arrival time, and π_1 is a parameter with $0 < \pi_1 < 1$. A constant π_0 is introduced to flexibly correct errors in level in the machine learning estimates. Under the assumption that ideal arrival time is close to actual arrival time, the update would certainly tend to reduce prediction errors.[13] The weighting of the two may be interpreted in the other direction: first randomizing ideal times around the actual arrival times, and the generated errors then being corrected by incorporating individuals' observed characteristics using machine learning.

The parameter values for π_1 and π_0 determine the distributions of estimated ideal times and schedule delays—and, thereby, our scheduling preference estimates.[14] We choose these parameter values to seek the best fit of the estimated specification to the data. We particularly ensure that the slope of the estimated quadratic utility is about zero around the ideal time by imposing the condition $\beta_1 = \gamma_1 = 0$ in specification (14). If one of them is positive, for instance, the scheduling utility profile is increasing locally around t^*, which would imply that the estimates of ideal times are systematically biased. While we report only estimates given our chosen parameter values of π_1 and π_0 below, we also tried other reasonable sets of these parameter values, but the preference parameter estimates are quite robust, so our conclusion does not largely rely on these parameter values.

[13] To see this point, let $e_i = t_i^* - \widehat{t_i^*}$ be the error from the machine learning. Plugging this expression to (17) to eliminate $\widetilde{t_i^*}$ and rearranging terms, we can write the error in prediction from using $\widehat{t_i^*}$ as the following (imposing π_0 for simple exposition):

$$t_i^* - \widehat{t_i^*} = (1 - \pi_1)(t_i^* - t_i) + \pi_1 e_i. \tag{18}$$

From the figure above, we can see that for earlier arrivals than ideal times with $t_i^* - t_i > 0$, e_i tends to be negative, and for late arrivals with $t_i^* - t_i < 0$, e_i tends to be positive. Since $0 < \pi_1 < 1$ is assumed, the first and the second term therefore would tend to cancel out, which would reduce the prediction errors. There is also the potential for a selection bias concern, because the example and the learning group may be systematically different in scheduling behaviors. The large set of predictors would contribute to addressing this concern. Also, as weighting effectively applies to both groups, weighting may help to mitigate this concern.

[14] In particular, as π_1 gets smaller, since we utilize less of the machine learning estimates and more of our conjecture, the distribution of the estimated ideal times get flattened and the distribution of schedule delays get sharpened.

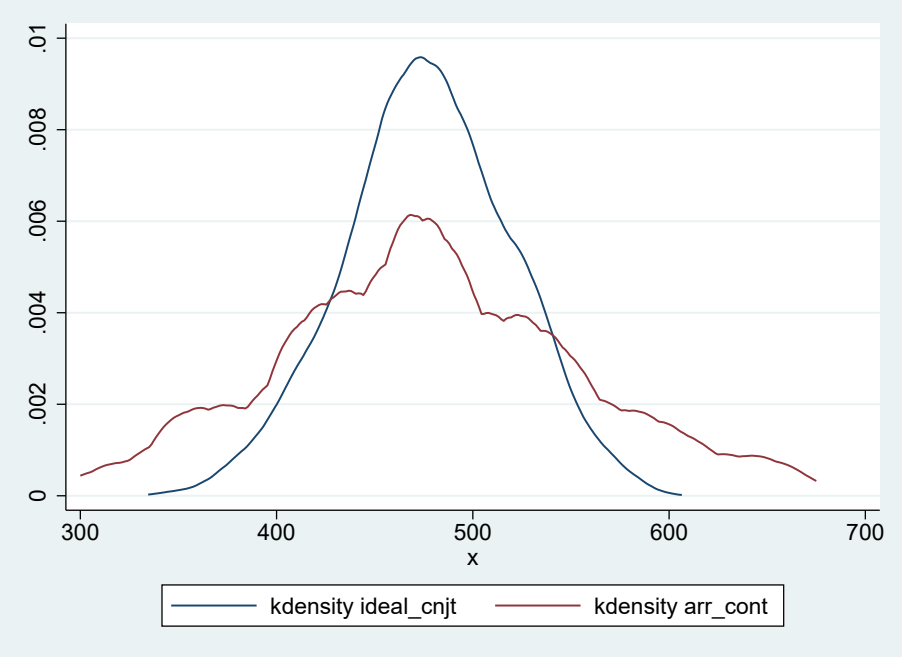
6.3 Scheduling Preference Parameters Estimates

We choose the parameter values of $\pi_1 = 0.6$ and $\pi_0 = -4.5$, which yields minimum standard errors of the estimates and ensure that the quadratic utilities have a zero slope at the ideal times. Figure 15 depicts the distribution of estimated ideal arrival times under the imposed parameters as well as the distribution of actual arrival times.[15]

In panel (a) of Figure 16, machine learning estimated ideal arrival times are plotted against actual arrival times. The large deviations of the points from the fitted line imply large prediction errors. In panel (b), we plot the conjecture-based adjusted ideal times against actual arrival times, which exhibits a stronger relationship between the two.

In Figure 17, we report the distribution of schedule delays, $t_{mi} - t_i^*$, calculated based on conjecture-adjusted estimates for ideal arrival times. In this figure, note that a negative value of the schedule delay means an earlier arrival. The median value is the schedule delay is 27.96 (minutes in absolute value) for arrivals earlier than the ideal time, and that of later arrivals is similar at 27.97.

Figure 15: Distribution of Ideal and Actual Arrival Times



[15] Note that the density of ideal times in this graph is relatively sharp compared to that in Figure 13, which we are using as a benchmark distribution of ideal arrival times. We could impose a small value for π_1 to flatten the curve. However, our goal is to ensure a good fit of each of the commuters rather the unidimensional distribution of ideal times. Our worry about having a small pi is that the ideal time estimates get too close to actual arrival times and generate overly small schedule delays, which would then bias the result.

Figure 16: Scatter Plots for Ideal and Actual Arrival Times

(a) Ideal times estimates not corrected (b) Ideal times corrected

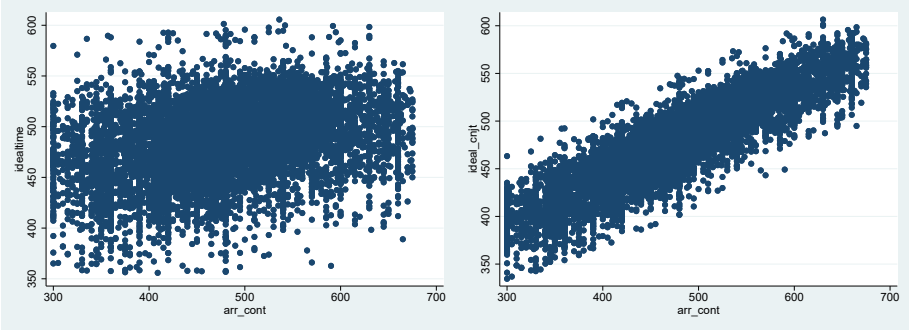


Figure 17: Distribution of Schedule Delays

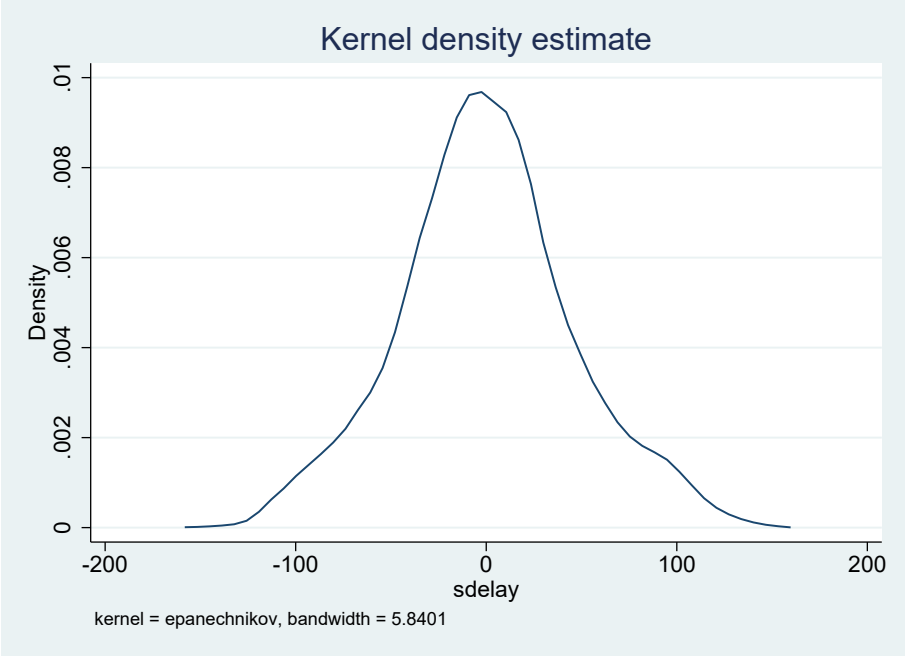


Table 5 reports our preference parameter estimates. In the simplest specification in column (1), we find that the coefficient of travel time (i.e., α_1) is -0.0695, and that of SDE is -0.0377, so the ratio of β_1 to α_1 is about 0.54. From the coefficient of SDL, the ratio γ_1/α_1 is 0.49. These numbers mean that commuters are willing to arrive early or late by an hour on average in order to reduce the travel time by 30 minutes. Or, the cost of staying at work early before one's shift starts (and the cost of being late) is about half of the time spent in the car.

In specification (2), we include the square terms of the schedule delays. As mentioned above, in this specification, we ensure that the coefficients the schedule delays (for early and late arrivals) are not positive, since otherwise there would be a locally increasing utility moving away from the ideal time. The large and significant coefficients on the quadratic terms supports the non-linearity of the utility, approximated by the quadratic form. In specification (3), we include the square of the travel time and the cubics of the schedule delays. From the positive coefficient on the square of the travel time, we can see that the marginal disutility from travel time diminishes. The coefficients on the cubics of the schedule delays are only weakly significant or not significant at all, so we prefer the quadratic specification.

In the specifications reported in columns (3)–(5), we include interaction terms to see how commuters in one group value the utility components differently from those in the other group. First, in column (3), we interact the utility components with the dummy for "rich" commuter, which indicates an annual income exceeding $100,000. We find that the higher-income group is more sensitive to all the utility components than the lower-income group. Especially, the rich group's unit cost of travel time is larger than that of the "poor" group by the largest amount (relative to the gap of the unit costs of the schedule delays between the two groups). The ratio of β_1 to α_1 is 0.479 for the high-income group and 0.644 for the low-income group, which makes sense if we consider the higher value of time for higher-income earners than lower-income earners, which is well-documented in the literature. In specification (4), we include the interaction terms with the female dummy, which shows that women are relatively more sensitive to schedule delays, or that they are more inflexible, than men. In specification (5), we define a "flexible" worker dummy which indicates workers in the CHTS sample who said that they are fairly flexible to adjust their work schedules or that their days at the primary workplace are fewer than 5. We can anticipate that these workers are also flexible to adjust their scheduling *within* a work day. As expected, flexible workers have relatively smaller schedule delay costs—evidence which may be regarded as validating our estimates.

Here is a short retrospective discussion on our assumption to use Eq. (18), which was made from our theory-based conjecture. A presumption behind this conjecture was a that the iso-cost curve is steeper than the travel-time profile, which may be unduly forcing the revelation of such situation. However, even with a higher value for π_1 (using less of the conjecture), or even with $\pi_1 = 1$ (which does not use the conjecture at all), our estimated parameters still imply steeper iso-cost curves. Thus, the conjecture itself does not drive the conclusion that the iso-cost curves are steeper, but rather, we use it to more accurately estimate the scheduling utility parameters.

Table 5: Scheduling Utility Estimates

	(1)	(2)	(3)	(4)	(5)	(6)
T	-0.0695***	-0.0764***	-0.145***	-0.0570***	-0.0682***	-0.0667***
	(0.00490)	(0.00517)	(0.0125)	(0.00735)	(0.00657)	(0.00560)
SDE	-0.0377***	-0.000419	0.00473	-0.0367***	-0.0360***	-0.0394***
	(0.000587)	(0.00201)	(0.00375)	(0.000871)	(0.000836)	(0.000701)
SDL	-0.0343***	-0.00121	0.00584*	-0.0313***	-0.0323***	-0.0354***
	(0.000477)	(0.00186)	(0.00352)	(0.000604)	(0.000633)	(0.000546)
SDE^2		-0.000529***	-0.000717***			
		(0.0000318)	(0.000117)			
SDL^2		-0.000419***	-0.000642***			
		(0.0000268)	(0.000101)			
T^2			0.000676***			
			(0.000113)			
SDE^3			0.00000152			
			(0.000000976)			
SDL^3			0.00000168**			
			(0.000000772)			
$T \times rich$				-0.0250**		
				(0.00984)		
$SDE \times rich$				-0.00258**		
				(0.00118)		
$SDL \times rich$				-0.00749***		
				(0.000965)		
$T \times female$					-0.00357	
					(0.00983)	
$SDE \times female$					-0.00359***	
					(0.00117)	
$SDL \times female$					-0.00430***	
					(0.000957)	
$T \times flex$						-0.0139
						(0.0116)
$SDE \times flex$						0.00619***
						(0.00125)
$SDL \times flex$						0.00475***
						(0.00114)
Interval ends with 15 or 45	-0.277***	-0.276***	-0.276***	-0.277***	-0.277***	-0.278***
	(0.0221)	(0.0220)	(0.0220)	(0.0221)	(0.0221)	(0.0221)
Sample size	118048	118048	118048	118048	118048	118048
Number of cases (trips)	7378	7378	7378	7378	7378	7378
Log likelihood	-18236.8	-17945.7	-17923.2	-18211.3	-18228.1	-18220.0
β/α	0.54			0.4790 (rich)	0.5516 (female)	0.4120 (flex)
				0.6439 (poor)	0.5279 (male)	0.5907 (non-)
γ/α	0.49			0.4730 (rich)	0.5100 (female)	0.3803 (flex)
				0.5491 (poor)	0.4736 (male)	0.5307 (non-)

Notes: a. Logit is used for estimation. The choice set is intervals 2 to 15 spanning from 6:22 to 9:52. b. Intervals of 1 and 16 are dropped so the individual is assumed to choose among 13 intervals with each interval width being 15 minutes. c. Parameter values for $\pi_1 = 0.6$ and $\pi_0 = -4.5$ are imposed. d. Standard errors are in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

7. The Implications of the Findings and Discussion

In this section, we provide some further insights on trip scheduling behavior based on our empirical evidence. We also discuss some issues around measuring the economic inefficiency resulting from congestion and the design of a congestion tolling policy.

7.1 Scheduling Choices and Ideal Arrival Times

In panel (a) of Figure 18, we use the parameter estimates on the linear-utility specification to draw an indifference curve, with the commuter's ideal arrival time assumed to be 8:15 at which the mean travel time is the longest. The average travel-time profile of Figure 7 is also placed to examine the commuter's arrival time choice. Given β/α (and γ/α) of about 0.5, travel time must fall by about half an hour to compensate for the increase in schedule-delay costs arising from arrivals one hour earlier or later than the ideal time. However, the much flatter travel-time profile means that travel time on average falls by at most 2–3 minutes during the one-hour interval. As a result, the commuter chooses to arrive at her ideal time, which is 8:15 in this example.

In panel (b), we use the quadratic utility estimates to draw an indifference curve, which shows that the commuter is quite flexible locally around the ideal time. However, travel time still falls too slowly to compensate for the increasing schedule delay costs from arrivals later or earlier than the ideal time, under which the optimal choice is, again, to arrive at her ideal time. If we assume a different ideal arrival time and a convex utility curve, then the commuter's optimal choice would be different from the ideal time. However, as long as the indifference curve is steeper than the travel-time profile, the chosen arrival timing would be close to the ideal time. Of course, there must be a large degree of heterogeneity in travel-time profiles and scheduling utilities among commuters. But the large difference between the average travel-time profile and the average scheduling utility suggests a strong tendency to arrive near one's ideal arrival time, which validates the conjecture made in the preceding section.

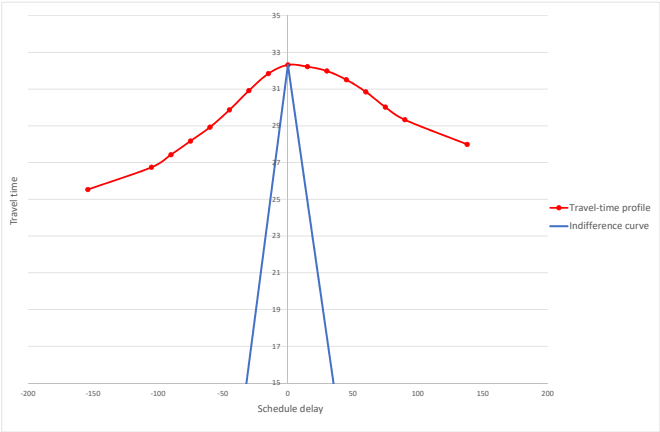
This situation would also have an implication for the distribution of ideal arrival times. Note that we confront a seemingly contradictory result: commuters exhibit a strong tendency to arrive around their desired arrival times while the resulting travel-time profiles are quite flat. This can be resolved if commuters' desired arrival times are quite dispersed.[16]

Finally, the large gap in slopes between travel-time profiles and schedule utility curves suggests that a small change in the level of congestion would not be enough to induce a choice of arrival at an inconvenient early or late time for the majority of commuters, which explains the the small adaptation tendency found in Section 5.

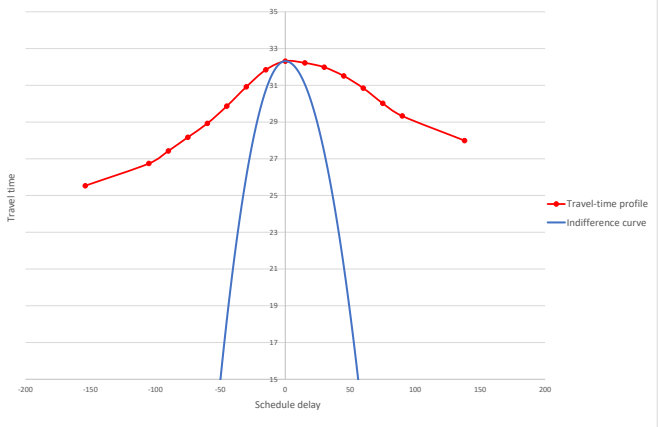
[16] The large dispersal of ideal arrival times is consistent with Hall (2021a). Small et al. (2005) also found a larger heterogeneity in the preference of road users. The relatively stable travel times caught researchers' attention in other literature (see Anas, 2015; Hall, 2021b).

Figure 18: Scheduling Utility and Travel-Time Profiles

(a) Linear utility



(b) Quadratic utility



7.2 Discussion on the Economic Cost of Congestion and the Policy of Congestion Tolling

We first discuss the quantification of the economic cost of congestion given our framework and empirical estimates. To narrow down the scope of our discussion, we rely on the bottleneck model tradition by conjecturing that the social optimum involves no queues, i.e., if all travelers could coordinate their choices, then they would arrive with no queues while fully utilizing the capacities of the roads.[17] Under this conjecture, the total queuing times quantifying the difference in total travel time costs between the social optimum and the laissez-faire condition estimated in this paper would be equivalent to the inefficiency cost arising from traffic congestion if the travel-time profiles were the same as the iso-cost curves (see Vickrey, 1969; Arnott et al., 1990, 1993).

However, given the large difference between the travel-time profiles faced by individuals and their scheduling utility curves discovered in this paper, measuring the cost of congestion becomes a non-trivial issue. The trade-offs implied in the two curves are quite different. In particular, the travel-time profile gives the schedule delays the commuter must accept to have a zero queuing time under the laissez-faire condition. The indifference curve gives the schedule delays she would be willing to accept for the removal of queuing she would otherwise experience under the laissez-faire condition. A difficult issue is which one to use when evaluating the economic benefit available from the removal of queues (or equivalently, when evaluating the cost from the presence of queues). Using the indifference curve may be best, given that it gives the reference utility level for comparison between the laissez-faire and the social optimum involving no queues. However, for the area under the indifference curve to correctly measure the economic inefficiency, it must ensure that all the commuters complete arrivals within a shorter morning time interval (since the interval implied in the indifference curve is much shorter than that of the travel-time profile as seen in Figure 18), which may not be possible due to the capacity of roads. If arrival times at the destination are instead preserved under the social optimum involving no queues, which is the result in the standard bottleneck models, then using the travel-time profile would give the correct inefficiency cost from congestion (and the benefit from the removal of queues).

Thus, to measure the economic cost of congestion more correctly, one would have to not only characterize the social optimum but also identify arrival times under the social optimum with heterogeneous travel-time profiles and endogenous scheduling choices in a similar framework to ours, which would then quantify the difference in total costs (not just total travel time costs) between the social optimum and the laissez-faire. This would be a difficult task, because each commuter is just one person contributing to the formulation of an individual travel-time profile that is shaped by others who in turn are shaping their own travel-time profiles.[18]

[17] Duranton and Venables (2018) argue that the congestion-free time may not be the reference point for identifying the social optimum. The model in their approach is based on Pigou (1920), where the externality is the source of drivers' failure to reach an outcome that is best for them collectively as a whole. Meanwhile, the framework that we use in this paper is the bottleneck model, where drivers' failure to coordinate their departure schedules lead to market failure.

[18] A similar analysis was carried out in Arnott et al. (1994), with an explicit consideration of the preference heterogeneity of commuters, but there is no heterogeneity in commuters' travel-time profiles in that paper.

The next issue is to how to design a tolling policy to achieve the social optimum in a decentralized setting. In the classical bottleneck model, in which the travel-time profile faced by individuals under the laissez-faire condition is the same as the iso-cost curve, the optimal time-varying toll schedule could be either. However, since the travel-time profile faced by an individual and her utility curve are largely different as discovered in this paper, neither may correspond to the socially optimal toll. Specifically, a toll resembling the travel-time profile (suggested in Kim, 2019) would not induce commuters to reschedule their trips, since the maximum benefit from rescheduling is the full removal of queuing times (7–8 minutes at most for typical commuters), which would not be enough to compensate the schedule-delay costs of several hours for those having average preferences. A utility-compensating toll (having the shape of the indifference curve) may impose an inefficiently large toll for an arrival around the peak time, given that arrivals are already quite dispersed. Thus, the conventional toll regime of the standard bottleneck model would be either ineffective or too harsh for commuters, and the large toll may face significant political resistance.

Our discussion highlights that because of the salient but often-ignored fact that commuters each face their own travel-time profiles, we are encountering challenges in the application of the key results of the standard bottleneck model to the quantification of the economic cost of congestion and the design of a congestion tolling policy. We may need a new or extended framework from the existing models that would incorporate the heterogeneity of travel-time profiles and the dispersal of ideal arrival times to revisit these fundamental issues in the economic models of congestion.

8. Conclusion

In this paper, we developed a simple model of trip timing choices faced by commuters facing their own travel-time profiles. We empirically quantified the key elements of the model such as the travel-time profiles and the scheduling utility function of car commuters in California. Using these parameters, we quantified the total congestion costs from morning commutes in California. We also empirically tested a hypothesis that commuters adapt to congestion dynamics by departing earlier or later in the morning if they would face a long congestion-related delay during the peak hour and found statistically significant evidence for it. Our estimated scheduling utility for commuters indicates that the unit cost of schedule delays is about half of the unit cost of travel time.

Perhaps the most important discovery is the large difference between travel-time profiles and indifference curves, from which we offer new insights on trip scheduling patterns and on the distribution of ideal arrival times. This finding raises some fundamental questions that cannot be easily answered in the classical bottleneck models. This paper therefore highlights the importance of considering the heterogeneity of location and routes when modeling congestion to solve these problems.

One limiting assumption of our model is the exogenous nature of ideal arrival times, so a possible extension would therefore be to endogenize the formulation of ideal arrival times. For example, we could distinguish the short-run and the long-run ideal arrival times, with the latter being determined by the long-run preference that shapes commuters' routines in the morning periods and determines the short-run ideal arrival time (see Peer et al., 2015; Verhoef, 2020, for more on this idea). We also need more research that investigates the topic in the same direction as our paper, using big data and new empirical tools to better apply economic models of traffic congestion and provide further insights into travel behavior.

Bibliography

Akbar, P., G. Duranton, V. Couture, and A. Storeygard (2019). Mobility and congestion in urban India. Unpublished paper.

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

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

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

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

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

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

Bento, A., J. D. Hall, and K. Heilmann (2021). Estimating the negative externalities of traffic congestion using big data. Unpublished paper.

Couture, V., G. Duranton, and M. Turner (2018). Speed. *Review of Economics and Statistics 100*(4), 725–739.

Duranton, G. and A. J. Venables (2018). Place-based policies for development. World Bank Policy Research Working Paper.

Fosgerau, M. and A. de Palma (2012). Congestion in a city with a central bottleneck. *Journal of Urban Economics 71*(3), 269–277.

Fosgerau, M. and J. Kim (2019). Commuting and land use in a city with bottlenecks: Theory and evidence. *Regional Science and Urban Economics 77*, 182–204.

Fosgerau, M., J. Kim, and A. Ranjan (2018). Vickrey meets Alonso: Commute scheduling and congestion in a monocentric city. *Journal of Urban Economics 105*, 40–53.

Glaeser, E. L., S. D. Kominers, M. Luca, and N. Naik (2018). Big data and big cities: The promises and limitations of improved measures of urban life. *Economic Inquiry 56*(1), 114–137.

Hall, J. D. (2021a). Can tolling help everyone? Estimating the aggregate and distributional consequences of congestion pricing. *Journal of the European Economic Association 19*(1), 441–474.

Hall, J. D. (2021b). Inframarginal travelers and transportation policy. Unpublished paper.

Hjorth, K., M. Börjesson, L. Engelson, and M. Fosgerau (2015). Estimating exponential scheduling preferences. *Transportation Research Part B: Methodological 81*(1), 230–251.

Hörcher, D., D. J. Graham, and R. J. Anderson (2017). Crowding cost estimation with large scale smart card and vehicle location data. *Transportation Research Part B: Methodological 95*, 105–125.

Kim, J. (2019). Estimating the social cost of congestion using the bottleneck model. *Economics of Transportation 19*, 100119.

Kreindler, G. E. (2020). Peak-hour road congestion pricing: Experimental evidence and equilibrium implications. Unpublished paper.

Kreindler, G. E. and Y. Miyauchi (2020). Measuring commuting and economic activity inside cities with cell phone records. Unpublished paper.

Peer, S., J. Knockaert, and E. T. Verhoef (2016). Train commuters' scheduling preferences: Evidence from a large-scale peak avoidance experiment. *Transportation Research Part B: Methodological 83*, 314–333.

Peer, S., E. T. Verhoef, J. Knockaert, P. Koster, and Y. Y. Tseng (2015). Long-run versus short-run perspectives on consumer scheduling: Evidence from a revealed-preference experiment among peak-hour road commuters. *International Economic Review 56*(1), 303–323.

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

Russo, A., M. W. Adler, F. Liberini, and J. N. van Ommeren (2021). Welfare losses of road congestion: Evidence from Rome. *Regional Science and Urban Economics 89*, 103692.

Selod, S. (2021). Big data in transportation. An economics perspective. Unpublished paper.

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

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

Small, K. A., C. Winston, and J. Yan (2005). Uncovering the distribution of motorists' preferences for travel time and reliability. *Econometrica 73*(4), 1367–1382.

Tang, C. K. (2021). The cost of traffic: Evidence from the london congestion charge. *Journal of Urban Economics 121*, 103302.

Tarduno, M. (2021). The congestion costs of Uber and Lyft. *Journal of Urban Economics 122*, 103318.

Verhoef, E. T. (2020). Optimal congestion pricing with diverging long-run and short-run scheduling preferences. *Transportation Research Part B 134*, 191–209.

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.

Wardman, M. (2001). A review of British evidence on time and service quality valuations. *Transportation Research Part E: Logistics and Transportation Review 37*(2–3), 107–128.

Yang, J., A. O. Purevjav, and S. Li (2020). The marginal cost of traffic congestion and road pricing: Evidence from a natural experiment in Beijing. *American Economic Journal: Economic Policy 12*(1), 418–453.

Appendix A. Further Notes on Data Construction

The first point pertains to how we deal with commuters traveling within the same zip code, since Google Maps would be unable to differentiate the locations between origin and destination for these commuters (with their expected travel time being always zero, since Google Maps would have the same zip code for home and work). Among the 9,127 different zip code pairs in the sample, 668 pairs have the same work and home zip code. For their home zip codes, we match the nearest home zip codes among other commuters traveling to the same work zip code as the commuter. Since the matched zip code is different and thus may overestimate the trip distance, we rescale the travel time outcomes using the ratio of the CHTS reported travel time and the Google Maps travel time predicted at the worker's chosen trip timing.

The second point pertains to the data collection algorithm we use to work with Google Maps data. We first set up the list of zip code pairs, and we then queried travel times for each of them at random timings between 5:00 and 11:00 AM on each weekday over the period from January 6, 2020, until July 28, 2021. The query time for each zip code pair is random since Google Maps randomly selects a zip code pair out of the set of listed zip code pairs at each timing and queries for it. The random selection of a route at each time is chosen to avoid the bias that may arise due to a systematically higher frequency of particular routes chosen at particular (peak) times within travel days; we try to collect full travel-time profiles for as many routes as possible.

The third point concerns the estimation of systematic travel time expectations that individuals have by alternate arrival time interval during days when congestion is anticipated. The notion in our framework is that each commuter would respond to her own systematic anticipation of travel times at her alternate choices, which would be formed from her own experiences of traveling over many days. Our goal is to estimate this anticipated menu of travel times, not one on a specific day, since it would involve unexpected delays due to a weather shock or accidents. This is the reason why we collected the data over a long horizon. However, out of the entire data collection period from January 6, 2020, to July 2021, we had to exclude the data collected subsequent to the lockdown policy enacted in California on March 19, 2020, in response to the soaring COVID-19 cases. Figure 6 shows a significant drop in travel times around this date and the trend has not fully recovered as of the summer of 2021. It is appropriate to exclude the travel time outcomes during these periods, because the commuters in the CHTS sample were subject to travel-time profiles similar to those of the pre-COVID-19 period and we want to explore scheduling choices under the congested situation. Thus, we limit our usage of Google Maps data to the "normal" period before March 19, 2020. We are left with 1,255,426 queried trips from January 6 to March 18, 2020, and we use these to construct the travel-time profiles during the congested (normal) time periods. When designing this research, we had not anticipated COVID-19, but we were fortunate to have begun data collection before the pandemic, as we obtained enough travel time predictions for each pair of home and work zip codes to construct the systematic travel-time profile for each commuter. Queried trips subsequent to COVID-19 lockdown measures were instead used to construct the congestion-free travel time for each route.

Finally, note that our final sample consists of 14,544 commute observations that have a "full" travel-time profile with no missing values on travel times in any interval or congestion-free travel period. This number is fewer than the 16,376 commuters in the raw CHTS sample: we lose 1,832 observations. The removed observations may have a missing value for the mean travel time in an interval for the route, since our query is random at each timing, so these zip code pairs failed to fill all the defined intervals. We drop the travel time menu for a route if it misses at least one value in the intervals, doing so in order to ensure that commuters in our sample face the "full" menu of travel times. Note however that the selection is effectively random, given the data collection algorithm for Google Maps.

Table A1: Ranking of Congestion Level by County

County	Population size[a]	Obs (CHTS)	Mean of $\widehat{Q_i^{peak}}$	Mean of per-mile $\widehat{Q_i^{peak}}_b$
San Francisco	881,549	322	17.52	1.59
Los Angeles	10,039,107	2839	14.64	1.08
Santa Clara	1,927,852	1159	13.02	0.91
Alameda	1,671,329	664	12.52	0.82
San Mateo	766,573	471	12.20	0.80
Marin	258,826	140	10.31	0.72
Orange	3,175,692	962	10.25	0.73
San Diego	3,338,330	598	9.70	0.68
Sacramento	1,552,058	359	8.74	0.65
Contra Costa	1,153,526	461	7.92	0.59
San Bernadino	2,180,085	547	6.25	0.41
Santa Cruz	273,213	188	5.65	0.53
Riverside	2,470,546	441	5.35	0.39
Monterey	434,061	365	4.88	0.50
Sonoma	494,336	305	4.36	0.38
Napa	137,744	130	4.34	0.28

Continued on next page

Table A1, continued

County	Population size[a]	Obs (CHTS)	Mean of $\widehat{Q_i^{peak}}$	Mean of per-mile $\widehat{Q_i^{peak}}$[b]
Solano	447,643	226	3.74	0.25
Ventura	846,006	378	3.56	0.33
Lassen	30,573	34	3.47	0.18
Stanislaus	550,660	166	3.38	0.35
Fresno	999,101	440	3.31	0.33
Yuba	78,668	61	3.20	0.18
San Joaquin	762,148	211	3.14	0.23
Yolo	220,500	86	2.94	0.20
Placer	398,329	151	2.94	0.31
Santa Barbara	446,499	218	2.86	0.24
Kern	900,202	369	2.56	0.18
El Dorado	192,843	98	2.56	0.27
Butte	219,186	122	2.51	0.23
Humboldt	135,558	95	2.36	0.28
Merced	277,680	144	2.32	0.18
San Benito	62,808	41	2.24	0.19
Madera	157,327	78	2.17	0.16
Sutter	96,971	53	2.05	0.12
Shasta	180,080	49	1.94	0.16
Tuolumne	54,478	62	1.90	0.17

Continued on next page

Table A1, continued

County	Population size[a]	Obs (CHTS)	Mean of $\widehat{Q_i^{peak}}$	Mean of per-mile $\widehat{Q_i^{peak}}_b$
Sierra	3,005	7	1.88	0.20
Siskiyou	43,539	46	1.87	0.12
Amador	39,752	32	1.80	0.13
Nevada	99,755	51	1.79	0.14
Tulare	466,195	270	1.53	0.15
San Luis Obispo	283,111	276	1.49	0.11
Lake	64,386	44	1.41	0.12
Calaveras	45,905	29	1.37	0.08
Kings	152,940	85	1.17	0.06
Mendocino	86,749	56	1.09	0.06
Mariposa	17,203	25	1.03	0.06
Glenn	28,393	44	0.96	0.13
Tehama	65,084	40	0.92	0.08
Imperial	181,215	205	0.87	0.07
Mono	14,444	29	0.80	0.04
Colusa	21,547	30	0.72	0.04
Plumas	18,807	47	0.56	0.04
Modoc	8,841	26	0.52	0.04
Trinity	12,285	27	0.50	0.06
Del Norte	27,812	74	0.20	0.02

Continued on next page

Table A1, continued

County	Population size[a]	Obs (CHTS)	Mean of $\widehat{Q_i^{peak}}$	Mean of per-mile $\widehat{Q_i^{peak}}$[b]
Inyo	18,039	68	0.08	0.00
All CA	**39,512,223**	**14,544**	**8.36**	**0.62**

Notes: This table shows the levels of congestion by work site county. Commuters are sorted by the county in which their workplaces are located, and the mean of peak queuing times and the mean of per-mile peak queuing time by the county groups are calculated. While the table sorts the counties (the mean of peak queuing times) from the highest to the lowest, note that the ranking based on the mean of per-mile queuing times is almost the same. a. The population size is of 2019 (Source for County Population Totals: 2010–2019, US Census Bureau). b. In calculating per-mile $\widehat{Q^{peak}}$, we use the Google Maps distance for the suggested route at each commuter's chosen arrival time.

Appendix B. Proof of Proposition 1

The proof of this proposition is as follows.

1. Points above the travel-time profile are not optimal, since the commuter could drive faster at any chosen trip timing and reduce commuting cost. Points below the travel-time profile are not feasible. Given the larger curvature of the iso-cost curves and $t_0 < t_* < t_1$, there exists at least one point at which an iso-cost curve and the travel-time profile meet, so it is sufficient to examine only the points on the travel-time profile. Given the curvature difference, the iso-cost curve is steeper than the travel-time profile at t_0, under which the commuter could have a lower-positioned iso-cost curve by arriving later than t_0, so t_0 is not optimal. The same reasoning is applied to show t_1 is not optimal. As another possibility, if an iso-cost curve crosses the travel-time profile at two points between t_0 and t_1, then neither of them can be optimal since the commuter could reduce the cost by choosing a time closer to t_*. The remaining possibility is that an iso-cost curve and the travel-time profile meet at a unique point, at which these two curves are tangent.

2. As an illustrating case, consider a convex travel-time profile and a linear iso-cost curve. If the two curves cross at any two points, then the commuter could shift the iso-cost curve downward by choosing a time closer to t_0, t_1, or t_*. Under the smaller curvature of the iso-cost curve, the two curves cross twice as long as the crossing points are not among t_0, t_1, or t_*, which implies that the only possible optimal choices are t_0, t_1, or t_*. This reasoning applies to any situation where the iso-cost curves have a smaller curvature than the travel-time profiles.

■

MTI FOUNDER

Hon. Norman Y. Mineta

MTI BOARD OF TRUSTEES

Directors