Peak-Hour Road Congestion Pricing: Experimental Evidence and Equilibrium Implications *

Gabriel E. Kreindler†

May 2, 2020

Abstract

Developing country megacities suffer from severe road traffic congestion, yet this is not directly informative about the magnitude of inefficiency in the current equilibrium. I study the peak-hour traffic congestion equilibrium in Bangalore. To measure travel preferences, I use a model of departure time choice to design a field experiment with congestion pricing policies and implement it using precise GPS driving behavior data. Commuter responses in the experiment reveal moderate schedule flexibility and a high value of time. I then show that in Bangalore, traffic volume has a moderate and linear impact on trip duration. Policy simulations with endogenous congestion reveal small travel time benefits and negligible commuter welfare gains from optimal congestion charges. This result is driven by the shape of the externality. Overall, these results suggest limited commuter welfare benefits from peak-spreading traffic policies in cities like Bangalore.

*I am deeply grateful to Ben Olken, Esther Duflo, Frank Schilbach, and Edward Glaeser for support throughout this project. I am indebted to Matt Turner for his discussion of the paper. I thank Alex Bartik, Nikhil Agarwal, Moshe Ben-Akiva, Peter Christiansen, Vikas Dimble, Benjamin Faber, John Firth, Andrew Foster, Chishio Furukawa, Nick Hagerty, Rachel Glennerster, Tetsuya Kaji, Myrto Kaloupsidi, Jing Li, Matt Lowe, Leslie Martin, Rachael Meager, Yuhei Miyauchi, Scott Nelson, Will Rafey, Otis Reid, Mahvish Shaukat, and Dan Waldinger for many helpful suggestions. Anupriya Khemka, Keerthana Jagadeesh, and Ashwin MB provided excellent research assistance. I also thank Mohannad Abunassar, Maryam Archie, Priya Chetri, Sasha Fleischman, Mahima Gupta, Aditi Sinha, Mamta Jat, Kristina Kelhofer, Michelle Nenciu, Sebastian Quinones, Sarvottam Salvi, Meghna Singh, Sahana Subramanyam, Tammy Tseng, Thuy Duong Vuong, Lantian Xiang, and Massieh Zare, who contributed valuable research assistance at various stages of the project. I gratefully acknowledge design and technical support for the smartphone app "Bangalore Traffic Research" from Adrian Drewett and Dharmendra Singh from Gridlocate Ltd. Funding for this project was generously provided by the Weiss Family Fund for Research in Development Economics, the IGC Cities Fund, the J-PAL Pilot Fund, and the J-PAL Urban Services Initiative Pilot Fund. This project has human subjects approval from MIT COUHES (protocol 1511312369A002), IFMR (IRB00007107), Harvard CUHS (IRB19-0456) and was registered in the AEA RCT Registry (AEARCTR-0002083).

Supplementary Materials available at `sites.google.com/site/gabrielkreindler/cp-sm`.

†Department of Economics, Harvard University. Email: `gkreindler@fas.harvard.edu`.

1 Introduction

Traffic congestion is a significant and pervasive cost in large cities, especially in developing countries, where urban population and private vehicle ownership are growing rapidly.[1] Long and unreliable travel times lower the effective agglomeration benefits of large cities, whether for accessing jobs, markets, services or amenities.

Commuters who drive in congested conditions impose externalities, notably by increasing travel times for the other commuters on the road. As congestion is highly concentrated in time and space in large cities, peak-hour traffic jams may be particularly inefficient. Reflecting this concern, several urban road traffic policies focus on reducing peak-hour congestion, either through pricing or quantity restrictions.[2]

However, high congestion *levels* do not automatically imply a significantly inefficient allocation of traffic. In cities in developing countries, the peak-hour road traffic congestion externality may be lower, for example due to road network properties or driving style. Alternatively, time in traffic may not be exceedingly costly, or relatively flexible schedules may allow commuters to spread traffic around to a sufficient extent.

Hence, while congestion pricing is the textbook policy response to traffic externalities, its quantitative relevance to traffic congestion problems in developing countries is an empirical question. In this paper, I study this problem in the context of the peak-hour traffic congestion equilibrium in Bangalore, India.

A quantitative analysis of optimal congestion charges and of the inefficiency in the current equilibrium requires two elements. First, estimating preferences requires precise data on urban travel behavior and plausible exogenous variation in the cost of traveling along the relevant dimensions. Second, the road traffic externality should be measured at the appropriate level at which agents make decisions and experience travel time.[3]

I set up an equilibrium model of peak-hour congestion, adapting the classic trip scheduling model (Vickrey, 1969; Small, 1982; Arnott et al., 1993; Noland and Small, 1995; Hall, 2018). In the model, commuters decide their trip departure time based on the distribution of travel times for different departure times. In an extension, they also choose between two routes that differ in travel duration.[4] The key preference parameters are the value of time spent driving, and the schedule costs of arriving

[1] Between 2005 and 2015, new private vehicle registrations have grown at 15% and 43% per year in India and China, compared to 0% in the United States and Europe (OICA, 2016).

[2] The congestion charge policy in Stockholm and Singapore's Electronic Road Pricing (ERP) policy have higher fees during the morning and evening peak hours. Jakarta's former "3-in-1" and the current "odd-even" policies are in effect during morning and evening peak hours only. Similarly, Manila's Unified Vehicular Volume Reduction Program (UVVRP) only applies during peak hours in certain parts of the city.

[3] A canonical result states that in a setting with a single measure of congestion, optimal congestion charges can be computed using only the externality and of the value of time. This relies on the assumption that average marginal value of time at the social optimum is the same (or can be derived from) the average marginal value of time at the current equilibrium. This assumption becomes significantly more restrictive with a multi-dimensional externality, as considered in this paper. In general, computing the deadweight loss depends on the demand elasticity parameter(s).

[4] The model makes explicit the substitution patterns between different times of the day. Other recent work estimating the deadweight loss of traffic congestion applies a static demand framework separately by time of day (Akbar and Duranton, 2017; Russo et al., 2019).

early or late, relative to an ideal arrival time that is unobserved to the researcher.[5] The profile of congestion is determined endogenously by aggregate departure rates.

I use the model to design a field experiment with congestion charge policies to estimate the preference parameters. I implement it within a sample of 497 commuters in Bangalore. Detailed travel behavior data was collected using a smartphone app that passively logged GPS location data from study participants. The two congestion pricing policies induce exogenous cost variation along the departure time and travel time dimensions. Under the "peak-hour" pricing policy, trips are charged according to a pay-per-km rate that depends on the trip departure time. The rate is piece-wise linear in departure time, so commuters sometimes have *marginal* incentives to change their departure times. Under the "route" or "area" pricing policy, commuters pay a flat fee for driving through a circular area located along their usual commute. The area is chosen individually for each commuter to induce a choice between a quick expensive route and a longer free detour route.[6] Intuitively, commuter departure time and route choices with and without each of these two pricing policies help identify the schedule cost parameters and the value of time, respectively.

Commuters respond to the two treatments by changing departure times and routes to avoid charges. Under "peak-hour" charges, in the morning before the peak-hour, commuters left around 4-6 minutes earlier on averge, with a smaller response after the peak-hour. In the evening, results are less precise and consistent with the reverse pattern (commuters leaving later after the evening peak-hour). In general, I find no impact on the number of trips. Under "route" charges, participants cross the congestion area around 20% less frequently, switching to longer routes that are not charged. Randomly varying the detour duration and the fee had imprecise and inconclusive results.

I next use the experimental price variation to structurally estimate the model of route and departure time choice for the morning commute. The main objective is to estimate the value of time and the schedule cost parameters. I estimate the model using two-step GMM with moments chosen to exploit the variation induced by the congestion charge experiments.

The estimated value of time spent driving is Rs. 1,085 per hour, and the estimated schedule costs of early and late arrival are Rs. 295 and Rs. 261, respectively. These results reveal two main points. First, the value of time is several times larger than the average self-reported hourly wage in this sample, indicating that commuters significantly dislike time spent in traffic in Bangalore. Estimates in the literature range between 50% and 100% of the wage. I discuss three potential reasons for this discrepancy. First, most of the literature relies on stated (hypothetical) preferences; results here are based on revealed preferences, which are typically larger (Small et al., 2005). Second, these results are relative to "Google Maps time," while commuters in this sample overestimate differences in travel times. Finally, these results load the entire choice between short and detour routes on the value of time. With a fixed cost of taking any detour route, the estimated value of time is smaller.

The second result is that commuters are moderately schedule flexible. The estimated costs of early and late arrival are roughly four times lower than the value of time, which indicates more

[5] In the baseline model, commuters are trip inelastic. I introduce an extensive margin decision in policy simulations.

[6] During the experiment, congestion charges were subtracted from a prepaid virtual account. The outstanding balance was transferred to study participant bank accounts periodically.

flexibility compared to the few previous estimates.[7] Schedule flexibility allows commuters to protect themselves from high peak-hour congestion by changing their departure times. However, the effect of flexibility on the magnitude of the peak-hour congestion inefficiency is ambiguous, because it affects both the equilibrium and the social optimum.

I next measure the road traffic externality. To measure traffic volumes, I use around 120,000 GPS trips collected using the smartphone app over six months in 2017. For speeds, results are similar with Google Maps data collected over time on a fixed sample of routes, and with speeds derived directly from the GPS data. To identify the causal impact of traffic volume on travel times, I use the large variation in traffic volume at different times of the day, including between peak- and off-peak-hours, similar to Akbar and Duranton (2017).

Citywide traffic volume has a moderate and linear impact on average travel times. Specifically, average normalized travel time for trips with a given departure time is approximately linear in the average volume of departures at that time. The linear relationship is robust for all calendar dates and when zooming in to major arteries. I find no evidence of convexity for high levels of traffic.[8] Quantitatively, a 33 minute peak-hour trip increases aggregate driving time for everyone else by approximately 16 minutes. I discuss additional robustness results and possible omitted variables.

I discuss two key reasons why these results differ from previous externality measurement. First, travel time and traffic volume are measured at the level of trips or trip segments, instead of at a single point in the road network. What may appear to be a local non-linearity may not be relevant at the level of entire trips, which is what commuters experience.[9] Second, I sketch two stylized models of traffic congestion in urban road networks. Unlike on (limited access) highways, stochastic obstructions such as traffic lights or cross traffic at junctions predict a positive slope relationship around zero traffic.

In the final part of the paper, I simulate an equilibrium model where commuters endowed with the preferences estimated from the experiment decide when to travel, and congestion is determined endogenously via the relationship described above. I compare the Nash equilibrium without tolls to the social optimum implemented with equilibrium optimal departure time "pigovian" congestion charges, assuming zero policy implementation costs and that revenue is redistributed lump-sum.[10]

Through the lens of the model, optimal departure time charges have a small effect on commuter welfare. Under the social optimum, the peak hour is lower and the average trip duration goes from 40.1 to 38.7 minutes on average, a 9% reduction in average travel time above free-flow. Commuter welfare – which also includes schedule costs – increases by Rs. 5 (7 US cents) or less than 1%. Low

[7] Small (1982) finds a ratio of 0.61. The moderate cost of late arrival in Bangalore is in contrast to the "value of urgency" from Bento et al. (2017).

[8] It is likely that at very high traffic volumes, travel time would become convex in traffic volume. However, this is outside the range of the current peak-hour traffic equilibrium.

[9] For example, Russo et al. (2019) estimate an exponential relationship between speed and density for (pointwise) loop detectors in Rome. This functional form predicts hypercongestion, and they find significant welfare losses from congestion.

[10] I continue to impose demanding simplifying assumptions. For example, I ignore driving externalities due to air pollution, I assume that firms do not change ideal arrival times in response to congestion pricing. I conduct a back of the envelope calculation of travel time benefits to bus passengers.

welfare gains are robust to varying commuter preferences and adding preference heterogeneity.[11]

These results depend strongly on the shape of the externality. Welfare gains would be significantly higher, and would depend more on preferences, with a more convex relationship between traffic volume and travel times. As the shape of the road traffic externality may be different in other cities, this result suggests that measuring city-wide road traffic externalities is of first-order importance.

The simulations so far ignore longer-term margins of adjustment. In the long term, commuter and firm location choices may be affected, which may have an ambiguous effect on welfare.[12] In principle, the departure time model can be embedded in a wider home- and workplace choice model with skill heterogeneity that incorporates general equilibrium effects, e.g. Tsivanidis (2019).

In the medium-run, commuters can cancel trips or switch to other modes, including public transport. To provide some insight into these issues, I run simulations where commuters can adjust along the extensive margin, a catch-all for using travel modes that have negligible externalities and for canceling trips. Welfare gains from optimal departure time congestion charges are increasing and approximately linear in the travel cost elasticity of the extensive margin trip decision. Gains range from 1.5% to 6.3% for elasticities of 0.10 to 1.4, respectively. These simulations show that commuters in Bangalore have to be very trip elastic, in order for peak-hour congestion pricing to lead to notable gains.

2 Traffic Congestion and Travel Behavior in Bangalore, India

Similar to other large cities in developing countries, Bangalore's fast-growing population and economy put stress on its transportation network. Akbar et al. (2018) rank Bangalore as the most highly congested city in India. In Bangalore, commuters essentially depend on the road in order to reach their destinations. Nearly all motorized transport, both private and public, travels on urban roads, so, to a first approximation, congestion affects all commuters.[13]

Figure 1 shows average travel delay in minutes per kilometer, collected from the Google Maps API on 30 routes (in both directions) in the study area in 2017. On average between 7 am and 10 pm on weekdays and across all routes, it takes 3.6 minutes to advance one kilometer. Travel delay is the inverse of speed, so this is equivalent to a speed of 10.3 miles per hour. This is extremely slow, but broadly in line with speeds in other heavily congested large cities in developing countries, such as downtown Jakarta, Indonesia (Hanna et al., 2017) and Delhi (Kreindler, 2016). Bangalore is much slower compared to cities in the U.S. For example, Anderson (2014) finds an average travel delay of 0.7 minutes per kilometer on urban highways in Los Angeles.

[11] Akbar and Duranton (2017) reach a similar conclusion of low deadweight loss due to traffic congestion for Bogotá, Colombia, using a representative household travel survey and Google Maps travel time data to estimate the demand for trips and supply of travel by time of day. The key contributions of this paper are to explicitly incorporate substitution between different times of the day in the equilibrium model, and using an experiment to estimate demand.

[12] Brinkman (2016) shows that "naive" optimal congestion pricing may lower agglomeration benefits by hampering connectivity within the city.

[13] The 2011 census reports that roads are used by 97% of all commuters – excluding those who do not travel, walk or use the bicycle. The main modal split is 33% using motorcycles, 15% cars, and 44% bus. Ridership on the Bangalore metro was below 100,000 per day in 2016, accounting for no more than 4% of all commuters.

Most of the day to day variation in traffic in Bangalore is explained by route by departure time cells. To see this, I compute the residual of the logarithm of delay, after removing fixed effects for each route and departure time combination, over 146 non-holiday weekdays. The standard deviation averages 0.11 over all routes and daytime hours, and reaches 0.14 around 9 am, indicating a moderate amount of travel time uncertainty. Residual delay is strongly auto-correlated in departure time on any given day and route, and 26% of the residual variation is explained by (date, route) fixed effects. Note that Google Maps may either underestimate or overestimate true travel time variation, due to prediction smoothing and prediction noise, respectively.

Figure 1 also shows strong predictable within day variation in traffic congestion. Between 7 am and 9 am, travel delay increases by 1.38 minutes per kilometer, or 57%. In other words, a trip that would take 40 minutes starting at 7 am would take more than an hour starting at 9 am. Similarly large changes in average travel delay occur around the evening peak.

This raises the possibility that commuters exert high externalities during the peak-hour, and a more efficient allocation would involve some of them leaving earlier before the peak, or later after the peak. Individual-level travel behavior data from GPS data (discussed in section 4.1) is consistent with high schedule flexibility. Most commuters vary their departure times significantly from day to day. For the median person, the standard deviation of the departure time for the first trip of the day from home to work is 29 minutes, which implies a 95% probability interval of almost two hours (Online Appendix Table A1, panel C). However, this daily variation does not automatically imply that commuters are schedule flexible. It is possible that desired travel times change from day to day (based on changes in work or other constraints) and commuters are inflexible around those times on any particular day. The model and experiment will help clarify and quantify these issues.

Separately, the peak-hour congestion profile is not directly informative about the magnitude of the traffic externality. Quantifying the externality requires measuring the underlying variation in traffic volume, which I do in section 7.

3 Theoretical Framework

In this section, I set up an equilibrium model of peak-hour traffic congestion. In the model, a commuter chooses their optimal departure time based on the profile of traffic congestion. The model builds on the canonical trip scheduling model (Vickrey, 1969; Arnott et al., 1993; Noland and Small, 1995; Hall, 2020) adapted to make easier to match to individual travel behavior data. On the supply side, the congestion profile is endogenous, determined by when commuters decide to travel. I discuss how the elasticities identified in partial-equilibrium pricing experiments relate to the key travel demand model parameters.

3.1 Travel Demand

An atomistic commuter i decides when to travel from home to work on day t, taking into account traffic conditions at different departure times. Departure time h takes discrete values, for example every minute. Travel time $T_{it}(h)$ is a random variable drawn from a distribution $\mathcal{T}_i(h)$. $T_{it}(h)$ is i.i.d. over days t. It may be correlated across individuals i and departure times h.

The decision to make the trip is inelastic. (I add an extensive margin decision in counterfactual simulations in section 8.) To begin, consider the case with a single route from home to work. Utility of leaving at departure time h and experiencing realized travel time T is quasi-linear in money:

$$u_i(h, T; h_{it}^A) = -\alpha T - \beta_E \left|h + T - h_{it}^A\right|_- - \beta_L \left|h + T - h_{it}^A\right|_+ - p(h) + \varepsilon_{it}(h) \tag{1}$$

Travel time cost is linear and α measures the marginal value of time spent driving. The second and third terms are the canonical way to measure scheduling preferences over arrival time $h + T$ (Arnott et al., 1993). Schedule heterogeneity is defined by an ideal arrival time h_{it}^A, and schedule costs by constant per-unit of time costs of arriving early β_E and of arriving late β_L.[14] $p(h)$ denotes monetary costs that depend on departure time, such as the congestion charges considered later.

The ideal arrival time h_{it}^A is drawn from an individual specific distribution $\mathcal{H}_i^A$, i.i.d. over days t. Ideal arrival times also take discrete values, for example every minute. This specification allows both for heterogeneity between commuters and within commuter, across days. This modeling choice is motivated by the high variance of within-commuter departure time documented in Online Appendix Table A1. The idiosyncratic shock $\varepsilon_{it}(h)$ affects the utility of leaving at h and is distributed according to a type-1 extreme value distribution with scale σ, which gives rise to logit choice probabilities conditional on h_{it}^A.

Timing is as follows. In the morning, before the earliest possible departure time, the commuter observes the ideal arrival time h_{it}^A and the idiosyncratic shocks $\varepsilon_{it}(h)$ for that day. The value of travel time $T_{it}(h)$ is realized after departure. The commuter chooses the departure time h_{it}^* to maximize expected utility over travel time realizations, given h_{it}^A and $(\varepsilon_{it}(h))_h$:

$$h_{it}^* \in \arg\max_h \mathrm{E}_{T_{it}(h) \sim \mathcal{T}_i(h)} \; u(h, T_{it}(h); h_{it}^A) \tag{2}$$

This model is designed to capture the main forces that determine departure time decisions within the peak-hour equilibrium, and it makes several specific parametric assumptions. It assumes linear travel time costs; more generally, this will capture a linear approximation of the value of time over the range of possible travel times. The schedule cost is also piece-wise linear, yet note that the uncertainty in travel time smooths out this profile. To further put the model in context, appendix A.1 presents a non-parametric model of departure time with time-varying flow values of time at home and at work. I show that the current model (without the idiosyncratic shocks ε and a fixed ideal arrival time) is the particular case when value of time at home is constant and the value of time at work is a step function around work start time. Finally, while otherwise straightforward to include, the current model does not include heterogeneity between commuters in the parameters α, β_E or β_L. I discuss this modeling choice when I structurally estimate the model. At that point, I will motivate and use a simple form of heterogeneity with two commuter types.

Model with route-choice. In one of the experimental treatments, participants will be charged for using their typical commute route, giving them to option to take (longer) detour routes to avoid

[14] The model also implicitly generates a value of travel time *reliability*; indeed, uncertainty in travel time may increase early and/or late arrival costs. $|x|_-$ and $|x|_+$ respectively denote the negative and positive parts of x, defined as non-negative numbers.

charges. To analyze these choices, which are informative for how commuters trade off time spent driving and money, I use a version of the model with a simple route choice decision. In the full equilibrium model, I focus on the departure time decision with a single route.

Assume that the commuter chooses between two routes in addition to departure time. Route $r = 0$ is the shorter, direct route from home to work, while the $r = 1$ (detour) route takes more time and is free. Travel time $T_{it}(h, r)$ on route r is drawn from a distribution $\mathcal{T}_i(h, r)$, i.i.d. over days t. I will assume that $ET_{it}(h, 1) > ET_{it}(h, 0)$ for any departure time. The utility of choosing departure time h, route r and experiencing travel time T is

$$u_i(h, r, T; h^A_{it}) = -\alpha T - \beta_E \left|h + T - h^A_{it}\right|_- - \beta_L \left|h + T - h^A_{it}\right|_+ - p(h, r) + \varepsilon_{it}(h, r) \qquad (3)$$

Monetary charges and the idiosyncratic shock now depend on both the departure time h and route r. I will assume that $\varepsilon_{it}(h, r)$ follows an extreme value distribution with correlation within each route, leading to a nested logit structure conditional on ideal arrival time. As before, the commuter chooses departure time and route to maximize expected utility over travel time realizations, given h^A_{it} and $(\varepsilon_{it}(h, r))_h$:

$$(h^*_{it}, r^*_{it}) \in \arg\max_{h,r} \mathrm{E}_{T_{it}(h,r)\sim\mathcal{T}_i(h,r)} u(h, r, T_{it}(h, r); h^A_{it}) \qquad (4)$$

3.2 Closing the model: road technology and equilibrium

The key technological constraint inherent to road travel is that higher traffic volume lowers speed. In this model, the peak-hour congestion profile depends on the profile of departures. City-wide, travel time for a given departure time h on day t is a random variable determined by the volume $Q_t(h)$ of departures at that time. There is a unit mass of atomistic commuters, and commuter i on day t faces a travel time profile proportional to the city-wide profile and to his typical trip length KM_i, namely travel time $T_{it}(h)$ is a random variable distributed according to $\mathcal{T}_{it}(h) = \mathrm{KM}_i \cdot \mathcal{T}(Q_t(h))$. Travel time may be correlated across commuters and departure times.

The relationship $\mathcal{T}$, or "road technology" for brevity, builds on the "Henderson" approach (Chu, 1995; Henderson, 1974). It can cover a wide range of magnitudes for the traffic congestion externality. For example, if expected travel time is convex and steep in travel volume, this can imply high peak-hour externalities.[15,16]

Equilibrium. In the equilibrium model with endogenous congestion, I focus on departure time as the only margin of choice, and abstract from physical space and multiple routes. Commuters choose optimal departure times based on (2) taking the travel time profile as given, and the travel

[15] For sufficiently convex relationship, this roughly corresponds to "hypercongestion," although the latter concept usually refers to the static relationship between outflow or speed and traffic *density* (Russo et al., 2019).

[16] The main theoretical limitation of the Henderson approach is that it may lead to "overtaking," whereby a commuter who leaves later than another arrives earlier. This will not happen if the volume of departures does not drop too suddenly relative to the slope of expected travel time as a function of volume, and if stochastic shocks are sufficiently correlated for nearby departure times. The main alternative model of traffic congestion is the bottleneck model with fixed capacity (Vickrey, 1969; Arnott et al., 1993), which has the advantage of being precise about the dynamics of travel delays over time. This model offers a poor fit to the citywide relationship between traffic volume and travel times in Bangalore (see section 7).

time profile is determined as above. Formally, the commuting environment on a given day t is a game of incomplete information with types $(h^A_{it}, (\varepsilon_{it}(h))_h)_i$. A Bayesian Nash equilibrium is fully determined by the equilibrium travel time profile $(\mathcal{T}(h))_h$. Commuters choose the optimal departure time, and the travel time profile must be consistent with aggregate choices.

In section 8, I compute the equilibrium in a simulated model using an asynchronous logit best-response dynamic, stopping when the profile of travel time converges. While a formal characterization of equilibrium uniqueness in this model is beyond the scope of the paper, in these numerical simulations, and for a range of starting conditions, I find a unique and stable equilibrium.

The second key object of interest is the social optimum implemented as a Nash equilibrium with "pigovian" departure time monetary charges $p(h)$ equal to the marginal social cost of a commuter traveling at h.

The goal of the empirical part of the paper will be to estimate the demand parameters and to calibrate the road technology. In a simulation model with these parameters, the key quantitative measures of interest are the equilibrium optimal congestion charge profile, and the magnitude of the inefficiency, namely the difference in commuter welfare (average utility) between the social optimum and the unpriced Nash equilibrium.

3.3 Pricing Experiments and the Key Demand Parameters

I now return to the key travel demand model parameters and discuss how they are related to the elasticities identified in partial-equilibrium congestion pricing experiments. I discuss additional advantages of pricing experiments, and potential challenges for model identification with observational data alone. Throughout, I assume that travel time distributions are observed by the researcher, while commuters' ideal arrival time distributions are not observed.

The goal of demand estimation in this setting is to separately identify the value of time spent commuting (α), schedule costs (parameters β_E and β_L), and schedule heterogeneity (ideal arrival times distributions $\mathcal{H}^A_i$). Separating these parameters is important for understanding the equilibrium properties of the model. Schedule costs determine the demand for peak-hour travel. The value of time also affects how commuters substitute between peak and non-peak travel, yet this depends on the slope of travel time as a function of departure time, an equilibrium object that will change in counterfactuals. Moreover, the value of time will quantify the value of lower overall congestion.[17]

Intuitively, there are two major model identification challenges: distinguishing schedule costs (β_E and β_L) from schedule heterogeneity ($\mathcal{H}^A_i$), and distinguishing value of time (α) from schedule costs. Both schedule costs and schedule heterogeneity affect the distribution of departure time choices. For example, recall that departure times vary significantly across- commuters and within a commuter, across days. This may be because commuters are roughly indifferent between departure times (i.e. low schedule costs β_E and β_L) or because ideal arrival times h^A_{it} vary across i and over t, while, in fact, schedule costs around ideal arrival times are high. The individual specific distribution of ideal

[17] I assume throughout that schedule heterogeneity does not change due to congestion pricing policies. This assumption would fail if, for example, firms changed work hours in response to peak-hour congestion policies. By allowing $\mathcal{H}^A_i$ to change, one could investigate the impact of "staggered hours" policies on commuter welfare.

arrival times $\mathcal{H}_i^A$ provides a significant amount of freedom to fit panel data on departure times.[18]

Consider a departure time monetary charge $p^D(h)$, which induces price variation in departure time. The first experiment in this project will implement a congestion pricing scheme that depends on departure time, according to a profile that broadly follows the peak-hour (see Online Appendix Figure A2). The induced pricing variation between different departure times helps estimate scheduling preferences. The key insight is that schedule heterogeneity and schedule costs have different implications for how commuters change behavior in response to this type of pricing. Intuitively, commuters are more elastic when schedule costs are lower.[19]

Now consider the problem of separating schedule costs from value of time. To describe the problem clearly, assume for a moment that there is no schedule heterogeneity, namely there is a unique ideal arrival time h^A. The expected utility of a departure time (excluding idiosyncratic shocks) is a combination of schedule costs and value of time, given the distribution of travel time at that departure time. Even if we know the utility of each departure time in monetary terms, we still need to separate these two components. For example, leaving earlier leads to higher schedule costs as well as shorter expected travel time.

An ideal experiment would vary travel time independently of departure time. To approximate this, the second experiment in this project will implement a route pricing scheme. Assume that the shortest route ($r = 0$) to the commuter's destination is charged, and that it can be avoided by using a longer detour route ($r = 1$). Route pricing identifies the price elasticity between the two routes, which depends on the value of time. In equation (3), the route-based congestion charge is modeled as a flat fee $p(h, r) = p^R \mathbb{1}(r = 0)$.[20]

I will use data from the two randomized experiments with departure time and route pricing to estimate the discrete choice model given by (4). The two experiments create price variation for departure time and travel time, hence intuitively commuter responses in the two cases are informative about scheduling costs and value of time, respectively. However, in the model, responses to each treatment depend on the entire vector of preferences, so the model parameters will be jointly estimated. A formal proof of model identification is beyond the scope of this paper. Instead, in section 6.1 I show numerically that estimated parameters correlate strongly with true parameters when I apply the estimation procedure to simulated data. To shed light on how the moments of the data affect estimated parameters, I report the sensitivity measure from Andrews et al. (2017).

---

[18] Characterizing formally under what conditions the model is identified is beyond the scope of this paper. Intuitively, the relative sizes of the finite grids for departure time and ideal arrival times matter, while the parametric nested logit specification helps identification.

[19] In the absence of pricing variation, previous work estimates or calibrates schedule preferences using assumptions or complementary data on the distribution of ideal arrival times (Small, 1982; Hall, 2020). For example, Hall (2020) splits commuters into "flexible" and "inflexible" categories based on survey responses, infers the arrival time distribution for the latter category based on realized arrival times, and assumes that "flexible" commuters have the same distribution of ideal arrival times.

[20] The choice between routes may also depend on factors other than the travel time difference between the two routes. For example, commuters may have a fixed cost γ_1 of taking the detour route $r = 1$, perhaps due to habit, lack of information on alternate routes, or a permanent route-specific preference shock for the short route. To separate any such fixed component from the value of time, the experiment included sub-treatments with random variation in the flat fee p^R and in the average duration difference $ET_{it}(h, 1) - ET_{it}(h, 0)$. As the empirical results from this experiment were imprecise, I do not use this variation when structurally estimating the model. I will explore how the results change for different assumed fixed costs.

By design, the experiment identifies preferences for travel holding the citywide travel time profile fixed. Equilibrium responses to congestion pricing will be different due to changes in this profile. The simulations in section 8 explicitly analyze this feedback. However, medium- and long-run travel elasticities may differ from those estimated here for other reasons. For example, firms may adjust by providing more flexible schedules, allowing commuters to more easily change their departure times. Measuring these elasticities and their full welfare implications are beyond the scope of this paper.[21]

The approach in this paper to design the experiment based on the model brings together two previous literatures. First, some papers use discrete choice models to estimate the value of time, of reliability, or of urgency from real-world driver decisions to use a faster tolled lane (Small et al., 2005; Bento et al., 2017). A separate group of papers analyzes reduced form impacts of road pricing experiments (Tillema et al., 2013; Martin and Thornton, 2017). Here, the randomized experiment is designed to allow to transparently recover the key commuter preference parameters in the model, value of time and scheduling preferences.[22]

4 Data Sources and Study Sample

The data backbone of the project is a data set of trips with precise GPS coordinates, collected using a newly developed smartphone app. This data was used both for measuring detailed travel behavior and for implementing the congestion charge policies in the experiment. This section describes how the app was designed, how the data was collected, and how it was automatically cleaned and classified. I also briefly describe several other data sources. The section ends with a description of the study participant recruitment procedure.

4.1 GPS Trip-level Data from Smartphone App

GPS traces. Travel behavior data was collected using a smartphone app that works in the background of a GPS-enabled Android smartphone and passively collects phone location data, without requiring any user input. To conserve battery power, updates were collected at variable time intervals, between every 30 seconds while traveling and every 6 minutes, when in stationary mode.[23] The phone location is identified by the phone operating system using GPS information, as well as cell phone network and WiFi information. (Henceforth, I will refer to this data simply as *GPS data.*)

[21]Note that some of these adjustments may already be happening in the current equilibrium. As in other large cities marred by high congestion, some large companies in Bangalore are already implementing this type of flexible work-hours policies (Merugu et al., 2009). Moreover, around 20% of the sample in this study are self-employed and may already have higher autonomy in deciding their own schedules. Finally, it is worth noting that the full welfare impacts of firm-level work-hours changes is ambiguous, due to complementarities between firms of having similar work hours (Henderson, 1981).

[22]A related empirical literature documents the impact of real-world traffic policies on volumes, travel times and air pollution, either for the aggregate impact of congestion pricing policies in London, Stockholm and Milan (TfL, 2006; Prud'homme and Bocarejo, 2005; Raux, 2005; Gibson and Carnovale, 2015; Karlström and Franklin, 2009), or for non-price, vehicle quantity restrictions in developing countries (Davis, 2008; Kreindler, 2016; Hanna et al., 2017; Gu et al., 2017).

[23]The app, called "Bangalore Traffic Research," was available from Play Store during the study period,. I worked together with GridLocate Ltd, a GPS tracking solutions company, to adapt one of their products to the specific needs of this project.

The app uploads data to a server at regular intervals using the phone's data or WiFi connection. The app has a simple interface that shows a map with the user's current location, and users could receive notifications in the phone notification panel.

Trip Data Processing. The raw GPS data for each user-day was automatically cleaned and classified into *trips* and *locations*, as well as segments corresponding to missing data. Consecutive trips with short stops of at most 15 minutes between them are linked together. 89% of trips in the sample have no stops (of more than 5 minutes) along the way. The algorithm tags trips outside Bangalore, defined as more than 18km away from the city center. This algorithm was used during the experiment to compute congestion charges for participants in the treatment groups.[24]

Measuring travel behavior using a smartphone-based app has several advantages over surveys. Self-reported behavior is affected by recall bias, rounding of departure times and trip duration, and tends to underestimate within-person temporal and route variation (Zhao et al., 2015). The study app solved these issues by collecting the relevant information completely automatically, without any user input at the beginning or end of a trip, and without requiring participants to later review and validate their trips. Using a smartphone as sensing device also improves over previous studies that required participants to carry a separate GPS device.[25]

The app sometimes did not collect GPS data, for either technical or human factors, if the phone or the location services are switched off, if app permissions are revoked, or if the phone is unable to determine its own location. I classify daily data into three quality categories based on the total duration without location data, and the total distance traveled without precise route information: good quality data, insufficient data, and no data. During the experiment, around 75% of days are good quality, which is the category used for analysis. I also define a measure of trip data quality, where good quality trips have precise departure and arrival times, are neither too short nor too long in terms of duration and length, do not have "jumps" without data, are not short loops, and are not very "swiggly." During the experiment, 64% of all non-weekend, non-holiday, daytime trips are good quality and thus included in the sample.[26] More details on the trip processing and day and trip data quality definitions are in Supplementary Material section SM.1.

Common Destinations and Regular Commuters. Travel behavior and preferences may differ on regular and non-regular trips. I identify home and common, recurring daytime destinations at the commuter level (such as a work or school) using a clustering algorithm to group locations into groups, followed by manual review of the location groups most frequently visited. Next, I compute the fraction of distance traveled between "home" and "work," as well as the fraction of days present at work. Using these two variables, I classify participants into regular and variable commuters. Around 75% of study participants have a regular destination, and the median regular commuter visits work on 91% of weekdays (Online Appendix Table A1 Panel B).

[24]The travel mode, including car, carpooling, motorcycle, walking or public transport, is not identified by the algorithm, yet short walking round trips are automatically excluded. The experiment did not incentivize mode change directly.

[25]In a phone survey performed after the experiment ended, only 2.5% of respondents said they left their phone at home "sometimes, for usual destinations."

[26]The largest category of low quality trips are due to imprecise departure time and short duration.

Google Maps data. I collected two types of Google Maps data on travel times that include information on traffic congestion. The first data set collected *real-time* travel time on 89 routes across Bangalore (in both directions), including 30 routes in the study area of South Bangalore, every 20 minutes throughout the day, for 207 days in 2017 (See Online Appendix Figure A1).[27] This data will be used to calibrate the distribution of travel times, holding route and departure time fixed, and in order to measure the road technology impact of traffic volume on speeds. The second data set is individual-level data on *typical* travel times between their home and work locations, at all departure times during the day. This data will be used to understand the choice set faced by individual commuters.

4.2 Study Sample and Survey Data

Study participants were recruited in a random sample of gas stations in South Bangalore.[28] Surveyors approached private vehicle drivers (commuters) who were using a car, motorcycle or scooter, excluding taxis and professional drivers, and invited them to participate in a study about understanding traffic congestion in Bangalore. After an eligibility filter,[29] the surveyor explained in broad terms the study purpose, mentioning monetary rewards for participation and the possibility to receive monetary incentives tied to changes in travel behavior. Respondents were invited to install the study smartphone app and answer a short survey. All respondents received a study kit including a branded study flyer and consent form. Some data was collected for everyone (including commuters who refused the survey): perceived age category, gender, and vehicle information.

Out of 16,911 persons approached, 43% refused to be interviewed. A further 28% were ineligible. 2,299 accepted to install the app, an estimated 27% of all eligible respondents.[30]

In the weeks after recruitment, the study collected travel data from the participant smartphone app. The study team monitored quality and contacted respondents in case data quality problems arose. Participants were also offered an incentive worth Rs. 300 in phone recharge for providing one week of quality data. Additional survey data collection related to the experiment is described in section 5.

---

[27] These queries return travel time mostly for exactly the same route. The route length in meters is identical to the daily median for that route for 97.1% of all observations, and for 95.6% of all morning peak-hour observations. The route length is 2.8 km on average, with 1.9 km and 3.4 km as 25th and 75th percentiles.

[28] Gas stations are ideal locations to meet commuters who regularly use their private vehicle. (During piloting our team attempted household visits, which suffered from a very low probability of finding respondents at home.) In gas stations, surveyors worked Monday–Saturday in one of two shifts, 8 am – 1 pm or 3 pm – 8 pm.

[29] A respondent was eligible if they reported being the owner or regular user of the private vehicle used on that day, traveling at least 20 Km in total per day, at least three days per week, owning a smartphone and not planing to leave Bangalore for more than two weeks over the following two months. Smartphone usage is very high: 76% of participants eligible based on other conditions owned a GPS Android smartphone (an additional 12% owned an iphone and were not included). Commuters with a personal driver were included if they were the ultimate decision-maker for commuting decisions. Professional drivers (delivery person, Uber/Ola/taxi driver, etc.) were not eligible.

[30] This percentage is calculated assuming the same fraction of ineligibles between those who answered the initial filter and those who refused.

4.3 Experimental Sample and Congestion Charge Platform

The experimental sample was selected based on app data quality and a second eligibility check.[31] All participants (including the control group) met in person with a surveyor at a location convenient for the respondent. After the meeting was scheduled and before it took place, participants were randomized into treatments. During the meeting, surveyors explained the treatment and (if applicable) how congestion charges function. Participants also received support materials such as a laminated rate card with information about congestion charges (see Online Appendix Figure A2). Overall, 497 or 22% of all app participants were enrolled in the experiment on a rolling basis.

During the experiment, congestion charges were deducted from a pre-paid virtual account. The outstanding balance at the end of each week during the experiment was transferred to the participant's bank account.[32] Participants were also charged a fee for severely incomplete GPS data, and if they did not make any trips on a given weekday. The "no trip" fee was designed to dissuade incentive gaming by leaving the smartphone at home for the entire day.[33] To establish trust, participants received a welcome bank transfer soon after the first meeting, and/or an external smartphone battery (power bank) as a gift during the meeting. A study call center was available if study participants had questions or complaints.

During the initial meeting, surveyors framed the account balance as the "respondent's money," and congestion charge as losses. During the experiment, charges were computed automatically and participants received daily account balance updates through SMS and app notifications. In addition, weekly phone calls reminded participants about their treatment group details. These features were designed to ensure salience of the congestion charges, which affects demand elasticities (Finkelstein, 2009).

Experimenter demand effects are an important concern in this setting. Commuters in Bangalore generally care deeply about traffic congestion, and study participants may be motivated to avoid congested times or areas by a sense of civic duty. While these responses may in principle be real, it is also possible that they are specific to this (short-term) experiment, where their participation was voluntary and compensated. I took several steps to guard the experimental results against this possibility. First, during the meeting surveyors were trained to present the options in a neutral light, and to emphasize at least twice that the study does not have a preference over whether participants change or do not change their behavior. Second, one of the treatment includes an "information/nudge" sub-treatment, where participants received flat payments and SMS and app notifications similar to those in the pricing sub-treatments. Finally, both treatments had sub-treatments with price variation, in principle allowing to estimate price responsiveness holding fixed the other features.

[31] Commuters with less than 5km of travel per day, and those who actually lived or spent considerable amount of time outside Bangalore, were dropped.

[32] During part of the experiment, logistical issues led to some delays in bank transfers for some participants.

[33] A maximum daily total charge and minimum account balance of Rs. 250 also applied. Account opening balances were chosen separately for each participant, based on a model that predicted expected charges given baseline travel behavior and a hypothesized responsiveness to treatment. The target final account balance was randomized to either Rs. 500 or Rs. 1,000 per week.

5 Experimental Design: Congestion Charge Policies

Departure Time Pay-per-Km Congestion Charge. Participants in this treatment were charged for each trip based on a per-km rate and the length of their trip. The rate was positive during a 3-hour interval during the morning and a 3-hour interval during the evening. Each charged interval had the same structure: a one hour increasing "shoulder" ramp when the rate grew linearly from zero to the peak rate, one hour of peak rate, and a one hour decreasing "shoulder" ramp when the rate fell linearly to 0. Online Appendix Figure A2 shows an example rate card (given to study participants) for the morning interval.[34]

Four sub-treatments were designed to separate the impact of prices from other features of the intervention: control, information, low rate, and high rate. Participants in the control group were monitored for 5 weeks, received regular updates about their data quality, and received a flat Rs. 300 payment per week for participation. Participants in the information group received daily messages about the trips they had completed the previous day, and information about how to reduce trip travel time by changing departure time (away from the morning and evening peak hours). They also participated for 5 weeks. The low and high rate groups had a maximum (peak) congestion rate of Rs. 12/Km and Rs. 24/Km, respectively. They received daily messages the trips they had completed the previous day, how much each was charged, and information about how to reduce trip charges by changing departure time (away from the morning and evening peak hours).

Participants received this treatment for three consecutive weeks out of four in total, either the first three or the last three, randomly chosen. Before the start of the congestion charge phase, participants underwent a three-day *trial phase* where they received congestion charge messages to understand how charging works.

Area or Route Congestion Charge. Participants in this treatment were charged for driving through a congestion area. The area was a disc with radius 250m, 500m or 1000m positioned along a route used frequently by the participant during the pre- period. The area induced an alternate non-intersecting detour route, which was between 3 and 14 minutes longer than the original route, based on Google Maps travel time data. Study participants never had a stable destination inside the congestion area. This procedure was possible for 254 out of the 497 experiment participants; the remaining 243 were not included in the area treatment.[35] During the meeting, surveyors carefully explained the area location, radius, boundaries, and one possible induced detour during the meeting before the experiment. This information was repeated in each daily reminder SMS. Surveyors explained that only routes that intersect the area will be charged, and that several detour routes may exist.[36]

The area treatment did not include a pure control group. Participants were randomized between being treated early (in the first week after the meeting, before the departure time treatment) or

[34] The start time of the charged interval differed by at most ±30 minutes between commuters, and was designed to maximize the overlap between the shoulder periods and typical departure times for that commuter based on baseline data. This procedure was implemented for *all* commuters before randomizing them to treatment groups.

[35] After the experiment, eleven area treatment participants were dropped because the chosen area overlapped with one of their baseline stable destinations.

[36] If asked, surveyors explained that the area location did not specifically target congested areas.

late (in the last week of the study, after the departure time treatment), which is the basis for the experimental comparison. The charge was in effect between 7 am and 9 pm, and applied at most once for the morning interval (7 am – 2 pm) and at most once for the evening interval (2 pm – 9 pm).

Sub-treatments were designed to identify the effect of price and detour time variation on choices. Low/high rate participants had a baseline charge of Rs. 80/Rs. 160. To create random variation in the induced detour, for each participant I selected multiple candidate congestion areas and computed the detour for each. Long (short) detour areas induced a predicted detour between 7 and 14 minutes (between 3 and 7 minutes) above the usual route. Participants were randomized into short/long detour groups, and the area to be implemented was randomly chosen from that group, if possible.[37] In an independent randomization, on two randomly chosen days, the congestion charge was 50% higher.

Experimental Design and Randomization. Online Appendix Table A2 explains the experimental strata, treatment probabilities, and treatment timing. The departure time treatment had four sub-treatments described above, with unequal probabilities given in panel A. The area treatment timing, the low/high rate, and the long/short detour sub-treatments were cross-randomized, giving 8 equal-probability treatment cells. The area timing randomization dictated the timing of the departure time treatment (panel B). The (four) departure time sub-treatments and the (eight) area sub-treatments were cross-randomized within each stratum.

Note that the area and departure time treatments never occurred at the same time.

Participants were added to the experiment on a rolling basis, and the allocation to treatments was pre-randomized for each stratum. Supplementary Materials section SM.4 explains the full procedure.

Experiment-Related Survey Data. A "stated preferences" phone survey was conducted with a random sub-sample of those who installed the app, and all respondents who eventually participated in the experiment. The survey covered questions about typical trip departure times, beliefs about travel times, and hypothetical choice questions that mapped to the two experiments described in the next sections. Responses to hypothetical questions were coded into a "stated preferences" value of time measure, and a stated preferences schedule flexibility measure. See Supplementary Materials section SM.3.1 for details.

During the experiment, after the in-person meeting, respondents were surveyed weekly by phone. In the treatment group, this survey included questions to check if respondents remembered the details of the treatment assigned to them; surveyors then reminded respondents these details.

5.1 Reduced-Form Responses to Congestion Charges

Sample Description and Experimental Integrity Checks. Appendix Table A3 describes the representativeness of the experimental sample. The 497 experimental participants represent 6% of all eligible commuters approached by surveyors in gas stations. Almost all commuters in this setting are men. Experiment participants appear around 2 years younger on average and are less likely to use a

[37] If a participant assigned to the long detour group only had a viable area with a short detour, that area was implemented.

car compared to a motorcycle or scooter (41% car users overall compared to 30% in the experiment). Conditional on vehicle type, there is no difference in vehicle value. Among survey respondents, experimental participants have similar self-reported income, they report traveling slightly more, and have slightly higher stated value of time and stated schedule flexibility.

Online Appendix Table A1 describes baseline travel behavior using GPS data for the experimental sample. Commuters make around 3 trips per day on aveargе, each around 25 minutes and 6 km long. Three quarters of the sample is regular commuters. Most commuters in this sample have variable departure times.

Online Appendix Table A4 reports the experimental balance check. The different treatment groups are similar along demographic and pre-period travel behavior variables. All coefficients are small, and joint significance tests cannot reject the null of no effect.

The Impact of Departure Time Pay-per-Km Charges. In principle, commuters may respond to charges by canceling trips with departure times during the charged period (the extensive margin), as well as by rescheduling trips to departure times with lower charges (intensive margin).

Figure 2 shows a first look at the causal impact of congestion charges on the distribution of trip departure times. To construct it, I consider four groups, each combination of before/during the experiment and the control/treatment group, pooling together the control and information sub-treatments. Within each group, I compute the kernel density of trip departure times (relative to the midpoint of the congestion charge for each commuter) and multiply it by the average number of trips per day in that group. I then plot the difference-in-difference of these four curves, as well as 200 commuter-level bootstraps. The Y axis measures the change in the number of trips per day in a one hour departure time window. This exercise does not take into account randomization strata.

There are three main results for the morning interval. First, there is no reduction in the total number of trips, as the integral of the curve is non-negative. Second, in the early morning, there is a strong shift in trips towards earlier departure times, which coincides with the linearly increasing "ramp" of the congestion charges. In this interval, there is a marginal incentive to advance the departure time. The results suggest that study participants understood this feature and decided to leave earlier and take advantage of lower charges. This shift in the mass of trips corresponds to commuters in this interval leaving around 4 minutes earlier on average due to the treatment.[38] Third, there is an increase around half an hour after the end of the congestion charge. The exact position of this increase does not map cleanly to the predicted response given marginal incentives, as in the case of the early morning. This increase is entirely concentrated among non-commuting trips; Online Appendix Figure A4 shows that this effect disappears for commuting trips of regular commuters. Instead, a slight substitution to later times during the decreasing "ramp" can be seen. This result suggests the possibility of discrete scheduling changes for some non-commuting trips.

In the evening, the figure shows a decrease in the number of trips in the later part of the congestion charge ramp. Unlike in the early morning, there is no evidence of an increase in the mass

[38] Supplementary Materials Table SM3 reports treatment impacts on departure time in specifications with individual fixed effects, and shows a significant 3.8 minutes earlier departure time for the early morning ramp. In specifications without commuter fixed effects (not shown), this effect is 6 minutes.

of trips during the latest part of the ramp. Overall, these results are consistent with a combination of extensive and intensive margin responses.

I run the following difference-in-difference specification:

$$y_{it} = \gamma^I T_i^I \times Post_t + \gamma^L T_i^L \times Post_t + \gamma^H T_i^H \times Post_t + \mu_t + \alpha_i + \varepsilon_{it}, \tag{5}$$

where y_{it} is an outcome of interest for commuter i on day t, $Post_t$ is a dummy for during the experiment, T_i^I, T_i^L and T_i^H are dummies for the information, low rate and high rate departure time sub-treatments, α_i is a commuter fixed effect, μ_t is a study cycle fixed effect whose categories are the period before the experiment, and each week in the experiment. The coefficients of interest, γ^I, γ^L and γ^H, respectively measure the impact of information, low rates and high rates relative to control, during the experiment relative to the period before.

For higher precision, I pool sub-treatments and run:

$$y_{it} = \gamma^{DT} T_i^{DT} \times Post_t + \mu_t + \alpha_i + \varepsilon_{it}, \tag{6}$$

where T_i^{DT} is a dummy of either low rate or high rate departure time sub-treatments.

The sample is all non-holiday weekdays when the respondent does not travel outside Bangalore. During the experiment, I include the three weeks when charges are in effect; in the control and information groups I also keep three weeks to make the timing comparable. Where necessary, the sample is restricted to days with "good quality" GPS data, as defined in Supplementary Materials section SM.2. Standard errors are clustered at the commuter level.

I first show that the departure time congestion charge treatments did not affect data quality (Online Appendix Table A5). Given that smartphone app data was used to implement the congestion charges, it is especially important to check that the fraction of days with quality GPS data was similar across treatments. During the experiment, participants provided good quality GPS data on approximately 75% of weekdays. Departure time sub-treatments do not have detectable differences on daily GPS data quality.[39]

Table 1 shows results on daily outcomes. Panels A and B show impacts from specifications (5) and (6). The first three columns show impacts on total daily "trip shadow rate." This outcome is the sum over all trips in the day of the commuter's own congestion rate profile (defined for all participants irrespective of treatment group) and using a normalized peak rate of 100. This outcome is computed uniformly across commuters and across time, and it is a summary statistic for whether the commuter's travel behavior is associated with high charges. It combines both intensive and extensive margin responses. In the last three columns, I focus exclusively on the extensive margin and use the total number of trips as outcome.

The high rate sub-treatment leads to a decrease of around 12 from a base of 96 total daily shadow rates in the control group, or a 13% decrease. The low rate sub-treatment coefficient is also negative but smaller in magnitude and not significant. When disaggregating by morning and evening in the second and third columns, these results remain similar but are less precise. The information group

[39] The experiment was generally successful in terms of *retaining* study participants: approximately 5% of participants dropped out right after the meeting, and this figure rose to 10% on the last day of the study. Drop outs are 2 percentage point more frequent in the treatment group, yet this difference is not statistically significant (p-value 0.20).

does not seem to have any effect on total daily shadow rates, with a point estimate very close to zero. This suggests that the daily SMS and app reminders did not, by themselves, affect departure time travel behavior as summarized by shadow rates. In panel B and for most of the following analysis, I pool together the control and information groups.

The last three columns show that departure time charges had small negative effects on the total number of trips, which most of the time are far from statistical significance. The point estimate on high rate in the fourth column implies a 2% decrease in the number of trips, and the point estimate on Charges in the same column implies a 4% decrease. However, there is a marginally significant decrease in the evening of 7%. This result mirrors the findings from Figure 2.[40]

Table 2 unpacks the effects on total daily shadow rates by finer time intervals, and for regular commuting trips. The reduction in charges is stronger early relative to late in the morning, and stronger late relative to early in the evening. At face value, these results are consistent with work hours acting as firm constraints for regular commuters and less asymmetric schedule costs for variable commuters.Supplementary Materials Table SM2 shows results for variable commuters.

In Online Appendix Figure A5 panel A, I investigate the heterogeneity in individual responses to the departure time treatments. The figure shows kernel density plots of the *change* in shadow rates (pre/post) at the individual level, separately in the control and treatment groups. Congestion charge treatment respondents have a nearly bi-modal distribution. This suggests that only a subset of commuters changed their behavior to take advantage of lower charges. This heterogeneity is difficult to assign to observable characteristics (Supplementary Materials Table SM5). For most variables I do not find evidence of differential response (e.g. car vs motorcycle users). The self-employed respond more strongly, while expensive vehicle owners (a category that includes all car drivers and expensive motorcycles) seem to respond more, although they also have a statistically significant, smaller, imapct on fewer trips (panel B). (Results not adjusted for multiple hypothesis testing.)

The Impact of Area or Route Charges. The area congestion charge induces a choice between a shorter, costlier route, and other longer but free routes. In this treatment, there is no pure control group. Instead, the empirical strategy is based on comparing commuters randomly assigned to be treated early or late. Specifically, in the first week I compare commuters treated early (treated group) to those treated late (control group). In the fourth week, these roles are reversed. The period before the experiment and the second and third week during the experiment are included to gain precision when estimating commuter fixed effects. Define $T_{it}^A = (1 - T_i^{Late})W_t^1 + T_i^{Late}W_t^4$ where T_i^{Late} is an indicator for being treated late, and W_t^s is an indicator for week s. I run the following specification:

$$y_{it} = \gamma^A \cdot T_{it}^A + \mu_t + \alpha_i + \varepsilon_{it} \tag{7}$$

The coefficient of interest is γ^A, which measures how the outcome y_{it} differs as a result of being exposed to area congestion charges, relative to similar commuters who are not treated that week.

I first check that the area congestion charge treatment did not reduce the number of days with

[40]The results are broadly similar when using total daily shadow *charges* (equal to the shadow rate multiplied by the trip length) or when using trips as observations, with stronger effects in the morning than in the evening.

high data quality (Online Appendix Table A5). Treated area respondents are in fact slightly *more* likely to provide good data, and this effect appears slightly larger in the fourth week, possibly because the control group was more likely to drop out of the study. This suggests that missing GPS data was mostly due to factors unrelated to the incentive to reduce charges by selectively disabling location tracking or by leaving the phone at home.

Panel A of Table 3 reports the impact on daily total "shadow" charges due to crossings of the congestion area. Shadow charges for a trip crossing is normalized to 100. The results show a large, precisely estimated decrease in the probability to cross the congestion area on a given day. The decrease is around 23% of the control mean, significant at the 1% level. The impact is slightly larger in the morning, and negative both for participants treated in the first and those treated in the last week (panel B). The last three columns show the impact on daily trips. Being treated results in more measured trips per day, with the effect concentrated in the morning and for participants treated in the last week. Overall, there is no evidence that commuters reduce the number of trips to avoid congestion charges.

Some of the respondents in the area control group in fact receive a departure time treatment. In Online Appendix Table A6 I restrict the entire analysis to 114 respondents in the control or information departure time sub-treatments. Results are similar and the impact is somewhat larger. The control mean is nearly identical to that Table 3, which supports simulation results showing that departure time charges have a negligible impact on route choice. These results are consistent with mental load: participants randomly selected to receive both the area and departure time treatments (at different times) responded less to the area treatment, compared to those who only received the area treatment. For consistency with the initial experimental design, I conduct the rest of the analysis on the pooled sample.[41]

I find imprecise effects from randomly varying the area charge amount and the detour duration. Appendix Table A7 shows that it is not possible to reject neither equal nor proportional effects, for both sub-treatments. In the case of detour, on a sub-sample with beliefs data, there is no evidence that the "long" detour sub-treatment affected beliefs of the detour duration differentially.

As for the departure time treatment, individual level response heterogeneity suggests that participants either responded strongly to the treatment or did not respond at all. In Online Appendix Figure A5 (panel B), for each area participant, I count the fraction of days crossing the congestion area, separately when treated and when in the control group. The distribution in the control group is concentrated near 1, as most commuters select the shortest route in the absence of charges (solid, gray bars). In the presence of charges, the distribution becomes bi-modal, with around 20 per cent of the population in the lowest bin, implying that some participants stopped crossing the congestion area at all (outline, red bars). It is difficult to assign this heterogeneity to observable sources (Supplementary Materials Table SM6).

[41] Supplementary Materials Table SM4 runs equation (7) at the trip level and shows similar results. The experimental impact on trip duration is imprecise. In most specifications it is not possible to reject 6.4 minutes, the average additional duration for trips that avoid the congestion area, as predicted from Google Maps data.

6 Structural Travel Demand Estimation

I now use the data and the experiment to estimate the key parameters in the travel demand model. This will provide monetary measures of individual preferences over schedule inflexibility and marginal changes in time spent driving.

Discrete Choice Model over Routes and Departure Times. I will estimate the discrete choice model over departure time h and route $r \in \{0, 1\}$ described in section 3. To recap equation (4), utility depends on travel time costs, schedule costs, monetary costs and idiosyncratic factors:

$$
\begin{aligned}
\mathrm{E}u_i(h, r; h^A_{it}) = & -\alpha \mathrm{E}T_{it}(h, r) \\
& - \beta_E \mathrm{E}\left|h + T_{it}(h, r) - h^A_{it}\right|_- - \beta_L \mathrm{E}\left|h + T_{it}(h, r) - h^A_{it}\right|_+ \\
& - p(h, r) + \varepsilon_{it}(h, r)
\end{aligned} \tag{8}
$$

where expectations are over travel time $T_{it}(h, r)$, distributed according to $\mathcal{T}_i(h, r)$. Monetary charges $p(h, r)$ will be zero, $p^D_{it}(h)$ or $p^R_{it}(r)$ for commuters respectively in the control group, in the departure time treatment, and in the area treatment. As before, h^A_{it} is the ideal arrival time, drawn from $\mathcal{H}^A_i$ i.i.d. over days, α is the value of time (in Indian rupees per hour), and β_E and β_L are schedule costs for early and late arrival. Commuters not included in the area treatment only choose departure time, as in equations (1) and (2).

There are three sources of heterogeneity. First, to allow different patterns of substitution between departure times and between routes, I assume that the random utility shocks $\varepsilon_{it}(h, r)$ follow an extreme value distribution with correlation within each route. This leads to a nested logit structure over routes and departure times, conditional on h^A_{it}. Route-level shocks intuitively capture idiosyncratic factors such as, for example, the need to make a quick stop along one of the routes. The variances of route-level and departure-time level utility shocks are determined by two parameters μ and σ. As costs in the utility function scale approximately linearly with route length, I normalize the logit parameter $\sigma_i = \frac{KM_i}{\overline{KM}}\sigma$ where $\overline{KM}$ is the sample average of pre-experiment trip length KM_i. This prevents commuters who travel far from having more precise choices mechanically, as would be implied by a constant $\sigma_i = \sigma$. Commuters not in the area treatment choose departure times according to a multinomial logit conditional on h^A_{it}, with σ_i as above. The formulas for choice probabilities are in Online Appendix A.3.1.[42]

Second, observed choices for commuter i are integrated over all potential ideal arrival times h^A_{it}. This is a version of allowing for correlated shocks across certain pairs of departure times. The distribution of ideal arrival times is individual-specific to help account for the variation of departure times both across and within commuters in the data. For comparison, I later also estimate a model with the same distribution $\mathcal{H}^A$ for all commuters.

Third, to capture the stark heterogeneity in responses to the congestion charge treatments, I

[42] Assuming independent utility shocks along the departure time grid may not seem attractive. However, the resulting choice probabilities have a familiar form. To see this, assume that expected utility is quadratic – which always holds as an approximation around the optimal departure time h^* – then as the grid becomes finer the multinomial logit model becomes equivalent to choosing h^* plus a normally distributed term with standard deviation related to the inverse curvature of the utility function at the optimum.

assume that there are two types of commuters that differ in how they respond to monetary charges $p(h, r)$. A fraction p is described by the preferences above, while the rest behave as if congestion charges $p(h, r)$ were always equal to zero. Commuter type is randomly assigned independently. This modeling assumption has two possible interpretations. First, the $1-p$ fraction may be infra-marginal because they have significantly larger coefficients α, β_E, β_L. In this case, the experimental variation does not allow to estimate their preferences. Second, these commuters may be inattentive or not remember the treatment details, consistent with the results in Supplementary Materials Table SM1. I refer to the first type as those who "respond" (to the treatment).[43]

The key preference parameters of interest are the marginal value of time spend driving α, and the schedule costs of arriving early (β_E) and late (β_L). The logit parameters σ and μ, the share p of commuters who respond, and the commuter-specific distribution of ideal arrival times $\mathcal{H}_i^A$, will also be estimated.

Data Sources and Estimation Sample. I use two types of Google Maps data to construct choice sets. For average driving times $\mathrm{ET}_{it}(h, 0)$ and short route length KM_i, for each study participant, I collected Google Maps predicted driving times on their home to work route at all departure times throughout the day. To calibrate travel time uncertainty, I use the real-time Google Maps data. Within route by departure time cells, driving time is approximately log-linearly distributed across the 145 weekdays in the data, with the standard deviation well explained by a quadratic in the average driving time (Supplementary Materials Figures SM3 and SM4). Thus, I assume that driving time follows a log-normal distribution around the measured Google Maps average driving time. For area treatment participants, I calibrate travel times on the alternate route ($r = 1$) as an individual-specific multiple of travel times on the short route ($r = 0$).[44]

The trip sample is all morning home to work trips. The commuter sample consists of 308 regular commuters who are either in the area treatment or who have at least one sample trip before the experiment and at least one during the experiment.

GMM Estimation and Moments. To estimate the model, I use a slightly modified version of generalized method of moments (GMM).[45]

Moments are chosen with three goals in mind: (a) the main dimensions of choice (departure time and route choice), (b) the variation induced by the experiment, and (c) both aggregate and heterogeneous responses. To leverage the experimental variation, each moment is either computed twice for control and treatment, or it measures the difference between treatment and control.

The first set of 31 moments measures the difference in difference in 10-minute departure time bin "market shares" (in departure time charges treatment minus control, during minus before the experiment of the probability to choose that departure bin). The second set of two moments quantifies

[43]A standard way to capture preference heterogeneity is by assuming a continuous, uni-modal distribution of random coefficients for α_i, β_{Ei} and β_{Li}. Attempts to fit models with normal and log-normal distributions were not successful. Due to the stark heterogeneity in responses to the treatment (see panels C-G in Online Appendix Figure A6), the variance terms of these distributions tends to diverge.

[44]Before the experiment, I obtained from Google Maps the driving time at 9 AM for the quickest route that does not intersect the congestion area. I use this travel time to construct the individual-specific detour route factor.

[45]Simulation is not required. The log-normal travel time distribution and the parametric utility function lead to closed form expressions for expected utility over travel time.

the variance of individual-level changes in shadow rates (during the experiment relative to before), in the departure time treatment and control groups. The third set of two moments measures the route choice "market shares" when treated and not treated with area charges. The last set of four moments measure the shares of commuters who choose the short route a fraction f of days, with $f \in [1/3, 2/3)$ and $f \in [2/3, 1)$, when treated and when not treated with area charges. (For formal expressions see Online Appendix A.3.2).

The non-standard aspect of the estimation concerns the commuter-specific distribution of ideal arrival times $\mathcal{H}_i^A$. Each time I simulate the model, I infer the distribution of ideal arrival times by approximately inverting the pre-experiment distribution of departure times, conditional on the other parameters $\theta = (\alpha, \beta_E, \beta_L, \sigma, \mu)$. The specific procedure is as follows. Before estimation, for each commuter, I approximate the distribution of pre-experiment departure times as a (truncated) normal distribution $\mathcal{N}_i^0$. For each simulation run, for each possible ideal arrival time h^A, I compute the departure time choice probabilities $\Pr(h \mid h^A, \theta)$ and approximate it by the *expected* departure time $\mathrm{E}h(h^A, \theta)$. I then use this function from arrival times to departure time to invert $\mathcal{N}_i^0$ to a distribution of ideal arrival times $\mathcal{H}_i^A$.[46] This entire procedure may introduce bias due to an incidental parameters problem. To address this, I will check that the entire estimation procedure can retrieve true parameters when applied to simulated data with exactly the same sample size.

Each stage of the two-step GMM optimization is repeated with 50 random parameter starting conditions, to make convergence to a global minimum more likely. Confidence intervals for estimated parameters are computed using 100 bootstrap iterations.[47]

6.1 Travel Demand Estimation Results

Table 4 shows the estimation results. The value of time is Rs. 1,085 per hour, early schedule costs are Rs. 295 and late schedule costs Rs. 261 per hour. In relative terms, schedule costs are moderate, indicating some degree of flexibility. To further put these values in context, consider a setting without travel time uncertainty where a commuter arrives early. With these preferences, he is indifferent between the current departure time and leaving 10 minutes earlier if that has a travel time around 3 minutes lower.[48]

The estimated value of time is significantly larger compared to the self-reported hourly wage

[46] By only considering expected departure time, this procedure may over-estimate the variance of $\mathcal{H}_i^A$. In practice, this is not a major concern because this variance is significantly larger than variance in choices. To check, I compute the implied aggregate distribution $\mathcal{N}_i^1$ of optimal departure times given by $\int_{h^A \sim \mathcal{H}_i^A} \Pr(h \mid h^A, \theta) dh^A$. $\mathcal{N}_i^0$ and $\mathcal{N}_i^1$ have very similar mean and standard deviation for the median commuter. To further address this concern, the full procedure includes an additional step where, for each i, I adjust the mean and standard deviation of $\mathcal{N}_i^0$ in the opposite direction (i.e. if $\mathcal{N}_i^1$ has larger variance, I shrink the standard deviation of $\mathcal{N}_i^0$) and repeat the above inversion. This second steps does not make a large difference for most commuters.

[47] Note that the model has an explicit likelihood function, recommending maximum likelihood on efficiency grounds. The reason for using GMM with the moments described above is that it exploits the variation induced by the experiment. In practice, running MLE has poor convergence properties. This may be related to the fact that most of the data is before the experiment, and the ideal time distribution gives the model significant degrees of freedom in fitting that part of the data.

[48] For further context, the slope between 6 am and 9 am in Figure 1 is between 2.5% and 4% every 10 minutes. Ignoring uncertainty, the minimum trip duration for which it is worth traveling earlier is between 70 and 110 minutes.

in this sample (Rs. 165).[49] Intuitively, for a given response probability p, the data moments that identify the value of time (and the logit parameter μ) are the fraction of commuters choosing the short route without and with charges (87% and 68% percent), and the average duration difference between the two (6.5 minutes).

Three additional factors are pertinent to how to interpret the estimated value of time α. First, the model uses Google Maps travel time, and study participants overestimate travel time differences relative to Google Maps. Hence, the value of *subjective* time will be lower.[50]

Second, the experiment captures the willingness to accept for a longer travel time. This may be larger than the willingness to pay for a quicker trip. The policy-relevant construct for measuring welfare effects of a hypothetical citywide congestion charge scheme depends on the secular trend of congestion. In Bangalore, willingness to pay to avoid an increase in congestion is plausible, because travel delay increased steadily at an annualized rate of 15% during the six month study period (not adjusted for seasonality).

Third, in reality commuters may have a fixed cost of switching to a new route, due to idiosyncratic factors, habit, or information. In Online Appendix Table A8 I re-estimate the model including a cost $\gamma_1 \in \{\text{Rs. } 12, 24, 48, 96\}$ of taking the long route ($r = 1$), in addition to the cost of additional travel time. The estimated the value of time α is decreasing and approximately linear in the fixed cost, reaching zero for a fixed cost roughly equal to the low area charge rate. A small fixed cost does not significantly alter the estimated value of time, yet larger fixed costs would matter.

The key result for schedule costs is that commuters are moderately flexible to change their departure times. In particular, this means that commuters have some ability to "self-insure" against congestion, in the sense that commuters will tend to change departure times in response to a localized increase in congestion, which will attenuate the negative welfare impact of the shock. Interestingly, I find similar early and late costs. These estimates are more precise than the reduced form results because the structural model is more precise about where to expect changes in trip timing for different values of β_E and β_L and because the analysis is restricted to regular commuters.[51]

In the current model, all commuters who respond have the same schedule costs. If in reality commuters have heterogeneous relative schedule costs, then intuitively the current estimates capture early costs β_E for commuters who already travel early (who are more likely to travel during the early "ramp" of the congestion charge profile), and late costs β_L for late travelers. Note that given the focus on policies that incentivize traveling away from the peak-hour, these are policy-relevant quantities.

The probability to respond to congestion charges is 42%, similar to the fraction of participants who were "attentive" to treatments (Supplementary Materials Table SM1). Intuitively, a value near

[49] Most previous estimates of value of time in the transportation economics literature are between 50% and 100% of the wage (Wardman, 1998). However, most studies use stated preferences, a method that underestimates the value of time (Small et al., 2005). In particular, Bento et al. (2017) find that commuter choices between tolled and un-tolled lanes imply very high values of time for some commuters, which they argue is due to a "value of urgency."

[50] The average detour is 6.5 minutes according to Google Maps, while in a phone survey, the median and average self-reported detour durations were 11.7 and 13.6 minutes. Similarly, study participants overestimate the time savings from earlier (later) departures for those who typicall travel before (after) the peak-hour by a factor of 1.5-3×.

[51] The lower amplitude of changes in the late morning is partly due to a lower *level* of trips then (Online Appendix Figure A6 panels A, B). Note also that the bootstrap confidence intervals show that late costs are less precisely estimated and it is not possible to reject very large late costs.

one half maximizes the variance of individual responses, emphasizing the stark response heterogeneity in the data. As an alternate method to explore preference heterogeneity, in Online Appendix Table A8, I show that forcing α_i, β_{Ei} and β_{Li} to be proportional to commuter i's wage increases the estimated average values. This is consistent with heterogeneity results in Supplementary Materials Tables SM5 and SM6 which do not find heterogeneity based on income.

The logit departure time parameter σ is smaller than the route parameter μ. The former is intuitively identified from the precision of the bunching "spikes" at ± 1.5 hours.

Model fit and sensitivity. Appendix Figure A6 shows the model fit graphically by plotting data and model prediction. The model does a good job of matching the experimental changes in volume of trips at the ± 1.5 hour "bunching" points (panel C), and it successfully replicates the heterogeneity in treatment responses in the data (panels D-G). In particular, note the U-shaped pattern in the area treatment group (panel F). The estimated model also does well matching the overall levels of departure time shares (panels A and B), which are not moments. Allowing the ideal arrival time distribution $\mathcal{H}_i^A$ to be individual-specific matters for this result. When I use a common distribution $\mathcal{H}^A$, model fit for departure market shared is significantly worse (results not shown) and estimated early schedule cost drops significantly (Online Appendix Table A8).

I use two empirical methods to shed light on model parameter identification. The first is to show numerically that the estimation procedure can recover the parameters using simulated data for various sets of random parameter. Specifically, I simulate data of exactly the same size using random model parameters, and re-run the entire estimation procedure on the synthetic data. I repeat this procedure 100 times. Scatter plots in Online Appendix Figure A8 and quantile regressions in Online Appendix Table A9 show that in general estimated parameters track the true parameters closely. The main exceptions are overestimates for high values of schedule costs, and systematic underestimates for σ.

The second exercise is to compute the sensitivity measure from Andrews et al. (2017). The (scaled) sensitivity matrix Λ captures how estimated parameters depend on the different moments of the data. Specifically, each entry $\Lambda_{\gamma k}$ measures the impact of a standard deviation increase in moment g^k, $1 \leq k \leq 39$, on estimated parameter $\gamma \in \{\alpha, \beta_E, \beta_L, \sigma, \mu, p\}$. Online Appendix Figure A7 shows that β_E depends most strongly on departure time moments around the bunching point at the -1.5 hour mark, as expected. Conversely, β_L depends most strongly on what happens around $+1.5$ hours. Second, Online Appendix Table A10 shows that the value of time depends much more strongly on the area treatment moments than on the departure time treatment moments. The probability to respond, p, is also affected more by the area moments. Finally, these results also emphasize that parameters are jointly estimated, with contributions from several moments.

Overall, the structural model offers a good fit to how commuters responded to the congestion charge experiments. The results indicate that commuters are fairly flexible to change their schedules by leaving earlier or later locally around their departure times, relative to how much they value time spent driving. However, in order to quantify the welfare impacts of congestion mitigating policies, it is also necessary to know how traffic responds to aggregate changes in driving patterns.

7 The Road Traffic Congestion Technology

Each additional vehicle on the road leads to slower road speeds. I now quantify this external cost in Banglaore using all the GPS trip data collected during the study and real-time Google Maps driving time data from the same period.[52]

The main empirical strategy is to use the significant within-day variation in city-wide traffic volume to measure the causal impact of traffic volume on speeds. I will argue that, to a first approximation, the large shifts in demand for travel at different times of the day trace out the technological supply of road speed.

To measure the *quantity* of driving, I use 117,527 trips coded from GPS data from 1,747 app users, covering 185 calendar dates and 44,034 user-days with travel information. (This sample includes but is not restricted to the experimental sample.) For road *speeds*, I use two different data sources that give similar results. My main data source is real-time Google Maps travel delay data collected every 20 minutes on 30 routes in the study area (in both directions) over the same calendar period. This data has the advantage of collecting speed data on a fixed set of routes. Recall that for 97.1% of all observations (and 95.6% of all morning peak-hour observations), the length in meters of the path indicated by Google Maps is exactly the same as the median for that day and route. I also compute trip-level travel delay directly from the GPS data. Both the quantity and speed measures focus on *trips* or trip segments, the level at which commuters make decisions and experience travel times. This is unlike the general approach in transportation engineering, which is to describe the relationship between speed and volume at measurement *points* along the road. See section 3.3.2 in Small et al. (2007) for a textbook review.[53]

I next summarize both variables for each departure time in a day, averaging over all non-holiday weekdays in the data. The volume of departures is a measure of traffic inflows into the urban road network. The measure of interest is the semi-elasticity of average travel time at departure time h with respect to the average volume of trip departures at that time, $Q(h)\frac{\partial T(h)}{\partial Q(h)}$. This measures the additional aggregate travel time imposed by an additional trip starting at time h.[54] This semi-elasticity is independent of the total number of trips in Bangalore, so I can estimate it consistently using quantity $Q^S(h)$ from a (representative) sample of trips. For interpretability, I show results on $\frac{\partial T(h)}{\partial Q^S(h)}$ with Q^S is normalized to mean 1.

Average travel delay is well explained by a linear function of traffic volume (Figure 3). The linear functional form is visible throughout the range of volumes, including at close to zero traffic. An increase in the number of vehicles equal to 10% of the mean is associated with an increase of 0.1 minutes per kilometer (Table 5). I can reject at the 95% level an exponent of 1.19 or greater on

[52]Driving also imposes other external costs: pollution emissions, pollution exposure, accidents, etc. Here I measure the impact on higher (and less reliable) driving times.

[53]Russo et al. (2019) use loop detector data in Rome, Italy to measure instantaneous speed and mention that this likely overestimates *trip-level* speed because it does not measure waiting at traffic lights.

[54]I cannot separate the impact of motorcycles and cars, and cannot account for vehicle occupancy. In the study sample, the share of trips made by car is roughly constant throughout the day, at around a third. Results should be interpreted as the average effect along these dimensions. In the GPS data, motorcycle speed is only slightly higher than car speed.

traffic volume (column 2).[55,56] Results using average delay computed from GPS trips are similar and slightly shallower (columns 4, 5). Panel A in Appendix Figure A9 shows that the entire distribution of travel delay computed from GPS trips, including the 90th percentile, scales linearly or sub-linearly as a function of traffic volume. This shows that the probability of severe traffic jams is not disproportionately higher during peak-hours.

Quantitatively, these results imply that an additional trip of average length increases average aggregate driving time by approximately 4 minutes for a 7 am departure time, and by approximately 16 minutes a peak-time departure time (9 am or 7 pm).[57] Given that the average peak-hour trip duration is 33 minutes, this congestion externality estimate is non-trivial yet small relative to many previous estimates. Notably, in the canonical fixed-capacity bottleneck model (Vickrey, 1969; Arnott et al., 1993),[58] it is easy to prove that a commuter's partial equilibrium marginal impact on aggregate driving time is equal to the duration of time after they pass through the bottleneck until the queue clears for the first time. This may be several times larger than the trip duration. For example, if a queue has strictly positive length for 4 hours, continuously, commuters at the start create approximately 4 hours of aggregate driving time, even if the queue is significantly shorter at any point in time. However, the bottleneck model predicts a non-linear increase in travel delay in the morning, and hence offers a poor fit for the *citywide* road technology in Bangalore. (The models with a sequence of bottlenecks in Online Appendix A.2 predict a linear relationship.)

Interpreting these results as the causal impact of driving on aggregate driving time raises several potential concerns. First, the sample of trips may over-represent peak-hour traffic volumes (and hence underestimate the semi-elasticity), for example due to recruitment times in gas stations. Note that recruitment time barely predicts trip departure time from GPS data.[59] In any case, recruitment bias may go the other way, as surveyors worked continuously between 8am-1pm and 3pm-8pm, yet commuter arrival rate in gas station is significantly higher during peak-hours. A related issue is if truck or other traffic is over-represented off-peak, meaning current results underestimate off-peak traffic volumes. I do not have the data to rule in nor rule out this possibility. Note that Bangalore does not have explicit daytime truck traffic restrictions, and it is not immediately clear that truck

[55] Travel time plausibly depends on the history of inflows, not only on contemporaneous inflow. Indeed, including lags in Table 5 increases the R^2 slightly. Another approach is to assume that speed depends on the *density* of all vehicles on the road at any given time. Using density as dependent variable gives very similar results, see panel B in Online Appendix Figure A9. Intuitively, the two variables are strongly correlated, because most trips are short relative to the scale of peak/off-peak fluctuations. As these approaches give similar results, I use the more parsimonious linear relationship.

[56] Results are similar and slightly shallower using variation across calendar dates, including holiday and weekends (Supplementary Materials Figure SM1). Note, there is less variation in average traffic volume at the daily level. In both cases, I normalize the dependent (volume) variable to mean 1, so the slopes are directly comparable.

[57] Aggregate driving time is $Q(h) \cdot T(h) \cdot \overline{KM}$ where $Q(h)$ is the volume of trips leaving at h, $T(h)$ their average travel delay (minutes/km) and $\overline{KM}$ the average trip length. The marginal change due to an additional trip is $Q(h) \cdot \frac{\partial T(h)}{\partial Q(h)} \cdot \overline{KM} \approx Q^S(h) \cdot \frac{\partial T(h)}{\partial Q^S(h)} \cdot \overline{KM} = Q^S(h) \cdot 0.99 \cdot 8.0$ km, where Q^S is in-sample traffic volume (the X axis in Figure 3). Quantity is $Q(h) \approx 0.5$ at 7 am and ≈ 2 at peak-hour.

[58] In this model, vehicles arrive at a bottleneck at a continuous time-varying flow. If the flow is below capacity, they pass through without delay. When the incoming flow exceeds capacity, a queue forms. The queue is cleared in a first-come-first-served order, at the bottleneck capacity rate per unit of time.

[59] The R^2 of the regression of trip departure time in the morning on morning recruiting time is below 0.04, and below 0.02 for the evening. See Panel C in Appendix Figure A9 for aggregate recruitment times and trip volumes.

traffic would reschedule operations to avoid traffic delays.

Another concern is that the true relationship is convex, yet averaging over time or over space "linearizes" it. For example, the exact moment of the peak-hour may vary slightly on different days, leading to attenuation bias in traffic volume. Online Appendix Figure A10 shows virtually identical results day by day, including on Sundays, when traffic volume is overall lower. These results bolster the case that the results trace out a technological relationship. In Online Appendix Figure A11, I count the number of trips by road artery and direction (arteries are shown in Online Appendix Figure A1). Even for individual arteries, delay appears linear in volume, with a slope that is sometimes different than the citywide slope.

A general concern is that this analysis misses an omitted variable correlated with high peak-hour demand and with faster peak-hour travel. However, compelling examples are not obvious for Bangalore. For example, the study period was overwhelmingly dry and mild, so weather events correlated with peak-hours are unlikely. An example would be if traffic police disproportionately alleviates traffic during peak-hours. (However, this may also be considered a "component" of the road technology.) Some omitted variables might instead bias the relationship upward. For example, higher pedestrian flows during peak-hours may further slow traffic, biasing the slope up.

Peak-hour trips may also be different, for example they may be longer and hence faster, as shown in Couture et al. (2018), or peak-hour drivers may drive faster in general. I run trip-level quantile (median) regressions of trip delay on the traffic volume measure used above (Online Appendix Table A11). Results remain similar and slightly shallower, including when controlling for trip length and commuter fixed effects.

The results in Bangalore are broadly similar to those reported by Akbar and Duranton (2017) in Bogotá, Colombia (Online Appendix Figure A9 panel D). Akbar and Duranton (2017) construct traffic density using entire trips using a representative transportation household survey. These results suggest a more general conclusion about a shallow road traffic externality in urban road networks in developing countries.[60] Russo et al. (2019) and Yang et al. (2019) use loop detector data to characterize road traffic externalities in Rome and Beijing.

A linear relationship between traffic volume and travel time, including at the road artery level, is a puzzle. Compared to canonical ways to model traffic externalities along highways or at bottlenecks, which are highly non-linear, the linear relationship I find is steeper at low traffic volumes and shallower for high traffic volumes.[61] In Online Appendix A.2 I describe two stylized models of traffic congestion that show that an approximately linear relationship between expected travel time and traffic volume can arise over a certain range of traffic volumes in road networks with unsynchronized traffic lights or where cars meet in unsignaled intersections.[62]

---

[60] However, note that the slope in Bangalore is several times shallower than the slope identified based on taxi trips in Geroliminis and Daganzo (2008) in Yokohama, Japan.

[61] A commonly used functional form to describe travel time T as a function of incoming flow V is given by $T = T_f \cdot \left(1 + a \cdot (V/V_k)^b\right)$, where T_f is time under free-flow, and V_k is the maximum road capacity. The parameter values for the exponent b vary considerably. For example, the Bureau of Public Roads (BPR) and the updated BPR functions use $b = 4$ and $b = 10$, respectively. See section 3.3.2 in Small et al. (2007).

[62] Akbar and Duranton (2017) propose a different model for this result. Each individual road segment has a convex relationship, and at peak-hour more of the traffic spills over to smaller side-roads.

It is likely that at some level of traffic, travel time in Bangalore would become highly convex in additional traffic volume, as streets overflow and vehicles block intersections. The results in this section suggest that these levels are not typically reached in the current equilibrium, even during peak-hours.

8 Policy Simulations

Equipped with preference estimates from the experiment and the calibrated road traffic congestion externality in Bangalore, in this section I simulate and compare the Nash equilibrium and social optimum in the model of peak-hour equilibrium from section 3.

The simulation model is populated with 3,080 agents. Each agent is a copy of one of the 308 real study participants, with the same route length and preferences as estimated in section 6, with a fixed ideal arrival time randomly drawn from the distribution estimated for that agent.[63] In the benchmark model, departure time is the only margin of choice. I later introduce an extensive margin decision.

Commuters select departure times probabilistically based on equation (2) with $\varepsilon_i(h)$ type-1 extreme value distributed with scale $\hat{\sigma}$. In the baseline simulations, I assume constant α, β_E and β_L.[64] Aggregate departure time choices affect travel times for a given departure time according to a linear relationship between expected travel time and traffic volume, as in Table 5. Travel time uncertainty is parametrized to be log-normal distributed with a standard deviation that is quadratic in the mean (see Supplementary Materials Figure SM4). Similar to the previous section, these simulations are scale-independent.

To compute Nash equilibria, I use an asynchronous logit best-response dynamic. Given (fixed) congestion charges that depend on the departure time, and an initial travel time profile, each period a 1% random sample of agents re-compute their logit choice probabilities. After each such round, the travel time profile is updated given aggregate traffic volumes at each departure time. Convergence is achieved when every agent is already close to best-responding, namely when the ℓ^2-norm of changes in choice probabilities, averaged over the entire population, is below a certain threshold. I find fast convergence to equilibrium; it takes around 9 revisions per capita to reach equilibrium. Moreover, this dynamic has a natural interpretation in terms of commuters revising their actions periodically.[65]

Social welfare is defined as average expected utility (over travel time uncertainty and idiosyncratic preference shocks $\varepsilon_{it}(h)$). To find the social optimal allocation, I compute a Nash equilibrium where departure time charges at time h are "pigovian," namely equal to the marginal social cost of departing at h.[66] I assume that all revenue is recycled in the form of lump sum transfers to commuters, and that

[63] I use a grid of departure times between 5am and 2pm, every 10 minutes. Results are robust to using 10× more agents and a finer departure time grid.

[64] In the structural estimation, commuters do not respond to congestion charges with probability $1 - p \approx 0.58$. The case here corresponds to interpreting these commuters as inattentive to the experiment and endowed with the same preferences. I later conduct simulations with preference heterogeneity.

[65] In practice, the simulation finds a unique equilibrium independently of starting travel time conditions.

[66] I assume that the social planner knows individual preferences but does not observe the exact realization of the random utility shocks. In this case, where also all commuters have the same externality conditional on departure

there are no implementation costs. This setup is conservative in favor of finding welfare benefits from optimal pricing. The marginal social cost must be calculated allowing other commuters to adjust, as the envelope theorem does not hold away from the social optimum.[67]

To find this fixed point, I use a "lazy" adjustment dynamic for congestion charges. Starting at the Nash equilibrium, at each iteration I compute the marginal social cost. This involves computing a new Nash equilibrium for each departure time. I then update charges with a 1/3 weight on the new marginal social cost and 2/3 on the current charges. This procedure converges in around 15 iterations with precision Rs. 0.1 for welfare.

The social optimum is described in Figure 4 and Table 6. There are four main findings. First, optimal congestion charges lead to small but notable improvements in travel times. The social optimum has a lower peak hour, and more commuters departing early, between 5:30 am and 7:15 am. At the peak, travel delay improves by around 0.15 minutes per kilometer. Average travel times improve by 1.5 minutes per commuter (Table 6). Second, trip congestion charges are high, over Rs. 300 (4 USD) at the peak. Third, the 10th and 90th percentiles of the change in average departure time (conditional on ideal arrival time) are leaving 8.8 minutes earlier, and 5.6 minutes later. These numbers are similar to the range of experimental responses to the departure time policy. Hence, these counterfactual results do not rely significantly on functional form extrapolations.

Fourth, through the lens of the model, the social optimum leads to a very small change in commuter welfare, an order of magnitude smaller than the effect on travel time alone. Welfare is Rs. 4.9 higher per commuter under the optimum (7 US cents), from a total trip cost of around Rs. 802, or a bit more than a half percentage improvement. The effect is larger relative to free-flow conditions, defined as the no-congestion speed, the intercept in Table 5. Overall, travel time benefits are nearly offset by schedule costs incurred by commuters who are induced to travel at privately inconvenient times. These results suggest that real-world, more forceful policies that attempt to cap peak-hour congestion may lower commuter welfare as defined here. Indeed, real-world policies are likely to include implementation costs, may not fully recycle the revenue, and are unlikely to charge the optimal time-varying fees.

The simulation model used here makes several strong assumptions. I do not take into account longer term preferences and adjustments, which may be different from the short-term responses measured in the experiment. As in the road technology estimation, I do not distinguish between the externalities generated by motorcycles and cars. Finally, this analysis ignores other traffic, including trips that are not between home and work, and taxi and bus passengers who would also benefit from reductions in travel times.[68] The simulations also ignore bus and truck traffic, which may respond differently to similar congestion charges and may affect traffic differently. These calculations also do account for other important social costs of congestion, such as pollution.

---

time, the planner can implement the social optimum with departure time congestion charges.

[67] Arnott et al. (1993) make this point in their model with identical agents.

[68] As a back of the envelope calculation, census data shows that in 2011 roughly the same fraction of workers used buses as private vehicles in Bangalore. Assuming a value of time twice as small for bus users, and assuming a similar distribution of departure times and distances, the 1.5 minute improvement in average travel time for bus users is worth approximately Rs. 13.5 on average per driver, almost three times larger than welfare gains among drivers.

Which model components drive these results? To find out, I conduct a series of counterfactual simulations. I first look at preference heterogeneity. In theory, the presence of commuters with lower value of time and/or lower schedule costs may increase the gains from congestion pricing. In Table 6 panel B I first assume that commuters have log-normal distributed preferences (α_i, β_{Ei} and β_{Li}) with the same variance as self-reported wages. Welfare gains are twice as large yet still small (a 1.3% increase in commuter welfare). Welfare gains are smaller when value of time and schedule costs parameters are independent.[69] I find similar results for a separate parametrization with binary heterogeneity.

I next vary relative preferences. Figure 5 shows very similar results for a value of time up to four times lower or higher than the estimated value. Figure 5 also shows that the shape of the externality is a crucial factor in the results. Assuming that average travel time is a cubic (instead of linear) function of traffic volume leads to an order of maginitude larger welfare gains from optimal congestion pricing (4.7% instead of 0.6%) and a stronger dependence on value of time. (See also Table 6 Panel C.)

Extensive Margin Trip Choice. In the final exercise, I introduce an extensive margin decision with a nested logit specification. The outer nest has two options, taking the trip ($x = 1$) and not taking the trip ($x = 0$). Trips are valuable: a commuter not making a trip incurs a cost proportional to trip length $\delta_i = \delta \cdot \text{KM}_i / \overline{\text{KM}}$.[70] Expected utility is given by:

$$\mathrm{E}u_i(x, h; h_i^A) = \begin{cases} -\alpha T - \beta_E \left| h + T - h_i^A \right|_- - \beta_L \left| h + T - h_i^A \right|_+ - p(h) + \varepsilon_i(1, h) & x = 1 \\ -\delta_i + \varepsilon_i(0, h) & x = 0 \end{cases} \tag{9}$$

where $\varepsilon_i(x, h)$ follow an extreme value distribution with correlation within each value of x, with parameter η for the choice between trip and no trip (higher η means "noisier" trip decision).

The congestion pricing experiment was not designed to estimate the extensive margin trip elasticity.[71] In Table 7, I simulate this model for six chosen combinations of (δ, η). In order to help assess which parameters are reasonable, each time, I compute the percent reduction in traffic (number of trips) due to a flat Rs. 200 (USD 2.7) fee imposed at the Nash equilibrium, and the implied elasticity with respect to total travel costs. (As a point of reference, in these simulations, average congestion charges at the social optimum vary between Rs. 240 and Rs. 264.)

[69] In these simulations, I use a higher shape parameter $\sigma = 80$, which makes choices "noisier." When using the estimated σ, the model finds a social optimum with "segregation" for nearby departure times, with large swings in traffic volume that sometimes lead to overtaking. I show in Panel B that results in the benchmark model are not sensitive to the choice of σ.

[70] The other components of utility scale approximately linearly with trip length, so this specification ensures that the probability of taking a trip is not mechanically higher for commuters with short trips.

[71] There are two reasons for this. The first is a measurement issue, as commuters would have strong incentives to leave their smartphones at home. A solution would be installing GPS devices in the private vehicle – as Singapore plans to implement its next generation Electronic Road Pricing policy, and as in the experiment in Martin and Thornton (2017). The second reasons is that extensive margin changes will plausibly take longer, as commuters need to find alternate travel arrangements or substitutes for canceling unessential trips. Note that Martin and Thornton (2017) do not find any reduced-form impact of distance-based congestion charges on trip extensive margin in Melbourne, even after two months.

Commuter welfare gains are increasing in the extensive margin elasticity. For small elasticities of around 0.10, welfare gains are around 1.5%. Achieving welfare gains of around 6% requires elasticities above 1, or a 30% reduction in traffic volume from the flat Rs. 200 fee. These results suggest that the linear road technology also limits potential gains from reducing the total volume of road traffic, for plausible values of the extensive margin elasticity.

9 Discussion

This paper shows that in cities with road networks similar to Bangalore, optimal time-varying peak-hour congestion pricing and similar quantity-based restrictions are unlikely to significantly improve commuter welfare as defined here. I make two main contributions. I design a field experiment guided by a travel demand model, with a particular focus on price variation that helps identify the key preference parameters. Second, I show empirically that Bangalore has a moderate peak-hour road speed externality and using equilibrium counterfactual simulations that the current equilibrium inefficiency is low.

Peak-hour pricing may be highly beneficial in other settings, such as where the externality is highly non-linear, and where differentiated pricing is feasible, e.g. tolling only certain lanes (Hall, 2018, 2020). In addition, peak-hour pricing may help correct other externalities, for example relieving prolonged exposure to air pollution from vehicle exhaust. A detailed measurement of those localized externalities, of awareness of and willingness to pay to avoid these harms is necessary to understand these issues quantitatively.

This paper also suggests that road infrastructure investment – including adding road capacity as well as investments to make road network flows more efficient – may currently be at an inefficiently low level. A particularly interesting possibility is that a more efficient urban road network and congestion pricing are complementary policies. This would be the case if, for example, using synchronized traffic lights would lead to a more convex relationship between traffic volume and speed, thereby making congestion pricing more valuable.

References

Akbar, P. A., Couture, V., Duranton, G., and Storeygard, A. (2018). Mobility and congestion in urban india. *working paper.*

Akbar, P. A. and Duranton, G. (2017). Measuring the cost of congestion in highly congested city: Bogotá. *CAF Development Bank of Latin America Working Paper 2017/04.*

Anderson, M. L. (2014). Subways, strikes, and slowdowns: The impacts of public transit on traffic congestion. *American Economic Review*, 104(9):2763–96.

Andrews, I., Gentzkow, M., and Shapiro, J. M. (2017). Measuring the sensitivity of parameter estimates to estimation moments*. *The Quarterly Journal of Economics*, 132(4):1553–1592.

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

Bento, A., Roth, K., and Waxman, A. (2017). Avoiding traffic congestion externalities? the value of urgency. *Working Paper*.

Brinkman, J. C. (2016). Congestion, agglomeration, and the structure of cities. *Journal of Urban Economics*, 94:13 – 31.

Chu, X. (1995). Endogenous trip scheduling: The henderson approach reformulated and compared with the vickrey approach. *Journal of Urban Economics*, 37(3):324 – 343.

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

Davis, L. (2008). The effect of driving restrictions on air quality in mexico city. *Journal of Political Economy*, 116(1):38–81.

Finkelstein, A. (2009). E-ZTAX: Tax Salience and Tax Rates. *The Quarterly Journal of Economics*, 124(3):969–1010.

Geroliminis, N. and Daganzo, C. F. (2008). Existence of urban-scale macroscopic fundamental diagrams: Some experimental findings. *Transportation Research Part B: Methodological*, 42(9):759 – 770.

Gibson, M. and Carnovale, M. (2015). The effects of road pricing on driver behavior and air pollution. *Journal of Urban Economics*, 89(Supplement C):62 – 73.

Gu, Y., Deakin, E., and Long, Y. (2017). The effects of driving restrictions on travel behavior evidence from beijing. *Journal of Urban Economics*, 102(Supplement C):106 – 122.

Hall, J. D. (2018). Pareto improvements from lexus lanes: the effects of pricing a portion of the lanes on congested highways. *Journal of Public Economics*.

Hall, J. D. (2020). Can Tolling Help Everyone? Estimating the Aggregate and Distributional Consequences of Congestion Pricing. *Journal of the European Economic Association*.

Hanna, R., Kreindler, G., and Olken, B. A. (2017). Citywide effects of high-occupancy vehicle restrictions: Evidence from three-in-one in jakarta. *Science*, 357(6346):89–93.

Henderson, J. (1974). Road congestion: A reconsideration of pricing theory. *Journal of Urban Economics*, 1(3):346 – 365.

Henderson, J. (1981). The economics of staggered work hours. *Journal of Urban Economics*, 9(3):349 – 364.

Karlström, A. and Franklin, J. P. (2009). Behavioral adjustments and equity effects of congestion pricing: Analysis of morning commutes during the stockholm trial. *Transportation Research Part A: Policy and Practice*, 43(3):283 – 296. Stockholm Congestion Charging Trial.

Kreindler, G. (2016). Driving delhi? behavioural responses to driving restrictions. *Working Paper*.

Martin, L. A. and Thornton, S. (2017). Can road charges alleviate congestion? *Working Paper SSRN*.

Merugu, D., Prabhakar, B. S., and Rama, N. (2009). An incentive mechanism for decongesting the roads: A pilot program in bangalore. In *Proc. of ACM NetEcon Workshop*.

Noland, R. B. and Small, K. A. (1995). Travel-time uncertainty, departure time choice, and the cost of the morning commute. *working paper Institute of Transportation Studies, University of California, Irvine*.

OICA (2016). Vehicle sales in the bric countries from 2005 to 2015 (in units). Technical report, International Organization of Motor Vehicle Manufacturers.

Prud'homme, R. and Bocarejo, J. P. (2005). The london congestion charge: a tentative economic appraisal. *Transport Policy*, 12(3):279 – 287.

Raux, C. (2005). Comments on "The London congestion charge: a tentative economic appraisal"; (Prud'homme and Bocajero, 2005). *Transport Policy*, 12.

Russo, A., Adler, M. W., Federica, L., and van Ommeren, J. N. (2019). Welfare Losses of Road Congestion. *Working Paper*.

Small, K. (1982). The scheduling of consumer activities: Work trips. *American Economic Review*, 72(3):467–79.

Small, K., Verhoef, E., and Lindsey, R. (2007). *The Economics of Urban Transportation*. Taylor & Francis.

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

TfL (2006). Central london congestion charging impacts monitoring fourth annual report. Technical report, Transport for London.

Tillema, T., Ben-Elia, E., Ettema, D., and van Delden, J. (2013). Charging versus rewarding: A comparison of road-pricing and rewarding peak avoidance in the netherlands. *Transport Policy*, 26(Supplement C):4 – 14.

Tsivanidis, N. (2019). Evaluating the impact of urban transit infrastructure: Evidence from bogota's transmilenio. *Working Paper*.

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

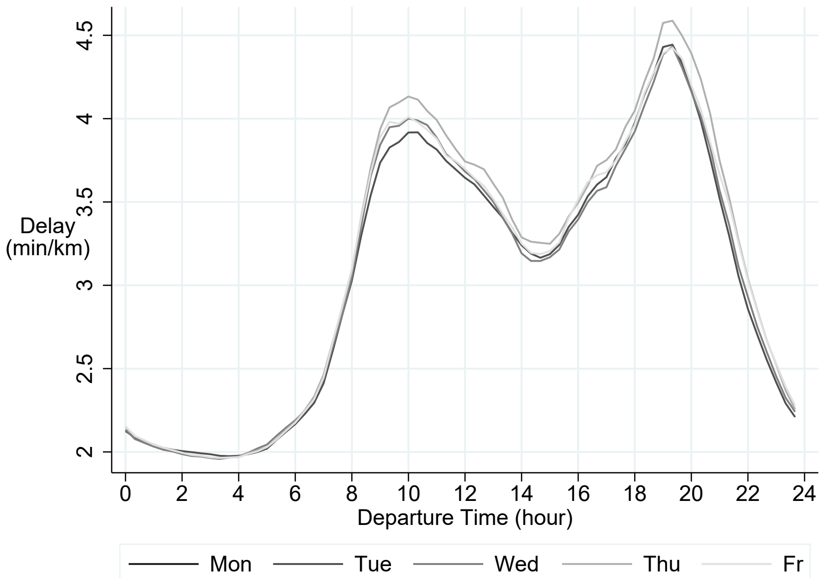
Wardman, M. (1998). The value of travel time: A review of british evidence. *Journal of Transport Economics and Policy*, 32(3):285–316.

Yang, J., Purevjav, A.-O., and Li, S. (2019). The marginal cost of traffic congestion and road pricing: Evidence from a natural experiment in beijing. *American Economic Journal: Applied Economics*, 12(1):418–53.

Zhao, F., Pereira, F. C., Ball, R., Kim, Y., Han, Y., Zegras, C., and Ben-Akiva, M. (2015). Exploratory analysis of a smartphone-based travel survey in singapore. *Transportation Research Record: Journal of the Transportation Research Board*, 2494:45–56.

Figures

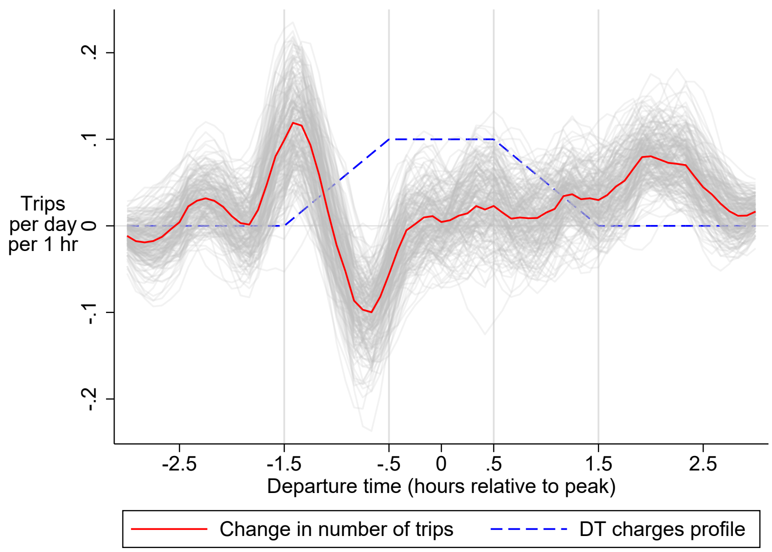
Figure 1: Average Travel Delay in the Study Region in Bangalore



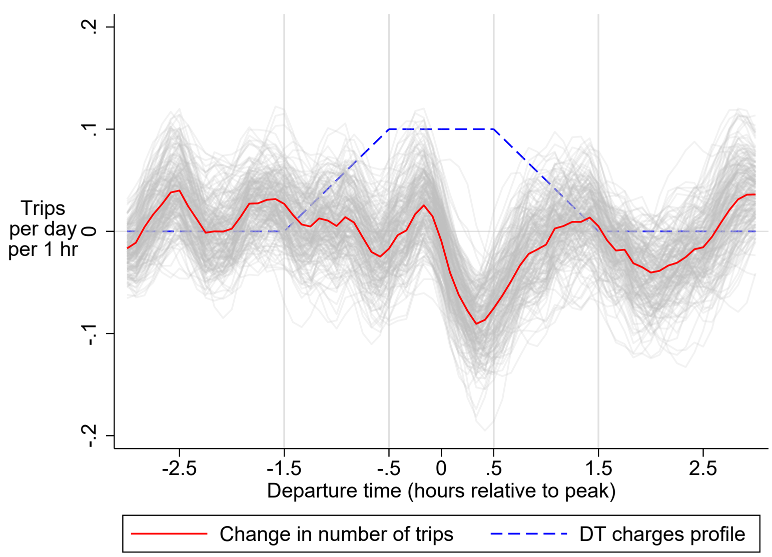
Notes: This graphs plots average travel delay as a function of departure time, on 30 major routes (in both directions) across the study area of South Bangalore, by day of the week. *Travel delay* is the number of minutes to cover one kilometer, i.e. the inverse of speed. (A travel delay of 2 minutes per kilometer corresponds to 18.6 miles per hour.) The travel time and route length data is obtained from the Google Maps API. For each route, I queried the "real-time" travel time (with traffic) as predicted by Google, every 20 minute between February 21st and September 14th, 2017. The sample excludes major holidays.

Figure 2: Impact of Departure Time (DT) Charges on the Distribution of Departure Times

Panel (A) Morning Peak

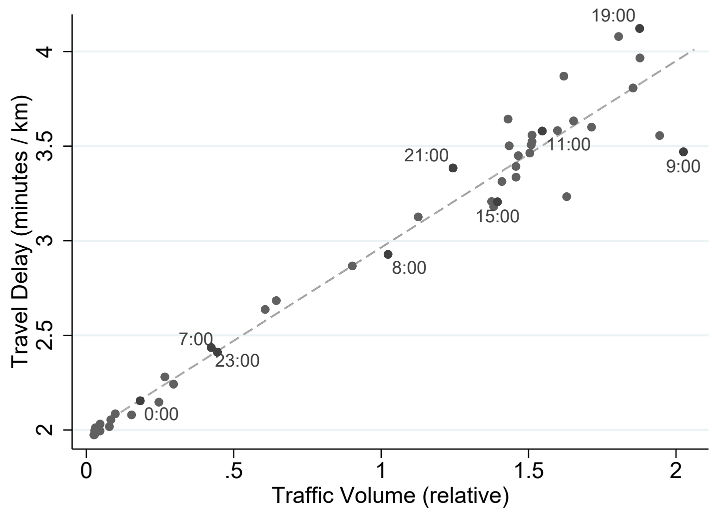


Panel (B) Evening Peak



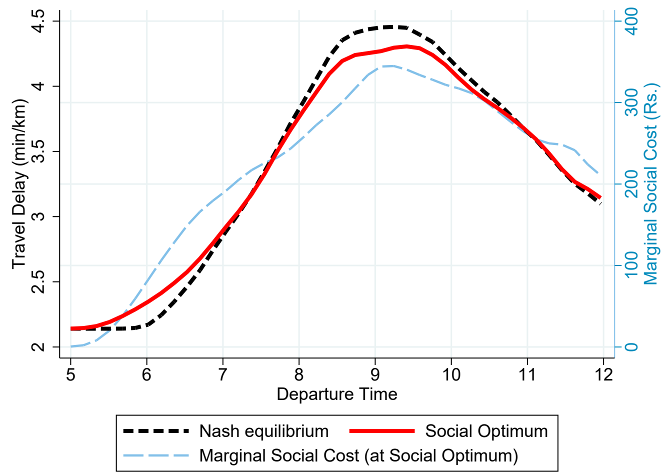
Notes: These graphs plot the impact of departure time charges on the distribution of departure times. The sample is all non-holiday weekdays with good quality GPS data, excluding days outside Bangalore. In the post period, the sample is restricted to the departure time treatment period, either the first or the last three weeks. To construct each figure, I consider four groups, each combination of before or during the experiment and the control or treatment group (pooling together the information and control sub-treatments). Within each group, I compute the kernel density of trip departure times (relative to the midpoint of the congestion charge for each commuter) and multiply it by the average number of trips per day in that group. I then plot the difference-in-difference of these four curves, as well as 200 commuter-level bootstraps. The Y axis measures the change in the number of trips per day in a one hour departure time window. This exercise does not take into account randomization strata. Online Appendix Figure A4 repeats the exercise restricting to commuting trips of regular commuters.

Figure 3: Road Technology: Travel Delay Linear in Traffic Volume



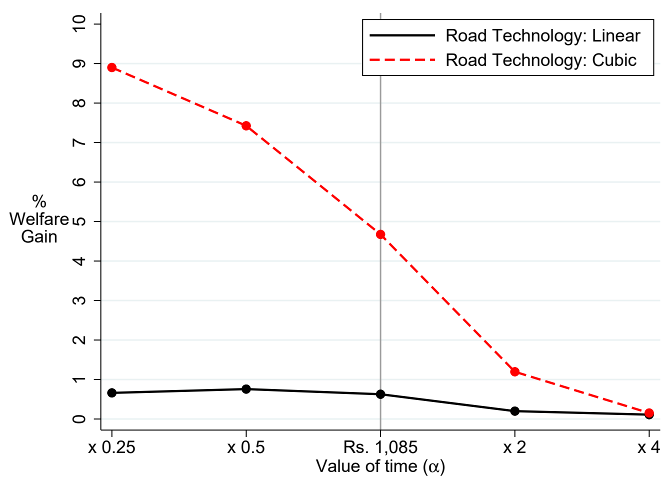
Notes: This graph shows that travel delay is approximately linear in the volume of traffic at the level of departure times. The traffic volume measure uses GPS data for 117,527 trips from 1,747 app users on 185 days and is computed as follows. For each minute in the day, I compute the total number of trips leaving at that time during weekdays, then smooth this value using a local linear regression with Epanechnikov kernel with 10 minutes bandwidth, and normalize to mean 1. Travel delay is derived from Google Maps data collected over 30 routes in South Bangalore (in both directions), every 20 minutes daily for 185 weekdays. I compute the average delay over all weekdays and routes for each departure time, interpolating at the minute level. Table 5 reports corresponding regressions. Supplementary Materials Figure SM1 repeats this exercise at the level of calendar dates instead of departure times. Online Appendix Figures A10 and A11 repeat this exercise separately for specific calendar dates, and for specific road arteries.

Figure 4: Policy Simulation: Unpriced Nash Equilibrium and Social Optimum



Notes: This graph shows the morning profile of travel delay under the simulated Nash equilibrium (black, dashed line, left axis) and under the social optimum (red, solid line, left axis). The social optimum is a Nash equilibrium implemented with (equilibrium-consistent) social marginal charges (blue, long dashed line, right axis). In the Nash equilibrium, departure time choice probabilities are given by multinomial logit based on the travel time profile, and the profile itself is determined by the aggregate rate of departures and the road technology. The social optimum is a Nash equilibrium with departure time charges with the fixed point property: charges are exactly the marginal social cost of adding a commuter at that departure time.

Figure 5: Policy Simulations: Varying Preference and Road Technology Parameters



Notes: This graph plots the improvement in commuter welfare of the social optimum relative to the Nash equilibrium, for different preference and road technology parameters. The black solid line corresponds to the linear road technology from Table 5 ($Delay = \lambda_0 + \lambda_1 Q$ where Q is relative volume), while the red dashed line corresponds to a cubic relationship ($Delay = \lambda_0 + \lambda_1 Q^3$). The X axis reports the value of time α used in the simulation, ranging from four times smaller to four times larger than the estimated value. All simulations in this table use a higher value of $\sigma = 32$. Supplementary Materials Figure SM2 shows the results for average travel time.

Tables

Table 1: Impact of Departure Time Charges on Daily Outcomes

	(1)	(2)	(3)	(4)	(5)	(6)
Outcome	*Total Shadow Rates Today*			*Number of Trips Today*		
Time of Day	AM & PM	AM	PM	AM & PM	AM	PM
Commuter FE	X	X	X	X	X	X
Panel A. All Departure Time Sub-Treatments						
High Rate × Post	-12.29**	-6.92*	-5.36	-0.07	-0.02	-0.04
	(6.04)	(3.77)	(3.41)	(0.14)	(0.07)	(0.07)
Low Rate × Post	-8.30	-3.30	-5.00	-0.08	-0.01	-0.08
	(6.14)	(3.58)	(3.80)	(0.14)	(0.07)	(0.07)
Information × Post	0.45	0.29	0.16	0.10	0.06	0.04
	(5.41)	(3.27)	(3.33)	(0.13)	(0.06)	(0.07)
Post	0.39	-1.36	1.76	0.02	-0.02	0.05
	(4.87)	(2.85)	(3.06)	(0.11)	(0.05)	(0.06)
Observations	15,585	15,585	15,585	15,585	15,585	15,585
Control Mean	96.33	48.08	48.25	3.04	1.15	1.29
Panel B. Any Departure Time Charge vs. Control or Information						
Charges × Post	-10.55**	-5.28**	-5.27**	-0.12	-0.04	-0.08*
	(4.18)	(2.51)	(2.57)	(0.10)	(0.04)	(0.04)
Post	0.65	-1.19	1.84	0.07	-0.01	0.04
	(3.94)	(2.38)	(2.54)	(0.09)	(0.04)	(0.04)
Observations	15,585	15,585	15,585	15,585	15,585	15,585
Control Mean	95.74	46.89	48.85	2.95	1.04	1.12

Notes: This table reports difference-in-difference impacts of the departure time sub-treatments on daily total shadow rates and total number of trips. In the first three columns, the outcome is the sum over trips that day of the trip shadow rate. The shadow rate for a given trip is between 0 and 100 and is computed based on the trip departure time, the respondent's rate profile, and a peak rate of 100 for all respondents. (See Online Appendix Figure A2 for an example of rate profile.) In the last three columns, the outcome is the number of trips that day. The sample is all non-holiday weekdays with good quality GPS data, excluding days outside Bangalore. Only good quality trips are included (See Supplementary Material section SM.2). In the post period, the sample is restricted to the departure time treatment period, either the first or the last three weeks. Column (2), (4) and (3), (6) restrict to trips in the morning interval (7am-1pm) and the evening interval (4-10pm), respectively. Panels A and B correspond to specifications 5 and 6, respectively. "Charges" is a dummy for either low rate or high rate. All specifications include respondent and study cycle fixed effects, and *Post* is an indicator for the experiment period. The mean of the outcome variable in the control group during the experiment is reported for each specification. Standard errors in parentheses are clustered at the respondent level. ${}^{*}p \leq 0.10$, ${}^{**}p \leq 0.05$, ${}^{***}p \leq 0.01$

Table 2: Impact of Departure Time Charges on Daily Total Shadow Rate: Commuting Trips

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Time of Day	AM & PM		AM			PM	
		all	pre peak	post peak	all	pre peak	post peak
Commuter FE	X	X	X	X	X	X	X
Sample:	*Regular Commuters, Home-Work and Work-Home Trips*						
Charges × Post	-7.94***	-3.76**	-3.00*	-0.76	-4.18**	-0.88	-3.30***
	(2.89)	(1.90)	(1.56)	(1.20)	(1.67)	(1.23)	(1.08)
Post	-1.74	-0.74	-1.29	0.55	-1.00	-0.69	-0.31
	(2.65)	(1.74)	(1.30)	(1.36)	(1.61)	(1.18)	(1.06)
Observations	12,115	12,115	12,115	12,115	12,115	12,115	12,115
Control Mean	40.80	23.37	14.27	9.10	17.44	9.15	8.29

Notes: This table reports the impact of departure time charges on daily total shadow rates for regular commuters and commuting trips, separately by time interval. The sample of users and days, and the specifications, are the same as in Table 1, panel B, further restricted to regular commuters and direct trips between their home and work locations (in either direction). Columns (3) and (6) restrict to trips before the peak, i.e. the mid-point of the rate profile. Columns (4) and (7) restrict to trips after the peak. Supplementary Materials Table SM2 reports these results for variable commuters. Standard errors in parentheses are clustered at the respondent level . $^{*}p \leq 0.10$, $^{**}p \leq 0.05$, $^{***}p \leq 0.01$

Table 3: Impact of Area Charges on Daily Outcomes

	(1)	(2)	(3)	(4)	(5)	(6)
Outcome	*Total Shadow Charges Today*			*Number of Trips Today*		
Time of Day	AM & PM	AM	PM	AM & PM	AM	PM
Commuter FE	X	X	X	X	X	X
Panel A. Pooled Treatment						
Treated	-23.34***	-13.43***	-9.90***	0.16*	0.11**	0.04
	(5.53)	(3.38)	(3.30)	(0.08)	(0.05)	(0.05)
Observations	8,827	8,827	8,827	8,827	8,827	8,827
Control Mean	108.43	54.56	53.87	2.52	1.14	1.38
Panel B. Treatment by Week						
Treated in Week 1	-19.69**	-13.64**	-6.05	0.01	0.06	-0.04
	(8.96)	(5.64)	(5.46)	(0.14)	(0.08)	(0.09)
Treated in Week 4	-29.40***	-14.47**	-14.93**	0.34**	0.19**	0.15
	(9.30)	(5.78)	(5.78)	(0.15)	(0.08)	(0.10)
Observations	3,652	3,652	3,652	3,652	3,652	3,652
Control Mean	108.43	54.56	53.87	2.52	1.14	1.38

Notes: This table reports difference-in-difference impacts of the Area or Route treatment on daily total shadow charges and total number of trips. In the first three columns, the outcome is the sum over all trips that day of the trip shadow charge. The shadow charge of a trip is equal to 100 if the trip intersects the respondent's congestion area, and 0 otherwise. In the last three columns, the outcome is the number of trips that day. The sample is all non-holiday weekdays with good quality GPS data, excluding days outside Bangalore. In the post period, all days except trial days are included; the sample in panel B only includes the post period. Column (2) and (5) restrict to the morning interval (7am-2pm), and columns (3) and (6) to the evening interval (2-10pm). The sample is restricted to 243 participants in the Area treatment. The Treated dummy is equal to one in the week when the individual is treated (first and fourth week of the experiment for "early area" and "late area" sub-treatment commuters, respectively) and zero otherwise. All specifications include respondent and study cycle fixed effects. The mean of the outcome variable in the control group in weeks one and four of the experiment is reported for each specification. Standard errors in parentheses are clustered at the respondent level. $^{*}p \leq 0.10$, $^{**}p \leq 0.05$, $^{***}p \leq 0.01$

Table 4: Structural Parameter Estimates

(1)	(2)	(3)	(4)	(5)	(6)
Value of time α (Rs/hr)	Schedule cost early β_E (Rs/hr)	Schedule cost late β_L (Rs/hr)	Logit inner σ (dep. time.)	Logit outer μ (route)	Probability to respond p
1,084.6	294.6	260.8	8.1	33.8	0.42
[792.9, 1,444.8]	[131.4, 467.5]	[155.0, 2,022.5]	[3.9, 26.2]	[25.0, 44.7]	[0.33, 0.60]

Notes: This table reports result from GMM estimation of the discrete choice travel demand model. The estimating equation, moments, data and procedure are described in section 6. The 10th and 90th percentiles from 100 bootstrap runs (with 20 random initial conditions) shown in parentheses.

Table 5: Road Technology: Travel Delay Linear in Traffic Volume

	(1)	(2)	(3)	(4)	(5)
Dependent Variable:	Travel Delay from Google Maps (min/km)			Travel Delay from GPS Data (min/km)	
Sample:	Departure Time		Dates	Departure Time	
Traffic Volume	0.99***	1.07***	0.86***	0.88***	1.18***
	(0.06)	(0.08)	(0.03)	(0.08)	(0.20)
Traffic Volume Exponent γ		0.88***			0.65***
		(0.16)			(0.24)
Constant	1.98***	1.93***	2.14***	2.74***	2.54***
	(0.03)	(0.06)	(0.03)	(0.09)	(0.14)
Observations	1,440	1,440	185	1,440	1,440
Traffic Volume Std.Dev.	0.69	0.69	0.16	0.69	0.69
R^2	0.96	0.96	0.61	0.96	0.96

Notes: This table shows the relationship between travel delay and traffic volume in Bangalore. Traffic volume in all columns is measured using GPS data. For each minute in the day, I compute the total number of trips leaving at that time during weekdays, then smooth this value using a local linear regression with Epanechnikov kernel with 10 minutes bandwidth, and normalize to mean 1. In column 3, for each date I compute the number of trips per capita (using the number of app users that day), and normalize to mean 1.

Travel delay is computed using Google Maps data or using GPS data. Google Maps real-time data is collected over 30 routes in South Bangalore (in both directions), every 20 minutes daily for 207 days, including weekends. Delay in columns 1-2 is the average Google Maps delay over all weekdays and routes, interpolated at the minute level. Delay in column 3 is daily average Google Maps delay over routes and departure times. (The overlap between GPS data and Google Maps data covers 185 days, including weekends.) Delay in columns 4-5 is computed using GPS trips. The sample is all weekday trips more than 2 kilometers long, without stops along the way, and with a trip diameter to total length ratio above 0.6 (the 25th percentile). I regress trip delay on trip length, and then smooth the residual using a local linear regression on departure time, with Epanechnikov kernel with 30 minutes bandwidth.

Columns 1, 3, and 4 report OLS regressions with Newey-West standard errors, with three-hour lag in columns 1 and 4, and 10 day lag in column 3. Columns 2 and 4 report results from nonlinear regressions $Delay_h = \lambda_0 + \lambda_1 Volume_h^{\gamma}$ with HAC standard errors with Newey-West kernel and three-hour lag. $^{*}p \leq 0.10$, $^{**}p \leq 0.05$, $^{***}p \leq 0.01$

Table 6: Policy Simulations: Commuter Welfare Gains from Optimal Peak-Hour Pricing

	(1) Nash	(2) Social Optimum	(3) Improvement	(4) Improvement (% of Nash)
Panel A. Benchmark Model: Travel Time and Commuter Welfare				
Travel Time (minutes)	40.1	38.7	-1.47	3.67%
Travel Time (minutes) above *Free Flow*	16.6	15.1	-1.47	8.86%
Welfare (Rupees)	-802.5	-797.5	4.92	0.61%
Welfare (Rupees) above *Free Flow*	-361.8	-356.9	4.92	1.36%
Panel B. Commuter Welfare (Rupees) With Preference Heterogeneity				
Benchmark model	-633.3	-628.8	4.45	0.70%
α_i, β_{Ei}, β_{Li} log-normal	-633.4	-625.5	7.95	1.26%
α_i, β_{Ei}, β_{Li} log-normal, independent	-578.5	-576.5	1.98	0.34%
α_i, β_{Ei}, β_{Li} binary (high/low)	-634.1	-628.6	5.53	0.87%
Panel C. Commuter Welfare (Rupees) Varying Preferences and Road Technology				
Low Value of Time (4× smaller α)	-202.2	-200.9	1.34	0.66%
High Value of Time (4× larger α)	-2688.2	-2685.2	2.98	0.11%
Cubic Road Technology	-882.9	-841.6	41.27	4.67%

Notes: This table compares the unpriced Nash equilibrium and the social optimum under different assumptions. Columns 3 and 4 report the improvement from the unpriced Nash to the social optimum, in levels and as a fraction of the baseline (Nash) value. Panel A describes the benchmark model with preferences as estimated from the experiment and the calibrated road technology. Travel times are calculated taking individual route length into account, and welfare is average expected utility, assuming charges are transferred lump-sum back to commuters. In rows 2 and 4 travel time and welfare are computed relative to "free-flow" benchmark, where delay is independent of traffic volume. Panel B reports simulation results where agents have heterogeneous preferences. Formally, in row 2, $\alpha_i = \alpha \cdot \xi_i$, $\beta_{Ei} = \beta_E \cdot \xi_i$ and $\beta_{Li} = \beta_L \cdot \xi_i$ where ξ_i is log-normal distributed with mean 1. The ratio of standard deviation to mean is the same for ξ_i as for self-reported wages. In row 3, $\alpha_i = \alpha \cdot \xi_i^A$, $\beta_{Ei} = \beta_E \cdot \xi_i^B$ and $\beta_{Li} = \beta_L \cdot \xi_i^B$ where ξ_i^A and ξ_i^B follow the same distribution as above and are independent. In row 4, the parameters are perfectly correlated and take one of two values with ratio 2, with equal probability: $\alpha \in \{\alpha_L, \alpha_H = 2\alpha_L\}$, etc. Throughout panel B, $\sigma = 80$. The first row establishes that results are similar to the benchmark model with this higher value of σ, for constant α, β_E and β_L. Panel C reports results that correspond to Figure 5.

Table 7: Policy Simulations: Commuter Welfare Gains with Extensive Margin Decision

Trip value δ	Logit trip parameter η	% Traffic Reduction from flat Rs. 200 fee	Implied elasticity at flat Rs. 200 fee	Nash Welfare (Rs.)	Improvement at Social Optimum (% of Nash)
Rs. 1,800	20	-2.9%	-0.10	-802.5	1.43%
Rs. 1,800	100	-3.5%	-0.12	-802.5	1.59%
Rs. 1,500	20	-8.5%	-0.31	-802.5	2.77%
Rs. 1,500	100	-11.4%	-0.42	-802.4	3.14%
Rs. 1,200	20	-28.6%	-1.18	-802.5	5.85%
Rs. 1,200	100	-33.8%	-1.43	-800.2	6.31%

Notes: This table reports commuter welfare and extensive margin elasticities for the model with trip decision, equation (9). The implied elasticity is computed with respect to total travel costs (equal to minus commuter welfare) at the Nash equilibrium. Trip probability is $>$ 99% at the Nash equilibrium for all parameter combinations. In these simulations, average congestion charges at the social optimum vary between Rs. 240 and Rs. 264.

A Online Appendix

A.1 Non-parametric Departure Time Choice Model

The model in section 3.1 makes several parametric assumptions. Here I describe a more general setup and show under what conditions is reduces to the main model. Consider the time t between 0 (early morning) and 1 (some time at work). The commuter leaves home at t_H and arrives at work at $t_W > t_H$. The commuter has a value for spending time t on any of three activities:

- The instantaneous value of time at home is $h(t)$.
- The instantaneous value of time at work is $w(t)$.
- The cost of time spent traveling is $-C(t_W - t_H)$.

The commuter chooses t_H, while $T = t_W - t_H$ is a random variable with compact support and with cumulative distribution function $F(T; t_H)$ that depends on t_H. (For consistency, there is a latest possible departure time $t_H^{\max}$ that ensures arrival before $t_W = 1$ with probability 1.) Travel demand is inelastic, namely the commuter always makes the trip.

The utility of leaving from home at t_H and arriving at work at t_W is:

$$U(t_H, t_W) = \int_0^{t_H} h(t)dt - C(t_W - t_H) + \int_{t_W}^1 w(t)dt \tag{10}$$

Taking expectations over travel time, expected utility is given by:

$$EU(t_H) = \int_0^{t_H} h(t)dt - \int_0^1 C(T)dF(T; t_H) + \int_0^1 w(t)dF(t - t_H; t_H) \tag{11}$$

The following special case is equivalent to the main model for a fixed ideal arrival time and no preference shocks:

Example 1. Assume $h(t) = h > 0$, $C(T) = c \cdot T$, while $w(t) = w^H$ for $t > t^*$ and $w(t) = w^L$ for $t \leq t^*$, with $w^L < w^H$. The commuter has a constant value at home, linear disutility of travel, and earns a small wage before the clock-in time t^*, and a larger wage afterwards. Rearranging (10):

$$U(t_H, t_W) = h \cdot t_H - c(t_W - t_H) + w^L |t_W - t^*|_- - w^H |t_W - t^*|_+$$

The model is isomorphic when subtracting h from the flow utilities and from the flow commuting cost, giving:

$$U(t_H, t_W) = -(c + h)(t_W - t_H) - (h - w^L)|t_W - t^*|_- - (w^H - h)|t_W - t^*|_+$$

This setup is equivalent to the main model with $\alpha = c + h$, $\beta_E = h - w^L$ and $\beta_L = w^H - h$. The canonical assumptions in the original parametric model formulation translate to the following restrictions in this environment:

- $\alpha, \beta_E, \beta_L > 0$ iff spending normal time at work is more valuable than being at home, and spending time at home is more valuable than early work time, or $w^H > h > w^L$,
- $\beta_L > \alpha$ always holds; commuters prefer to arrive early rather than driving more.

A.2 Two Models of Traffic Congestion in Urban Road Networks

I introduce two stylized models of traffic congestion in an urban road network. The key result in both cases is that a marginal increase in traffic volume starting at zero has an approximately linear effect on travel delay. This result is in contrast with the bottleneck model and traditional highway specifications (the Bureau of Public Roads (BPR) functions (Small et al., 2007)), where the marginal effect at zero is zero.

Model 1. A Sequence of un-synchronized Traffic Lights. The road network is a long straight line with many equally spaced, unsynchronized traffic lights. Each traffic light alternates between "red" and "green" for equal durations of 0.5 time units. When "green," the traffic light is a bottleneck with capacity $c > 0$, while during "red" the capacity drops to zero. The space between traffic lights is sufficiently long so that queues never spill back to the next traffic light. All commuters travel from one end of the line to the other, and they start in periodic bursts of constant flow c that last Δt and are one unit of time apart, with $\Delta t \ll 1$.

There are two main cases to consider. If the burst of traffic arrives during the red cycle, it immediately forms a queue of length $c\Delta t$. The commuters wait on average $1/4$ until the traffic light turns green, and then $\Delta t/2$ on average to go through the bottleneck after the light turns green (because the traffic light has capacity c it takes Δt until the entire queue dissipates, so $\Delta t/2$ for the average commuter). Second, if the burst of traffic arrives at the traffic light during the "green" cycle, it passes through without delay.[72]

Hence, average delay for traffic volume $c\Delta t$ is equal to $1/8 + \Delta t/4$. As traffic lights are assumed to be *numerous* and *unsynchronized*, travel delay for each commuter is given approximately by a multiple of the travel delay expectation.

Overall, average travel delay is linear in traffic volume for traffic volume close to zero.

Model 2. Cross Traffic with un-signaled Intersections.[73] The traffic network consists of two perpendicular, single-lane and one-way streets AB and CD of length 1. Each car has length $\ell \ll 1$, it has the width of the entire lane, and normally travels with constant speed $1/\Delta t$ (hence, it takes Δt to go through the entire segment without delay). The two streets intersect. When a car C arrives at the intersection, if there is no cross-traffic it goes through without delay. If a car C' from the other street is going through the intersection, car C has to wait until the tail of C' exits the intersection. It takes car C' exactly $\ell\Delta t$ time to pass through. Hence, the expected delay for C if it intersects with C' is $\ell\Delta t/2$. We ignore cases where cars going in the same direction bump into each other. (The model can easily be extended to include this.)

Assuming k cars per unit of time traveling from A to B and k cars traveling from C to D, the expected travel time for each car is $\Delta t + (1 - k\ell) \cdot 0 + k\ell \cdot \ell\Delta t/2$, which is linear in traffic volume k.

Once again, average travel delay is linear in traffic volume for traffic volume close to zero.[74]

[72]An additional case arises if the burst arrives at the green light less than Δt before it turns red. This means part of the queue goes through and the rest waits a full red cycle. I assume that the entire burst goes through if the light is initially green. This leads to a second order term in average delay in terms of Δt.

[73]I thank Andrew Foster for the idea for this model.

[74]Travel delay is either exactly Δt or uniformly distributed between Δt and $\Delta t + \ell\Delta t/2$. In a larger network made up of $2M$ streets that form a square grid with M^2 intersections, travel time will be closer to its expectation.

A.3 Travel Demand Estimation

A.3.1 Nested Logit Choice Probabilities

In the model with two routes from section 6, the probability to choose a given departure time and route can be decomposed as $\Pr(h, r \mid h_{it}^A) = \Pr(h \mid r, h_{it}^A) \Pr(r \mid h_{it}^A)$. Denote by $V_{it}(h, r, h_{it}^A)$ the constant part of utility in (8) (i.e. without the utility shock $\varepsilon_{it}(h, r)$), then

$$\Pr(h \mid r, h_{it}^A) = \frac{\exp\left(\frac{1}{\sigma_i} V_{it}(h, r, h_{it}^A)\right)}{\sum_{h'} \exp\left(\frac{1}{\sigma_i} V_{it}(h', r, h_{it}^A)\right)} \tag{12}$$

where $\sigma_i = \frac{KM_i}{\overline{KM}}\sigma$, KM_i is the average (pre-experiment) trip length for commuter i, and $\overline{KM}$ is the sample average of KM_i. (The dependence on congestion charges $p(h, r)$ and on the travel time profile is implicit throughout.) The route choice probability is

$$\Pr(r \mid h_{it}^A) = \frac{\exp\left(\frac{1}{\mu} V_{it}(r, h_{it}^A)\right)}{\exp\left(\frac{1}{\mu} V_{it}(0, h_{it}^A)\right) + \exp\left(\frac{1}{\mu} V_{it}(1, h_{it}^A)\right)} \tag{13}$$

where $V_{it}(r, h_{it}^A) = \sigma_i \log\left(\sum_h \exp\left(\frac{1}{\sigma_i} V_{it}(h, r, h_{it}^A)\right)\right)$ is the "logsum" or "inclusive value" term, i.e. the expected utility assuming i chooses route r.

Commuters not in the area treatment only choose departure time, according to multinomial logit. Their choice probabilities are given by

$$\Pr(h \mid h_{it}^A) = \frac{\exp\left(\frac{1}{\sigma_i} V_{it}(h, h_{it}^A)\right)}{\sum_{h'} \exp\left(\frac{1}{\sigma_i} V_{it}(h', h_{it}^A)\right)}$$

where $V_{it}(h, h_{it}^A)$ is given by the constant part of utility in (1).

Ideal arrival time on a given day and the "response" type of a commuter are not observed. Hence, overall choice probabilities are obtained by first integrating over the (individual-specific) distribution of ideal arrival times $h_{it}^A \sim \mathcal{H}_i^A$. Second, choice probabilities are different according to the response type of the commuter. For the latter, the observed choice probabilities in the presence of monetary charges $p(h, r) \neq 0$ is an average between choice probabilities computed including the monetary charges in the utility function, with weight p (the share of commuters who "respond"), and choice probabilities computed without monetary charges, with weight $1 - p$.

A.3.2 GMM Moments That Exploit Experimental Variation

This section describes the moments used in the GMM estimation. Optimal departure time and route choices (random variables) are denoted by $h_{it}^*(\theta, p_{it})$ and $r_{it}^*(\theta, p_{it})$, where $\theta = \{\alpha, \beta_E, \beta_L, \sigma, \mu\}$ is the vector of preference parameters, and $(p_{it}(h, r))_{h,r}$ is the profile of congestion charges. The dependence on the travel time profile and the ideal arrival time h_{it}^A is implicit. I denote by $p_{it}^D(h)$ and $p_{it}^R(r)$ the congestion charges in the experiment. Throughout, $\sim$ indicates quantities in the data.

Departure Time "Market Share" Moments. The first 31 moments match the difference in difference in departure time market shares, between the departure time treatment and control groups, during the experiment relative to before (see Online Appendix Figure A6 panel C). Formally, define 10-minute departure time bins H^k between -155 and $+155$ minutes relative to the rate profile peak, with midpoint h_k between -2.5 and $+2.5$ hours. Denote $P_{ik}^{DT}(\theta, p_{it}) = \Pr(h_{it}^*(\theta, p_{it}) = h_k)$. (Recall, departure time is discrete in the model.)

In the data, define $\widetilde{P}_{ik}^{DT}(pre)$ and $\widetilde{P}_{ik}^{DT}(post)$ the fractions of trips starting in bin H^k for i in pre- and post- periods, respectively. The k-th moment is ($k \in \{1, \dots, 31\}$):

$$\begin{aligned} g_i^k(\theta, p) = &((\widetilde{P}_{ik}^{DT}(post) - \widetilde{P}_{ik}^{DT}(pre))T_i^{DT} - (\widetilde{P}_{ik}^{DT}(post) - \widetilde{P}_{ik}^{DT}(pre))(1 - T_i^{DT})) \\ &- p \cdot ((P_{ik}^{DT}(p_{it}^D) - P_{ik}^{DT}(0))T_i^{DT} - (\underbrace{P_{ik}^{DT}(0) - P_{ik}^{DT}(0)}_{=0})(1 - T_i^{DT})) \end{aligned}$$

where T_i^{DT} is an indicator for departure time charges (low or high rate). Note: the heterogeneity parameter p enters by attenuating the model term.

Departure Time Heterogeneity Moments. The next two moments target the variance of how much average shadow rates change for an individual, before the experiment versus during the experiment (see Online Appendix Figure A6 panels D-E). Here we focus on "early-morning" trips, namely trips with departure time before the mid-point of the congestion charge profile, $h_{it}^* < 0$. Denote by h_{it}^{E*} this conditional variable. For these moments, it is important to take sampling variation into account when simulating the model. Denote by N_i^0 and N_i^1 the number of days in the data in the pre and post periods when i travels in the early morning. Denote by $ch\,(h)$ the shadow rate of departure time h. Model-predicted individual-level change in shadow rate given congestion charges p_{it} is:

$$ch_i^{DT}(\theta, p_{it}) = \frac{1}{N_i^1} \sum_{t=N_i^0+1}^{N_i^0+N_i^1} ch(h_{it}^{E*}(\theta, p_{it})) - \frac{1}{N_i^0} \sum_{t=1}^{N_i^0} ch(h_{it}^{E*}(\theta, 0))$$

Choices on different days are assumed to be independent (including the draw of ideal departure time h_{it}^A). Denote the individual effect in the data by $\widetilde{ch}_i^{DT}$.

The two departure time heterogeneity moments match the variance (over commuters and within commuter) of ch_i^{DT} in the treatment and control groups:

$$\begin{aligned} g_i^{32}(\theta, p) &= \left(var\left(p \cdot ch_i^{DT}(\theta, p_{it}^D) + (1-p) \cdot ch_i^{DT}(\theta, 0)\right) - \widehat{var}\left(\widetilde{ch}_i^{DT}\right)\right) \cdot T_i^{DT} \\ g_i^{33}(\theta, p) &= \left(var\left(ch_i^{DT}(\theta, 0)\right) - \widehat{var}\left(\widetilde{ch}_i^{DT}\right)\right) \cdot \left(1 - T_i^{DT}\right) \end{aligned}$$

Note, moment 33 essentially ensures that the model can replicate the sampling variation in the data.

Area or Route "Market Share" Moments. The next two moments match route choice market shares, namely the probability to intersect the congestion area when treated and when not treated, for commuters in the area congestion charge treatment.

Formally, define $P_i^A(\theta, p_{it}) = \Pr(r_{it}^*(\theta, p_{it}) = 0)$ the probability to take the short route (intersect the congestion area), given charges $(p_{it}(h, r))_{h,r}$. In the data, define $\widetilde{P}_i^A(treat)$ and $\widetilde{P}_i^A(control)$ the fraction of days (mornings) when the commuter intersects the congestion area, when treated and when not treated, respectively. The area moments are:

$$g_i^{34}(\theta,p) = p\cdot P_i^A(\theta,p_{it}^R) + (1-p)\cdot P_i^A(\theta,0) - \widetilde{P}_i^A(treat)$$
$$g_i^{35}(\theta,p) = P_i^A(\theta,0) - \widetilde{P}_i^A(control)$$

for commuters i in the area treatment, and zero otherwise.

Area or Route Heterogeneity Moments. The area heterogeneity moments target the distribution of the fraction of days when an individual intersects the area (see Online Appendix Figure A6 panels F-G). Once again, it is important to take sampling variation into account. Define M_i^1 and M_i^0 the number of days observed in the data when i is treated and not treated, respectively.

Define the (random) sample frequency of intersecting the area in control and treatment as

$$ch_i^{A,0}(\theta,p_{it}) = \frac{1}{M_i^0}\sum_{t=1}^{M_i^0} r_{it}^*(\theta,p_{it}) \quad \text{and} \quad ch_i^{A,1}(\theta,p_{it}) = \frac{1}{M_i^1}\sum_{t=M_i^0+1}^{M_i^0+M_i^1} r_{it}^*(\theta,p_{it})$$

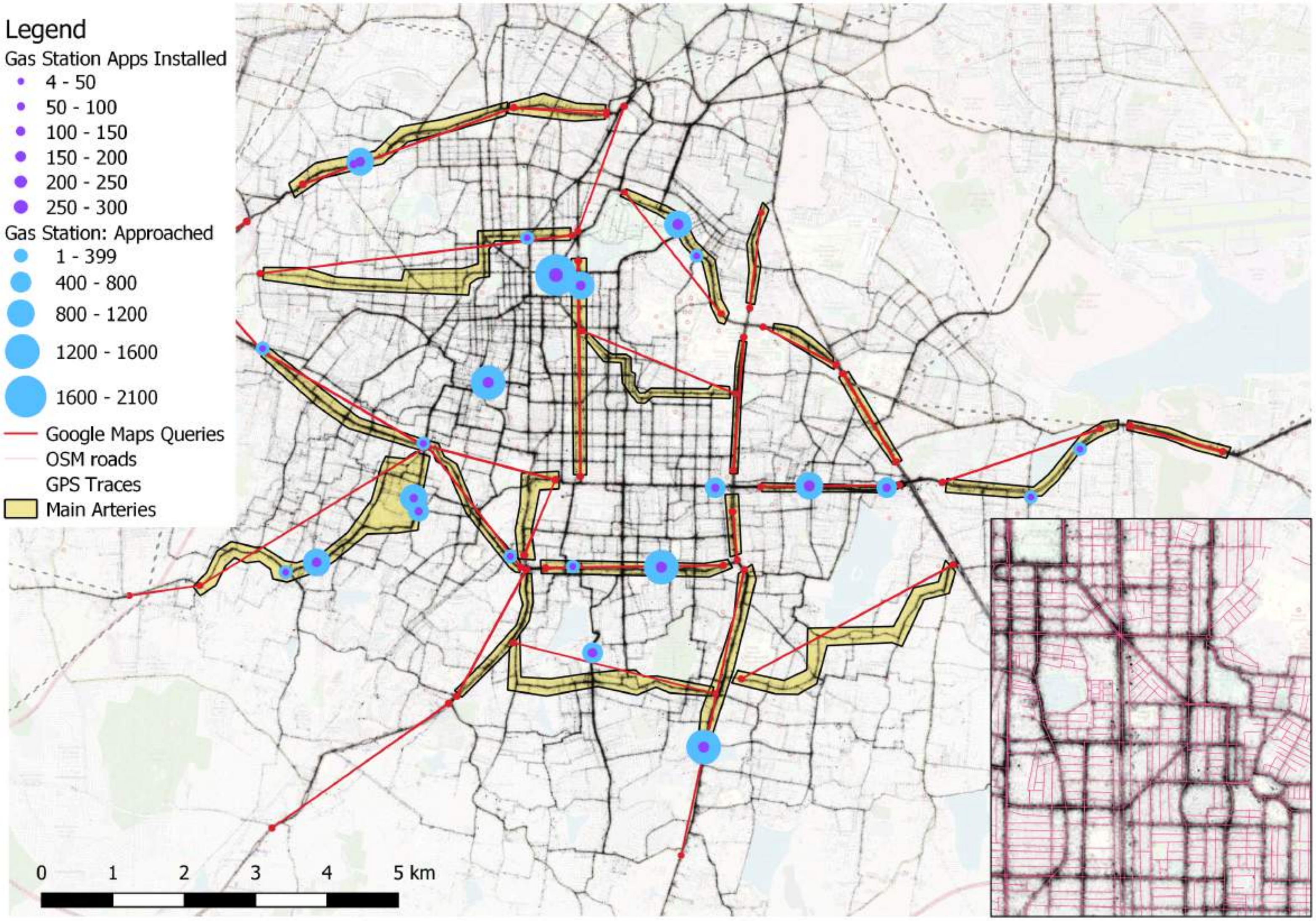
In the data, denote the corresponding quantities by $\widetilde{ch}_i^{A,0}$ and $\widetilde{ch}_i^{A,1}$, respectively. We are interested in the distribution of these variables. The four area heterogeneity moments match the probability that these variables are the middle or top third of the unit interval (the moment for the bottom third is omitted because it is colinear with the others), when treated and not treated.

$$\begin{aligned}
g_i^{36}(\theta,p) &= p\Pr\left(ch_i^{A,1}(\theta,p_{it}^R)\in[1/3,2/3)\right) + (1-p)\Pr\left(ch_i^{A,1}(\theta,0)\in[1/3,2/3)\right) \\
&\quad - \mathbb{1}\left(\widetilde{ch}_i^{A,1}\in[1/3,2/3)\right) \\
g_i^{37}(\theta,p) &= p\Pr\left(ch_i^{A,1}(\theta,p_{it}^R)\in[2/3,1]\right) + (1-p)\Pr\left(ch_i^{A,1}(\theta,0)\in[2/3,1]\right) \\
&\quad - \mathbb{1}\left(\widetilde{ch}_i^{A,1}\in[2/3,1]\right) \\
g_i^{38}(\theta,p) &= \Pr\left(ch_i^{A,0}(\theta,0)\in[1/3,2/3)\right) - \mathbb{1}\left(\widetilde{ch}_i^{A,0}\in[1/3,2/3)\right) \\
g_i^{39}(\theta,p) &= \Pr\left(ch_i^{A,0}(\theta,0)\in[2/3,1]\right) - \mathbb{1}\left(\widetilde{ch}_i^{A,0}\in[2/3,1]\right)
\end{aligned}$$

for commuters i in the area treatment, and zero otherwise. The first moment matches the fraction of commuters who take the detour route between $1/3$ and $2/3$ of the time when experiencing area charges. The second moment repeats this for the interval $[2/3,1]$. Note the parameter p attenuates the model predicted responsiveness. The last two moments repeat this for the control group.

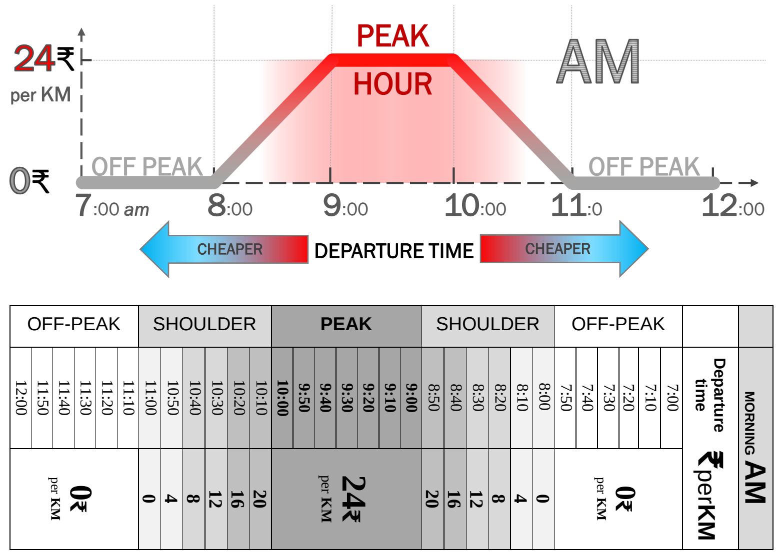
B Online Appendix: Figures

Figure A1: Study Area and Recruitment Locations



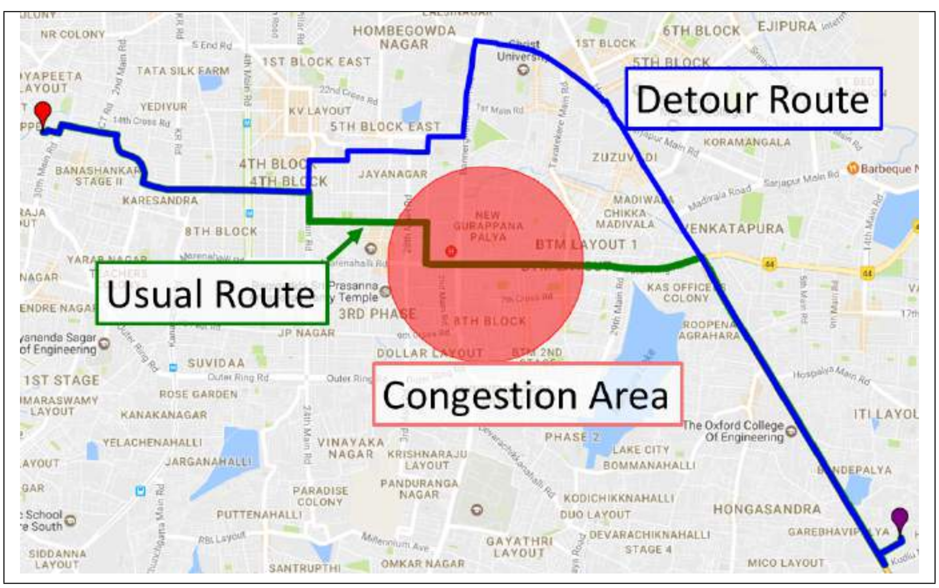
Notes. This figure shows the study area in South Bangalore. Blue and purple discs represent the randomly chosen gas stations where study participants were recruited (the blue diameter indicates the number of commuters approached, and the purple diameter the number of apps successfully installed). The black points in the background represent all the GPS data collected during the study. The red straight lines connect origin and destination pairs corresponding to the Google Maps queries used in Figure 1 (queries are done in both directions). (The dashed gray lines correspond to queries in other parts of the city, which are not used here.) The light yellow areas correspond to major arteries used in the road technology analysis in Appendix Figure A11.

Figure A2: Departure Time Congestion Charge (AM) Rate Profile Card Example



Notes: This figure shows an example of Rate Profile card that study participants in the departure time charge sub-treatments received. The cards for different participants differed in the value of the *peak rate* (Rs. 12/Km and Rs. 24/Km in the Low Rate and High Rate sub-treatments, respectively), and in the starting time of the profile (between 8 am and 9 am for AM, and between 5 pm and 6 pm for PM).

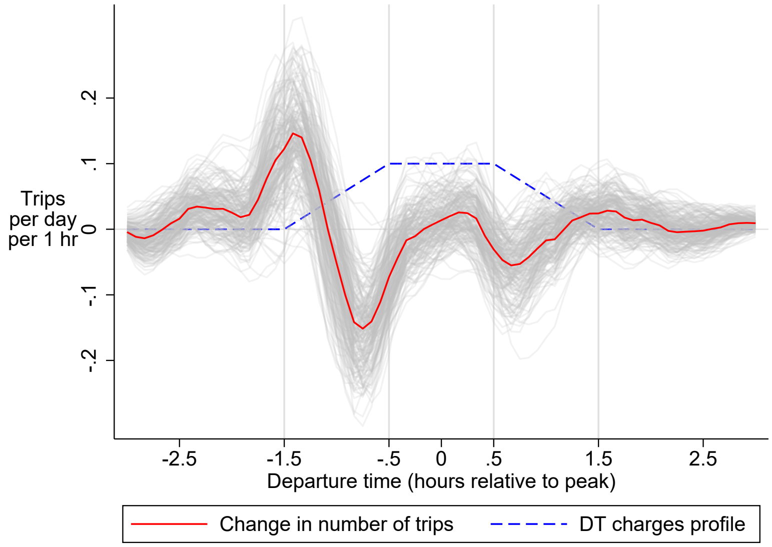
Figure A3: Area Congestion Charge Map Example



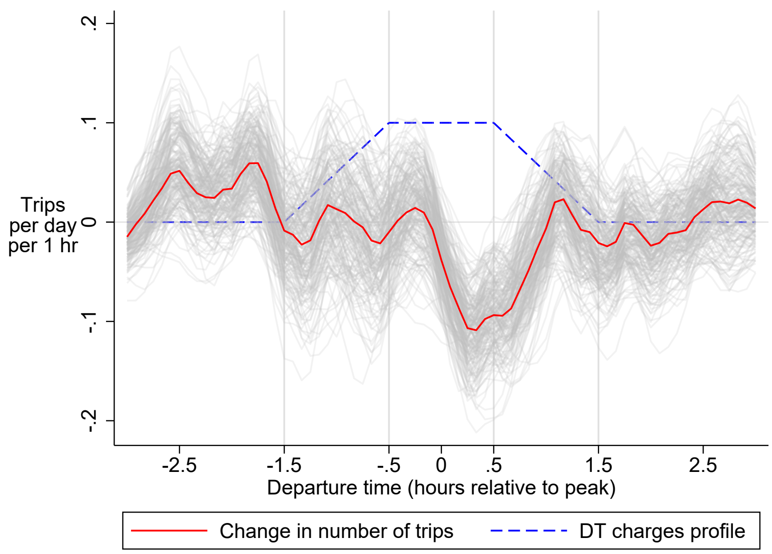
Notes: This figure shows an example of congestion area map that participants received. Congestion areas were selected as follows. Given a regular route between home and work (in green), several "candidate" areas were selected along the route, with a radius of 250m, 500m or 1000m. For each candidate area, the quickest detour route that avoids the congestion area was found using a custom algorithm using multiple Google Maps API queries. Candidate areas with detours between 3 and 14 minutes longer were manually reviewed, and the final area was randomly selected from within this group.

Figure A4: Impact of Departure Time Charges on Departure Times (Commuting Trips)

Panel (A) Morning Peak

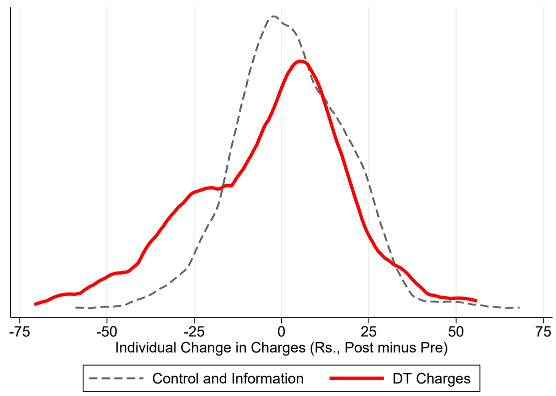


Panel (B) Evening Peak

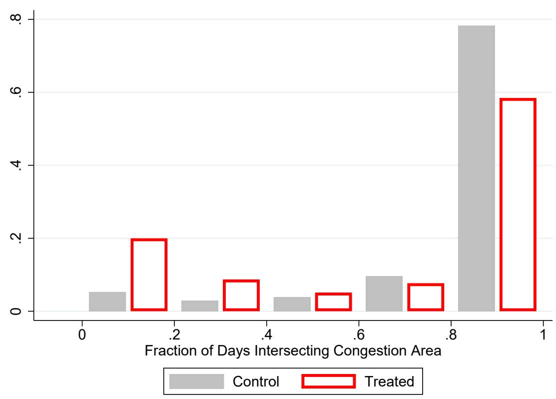


Notes: Version of Figure 2 restricting to regular commuters and home-work and work-home commuting trips.

Figure A5: Treatment Heterogeneity for Departure Time and Area Treatments



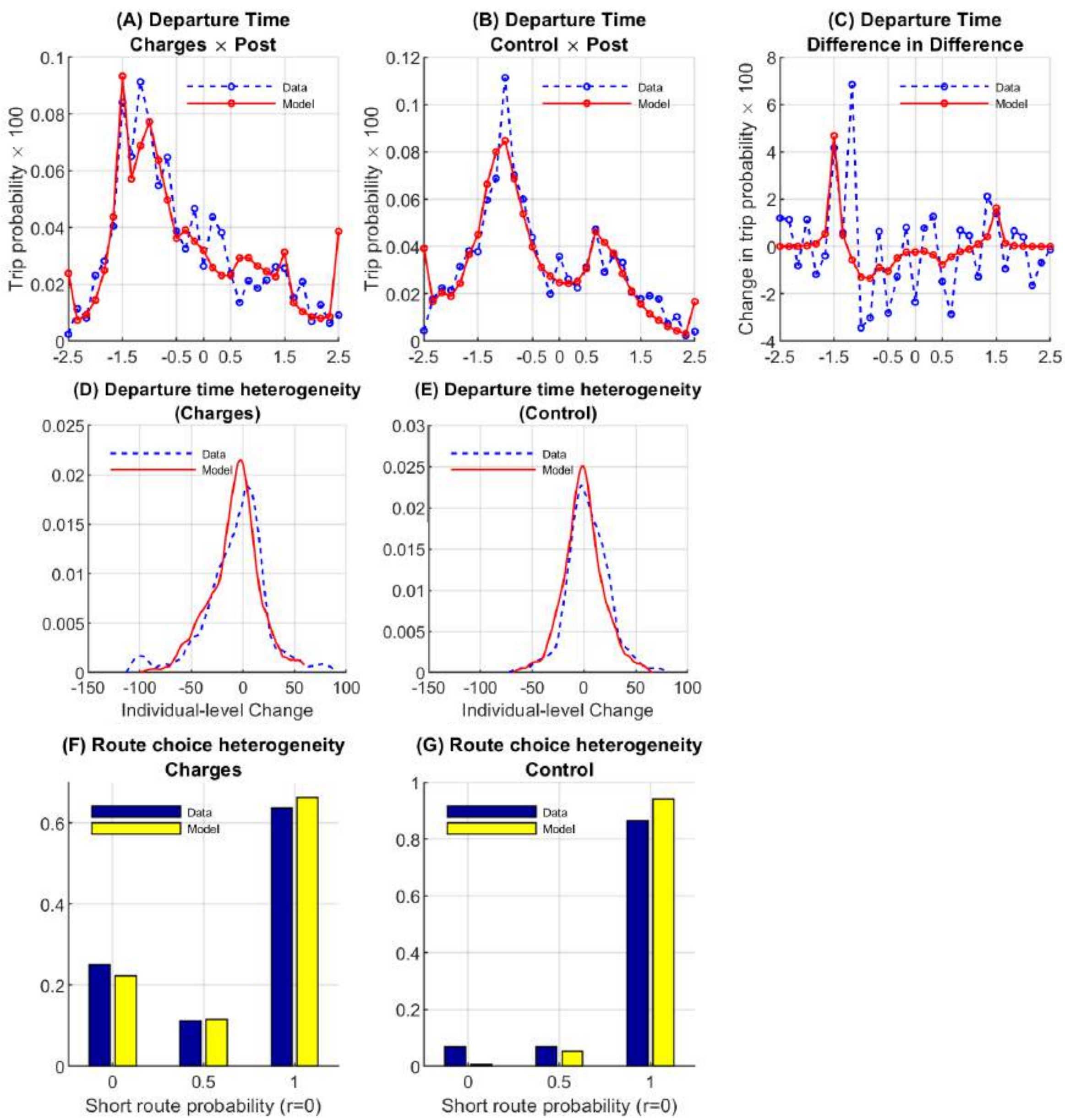
Panel (A) Individual-Level Change in Shadow Trip Rate, by Departure Time Treatment



Panel (B) Frequency of Avoiding the Congestion Area, by Area Treatment Status

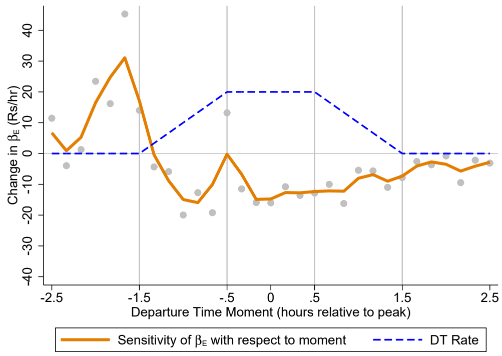
Notes. This figure shows the distribution of individual changes in shadow rates in the departure time treatment and control (panel A) and the distribution of individual frequency of intersecting the area when treated and not treated in the area treatment (panel B). In panel A, the sample is all regular commuters and all trips between the morning profile peak and 2 hours earlier. The graph plots the pre-post change in shadow trip rates for each participant, separately for participants with charges (low or high rate) versus those without charges. The sample for panel B is all days with trips in the morning between home and work for regular commuters. The graph shows the histogram of the fraction of days when a participant intersects the congestion area, separately for (area) treated and not treated participants.

Figure A6: Structural Model Fit

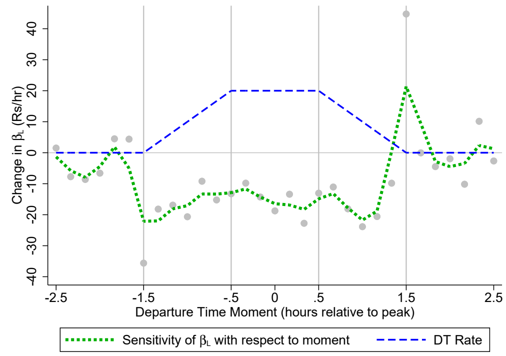


Notes: This figure shows the in- an out-of-sample fit of the estimated structural model. Panels A and B show the densities of departure time in the treated and control groups, during the experiment (Post); these two moments were not directly targeted in the estimation. Panel C plots the 31 departure time moments that correspond to the difference in difference (treated vs. control, during vs. before). Panels D and E show the distributions of individual effects in the departure time treatments (changes in shadow charges between pre- and post-). Panels F and G show the distributions of individual effects in the area treatment (fraction of days intersecting the congestion area when treated and not treated). For the last 4 panels, the model prediction is computed for the same sample size as in the data (i.e. taking sampling variation into account).

Figure A7: Structural Model Diagnostics



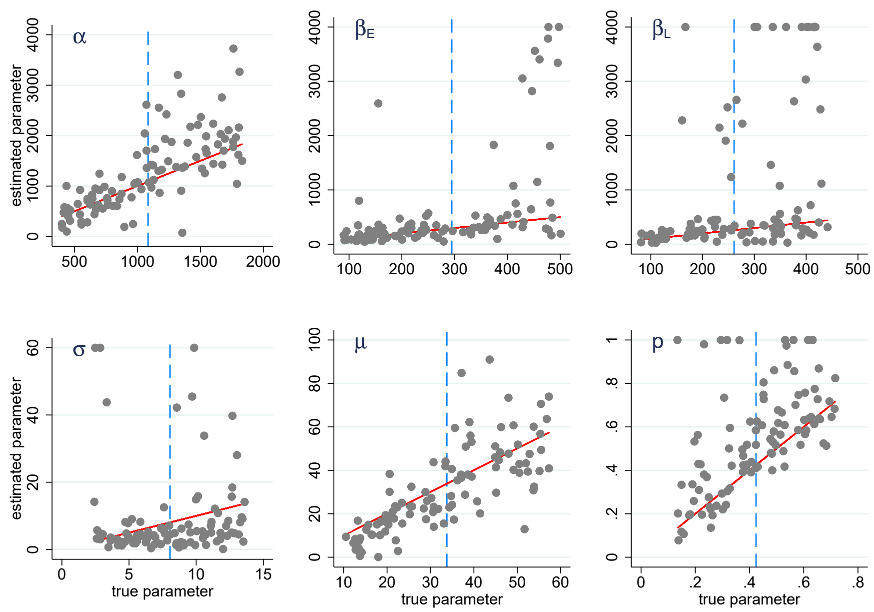
Panel (A) Sensitivity of **Early** Schedule Cost β_E with Respect to Departure Time Moments



Panel (B) Sensitivity of **Late** Schedule Cost β_L with Respect to Departure Time Moments

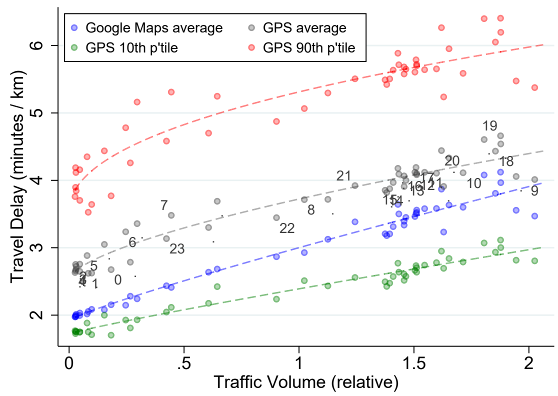
Notes: This figure plots the scaled sensitivity measure from Andrews et al. (2017), quantifying the change in the estimated early and late schedule cost parameters $\widehat{\beta}_E$ and $\widehat{\beta}_L$ given by one standard deviation change in each of the 31 departure time moments. (See Online Appendix Table A10 for the full definition.)

Figure A8: Structural Model Numerical Identification Check

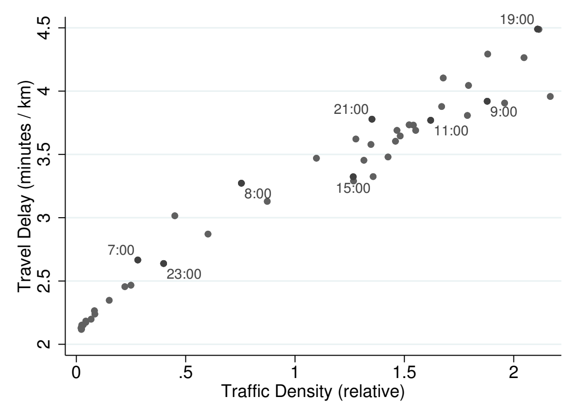


Notes: This figure shows that the estimation method is generally successful in recovering underlying model parameters. To construct it, I select 100 sets of randomly chosen "true" parameters. For each set of parameters, and each agent i, I first invert the $\mathcal{H}_i^A$ distribution from pre-experiment (real) data. I then simulate choice data, using the same number of agents and number of days per agent. I then estimate the model on the simulated data using 20 random initial parameters (that do not depend on true parameters). Each graph shows the estimated parameter on the Y axis, and the true parameter on the X axis. Outlier values are censored to $4,000$ for β_E and β_L and 60 for σ. The red diagonal line is identity, and the blue vertical line indicates the parameter from Table 4. See also Appendix Table A9.

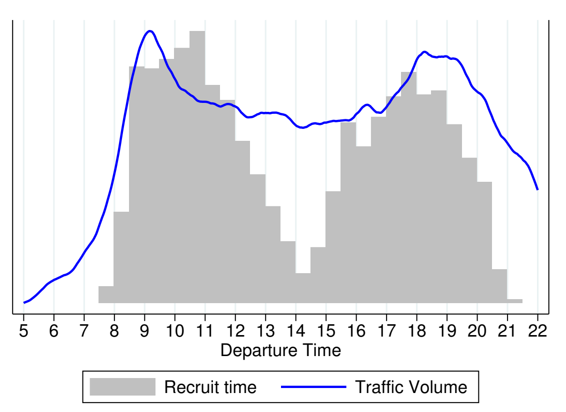
Figure A9: Road Technology Estimation Robustness Checks



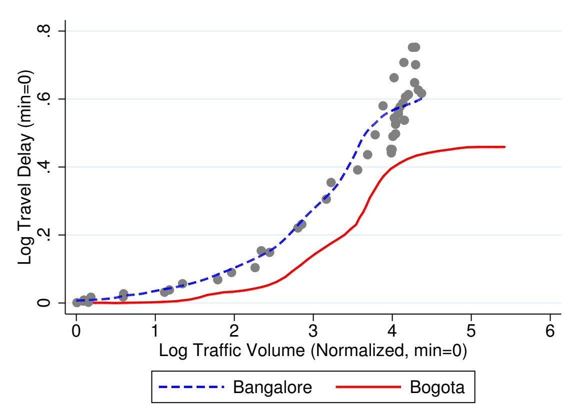
Panel (A) Travel Delay from GPS Data and Google Maps



Panel (B) Travel Delay and Traffic Density



Panel (C) Recruitment Time and Trip Time Distributions



Panel (D) Comparison with (Akbar and Duranton, 2017)

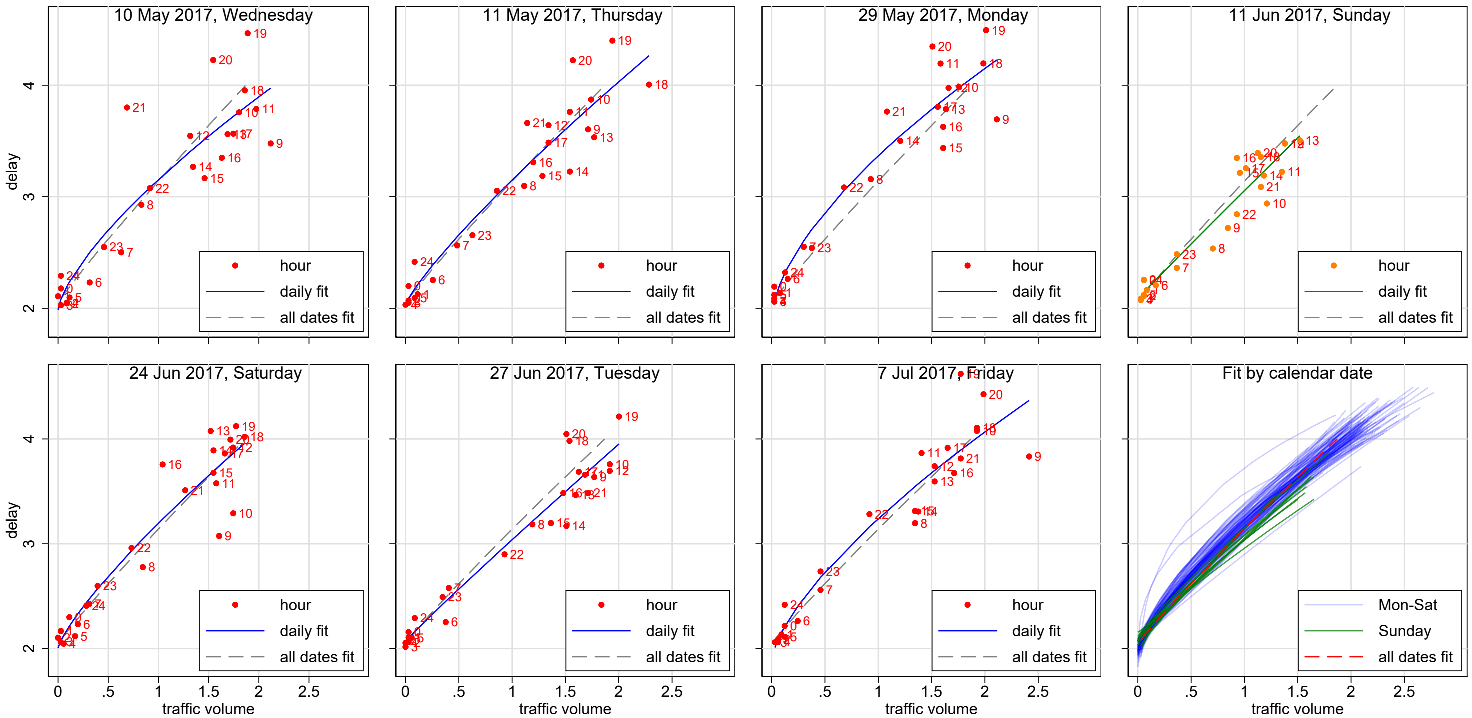
Notes: Panel A uses travel delay from GPS trips to replicate Figure 3. The notes for Table 5 describe the sample and variable construction for traffic volume and average delay from GPS data. In addition, for each departure time, I compute the 10th and 90th percentile of delay for trips from the GPS data with nearby departure times, weighted with an Epanechnikov kernel with 30 minutes bandwidth. The blue line and dots use Google Maps average delay. The gray line and dots use the average delay using GPS trips from smartphone app users (for all trips starting around that departure time). The 10th and 90th percentiles of the travel delay around a given departure time are shown in green and red, respectively. The lines are non-linear regression fits as in column (2) in Table 5.

Panel B replicates Figure 3 with traffic density instead of volume of departures on the X axis. Road density at a certain time is defined as the number of ongoing trips.

Panel C plots the distribution of participant recruitment times (histogram in solid gray) and the distribution of trip departure times (kernel density plot in solid blue line). Both Y axes start at zero.

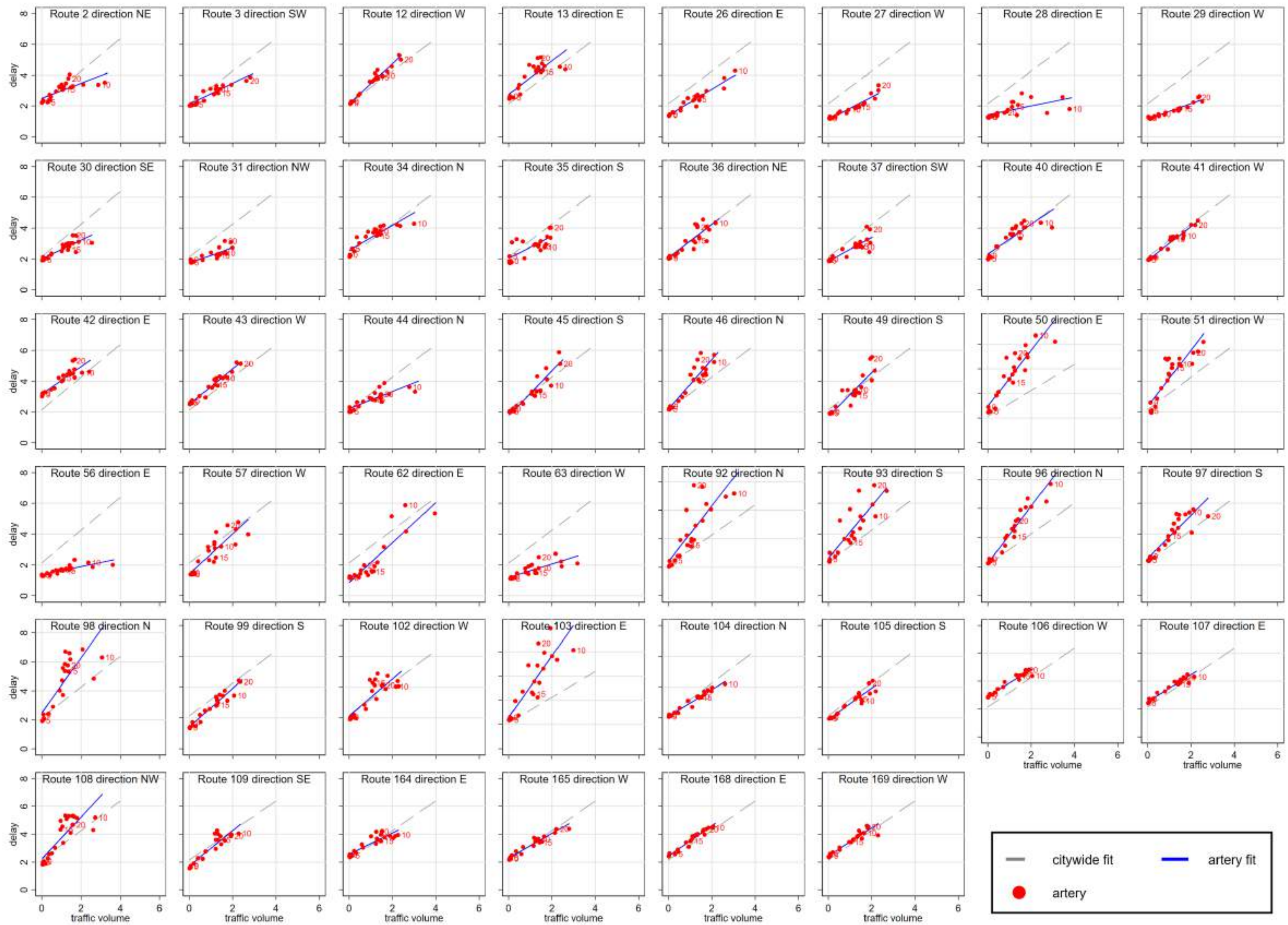
Panel D compares log-log road technology estimates from this paper (gray dots, dashed blue line) with those from Akbar and Duranton (2017) in Bogotá (red solid line). (Their estimate is computed from Figure 4 panel C.)

Figure A10: Road Technology at the Daily Level



Notes. These graphs replicate Figure 3 panel A by date. The first 7 panels show the relationship between hourly GPS traffic volume and Google Maps travel delay for 7 randomly chosen calendar dates (one for each day of the week). The last panel overlays the predicted fit for all calendar dates in the sample. The sample is calendar dates with above-median number of GPS trips (at least 571 trips per day). The measure of volume on day d at hour h is the normalized number of trips: $V_{dh} = T_{dh}/\overline{T} \cdot \overline{C}/C_d$, where T_{dh} is the number of trips on d at h, with average $\overline{T}$, C_d is the number of commuters with at least a trip on day d, with average $\overline{C}$. Google Maps delay is averaged at the hourly level on each calendar date. Each fit is a power fit as in column 2 in Table 5. In a regression of delay on date and hour fixed effects, the coefficient on traffic volume V_{dh} is 0.31 and highly statistically significant. The fixed effects explain 95% of the variation, and traffic volume explains 12% of the remaining variation.

Figure A11: Road Technology on Major Arteries



Notes. These graphs replicate Figure 3 panel A for major arteries depicted in Appendix Figure A1, separately by direction. The Y axis is average Google Maps travel delay for that road segment. To compute traffic volume, I define a buffer area around each artery. I then count the number of GPS trips that travel along the artery in each direction for each time of day, excluding short trips that intersect the artery for less than 200 meters.

C Online Appendix: Tables

Table A1: Descriptive Statistics about Travel Behavior

	Median	Mean	Std. Dev.	10 Perc.	90 Perc.	Obs.
Panel A. Trip Characteristics						
	Median	Mean	Std. Dev.	10 Perc.	90 Perc.	Obs.
Total Number of Trips						51,473
Number of Trips per Day	2.86	3.15	[1.16]	1.90	4.86	497
Median trip duration (minutes)	24.50	27.43	[12.83]	15.20	42.60	497
Median trip length (Km.)	5.91	7.17	[4.67]	2.92	13.36	497
Panel B. Commute Destination Variability						
	Median	Mean	Std. Dev.	10 Perc.	90 Perc.	Obs.
Regular Commuter	1.00	0.76	[0.43]	0.00	1.00	497
Frac. trips Home-Work, Work-Home	0.38	0.39	[0.21]	0.13	0.67	378
Frac. of trips Work-Work	0.03	0.06	[0.08]	0.00	0.15	378
Frac. of days present at Work	0.92	0.86	[0.16]	0.60	1.00	378
Panel C. Departure Time Variability						
(Standard Deviation of the Departure Time in hours)						
	Median	Mean	Std. Dev.	10 Perc.	90 Perc.	Obs.
First Trip (AM)	1.28	1.24	[0.50]	0.54	1.83	496
Last Trip (PM)	1.72	1.71	[0.50]	1.04	2.36	497
First Home to Work Trip (AM)	0.48	0.62	[0.52]	0.15	1.28	332
Last Work to Home Trip (PM)	0.80	0.95	[0.63]	0.28	1.78	322

Notes: This table reports summary travel behavior statistics for the experimental sample of 497 commuters. See section 4.1 for the definition of home and work locations and of regular commuter. In panel C, I compute the within-commuter variation in departure times for different classes of trips.

Table A2: Experimental Design

Panel A. Treatment Strata

Strata			*Departure Time Sub-treatment*			
Area Eligibility	Car or Moto	Daily KM	High Rate	Low Rate	Info	Control
Eligible	Car	Low	3/8	1/8	2/8	2/8
Eligible	Car	High	1/8	3/8	2/8	2/8
Eligible	Moto	Low	3/8	1/8	2/8	2/8
Eligible	Moto	High	1/8	3/8	2/8	2/8
Ineligible	Car	Low	1/12	3/12	4/12	4/12
Ineligible	Car	High	3/12	1/12	4/12	4/12
Ineligible	Moto	Low	1/12	3/12	4/12	4/12
Ineligible	Moto	High	3/12	1/12	4/12	4/12

Panel B. Treatment Timing

			Treatment by Week in Experiment			
Area Eligibility	*Dep. Time Timing*	*Dep. Time Sub-Treatment*	1	2	3	4
Eligible	Late	High rate	A	H	H	H
		Low rate	A	L	L	L
		Information	A	I	I	I
		Control	A	C	C	C
Eligible	Early	High rate	H	H	H	A
		Low rate	L	L	L	A
		Information	I	I	I	A
		Control	C	C	C	A
Ineligible	Late	High rate	I	H	H	H
		Low rate	I	L	L	L
		Information	I	I	I	I
		Control	C	C	C	C
Ineligible	Early	High rate	H	H	H	I
		Low rate	L	L	L	I
		Information	I	I	I	I
		Control	C	C	C	C

Notes. There were eight strata in the experiment, all combinations of participants eligible or ineligible for the area charge, car or non-car (motorcycle or scooter) users, and participants with high or low daily travel distance in the baseline period. Departure time sub-treatment probabilities are given in panel A. There are eight area sub-treatments: all combinations of high/low charges, short/long detour, and early/late. All have equal probabilities. Sub-treatment are cross-randomized (see Supplementary Materials section SM.4). Treatment timing is presented in panel B. The letter A corresponds to the area treatment. The letters H, L, I and C respectively correspond to high-rate, low-rate, information and control in the departure time treatment.

Table A3: Experimental Participant Sample Representativeness

	(1)	(2)	(3)	(4)	(5)	(6)
	In Experiment		Not in Experiment		Difference	
	Mean	[SD]	Mean	[SD]	in SD units	N
Panel A. All Respondents Approached						
Male respondent	0.98	[0.13]	0.97	[0.17]	0.09**	8,231
Age	33.3	[8.2]	35.2	[8.7]	-0.21***	8,231
Car driver	0.30	[0.46]	0.41	[0.49]	-0.24***	8,227
Log vehicle price (residual)	10.5	[0.4]	10.5	[0.4]	-0.00	7,188
Panel B. Survey Respondents						
Log income	9.96	[0.71]	9.91	[0.73]	0.07	2,656
Stated Daily Travel (Km/day)	47.1	[24.0]	45.1	[25.1]	0.08*	4,427
Stated Value of Time (Rs/hr)	206.0	[138.9]	189.0	[151.3]	0.11*	1,001
Stated Schedule Flexibility (min)	20.0	[10.9]	18.7	[12.0]	0.11*	952

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Business owner or manager	Accountant, Teacher, Doctor	Software and IT	Engineers, Technical	Office staff	Manual jobs	Mobile professions	Student	Others, Retired	Total
Panel C. Survey Respondents										
In Experiment	16.7	7.5	10.3	14.3	15.4	8.4	15.6	9.0	2.9	455
Not in Experiment	15.6	6.2	10.1	11.2	18.1	9.5	12.0	13.4	3.9	2,458

Notes. These results describe respondent selection into experiment by comparing the experimental sample (497 respondents) to the entire sample of eligible commuters approached in gas stations by the survey team (panel A) and to the full survey sample (panels B and C). The sample in Panel A is all respondents approached in gas stations, excluding ineligible respondents. Weights are used to (a) account for missing data for each variable, and (b) to adjust for the estimated $\sim 52\%$ ineligible respondents among survey refusals (for refusals, $7,218$ respondents did not complete the eligibility filter, and I assume the same proportion were ineligible). Gender, age and car driver variables are visually assessed by the surveyor for all respondents. Vehicle value (residual) is imputed based on vehicle type (car/motorcycle), make and model, using pricing data scrapped from a used-vehicles website in Bangalore, residualized on a "car" dummy. Monthly income is self-reported during the recruitment survey (the respondent is handed the tablet to enter the amount confidentially – the surveyor never sees the amount), truncated at Rs. $100,000$ ($\sim 1,300$ USD). Occupation is self-reported during the recruitment survey. Value of time and schedule flexibility are based on choices in hypothetical scenarios in a follow-up phone survey; see Supplementary Materials section SM.3.1 for details. The difference in SD units includes significance levels from a (weighted) regression of the row outcome variable on an indicator for being in the experiment. $^{*}p \leq 0.10$, $^{**}p \leq 0.05$, $^{***}p \leq 0.01$

Table A4: Experimental Balance Checks

		Departure Time Treatments								Area Treatment			
		Information	(S.E.)	Low Rate	(S.E.)	High Rate	(S.E.)	Obs.	Control Mean	Area Early	(S.E.)	Obs.	Control Mean
(1)	Car user	0.01	(0.02)	0.01	(0.01)	0.01	(0.02)	497	0.28	-0.01	(0.01)	254	0.28
(2)	Regular destination	-0.05	(0.05)	0.00	(0.05)	-0.09*	(0.05)	497	0.77	-0.05	(0.03)	254	0.95
(3)	Age	-0.93	(0.93)	1.26	(1.01)	-0.09	(1.07)	497	33.20	-1.35	(0.94)	254	34.30
(4)	Log vehicle price	0.11**	(0.05)	0.09	(0.05)	0.03	(0.06)	453	11.06	0.00	(0.05)	231	11.17
(5)	Log income	0.01	(0.10)	-0.02	(0.14)	-0.07	(0.14)	411	10.11	-0.09	(0.12)	211	10.24
(6)	Frac days with good GPS data	-0.01	(0.03)	-0.02	(0.03)	-0.00	(0.03)	497	0.41	0.00	(0.03)	254	0.42
(7)	Frac days present at work	0.01	(0.03)	0.00	(0.03)	-0.03	(0.04)	497	0.70	-0.03	(0.03)	254	0.79
(8)	Number of trips per day	-0.12	(0.11)	-0.05	(0.13)	-0.00	(0.14)	497	1.25	-0.00	(0.12)	254	1.15
(9)	Total distance per day (Km.)	-0.46	(0.69)	-0.22	(0.83)	0.32	(0.88)	497	8.29	0.19	(0.83)	254	8.79
(10)	Total duration per day (min)	-2.99	(3.04)	-1.51	(3.55)	1.23	(3.84)	497	35.52	0.49	(3.50)	254	35.49
(11)	Total D.T. shadow rate per day	-0.84	(3.82)	-0.83	(3.90)	-0.32	(4.20)	497	38.82	-0.51	(3.79)	254	37.92
(12)	Total Area shadow rate per day	-2.27	(3.18)	-2.85	(4.23)	-0.40	(4.88)	497	23.83	0.69	(5.61)	254	50.73
(13)	Joint Significance Test F-stat	0.26								0.02			
(14)	Joint Significance Test P-value	0.86								0.90			

Notes. This table shows experimental balance checks for the departure time and area treatments. Variables 1,3,4, and 5 are from the recruitment survey, while the remaining eight variables are calculated from the GPS trips data before the experiment. Each row and group of columns combination reports coefficients from a regressions with the row header as outcome. In the "Area Treatment" columns, the sample is restricted to 254 participants who receive the area treatment, and the dependent variable is whether the respondent was assigned to the "early area" sub-treatment (to receive the area charges in week 1 as opposed to week 4). All regressions include randomization strata dummies. Rows 13 and 14 report the F-statistic and p-value from column-wise joint significance tests. Robust standard errors are shown in parentheses. $^{*}p \leq 0.10$, $^{**}p \leq 0.05$, $^{***}p \leq 0.01$

Table A5: GPS Data Quality at Daily Level (Attrition Check)

	(1)	(2)	(3)	(4)
Commuter FE	X	X	X	X
Panel A. Departure Time Treatment				
High Rate × Post	0.02			
	(0.05)			
Low Rate × Post	-0.00			
	(0.05)			
Information × Post	-0.00			
	(0.04)			
Post	0.08**			
	(0.03)			
Observations	24,779			
Control Mean	0.76			
Panel B. Area Treatment and Sub-treatments				
Treated	0.05**		0.04	0.05
	(0.02)		(0.03)	(0.03)
Treated in Week 1		0.02		
		(0.04)		
Treated in Week 4		0.07		
		(0.04)		
Treated × High Rate			0.01	
			(0.04)	
Treated × Short Detour				-0.05
				(0.05)
Post	0.05*	0.04	0.05*	0.07*
	(0.03)	(0.03)	(0.03)	(0.04)
Observations	13,433	13,433	13,433	7,986
Control Mean	0.73	0.73	0.73	0.75

Notes. This table shows experimental impacts on the quality of the GPS data received from study participants. The outcome is a dummy for good quality GPS data on a given day (see Supplementary Materials section SM.2 for the definition). The sample covers all non-holiday weekdays for all experiment participants, excluding days outside Bangalore. In the post period, the sample in Panel A is restricted to the departure time treatment period, either the first or the last three weeks. The sample in panel B is restricted to the area treatment period, either the last of the first week. Panel B restricts to 243 participants in the Area treatment, except in column (4) where the sample consists of the 148 Area participants for whom candidate areas included at least one with short detour (3-7 minutes) and at least one with long detour (7-14 minutes). All specifications include respondent and study cycle fixed effects. The mean of the outcome variable in the control group during the experiment is reported for each specification. Standard errors in parentheses are clustered at the respondent level. $^{*}p \leq 0.10$, $^{**}p \leq 0.05$, $^{***}p \leq 0.01$

Table A6: Impact of Area Charges on Daily Outcomes (D.T. Control and Info Sample)

	(1)	(2)	(3)	(4)	(5)	(6)
Outcome	*Total Shadow Charges Today*			*Number of Trips Today*		
Time of Day	AM & PM	AM	PM	AM & PM	AM	PM
Commuter FE	X	X	X	X	X	X
Panel A. Pooled Treatment						
Treated	-32.24***	-20.67***	-11.56***	0.09	0.10	-0.01
	(7.94)	(4.80)	(4.34)	(0.13)	(0.08)	(0.08)
Observations	4,390	4,390	4,390	4,390	4,390	4,390
Control Mean	107.11	54.36	52.75	2.59	1.15	1.44
Panel B. Treatment by Week						
Treated in Week 1	-28.99**	-20.36***	-8.63	-0.05	0.11	-0.16
	(12.46)	(7.53)	(7.72)	(0.21)	(0.13)	(0.13)
Treated in Week 4	-34.69**	-21.07**	-13.62	0.32	0.13	0.18
	(13.85)	(8.55)	(8.27)	(0.22)	(0.12)	(0.14)
Observations	1,772	1,772	1,772	1,772	1,772	1,772
Control Mean	107.11	54.36	52.75	2.59	1.15	1.44

Notes: This table replicates Table 3 restricting the sample to respondents in the control or information groups in the departure time treatment. The sample includes 114 respondents.

Table A7: Impact of Area Charge Sub-Treatments on Daily Outcomes

	(1) Shadow Charges	(2) Beliefs (Rs.)	(3) Beliefs (minutes)	(4) Google (minutes)
Panel A. High/Low Area Charge Sub-Treatments				
Treated × High Rate	-26.8***	195.0***		
	(7.9)	(5.4)		
Treated × Low Rate	-20.1**	105.2***		
	(7.8)	(5.1)		
Observations	8,827	99		
Commuters	243	99		
Control Mean	110.2			
P-val High=Low	0.55	0.00		
P-val High=2 × Low	0.44	0.06		
Panel B. Short/Long Detour Sub-Treatments				
Treated × Short Detour	-20.6***		22.7***	5.0***
	(7.4)		(3.3)	(0.5)
Treated × Long Detour	-24.0**		23.9***	8.6***
	(12.1)		(3.1)	(0.5)
Observations	5,358		67	67
Commuters	148		67	67
Control Mean	111.7			
P-val Short=Long	0.82		0.65	0.00
P-val Short=1.8 × Long	0.33			
P-val Short=Long / 1.8			0.00	0.71

Notes: This table reports difference-in-difference impacts of Area sub-treatments on daily total shadow charges and beliefs. The sample in panel A is the same as in Table 3. In panel B the sample consists of 148 area treatment participants for whom candidate areas included at least one with short detour (3-7 minutes) and at least one with long detour (7-14 minutes). In columns 2-4, the sample is restricted to participants surveyed by phone during the experiment on the week before the area treatment started. The outcome in column 1 is the same as in Table 3. Columns 2 and 3 use self-reported beliefs about the magnitude of the area charge (in Rupees) and of the detour duration (in minutes), respectively. In column 4 the outcome is Google Maps' prediction detour duration (in minutes). Standard errors in parentheses are clustered at the respondent level. $^{*}p \leq 0.10$, $^{**}p \leq 0.05$, $^{***}p \leq 0.01$

Table A8: Structural Estimation Robustness

	(1) Value of time α (Rs/hr)	(2) Schedule cost early β_E (Rs/hr)	(3) Schedule cost late β_L (Rs/hr)	(4) Logit inner σ (dep. time.)	(5) Logit outer μ (route)	(6) Probability to respond p
Baseline	1,084.5	294.6	260.8	8.1	33.8	0.42
Detour fixed cost 12 INR	931.8	278.5	257.7	7.7	34.3	0.41
Detour fixed cost 24 INR	785.4	262.1	255.3	7.3	34.7	0.40
Detour fixed cost 48 INR	497.6	235.8	249.9	6.9	34.9	0.38
Detour fixed cost 96 INR	25.0	210.0	227.7	6.7	35.3	0.38
Parameters $\propto$ wage	2,295.3	814.2	879.6	5.1	36.7	0.68
Common ideal arrival time dist. $\mathcal{H}^A$	1,199.1	81.9	257.3	5.9	35.2	0.39

Notes: This table reports GMM estimation results using different assumptions. The first row replicates Table 4. Rows 2-5 include a fixed cost of using the detour route $r = 1$. In row 6, parameters are proportional to the person's wage, $\theta_i = w_i/\bar{w} \cdot \theta$ for $\theta \in \{\alpha, \beta_E, \beta_L\}$ where $\bar{w}$ is the average wage; the table reports average parameters (θ). In row 7, the distribution of ideal arrival times $\mathcal{H}^A$ is the same for all commuters. During estimation, it is computed by inverting the *aggregate* distribution of realized departure times in the pre-period.

Table A9: Structural Model Numerical Identification Check

	(1)	(2)	(3)	(4)	(5)	(6)
			Estimated Parameter			
	$\hat{\alpha}$	$\hat{\beta}_E$	$\hat{\beta}_L$	$\hat{\sigma}$	$\hat{\mu}$	$\hat{p}$
(True) Value of time α	**1.16*** **	0.03	0.65	0.12	-0.04	0.21
	(0.13)	(0.42)	(1.64)	(0.16)	(0.10)	(0.14)
(True) Penalty early β_E	-0.28**	**1.16*** **	0.12	-0.10	-0.17	-0.18
	(0.13)	(0.43)	(1.67)	(0.17)	(0.10)	(0.14)
(True) Penalty late β_L	-0.15	0.08	**1.66**	0.06	-0.03	0.01
	(0.13)	(0.42)	(1.64)	(0.16)	(0.10)	(0.14)
(True) Logit inner σ	0.05	-0.17	-0.03	**0.40** **	0.10	-0.00
	(0.13)	(0.43)	(1.67)	(0.17)	(0.10)	(0.14)
(True) Logit outer μ	0.03	-0.03	-0.02	0.31*	**1.09*** **	0.08
	(0.12)	(0.42)	(1.64)	(0.16)	(0.10)	(0.14)
(True) Probability to respond p	-0.09	-0.53	-0.23	0.16	-0.06	**0.85*** **
	(0.13)	(0.42)	(1.64)	(0.16)	(0.10)	(0.14)
Observations	100	100	100	100	100	100

Notes: This table shows that the estimation method recovers underlying model parameters. See notes for Online Appendix Figure A8. Each column shows results from a quantile (median) regression of the estimated parameter on the vector of true parameters. For each type of parameter θ, the true θ_j and estimated $\hat{\theta}_j$ parameters are normalized by the standard deviation of the the true parameter over the simulation iterations j ($std(\theta_j)$). Diagonal coefficients in bold. $^{*}p \leq 0.10$, $^{**}p \leq 0.05$, $^{***}p \leq 0.01$

Table A10: Structural Estimation Sensitivity Measure

	(1) Value of time α (Rs/hr)	(2) Schedule cost early β_E (Rs/hr)	(3) Schedule cost late β_L (Rs/hr)	(4) Logit inner σ (dep. time.)	(5) Logit outer μ (route)	(6) Probability to respond p
Panel A. Departure Time Moments						
(1-31) Average absolute value for departure time bin moments	9.5	11.3	13.3	0.87	0.30	0.004
(32) Variance individual effects departure time treatment	7.1	-29.9	-24.9	2.85	0.09	-0.003
(33) Variance individual effects departure time control	-12.3	-20.2	-10.0	1.19	-0.44	-0.006
Panel B. Area moments						
(34) With charge: probability to intersect area	-19.1	-15.0	-18.3	0.02	-1.37	-0.028
(36) With charge: sample frequency to intersect area $\in [1/3, 2/3)$	122.6	28.7	18.2	0.04	4.02	0.045
(37) With charge: sample frequency to intersect area $\in [2/3, 1]$	71.7	15.9	20.3	-0.57	3.09	0.011
(35) Without charge: probability to intersect area	178.7	46.8	26.7	-0.01	0.18	0.075
(38) Without charge: sample frequency to intersect area $\in [1/3, 2/3)$	-101.1	-19.2	-1.4	0.27	0.99	-0.033
(39) Without charge: sample frequency to intersect area $\in [2/3, 1]$	-248.7	-57.0	-22.2	0.74	-0.62	-0.082

Notes: This table reports the estimated sensitivity measure Λ from Andrews et al. (2017), scaled by the (bootstrap) standard deviation of each moment. Each entry Λ_{pj} measures the change in estimated parameter θ_p due to a one standard deviation change in moment m_j. The matrix Λ is estimated by $\hat{\Lambda} = \left(\hat{S}'\hat{W}\hat{S}\right)^{-1}\hat{S}'\hat{W}\text{diag}\left(\hat{\sigma}_j\right)$ where $\hat{S}$ is the Jacobian of the moments with respect to parameters evaluated at the estimated parameters, $\hat{W}$ is the estimated optimal weighting matrix, and $\text{diag}\left(\hat{\sigma}_j\right)$ is the diagonal matrix with jth entry given by the (bootstrap) estimated standard error of moment j.

Table A11: Road Technology Trip Level Regressions

	(1)	(2)	(3)	(4)
Dependent Variable:	Trip Delay (min/km)			
Commuter FE			X	X
Traffic Volume at Trip Departure Time	0.86***	0.84***	0.70***	0.70***
	(0.04)	(0.03)	(0.04)	(0.03)
Trip Length (km)		-0.05***		-0.01**
		(0.00)		(0.00)
Constant	2.48***	2.97***	3.10***	3.16***
	(0.06)	(0.05)	(0.06)	(0.06)
Observations	61,893	61,893	61,893	61,893

Notes: This table reports trip-level quantile (median) regressions of the trip delay, defined as trip duration in minutes divided by trip length in kilometers, on the average traffic volume at the trip departure time and trip length. The sample is all weekday trips more than 2km in length, without any stops along the way, and with a trip diameter to total length ratio above 0.6 (the 25th percentile). Columns 3 and 4 first residualize the trip delay on commuter fixed effects. Standard errors in parentheses are clustered at the commuter level. $^{*}p \leq 0.10$, $^{**}p \leq 0.05$, $^{***}p \leq 0.01$

SM Supplementary Material for "Peak-Hour Road Congestion Pricing: Experimental Evidence and Equilibrium Implications"

Date: May 2, 2020

SM.1 Trip Classification Algorithm

This section describes the algorithm that processes raw GPS data obtained from the study smartphone app into trips. This algorithm was used to process 74,059 days of data, covering 22,434,466 unique locations points from 2,300 devices.

The algorithm has five main parts:

1. Remove outliers from raw data.
2. Segment each day into "trip" "location" segments, as well as "gap" and "jump" (missing data) segments.
3. Classify home and work locations for experiment participants. (This step was not used during the experiment, only for final data analysis.)
4. Combine segments into (final) locations and (final) trips.
5. Compute quality measures for each trip and for each day.

I now describe each part in more detail.

1. Removing GPS Data Outliers. Smartphones sometimes temporarily report an erroneous location, which the algorithm identifies and drops. The main cases are:

- *imprecise location.* When location is determined solely based on the cell phone tower, the accuracy is in the range of 600-800 meters. Points with accuracy radius above 400 meters are initially dropped.[75]
- *sharp angle jump.* This occurs when the location jumps to a precise but distant location, before returning to the current location. Such points are identified when a speed threshold is exceeded and the location returns close to the original location. These points are dropped.
- *multiple jumps.* This case occurs when several location jumps take place in succession. This case is identified similarly to the previous case. These points are dropped.
- *lazy location.* This occurs when the smartphone is moving continuously yet the location remains stuck for a period of time and then jumps suddenly. A point is labeled as "lazy" if the speed to the next point exceeds a certain threshold, and removing it leads to speed below the threshold. These points are dropped.

2. Day segmenting. This algorithm has the following steps for each day of data (for a given participant).

1. **Location score.** For each GPS point X, this is a number between 0 and 1 that roughly captures the fraction of points that occur a short while after X and that are near to X. The score for a point Y is 1 if $\text{dist}(X, Y) \leq 100$ meters, and linearly decreasing from 1 to 0 if $\text{dist}(X, Y)$ is between 100 and 200 meters, and zero otherwise. If there exists at least one point Y between 1.5 and 5 minutes after X, then the sample is all points between 1.5 and 15 minutes after X, and each point is weighted using a triangular kernel (with points at 1.5, 5 and 15 minutes). If there are no points Y between 1.5 and 5 minutes after X, the sample is all

[75] The accuracy radius distribution is highly bimodal. Among points below 400 meters, the vast majority are below 100 meters and most below 50 meters.

points between 0 and 30 minutes after X, with triangular weights (with points at 0, 15, and 30 minutes). If there are no points Y in the 30 minutes after X, the default location score is 0.5. The location score is also calculated looking *backward* in time, with the same parameters.

2. **Initial location and trip segments.** In this step, the algorithm iterates in reverse order through GPS points starting from the last point of the day, with a state initialized to "location." Moving from a point N to point $N-1$, the state evolves as follows. If the state of N is "location," the state of $N-1$ changes to "trip" if the (backward-looking) location score of point $N-1$ is below 0.3. If the state of N is "trip," the state of $N-1$ changes to "location" if the (backward-looking) location score of point $N-1$ is above 0.7. Otherwise, the state of $N-1$ is the same as that of N. This algorithm is relatively conservative and only switches the state if we have information consistent with the other state. A trip (location) *segment* is a sequence of points all coded "trip" ("location").
3. **Define missing data segments.** First, at this point, low-accuracy points are added back to location segments as long as the location centroid is within the accuracy radius of the low-accuracy point. (The reason is that they offer additional confirmation that the smartphone was at that location.) There are three types of missing data segments:

   - a "gap" is the period between two consecutive points that are within a "location" segment and more than 60 minutes apart.
   - a "jump" is the period between two consecutive points, the first on a "location" and the second on a "trip," which are far away: more than 1,000 meters apart, or more than 500 meters apart and 10 minutes apart. (If these conditions are false but the next point is still more than 20 minutes away, the period is labeled as a "gap.")
   - a "jumptrip" is the period between two consecutive points, the first one on a "trip" segment, that are far away: more than 2,000 meters apart, or more than 500 meters apart and more than 15 minutes apart, or more than 300 meters apart and more than 20 minutes apart.

   For each missing data segment I compute the duration and (geodesic) distance between the endpoints.

3. Classify main common locations. This proceeds in two steps. First, the algorithm clusters locations using the entire data for a given person. It uses the Density-Based Spatial Clustering of Applications with Noise algorithms, using the command `DBSCAN` in the `sklearn.cluster` package. The second step is to manually identify home and work locations starting from the most commonly visited location clusters.

Note, the main location clustering step was not used during the experiment, only for final data analysis.

4. Final locations and trip segments. In this last step, segments are combined into two main types: (final) location and (final) trip. First, each spell of consecutive location and gap segments is combined into a single location segment, and each spell of consecutive trip, jump and jumptrip segments is combined in a single trip segment. Second, the algorithm iterates over trips, starting in the morning. Each trip is expanded to include later segments one by one, until one of the following conditions holds:

- the last added segment is followed by a location with duration $M = 15$ minutes or more, or
- the destination of the last added segment is the home or the work location, or
- the destination of the last added segment is the same as the (original) trip origin.

In particular, two consecutive trips separated by a stop of 15 minutes or less will be merged into a single trip, except if the location is the home or the work location of that commuter, or if the second trip is a "return" trip.

SM.2 Data Quality Measures and Analysis Sample

SM.2.1 Day Data Quality Measures

To define data quality for a given day, I first compute the total duration and distance corresponding to missing data. Specifically, I compute the *total gap duration* as the total duration of gap, jump and jumptrip segments in the day, between 7 am and 9 pm. In order not overly penalize missing data during periods when the smartphone was most likely stationary, for "gap" segments, I use the segment duration minus 45 minutes, or zero, whichever is larger. The *total jump distance* is the sum of distances for all jump and jumptrip segments in the day between 7 am and 9 pm.

I define three categories of day data quality. Quality is *good* if the total gap is below 3.5 hours and the total jump below 1.5 kilometers. Quality is *medium* or *inferior* otherwise. Quality is *inferior* if the total gap is over 6.5 hours.

The analysis sample is good and medium quality days. Table A12 shows the number and fraction of days in the sample by quality level, for three samples. In the first two columns, it considers weekdays during the experiment for the experimental sample. The middle two columns it considers the same sample of users, and all weekdays in the study (from the day when a user joined the study until the end of the study on September 11, 2017). The last two columns include the full sample of users and full sample of weekdays in the study.

Table A12: Date data quality measures

	(1)	(2)	(3)	(4)	(5)	(6)
Sample:	Experimental Sample Experimental Period		Experimental Sample Full Period		Full Sample Full Period	
Good quality:	7,528	0.58	18,863	0.35	36,164	0.10
Medium quality:	2,251	0.17	6,578	0.12	16,152	0.05
Inferior quality:	1,455	0.11	5,462	0.10	21,743	0.06
No data:	1,745	0.13	23,478	0.43	273,815	0.79
Total days:	12,979		54,381		347,874	
Unique users:	497		497		2301	

SM.2.2 Trip Analysis Sample

The analysis sample for trips is defined according to the criteria listed in Table A13. Overall, in the experimental sample, 64% of all non-weekend, non-holiday, daytime trips are included in the sample.

The sample and quality criteria used to construct the sample of trips are the following:

- Exclude from the sample weekends, holidays and trips before 7 am and after 10 pm.
- Exclude from the sample trips outside Bangalore. A trip is defined as outside the city if $\geq 70\%$ of the trip duration is at least 18 km from Bangalore's city center.
- Drop trips with imprecise departure time, defined as at least 15 minutes between the last point in a location and the first point on the trip. Similarly, drop trips with imprecise arrival time.
- Drop unusually short or long trips, both in terms of distance and duration.
- Drop trips for which "jumps" constitute a large part of the trip (that last at least 30 minutes or are at least 30% of the trip duration).
- Drop "swiggly" trips with a diameter (largest distance between any two points on the trip) to distance (path length) ratio of less than 0.3. These trips follow a very sinuous path.
- Drop "short loop" trips, namely trips that are overall less than 2 km and have a ratio between the origin-destination distance and the distance (path length) of less than 0.3.

Table A13: Trip data quality measures and analysis sample

	(1)	(2)	(3)	(4)
Sample:	Experimental		Full	
1. Weekend:	51,169	0.27	0	0.00
2. Outside Bangalore:	11,423	0.06	30,966	0.11
3. Nighttime:	10,014	0.05	0	0.00
4. Imprecise departure time:	12,736	0.07	34,019	0.12
5. Imprecise arrival time:	3,746	0.02	9,358	0.03
6. Long trip (> 3 hours):	360	0.00	771	0.00
7. Long trip (> 35 km):	510	0.00	1,145	0.00
8. Short trip (< 5 minutes):	12,078	0.06	26,011	0.09
9. Short trip (< 500 meters):	1,278	0.01	3,059	0.01
10. Has jumps:	5,529	0.03	13,568	0.05
11. Swiggly:	1,976	0.01	4,220	0.02
12. Short loop trips:	4,007	0.02	8,980	0.03
13. In the analysis sample:	74,762	0.39	145,325	0.52
Total trips:	189,588		277,422	
Total days:	40,178		64,832	

SM.3 Survey Data Collection Details

SM.3.1 Hypothetical Choice Questions

The "stated preferences" phone survey collected data on typical departure times and travel times, beliefs about travel times at earlier and later times, as well as hypothetical choice questions for route choice and departure time in the presence of congestion charges. These two types of questions were designed to be similar to the two experiments:

> ***Value of Time Questions.*** *Imagine there were two different routes to go from home to work. The routes are identical: same distance, same road quality, etc. The only difference is that one route is faster, but it has a charge (toll) that you must pay. The other route is slower but it's free.*
>
> *One route takes* T_0 *minutes and has a charge. The other route takes* $T_1 = T_0 + \Delta T$ *minutes and no charge. This route takes* ΔT *minutes more.*
>
> *[Surveyor asks for each of* $p^R \in \{\text{Rs.}100, 90, \ldots, 30, 25, 20, 15, 10, \}$*]*
>
> *Please tell me, what do you prefer:* T_0 *minutes and paying* p^R *or* T_1 *minutes for free?*

> ***Schedule Flexibility Questions.*** *Now please imagine that there is a toll on your usual route, and the toll is different at different times. Imagine that everything else on the route is the same as your usual route. We would like to know if you would change your departure time to avail of a lower toll.*
>
> *The toll is* p^D *Rs for leaving at your usual departure time, h. If you leave [**direction**] it is less expensive. For example:*
>
> - *If you leave* 5 *minutes [**direction**], you save* $\Delta p^D \cdot 5$ *Rs (toll is* $p^D + \Delta p^D \cdot 5$ *Rs).*
> - *If you leave* 10 *minutes [**direction**], you save* $\Delta p^D \cdot 10$ *Rs (toll is* $p^D + \Delta p^D \cdot 10$ *Rs).*

- *If you leave 20 minutes [**direction**], you save $\Delta p^D \cdot 20$ Rs (toll is $p^D + \Delta p^D \cdot 20$ Rs), and so on.*

*Leaving [**other-direction**] is more expensive, with the same amounts. For every minute that you leave [**direction**], you pay less.*

Q1. Based on this information, would you change your departure time?

- *Yes: I would leave earlier*
- *No: I would leave at the same time*
- *Yes: I would leave later*

Q2. How much earlier/later would you leave, on average? (in minutes) [integer]

All questions had specific numbers, partly based on previous responses (e.g. T_0, h). [**direction**] was randomly chosen to be either "earlier" or "later."

SM.4 Randomization Details

The randomization strata were all eight combinations of area eligible/ineligible, car/motorcycle or scooter, high/low kilometer travel at baseline. Participants were assigned to treatment on a rolling basis, and the treatment allocation was pre-randomized within each stratum. This was done differently for strata that were area eligible than those that were area ineligible.

In each of the four strata for area ineligible participants, blocks of 24 consecutive positions were perfectly balanced for all (8) combinations of 4 departure time sub-treatments and early/late timing. The departure time sub-treatment probabilities are shown in panel A in Appendix Table A2. The early/late groups were equal probability. Randomization was implemented by choosing a random permutation of $\{1, 2, \ldots, 24\}$ for each block of 24 positions.

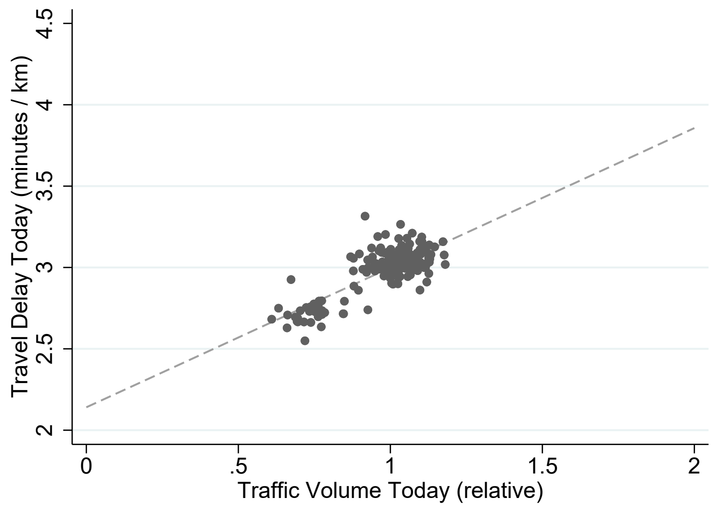
In each of the four strata for area eligible participants, blocks of 64 consecutive positions were perfectly balanced for all (32) combinations of 4 departure time sub-treatments and 8 area sub-treatments (all combinations of early/late, high/low rate, long/short detour). The departure time sub-treatment probabilities are shown in panel A in Appendix Table A2. The area sub-treatments were equal probability. In addition, within the 64 positions, blocks of 8 consecutive positions were balanced on the "marginals" for the (8) area sub-treatments, and for the (4) departure time sub-treatments. This restriction generates approximate stratification by enrollment time. Randomization was implemented for each block of 64 positions by randomly choosing a permutation of $\{1, 2, \ldots, 64\}$ that satisfies the "marginal" balance condition.

The restriction to stratify "marginals" by time is equivalent to covering the complete 8×8 bipartite graph with 8 disjoint perfect matchings. I implemented an algorithm to generate a random covering based on the proof of König's 1931 theorem.[76] The intuition is the result that in a general bipartite graph, a maximal matching can be achieved with a modified greedy algorithm. Imagine that the greedy algorithm is stuck, yet there exist two vertices in the two sides of the bipartite graph that are "exposed," namely each is connected to an edge whose other node is not covered by the current matching. Then it is possible to prove that the matching can be modified and grown by one edge via an "alternating path." This constructive proof can be used to sequentially add matchings until a full covering is found. I modified this algorithm to ensure random sampling from the set of all possible coverings.

[76] I acknowledge Michel X. Goemans' lecture notes on combinatorial optimization (MIT course 18.433, 2009).

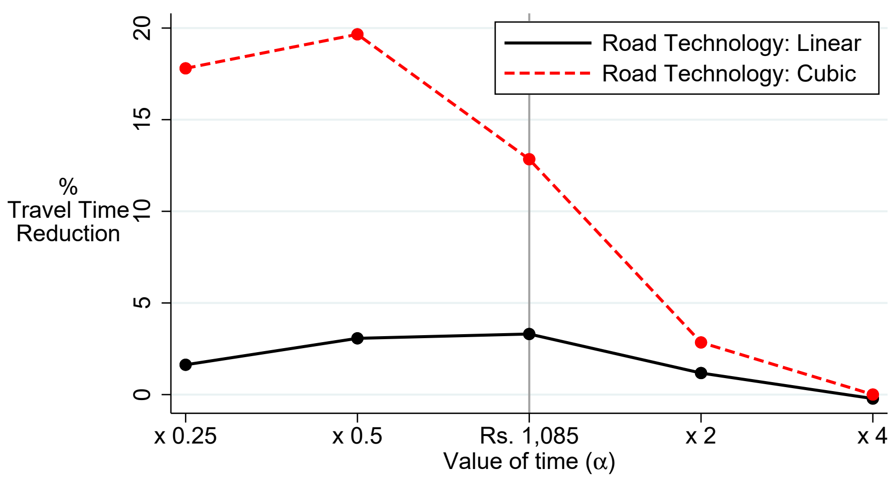
SM.5 Supplementary Material: Figures

Figure SM1: Road Technology: Travel Delay and Traffic Volume over Dates



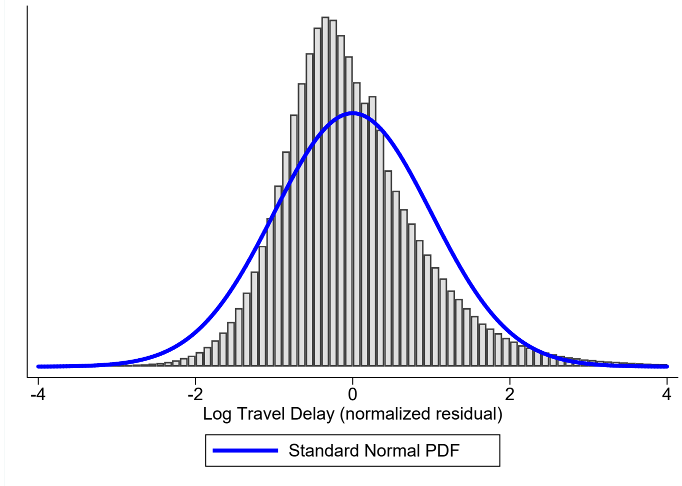
Notes: This graph shows how calendar date-level travel delay and volume of traffic are related. Data is as in Figure 3, and includes weekends. For each date I compute the number of trips per capita (using the number of app users that day), and normalize this variable to have mean 1. I compute the average delay over all routes and departure times, for each day in the data. Table 5 column 3 reports the regression version of this relationship.

Figure SM2: Policy Simulations: Varying Preference and Road Technology Parameters



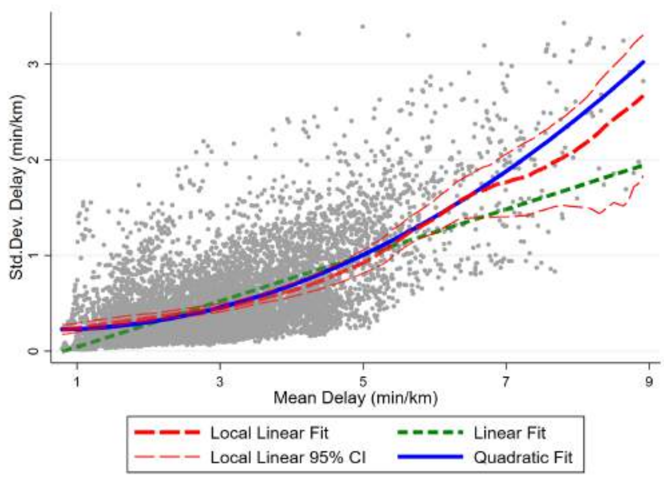
Notes: This graph replicates Figure 5 for travel time.

Figure SM3: Google Maps Travel Time is Approximately Log-Linear Distributed



Notes: This figure shows the shape of the day-to-day variation of log normalized travel time. For each route and departure time cell, I consider the distribution of travel times over 145 weekdays. Within each cell, I compute the normalized residual by subtracting the mean and dividing by the standard deviation for that cell. The graph shows the distribution of the log residuals for all cells, and a standard normal (solid blue line).

Figure SM4: Travel Time Standard Deviation is Approximately Quadratic in Travel Time Mean



Notes: This figure shows the relationship between the mean and standard deviation of travel time. Each dot represents a route and departure time cell, and the two axes measure the mean and standard deviation in that cell over over weekdays. The local linear, linear and quadratic fits are respectively shown in red (long dash), green (dash) and blue (solid). The local linear fit uses an Epanechnikov kernel with 0.5 minute per kilometer bandwidth and 95% confidence intervals bootstrapped by route, are also shown (thin red dashed line). The estimated quadratic equation is:

$$\text{StdDevDelay} = \underset{(0.02)}{0.24} - \underset{(0.01)}{0.05} \cdot \text{MeanDelay} + \underset{(0.002)}{0.04} \cdot \text{MeanDelay}^2$$

SM.6 Supplementary Material: Tables

Table SM1: Measures of "attention" to the experiment

	(1) Fraction Correct	(2) Observations
Panel A. Departure Time Treatment		
Charges are per-KM	61.8%	133
Rate fn of departure time	57.8%	133
Peak rate correct	55.1%	133
Two out of three correct	55.4%	133
Panel B. Area Treatment		
Knows area location	56.1%	132
Daily charges correct ($\geq 4/5$ days)	49.2%	132

Notes: This table report results from a phone survey with treatment group study participants. It took place after the initial in-person meeting (when a surveyor explained the treatments), either during or up to a week before the respective treatment started. The table reports the fraction of respondents who correctly identified certain aspects of their treatment. Panel A reports the fraction of respondents who correctly and unprompted identified that congestion charges are proportional to trip length, that they depend on when the trip starts, and the maximum rate (Rs. 12 or Rs. 24). Panel reports the fraction of respondents who correctly described the congestion area location, and correctly identified the area charges for at least four out of five days.

Table SM2: Impact of Departure Time Charges on Daily Total Shadow Rate

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Time of Day	AM & PM		AM			PM	
		all	pre peak	post peak	all	pre peak	post peak
Commuter FE	X	X	X	X	X	X	X
Panel A. Full Sample							
Charges × Post	-10.55**	-5.28**	-3.12*	-2.17	-5.27**	-2.42	-2.84*
	(4.18)	(2.51)	(1.73)	(1.49)	(2.57)	(1.78)	(1.57)
Post	0.65	-1.19	0.29	-1.48	1.84	1.61	0.23
	(3.94)	(2.38)	(1.65)	(1.50)	(2.54)	(1.79)	(1.57)
Observations	15,585	15,585	15,585	15,585	15,585	15,585	15,585
Control Mean	95.74	46.89	23.81	23.08	48.85	24.60	24.25
Panel B. Regular Commuters, Home-Work and Work-Home Trips							
Charges × Post	-7.94***	-3.76**	-3.00*	-0.76	-4.18**	-0.88	-3.30***
	(2.89)	(1.90)	(1.56)	(1.20)	(1.67)	(1.23)	(1.08)
Post	-1.74	-0.74	-1.29	0.55	-1.00	-0.69	-0.31
	(2.65)	(1.74)	(1.30)	(1.36)	(1.61)	(1.18)	(1.06)
Observations	12,115	12,115	12,115	12,115	12,115	12,115	12,115
Control Mean	40.80	23.37	14.27	9.10	17.44	9.15	8.29
Panel C. Variable Commuters, All Trips							
Charges × Post	-5.17	-2.18	1.63	-3.81	-2.98	-5.36	2.37
	(8.50)	(5.18)	(3.10)	(3.36)	(5.78)	(3.83)	(3.78)
Post	1.67	1.10	1.52	-0.41	0.57	2.62	-2.04
	(7.67)	(4.67)	(3.00)	(3.75)	(5.13)	(3.91)	(3.23)
Observations	2,917	2,917	2,917	2,917	2,917	2,917	2,917
Control Mean	87.17	38.21	16.87	21.35	48.95	25.99	22.96

Notes: This table reports difference-in-difference impacts of the departure time congestion charges on daily total shadow rates. Overall, the sample of users and days, and the specifications, are the same as in Table 1, panel B. Columns (3) and (6) restrict to trips before the peak, i.e. the mid-point of the rate profile. Columns (4) and (7) restrict to trips after the peak. Panel B restricts to regular commuters and direct trips between their home and work locations (in either direction), and panel C restricts to variable commuters. Standard errors in parentheses are clustered at the respondent level . $^{*}p \leq 0.10$, $^{**}p \leq 0.05$, $^{***}p \leq 0.01$

Table SM3: Impact of Departure Time Charges on Trip Departure Times

	(1)	(2)	(3)	(4)
Time of Day	AM		PM	
	pre peak	post peak	pre peak	post peak
Commuter FE	X	X	X	X
Panel A. Departure Time (minutes)				
Charges × Post	-3.0*	2.8*	-0.3	2.2
	(1.7)	(1.6)	(1.6)	(1.5)
Observations	5,939	5,047	5,596	5,478
Panel B. Number of Trips				
Charges × Post	-0.01	-0.01	-0.02	-0.01
	(0.02)	(0.02)	(0.02)	(0.02)
Observations	15,585	15,585	15,585	15,585
Control Mean	0.28	0.25	0.27	0.24
Panel C. Departure Time (minutes)				
Charges × Post	-3.8*	5.6	4.2	7.1**
	(2.0)	(3.4)	(3.6)	(3.0)
Commuter Sample	X	X	X	X
Observations	3,003	1,328	1,613	1,423
Panel D. Number of Trips				
Charges × Post	-0.01	-0.00	-0.01	-0.02
	(0.02)	(0.01)	(0.01)	(0.01)
Commuter Sample	X	X	X	X
Observations	12,115	12,115	12,115	12,115
Control Mean	0.21	0.10	0.11	0.09

Notes: This table reports the impact of the departure time congestion charges treatment on departure times (panels A and C) and the number of trips, for trips taking place during the pre- and post-peak "ramps" in the morning and evening. Panels A and B use all (good quality) trips and users, while panels C and D restrict to regular commuters and trips between home and work or vice-versa. The sample in column 1 is all trips between 1.5 and 0.5 hours before the ramp midpoint for that commuter, in the morning. This interval corresponds to the early AM "ramp" when the congestion rate is linearly increasing. The outcomes in the other columns are defined analogously.

Table SM4: Impact of Area Charges on Trip Duration and Trip Shadow Charge

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Outcome	*Trip Shadow Charge*	*Trip Duration (minutes)*			*Trip Duration truncated at 99% (minutes)*		
Route FE	X	X	X	X	X	X	X
Specification			IV	IV		IV	IV
Panel A. Pooled Treatment							
Treated	-22.46***	0.43			1.13*		
	(3.43)	(0.74)			(0.66)		
Avoids Area			1.90			5.03*	
			(3.10)			(2.66)	
Observations	7,489	7,489	7,483	.	7,415	7,409	
Control Mean	83.38	40.93			39.58		
P-val $\neq$ avg. detour			0.15			0.60	
Panel B. Treatment by Week							
Treated in Week 1	-18.52***	-0.60			0.31		
	(5.04)	(1.27)			(1.10)		
Treated in Week 4	-27.15***	1.18			1.70		
	(6.16)	(1.14)			(1.08)		
Avoids Area			-3.20	4.68		1.37	6.23*
			(6.70)	(3.68)		(5.55)	(3.44)
Observations	3,104	3,104	2,375	2,205	3,074	2,351	2,186
Control Mean	83.38	40.93			39.58		
P-val $\neq$ avg. detour			0.15	0.63		0.36	0.95

Notes: This table reports difference-in-difference impacts of the Area treatment on trip shadow charge (column 1) and on trip duration (columns 2-7). The shadow charge of a trip is equal to 100 if the trip intersects the respondent's congestion area, and 0 otherwise. "Avoids Area" is a dummy for trips that do not intersect the area (shadow charge of zero). In columns 5-7 trips with duration above the 99th percentile (112 minutes) are dropped. The sample of users and days are the same as in Table 3, except that we restrict to regular commuters and good quality, home to work or work to home trips. All specifications include route fixed effects. Columns 3,4,6, and 7 are 2SLS where "Avoid Area" is instrumented by the area treatment. In panel B, columns 3 and 6, week 4 is dropped from the sample, so the comparison is entirely across commuters. In panel B, columns 4 and 7, week 1 is dropped from the sample. The table also reports the p-value of equality between the coefficient on "Avoids Area" and the average extra travel time on the quickest detour route (6.4 minutes according to Google Maps). Standard errors in parentheses are clustered at the respondent level . $^{*}p \leq 0.10$, $^{**}p \leq 0.05$, $^{***}p \leq 0.01$

Table SM5: Departure Time Charges Treatment Heterogeneity

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9))
Heterogeneity Dummy Variable K	Regular Destination	Self Employed	Low Income	Car Driver	Low Vehicle Value	Older	Small Stated α	Small Stated β	Short Route
Panel A. *Departure Time Treatment: Total Shadow Rates Today*									
Charges×Post×($K = 0$)	-11.2	-8.9	-8.9	-11.7*	-25.9***	-3.9	-10.5	-17.7**	-15.0**
	(9.2)	(5.5)	(7.4)	(6.3)	(7.4)	(7.4)	(7.8)	(7.0)	(7.2)
Charges×Post×($K = 1$)	-13.8**	-33.0**	-13.3	-16.5*	0.7	-18.4***	-15.0*	-11.2	-13.0*
	(6.0)	(12.9)	(9.1)	(8.5)	(7.4)	(6.7)	(7.7)	(8.7)	(7.0)
Observations	15,585	15,341	12,944	15,585	14,321	15,585	13,422	13,019	15,585
Participants $K = 0$	119	407	228	350	236	174	221	211	249
Participants $K = 1$	378	82	183	147	217	323	204	201	248
Control Mean $K = 0$	87.56	90.09	90.94	96.77	93.07	88.86	91.47	89.47	93.82
Control Mean $K = 1$	98.48	121.08	104.25	93.50	99.59	99.90	102.15	101.88	97.69
P-value interaction	0.81	0.09	0.71	0.65	0.01	0.15	0.68	0.56	0.85
Panel B. *Departure Time Treatment: Number of Trips Today*									
Charges×Post×($K = 0$)	-0.35	-0.06	0.01	-0.12	-0.29*	-0.08	-0.06	-0.16	-0.11
	(0.25)	(0.12)	(0.15)	(0.15)	(0.16)	(0.19)	(0.19)	(0.15)	(0.16)
Charges×Post×($K = 1$)	-0.07	-0.37	-0.27	-0.17	0.12	-0.16	-0.04	-0.09	-0.19
	(0.14)	(0.35)	(0.24)	(0.18)	(0.19)	(0.15)	(0.17)	(0.20)	(0.17)
Observations	15,585	15,341	12,944	15,585	14,321	15,585	13,422	13,019	15,585
Participants $K = 0$	119	407	228	350	236	174	221	211	249
Participants $K = 1$	378	82	183	147	217	323	204	201	248
Control Mean $K = 0$	2.95	2.78	2.83	3.01	2.78	2.88	2.89	2.81	2.76
Control Mean $K = 1$	2.95	3.70	3.23	2.82	3.09	2.99	3.08	3.11	3.14
P-value interaction	0.32	0.40	0.32	0.86	0.09	0.72	0.96	0.79	0.74

Notes: This table reports heterogeneous experimental response by observable characteristics. All heterogeneity variables K are dummy variables. They are whether the commuter:

1. has a stable destination (is a regular commuter as defined in section 4)
2. is self-employed
3. has below-median self-reported income
4. is a car driver at time of recruitment
5. has a vehicle value below median (vehicle value above median includes all cars and some motorcycles)
6. is at least 35 years old
7. has stated preference value of time (α) below median
8. has stated preference schedule cost (β) below median
9. has pre-experiment home to work route length below median

Data. Average vehicle values are scrapped from an online marketplace in Bangalore and matched by vehicle type, brand and model. Stated preferences are from a phone survey, see Supplementary Material SM.3.1. *Specification.* Each regression includes commuter fixed effects, study period fixed effects interacted with each group. The last line in each panel reports the p-value from the test of whether the two groups ($K = 0$ and $K = 1$) responded identically to the experiment. Inference is not adjusted for multiple hypothesis testing.

Table SM6: Area Congestion Charge Treatment Heterogeneity

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Heterogeneity Dummy Variable K	Self Employed	Low Income	Car Driver	Low Vehicle Value	Older	Small Stated α	Small Stated β	Short Route	Seldom Avoid Area
Panel C. *Area Treatment: Total Shadow Rates Today*									
Treated$\times(K=0)$	-21.5***	-28.0***	-26.6***	-20.8***	-13.7	-24.0***	-13.0	-28.9***	-12.2
	(6.0)	(8.1)	(6.5)	(7.7)	(9.4)	(7.8)	(9.2)	(6.6)	(7.6)
Treated$\times(K=1)$	-24.2*	-27.9***	-15.7	-25.7***	-28.2***	-24.8***	-32.9***	-14.8	-35.2***
	(14.4)	(8.9)	(10.3)	(8.3)	(6.8)	(8.8)	(7.7)	(10.5)	(7.9)
Observations	8,693	7,289	8,827	8,043	8,827	7,362	7,133	8,827	8,827
Participants $K=0$	204	114	174	118	79	94	92	160	117
Participants $K=1$	35	87	69	102	164	106	101	83	126
Control Mean $K=0$	103.56	116.36	108.48	106.65	97.93	94.68	93.37	113.69	85.64
Control Mean $K=1$	137.78	106.67	108.30	104.86	113.61	116.37	114.78	98.34	128.48
P-value interaction	0.86	0.99	0.37	0.66	0.21	0.95	0.10	0.26	0.04
Panel D. *Area Treatment: Number of Trips Today*									
Treated$\times(K=0)$	0.18**	0.15	0.06	0.17	0.15	0.17	0.18	0.10	0.21
	(0.09)	(0.12)	(0.10)	(0.11)	(0.15)	(0.12)	(0.12)	(0.09)	(0.14)
Treated$\times(K=1)$	-0.05	0.19	0.39**	0.20	0.16	0.16	0.18	0.22	0.07
	(0.23)	(0.16)	(0.16)	(0.14)	(0.10)	(0.12)	(0.13)	(0.17)	(0.10)
Observations	8,693	7,289	8,827	8,043	8,827	7,362	7,133	8,827	8,827
Participants $K=0$	204	114	174	118	79	94	92	160	117
Participants $K=1$	35	87	69	102	164	106	101	83	126
Control Mean $K=0$	2.30	2.59	2.57	2.40	2.44	2.44	2.31	2.34	2.57
Control Mean $K=1$	3.81	2.60	2.40	2.60	2.56	2.50	2.57	2.86	2.48
P-value interaction	0.34	0.86	0.08	0.88	0.96	0.95	0.99	0.57	0.41

Notes: This table reports heterogeneous experimental response by observable characteristics. All heterogeneity variables K are dummy variables. See table notes for SM5. The dummy variable in the last column is whether the frequency of intersecting the congestion area pre-experiment is below median. Inference is not adjusted for multiple hypothesis testing.