

The Value of Waiting Time in the United States: Estimates from Nationwide Natural Field Experiments*

Ariel Goldszmidt[†] John A. List[‡] Robert D. Metcalfe[§]
Ian Muir[¶] V. Kerry Smith^{||} Jenny Wang^{**}

Friday 16th October, 2026

Abstract

Values of time contribute to the determination of individuals' relative shadow prices of goods and services, investments, productivity, economic growth, and measurements of income inequality. By leveraging key features of the classic time allocation theories together with preference restrictions commonly used in nonmarket valuation, we estimate the value of waiting time (VOT) via two large-scale natural field experiments run by the ridesharing company Lyft. We use randomized variation in wait times and prices to estimate consumers' VOT in a data set of nearly 15 million observations across 13 large United States metropolitan areas. We find that the VOT is approximately \$19 per hour (or 67% (90%) of the pre-tax mean (median) wage rate in 2015 dollars), varies predictably with choice circumstances, and is increasing in the amount of waiting time. Our VOT estimate is larger than what is currently used by the U.S. Government for travel time, but by less than the typical multiplier used for waiting time.

Keywords: Value of time; natural field experiments; ridesharing; transportation; benefit-cost analysis.

*We thank the editor Kory Kroft and three anonymous referees for detailed and constructive comments on an earlier draft of this paper. Michael Anderson, Vittorio Bassi, Antonio Bento, Alec Brandon, Jonathan D. Hall, Nathan Hendren, Justin Holz, Matthew Kahn, Shanjun Li, Paulina Oliva, Joseph Shapiro, Ken Small, Cliff Winston, and numerous discussants at Lyft also provided excellent comments.

[†]University of Chicago

[‡]University of Chicago and NBER

[§]Columbia University and NBER

[¶]Walmart

^{||}Arizona State University and NBER

^{**}Massachusetts Institute of Technology

1 Introduction

Until the 1960s, economists used a person’s wage rate to approximate their value of time (VOT) and assumed it was approximately constant across contexts and times of the day or week. Later, a “*rule-of-thumb*” was devised for valuing time. It maintained the VOT was a constant fraction of the wage rate. [Mohring \(1965\)](#) characterized the fraction as being picked “*out of the air*”. Around the same time, there was an effort to explore the theoretical implications of different allocations of non-working time ([Becker, 1965](#); [DeSerpa, 1971](#)). Two assumptions in these classic studies transformed the framework for valuing time: (1) agents were assumed to be able to allocate their non-working time to different tasks; and (2) the technology for producing household services allowed for complementarity between time and purchased commodities. The first allows an analyst to estimate VOTs from observed choices. The second implies the estimated marginal value of time is welfare consistent. As a result, there is now a sizable literature estimating the values of time from discrete or continuous time-money trade-offs in surveys, small-scale trials, real-world transportation settings ([Deacon and Sonstelie, 1985](#); [Smith and Mansfield, 1998](#); [Small et al., 2005](#); [Wolff, 2014](#)), and from life-cycle trends in time use ([Aguilar and Hurst, 2007](#); [Ramey and Francis, 2009](#); [Ghez et al., 1975](#); [Juster and Stafford, 1991](#); [Robinson and Godbey, 1999](#)).

None of these studies or the subsequent research has directly addressed the validity of Mohring’s “*out of the air*” characterization of the rule of thumb for evaluating the costs (or gains in the case of savings) of changes in time in the transportation sector. In this paper, we estimate the value of time using two large-scale field experiments run on the Lyft rideshare platform. The experiments provide exogenous variation in passengers’ waiting time and price quotes. Passengers’ responses to these variations in waiting time and prices allow estimation of the marginal value of waiting time. The field experiments we study span 13 large U.S. metropolitan areas, 3.7 million unique passengers, and 14.8 million passenger sessions (periods of interaction with the app in which a user may request a ride).

Our results have at least two distinctive features and, as a result, add an important contextual detail to the existing literature. First, the observed choices are made in response to randomly-assigned variations in prices and waiting times. Other papers estimating rideshare demand have instead had to assume that certain determinants of prices and wait times are exogenous with respect to unobserved circumstances of choice. For prices, other studies consider variation due to discontinuities in surge pricing ([Castillo, 2024](#)), persistence in driver-set rates ([Buchholz et al., 2025](#)), and discontinuities in surcharges on downtown trips (?); for wait times, other studies consider variation in the location and characteristics of

nearby drivers (Castillo, 2024; Buchholz et al., 2025) and variation in travel times attributable to road infrastructure (?).¹ Since our results are based on *randomly assigned* wait times and prices, our value of time estimates do not require strong assumptions about how the context for users’ choices influences the variation in wait times or prices. Our overall average VOT estimates are close to those found in these other studies, helping to reduce concerns about the exogeneity assumptions they maintain to identify VOT.

A second distinction arises from the scale of our experiments, which include consumer choices across 13 metropolitan areas. Each area has a different set of public transit alternatives; population densities; spatial distributions of residential, commercial, and industrial activities; as well as different weather conditions during the time period associated with our experiments. The design of our experiments allows us to independently estimate VOTs across these different settings to consider how the context of users’ choices affect their demand. More specifically, we can consider how time of the day, weekday or weekend, location of a user, and weather influence the VOT estimates we recover from people’s decisions.

We find that the overall average estimate for the VOT is quite stable across the model specifications and the estimation decisions (e.g. selection of instruments) considered as part of evaluating the robustness of our results.

The random assignment together with differences in cities considered provides clear evidence that context matters. The “timing” of time use affects our VOT estimates in ways that are consistent with what we would expect. Weekends generally have lower VOT estimates than weekdays. Peak commute times matter. The availability of public transit also matters. For example, when a public alternative is close, the implied VOT is higher, suggesting that the trip requires a faster arrival time. When a public alternative is more distant or not available, the VOT estimate is generally lower. Snow matters: riders do not like to wait in poor weather conditions. Business users have higher estimated VOTs, perhaps due to the ability of these users to expense part or all of the (monetary) cost.

Thus, our approach addresses the key challenge faced in much of the empirical VOT literature: the research design implies the time-use component of a mode of transport is often correlated with other determinants of passenger demand. Our design ensures that differences in passenger demand are driven by exogenous differences in wait times or prices.

The two field experiments differ in their assignment mechanisms. The first field experiment (Experiment 1) randomized consumers to wait longer for a ride or not, and maintained

¹Rosaia (2025) differs in estimating preferences by regressing the *relative* shares of two competing platforms on the difference in their prices and wait times within location-hour blocks. He argues in effect that any endogenous demand shocks correlated with prices and wait times should cancel out between the two platforms.

treatment assignments for eight weeks. The waiting-longer treatment was implemented by removing cars from the dispatch queue for the treated users, generating variation in wait time. The second field experiment (Experiment 2) randomized location-hour blocks between multiple treatment levels: waiting an extra 60 seconds, 150 seconds, or 240 seconds, irrespective of local market conditions. This multiple treatment levels allow for an assessment of whether the VOT estimate is related to the length of waiting time that is exogenous to the circumstances of choice. To the best of our knowledge, this is the first field experiment to impose exogenous variation in the time the consumer faces to buy a good.

We draw several important conclusions from these experiments. First, we find that consumers respond to both wait times and prices. The response elasticities for prices are significantly larger than wait time elasticities in absolute magnitude. The VOT for waiting time implied by these two elasticities is \$19.38 per hour² (s.e. 1.39; all prices in 2015 dollars). However, there is clear evidence that the VOT is not constant and varies predictably with the context for users' choices.

Our second set of results concern the shape of the marginal VOT curve over different magnitudes of changes in wait time. In Experiment 2, the level of wait time varies orthogonally to all other customer and market characteristics. We find that on average, the wait-time elasticity is identical to that estimated in Experiment 1. However, the wait-time elasticity and the VOT estimates are significantly larger when the estimation strategy uses larger increments in waiting time. For example, waiting an extra 240 seconds generates a VOT estimate that is double the value of that generated from waiting an extra 60 seconds. This finding suggests that the VOT is convex over time (for the range of ETA variation in our experiments).

The remainder of the paper is structured as follows. Section 2 describes the market context for the Lyft rideshare service and the design of our field experiments. Sections 3 and 4 report empirical results from the two natural field experiments and detail our framework for exploring heterogeneity. Section 5 addresses potential issues in identification and external validity of our estimates, and Section 6 discusses the policy implications of our findings. Section 7 concludes. The online appendix includes supplemental analysis.

²We stress that while we measure the value of waiting time in the conventional dollars-per-hour units, our estimates are only representative of valuations for shorter (<15 minutes) stretches of time, and should not be extrapolated for valuing waiting time on the order of hours.

2 Background and Design of the Field Experiments

Two large-scale field experiments conducted in cooperation with Lyft provide the data for our analysis. Lyft is a ridesharing platform that matches consumers (passengers) searching for motor vehicle transportation with independent contractors (drivers) providing the service.³ Passengers access Lyft through a smartphone app. Figure 1 shows the user interface at the time of the experiment. Each opening of the Lyft passenger app starts a *session*, which is the unit of observation for our analysis. A session ends either when the passenger completes a trip or after 30 minutes of inactivity (whichever comes first).

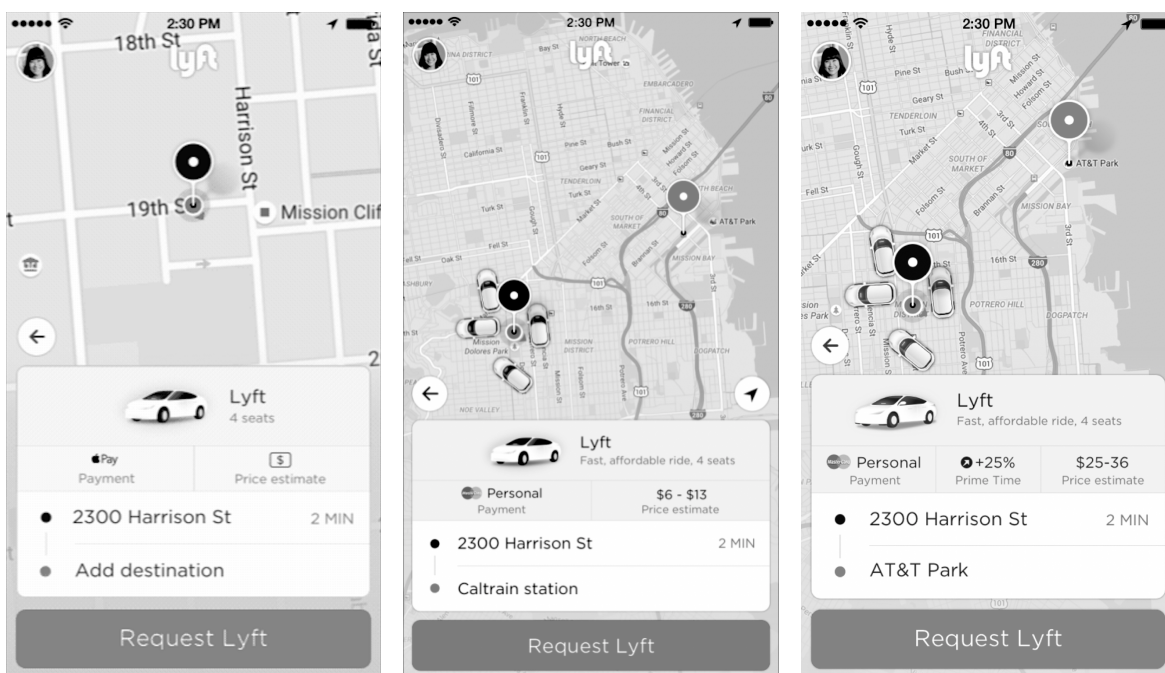


Figure 1
Screenshots of the passenger Lyft app’s user interface, at the time of the experiment. Left: no destination entered and no Prime Time active. Middle: destination entered and no Prime Time active. Right: destination entered and +25% Prime Time active.

The app shows the estimated wait time (“ETA”) and the price multiplier active for the user’s current location (“Prime Time” or “PT”). ETA is an estimate of the time, in minutes, that it would take a dispatched driver to reach the passenger’s location from the moment their request is accepted by that driver. PT is a dynamic mechanism that increases prices from a base level to balance the local demand (volume of Lyft ride requests) and supply

³Our discussion of the Lyft platform describes what existed in late 2015 through early 2016, when Experiment 1 was conducted. Today, Lyft also provides options for bikes, scooters, transit, and rental cars on its platform.

(available pool of Lyft drivers); a PT multiplier of +25% (1.25) means that a ride will cost 25% more than the usual base fare (which is a function of ride type, distance, and travel time for each service area). If the passenger enters a destination, they are also shown an estimated range for the base price of the trip (see middle and right screenshots in Figure 1). The passenger can hit the “Request Lyft” button to have a driver dispatched to them; at the time of the experiment, a request could be made without a destination entered.⁴

For each session, Lyft records: the user’s unique ID code; the session start time; the user’s current location, as a latitude/longitude pair and a geohash12;⁵ whether the user has a business account; the ETA and PT shown in the session;⁶ and counts of how many requests the user makes and rides the user completes. These data allow a number of variables to be defined that characterize the circumstances facing the potential passengers as they made their choices, such as the time of day, their location, and their past pattern of Lyft usage.⁷

2.1 Design of Experiment 1⁸

Experiment 1 was conducted in nine U.S. metro regions⁹ for eight weeks between December 2015 and January 2016 and involved 720,059 customers and 5,177,358 sessions. Random samples of all Lyft users in each city were randomly assigned to six experimental treatment

⁴Our primary model of passenger demand estimates sensitivity to price using experimental variation in the PT multiplier, and does *not* require information about the base price. This feature is important because many passengers request trips without entering a destination and receiving a base price estimate: during the first experiment, over 60% of sessions do not enter a destination and more than half of trip requests are made without a destination entered. As a robustness check, we estimate our model separately for users who do and do not enter a destination in Table B.14. The estimated value of time on the subpopulation entering a destination (\$8.20) is significantly lower than our main estimate. However, we also find that the experimental treatments themselves affect the decision to input a destination (Table B.15), so this estimate may be subject to selection bias.

⁵Geohash is a hierarchical spatial indexing system that assigns each location on Earth a twelve-digit code. The two levels of granularity we consider in this paper are the geohash5, which has a width and height of at most 4.89 kilometers, and the geohash7, which has a width and height of at most 153 meters.

⁶ETA and PT may vary over the course of a session due to real-time changes in local supply and demand. Unless otherwise noted, our analysis always focuses on the last shown ETA and PT in the session. ETAs vary over the course of a session in 51.3% of sessions, though the difference between the maximum and minimum ETAs shown is one minute or less in 76.4% of sessions. PT varies in only 11.9% of sessions.

⁷Lyft offers various ride modes, including Lyft (the standard mode); Lyft Line, now called Shared (in which several passengers share a single car with multiple pickups and dropoffs); and Lyft XL (which offers larger vehicles). At the time of the experiment, the majority of rides (about three-quarters) were standard Lyft rides. We include all ride types in our analysis, since treatment has some effect on ride types considered (see Table B.16).

⁸Experiment 1 was designed and run by Lyft to estimate long-term passenger sensitivity to variations in the price and waiting time. Lyft made available to us all the details of the experiment’s implementation.

⁹San Francisco, Austin, Atlanta, Miami, Los Angeles, San Diego, Boston, Seattle, and New York City.

groups.¹⁰ Users were assigned to one of two wait time treatments (normal or high) and one of three price treatments (low, normal, or high). The normal ETA, normal price combination constitutes the control group. In total, 37% of users in each city were included in the experiment: 15% of users were assigned to the control group, 2% to the high ETA high price group, and 5% each to each of the remaining four groups. The remaining 63% of users are excluded from our analysis because they were subject to other experiments conducted by Lyft concurrently with the one studied here. Throughout the eight weeks of the experiment, each individual remained in the same treatment.

The treatments varied the prices and wait times faced by users by modifying Lyft’s pricing and matching algorithms. Price variation was achieved by modifying the pricing algorithm for each user to make it more (for high-price treatment groups) or less (for low-price treatment groups) sensitive to local market conditions. This modification weakly raises or lowers a user’s PT and hence their effective price.¹¹ Waiting time variation was achieved by removing drivers from the nearest driver queue of each affected passenger: the nearest driver, all drivers whose ETA was within 30 seconds of that of the nearest driver, and one additional driver were removed from the queue. This removal has the effect of increasing wait times by at least 30 seconds for all passengers subject to the high wait time treatment.¹²

2.2 Design of Experiment 2

Experiment 2 ran for eight weeks from April 2017 to June 2017, and involved 3.3 million passengers and 9.7 million sessions across 10 metropolitan areas.¹³ This experiment randomized location (geohash7; roughly 153 by 153 meter squares)-hour blocks into one of four groups: (1) Control (10% of the sample); (2) ETA plus at least 60 seconds (5%); (3) ETA plus at least 150 seconds (3%); and (4) ETA plus at least 240 seconds (2%). To increase ETAs, Lyft removed the closest drivers from the dispatch queue until the ETA was increased

¹⁰Treatment groups were assigned by hashing each user’s unique Lyft ID, so that each group constitutes a representative sample of Lyft users in the affected cities during the time of the experiment. Table B.1 demonstrates balance on pre-experiment user behavior.

¹¹The change is only weak because PT takes values in a discrete set, and the sensitivity adjustment will not always be sufficient to push PT to a higher or lower level.

¹²Because the experiment involves only waiting time increases, it identifies a willingness to accept higher waiting times, which may differ from the willingness to pay for lower waiting times if users’ preferences are reference-dependent (De Borger and Fosgerau, 2008). Reference-dependence may be amplified by the presence of a stable outside alternative in the form of a competing rideshare service. However, when we estimate values of time by experience on Lyft (Table B.10), a potential proxy for reference-dependence, we do not find significant variation. A more complete test of reference dependence would require ETA variation in both directions over a longer duration.

¹³Los Angeles, San Francisco, New York City, Chicago, New Jersey, Boston, Philadelphia, D.C., Miami, and Atlanta.

by at least the desired amount for the treatment (60, 150, or 240 seconds). This multiple-treatment-level design allows examination of the curvature of the request demand in wait time, which is not possible with the single wait time treatment level of Experiment 1. By design, the increases in wait time were orthogonal to all other features of the location-hour block, such as availability of public transit. Because price multipliers were not randomized in this experiment, we focus on waiting time elasticities rather than VOTs in our analysis.

3 Summary Statistics

3.1 Experiment 1

Table 1 displays average ETA, PT, completed ride price, request rates, and session counts by treatment group. The high ETA treatment increased average wait times from 3.1 minutes to 4.6 minutes. The high-price treatment increased average PT levels from 10.0% to 16.3%, while the low-price treatment decreased the average PT levels to 3.7%. These PT differences resulted in completed ride prices that were 2.5% higher for passengers receiving high-price treatments and 2.0% lower for passengers receiving low-price treatments. Customers in the high wait time and high price treatments had fewer sessions, requests, and completed rides than control passengers (and vice versa for passengers in the low-price group).

Table 1

Average ETA, Prime Time, request rate, and completed ride price by treatment group.

	ETA (minutes)	Avg. PT (%)	Request Rate (%)	Completed Ride Price (\$)	Sessions (#)
Control	3.082 (0.004)	10.01 (0.03)	61.99 (0.08)	13.83 (0.02)	2,123,671
Normal ETA High Price	3.103 (0.008)	16.36 (0.07)	60.12 (0.14)	14.20 (0.04)	689,325
Normal ETA Low Price	3.085 (0.008)	3.65 (0.02)	63.36 (0.13)	13.54 (0.04)	722,367
High ETA Low Price	4.640 (0.011)	3.65 (0.02)	62.36 (0.13)	13.53 (0.04)	693,141
High ETA Normal Price	4.649 (0.011)	10.14 (0.06)	60.87 (0.14)	13.79 (0.04)	683,159
High ETA High Price	4.661 (0.017)	16.27 (0.11)	58.99 (0.21)	14.06 (0.06)	265,695

Notes: Standard errors clustered at the user level in parentheses.

3.2 Experiment 2

Table 2 provides the average ETAs and request rates for Experiment 2. As noted above, randomization of the three wait time variants was at the location-hour block level, rather than at the individual level. We report the number of location-hour blocks for each variant

along with the control values and the total number of sessions in each group. The request rate is highest in the control blocks and lowest in the plus 240 blocks.

Table 2
ETAs, sample sizes, and request rates in Experiment 2.

	Avg. ETA (minutes)	# of Location- Hour Blocks	# of Sessions	Req. Rate (%)
Control	3.433 (0.002)	3562926	4861532	62.02 (0.02)
Plus 60+ seconds	5.099 (0.003)	1771883	2410230	61.23 (0.03)
Plus 150+ seconds	6.332 (0.003)	1060521	1441839	60.31 (0.04)
Plus 240+ seconds	7.673 (0.004)	703933	955219	58.86 (0.05)

Notes: Standard errors clustered at the location-hour level in parentheses.

4 An Empirical Model of the Rideshare Decision

We hypothesize a simplified model for the rideshare request decision. We assume a passenger opens the app and receives a price and waiting time quote for a potential trip. Individual responses are indexed by passenger (i) and session (j). Equation (1) describes a reduced form model for the indirect utility difference dV_{ij} from requesting a ride as a function of the price P_{ij} , which is the product of a base price B_{ij} and a PT price multiplier $\mu_{ij} \equiv (1 + \text{PT}_{ij})$; the wait time ETA_{ij} ; and an unobserved shock ε_{ij} :

$$dV_{ij} = v(P_{ij}, \text{ETA}_{ij}) + \varepsilon_{ij} = v(B_{ij}\mu_{ij}, \text{ETA}_{ij}) + \varepsilon_{ij} \quad (1)$$

Each individual decides whether to request a trip based on whether their expectation (over the base price B_{ij} , which depends on the realization of traffic conditions) of this utility difference is positive.

The option to *not* request a trip is a composite of several alternative transit options (driving, public transit, walking, etc.), the characteristics of which we neither observe nor explicitly model. Though the characteristics of these options vary across sessions, randomization ensures that our experimental treatments are orthogonal to any such variation (at least at the start of the experiment). This orthogonality allows us to estimate price and wait time preferences without modeling the full choice set of each passenger, since the unobserved

choice attributes are on average the same between passengers in different treatments.¹⁴

To parameterize the model in (1), we assume a linear (in parameters) probability model of the request decision¹⁵ $\text{Request}_{ij} \in \{0, 1\}$ with price and wait time entering in logs:

$$\text{Request}_{ij} = \beta_1 \ln(\mathbb{E}_{ij}[P_{ij}]) + \beta_2 \ln \text{ETA}_{ij} + \varepsilon_{ij} = \beta_1 \ln(\mathbb{E}_{ij}[B_{ij}]) + \beta_1 \ln \mu_{ij} + \beta_2 \ln \text{ETA}_{ij} + \varepsilon_{ij} \quad (2)$$

Here $\mathbb{E}_{ij}$ denotes the individual’s (subjective) expectation; the individual is assumed to take the quoted ETA and PT multiplier as given, but must form an expectation about the base price. The logarithmic functional form in equation (2) allows observed choices to be used to estimate our measure for the price equivalent of waiting time without observing the expected base price $\mathbb{E}_{ij}[B_{ij}]$ for each session.¹⁶

The logarithmic formulation in (2) allows a separation of the base price and the price multiplier as determinants of the request, and identification of the price effect through experimental variation in the price multiplier. Since the expected base price is unobserved, we include this effect as one of the components of the model’s error. The marginal value of waiting time is (minus) the marginal rate of substitution of ETA_{ij} and P_{ij} , which depends

¹⁴Variation in the characteristics of the alternative options may correlate with Lyft prices and wait times, but our instrumental variable strategy allows us to isolate the variation in prices and wait times driven solely by the experimental randomization. Moreover, our main specification include controls for location (geohash5 dummies) and time (week of year and hour of week dummies) in our main specification. These controls proxy for the characteristics of alternative transit options, and their presence in the specification adjusts our estimates for any chance imbalances in the alternative sets across experimental treatment arms.

¹⁵Our dependent variable of interest is an indicator of whether a request is made, *not* whether the requested ride is successfully completed. Some requests are not completed because a passenger is not matched to any driver or the passenger or driver cancels a ride before it finishes; these situations amount to 0.6% and 10.0% of the total requests in Experiment 1, respectively. If some passengers always ignore the upfront ETA quote, make a request, and then cancel if the matched driver is too far away, focusing on the request indicator will lead to underestimates of the wait time elasticity and the VOT. Indeed, we estimate a significantly larger VOT when we use an indicator of ride completion as a dependent variable in the demand model. Nonetheless, our preferred specification focuses on requests because the request decision is made solely by the passenger, while the completion of a ride is a complex and dynamic function of both rider and driver(s) preferences and behaviors.

¹⁶Our model treats the decision to request each trip separately as a discrete choice. This formulation allows us to exploit a restriction to preferences introduced by [Mäler \(1971\)](#). He argued that a quality feature of a good or service (in our case, waiting time for a rideshare trip) is only relevant to people that consume that good or service; changes in the amount of that quality feature do not affect the utility of non-consumers of the good. This relationship was labeled weak complementarity. [Smith and Banzhaf \(2007\)](#) demonstrated that weak complementarity together with the [Willig \(1978\)](#) condition assure there is a price equivalent of a quality change. They concluded that if we can assume “weak complementarity and the Willig condition are satisfied, then the price index for quality is defined by the Marshallian consumer surplus per unit of the good experiencing the change and will be invariant with income” (p. 46).

on the values of P_{ij} and ETA_{ij} as well as β_1 and β_2 :

$$\text{VOT} = -\frac{dP_{ij}}{d\text{ETA}_{ij}} = \frac{\beta_2}{\beta_1} \frac{P_{ij}}{\text{ETA}_{ij}}. \quad (3)$$

Given that ε_{ij} in equation (2) includes the outside option for each session as well as unobserved factors that may simultaneously affect both the utility difference between rideshare trips and the unknown alternatives to these decisions as well as market conditions, it will not be independent of ETA_{ij} and μ_{ij} . We therefore use two-stage least squares (2SLS) with experimental treatment indicators as instruments to estimate the model. Our 2SLS system is:

$$\ln \text{ETA}_{ij} = \gamma_0 + \gamma_1 \text{T1}_{ij} + \gamma_2 \text{T2}_{ij} + \gamma_3 \text{T3}_{ij} + \gamma_4 \text{T4}_{ij} + \gamma_5 \text{T5}_{ij} + \gamma'_X X_{ij} + \eta_{ij} \quad (4)$$

$$\ln \mu_{ij} = \delta_0 + \delta_1 \text{T1}_{ij} + \delta_2 \text{T2}_{ij} + \delta_3 \text{T3}_{ij} + \delta_4 \text{T4}_{ij} + \delta_5 \text{T5}_{ij} + \delta'_X X_{ij} + \zeta_{ij} \quad (5)$$

$$\text{Request}_{ij} = \beta_0 + \beta_1 \ln \mu_{ij} + \beta_2 \ln \text{ETA}_{ij} + \beta'_X X_{ij} + \varepsilon_{ij}. \quad (6)$$

Here T1 through T5 are dummy indicators of the five non-control treatment assignments in Experiment 1 and X is a set of passenger and session characteristics, included solely to increase precision by absorbing variation in the outside option utility and the base price. Note that the treatments are randomized *unconditionally* of the covariates X ; as such, though our estimating equations feature several mutually-exclusive treatments, they are not subject to the contamination bias described by [Goldsmith-Pinkham et al. \(2024\)](#).¹⁷

Estimates of β_1 and β_2 together with assumed values for P and ETA yield an estimate of the marginal value of a unit of waiting time. We evaluate our marginal value at the control average waiting time $\overline{\text{ETA}}$ and completed ride price $\overline{P}$ during the experiment:

$$\widehat{\text{VOT}} = \frac{\hat{\beta}_2}{\hat{\beta}_1} \frac{\overline{P}}{\overline{\text{ETA}}}. \quad (7)$$

Though we focus on a linear probability model for our main results, we consider discrete-choice specifications in Section 5 below. Our main model also assumes $\ln P$ and $\ln \text{ETA}$ have constant effects on the probability of a request. If these effects are heterogeneous, our 2SLS estimates will be consistent for local averages of these effects. Due to non-linearity, however, the VOT expression in (7) evaluated at average price and ETA sensitivities will in general differ from the average VOT in the population. As a robustness check, we consider a more

¹⁷The unconditional exogeneity of our price and wait time variation is another advantage of our experimentally-randomized instruments over the quasi-experimental instruments employed in other recent studies.

flexible mixed-logit model of requests in Section 5.

4.1 Estimating the Choice Model

We estimate our parameters with 2SLS. Equations (4) and (5) define the first stage models and equation (6) the second stage. Our VOT estimator is (7). The covariate vector X consists of dummy variables for the session location, local hour of week and week of year; user experience with Lyft; and whether the user has a business profile. We also report demand elasticities, computed by dividing the semi-elasticities $\hat{\beta}_1, \hat{\beta}_2$ by the average request rate.

The necessary conditions to obtain consistent estimates of $\hat{\beta}_1, \hat{\beta}_2$ in (6) via 2SLS are that our instruments are (1) correlated with the endogenous variables $\ln(\text{ETA}_{ij})$ and $\ln(\mu_{ij})$ (conditional on controls) and (2) orthogonal to the error term ε_{ij} . F statistics from the first-stage regressions confirm that our instruments are relevant and strong by conventional standards (Staiger and Stock, 1997), satisfying the first condition. Exogeneity should follow directly from treatment randomization, but may fail if exposure to the treatment has cumulative effects on passengers, influencing their future behavior beyond the effect of the ETA and PT assigned for each session. However, when we estimate the model separately for each week or for each passenger’s j th session, we find little variation in VOT estimates over the course of the experiment (Table B.10). This result suggests that selection due to customers deciding not to use the app over time does not affect the estimated time/price trade-offs, and that our VOT estimates primarily reflect short-run effects of wait time and price on demand.¹⁸

We use the standard cluster-robust 2SLS variance-covariance estimator for $\hat{\beta}$ with clustering at the passenger level to match the experiment’s randomization. Standard errors for the VOT estimate are derived from the delta method.¹⁹

¹⁸Additionally, Experiment 2 randomizes at the location-time-level rather than the user-level, and finds similar ETA sensitivities; see Section 4.3 below.

¹⁹Treating average price and ETA as fixed, we define a nonlinear function $g : \mathbb{R}^3 \rightarrow \mathbb{R}$ so that our estimator of the VOT and its variance are given by:

$$g(\hat{\beta}) = \frac{\hat{\beta}_2}{\hat{\beta}_1} \frac{\bar{P}}{\bar{\text{ETA}}}, \quad \widehat{\text{Var}}[g(\hat{\beta})] = [\nabla g(\hat{\beta})]^T (\hat{\Sigma}/n) [\nabla g(\hat{\beta})],$$

where $\hat{\Sigma}/n$ is the estimated variance-covariance of $\hat{\beta}$.

4.2 Results: Experiment 1

Tables 3 and 4 provide the first- and second-stage results for the main model, estimated for the Experiment 1 sample.²⁰ We find that the experimental variants shifted wait times and prices in the intended directions, and that our instruments are strong (Staiger and Stock, 1997). In the second-stage results, we find that customers respond to higher wait times and higher prices by decreasing demand.

To calculate wait time and price elasticities, we divide the estimated semi-elasticities by the full-sample average request rate of 61.7%. We find that the request probability is much more sensitive to a change in price than to the same percent change in wait times: a 10% increase in price decreases the request rate by 5.9% while a 10% increase in the wait time decreases the rate by only 0.43%.²¹ Taken together, our estimated coefficients imply a VOT of \$19.38 per hour²² at baseline waiting time of 3.08 minutes and price of \$13.83 (the average wait times and actual fares for control group).

Our data allow us to examine how the price equivalent of wait time (see footnote 16) varies across different locations, times of day, weather conditions, and other features. We focus on heterogeneity in three types of features:

1. Individual trip characteristics
2. Market attributes
3. Non-price or time trip characteristics

Since session context is *not* exogenously assigned, differences in estimated VOTs between different contexts cannot be interpreted as “causal effects” of context. Differences in estimated VOTs between different locations, for example, may be driven by a variety of other contextual features that correlate with location, or by differences in the attributes of users who select into the market in different locations. We explore these selection/correlation issues with two reweighting exercises. In Appendix Section D, we reweight Experiment 1 data

²⁰The results of OLS estimation of equation (6) as well as a Durbin-Wu-Hausman exogeneity test are reported in Table B.3 in the Appendix. We find significant evidence that ln ETA is endogenous in (6). Table B.4 provides support for one possible explanation for this endogeneity: the level of and variation between ETAs are both correlated with exogenous demand shocks. In Table B.5, we show that flexible location-time controls do not suffice to remove this endogeneity.

²¹Our price elasticity estimate is consistent with Cohen et al. (2016), whose mean point estimate for the price elasticity of demand using 2015 Uber data is -0.57 .

²²Delta method standard error of \$1.39. A nonparametric pairs-cluster bootstrap (Cameron and Miller, 2015) of the full estimation procedure with 999 replications produced a standard error of \$1.40 and a bootstrap- t 95% confidence interval (Efron and Tibshirani, 1994) of (\$16.71, \$22.19).

Table 3
First-stage regression results for Experiment 1.

	ln ETA	ln ETA	ln μ	ln μ
High ETA High Price	0.474*** (0.004)	0.480*** (0.001)	0.049*** (0.001)	0.049*** (0.001)
High ETA Normal Price	0.473*** (0.003)	0.479*** (0.001)	0.001** (0.000)	0.001** (0.000)
High ETA Low Price	0.472*** (0.003)	0.479*** (0.001)	-0.041*** (0.000)	-0.042*** (0.000)
Normal ETA High Price	0.004 (0.003)	0.001 (0.001)	0.050*** (0.001)	0.050*** (0.000)
Normal ETA Low Price	0.001 (0.003)	0.001 (0.001)	-0.041*** (0.000)	-0.041*** (0.000)
Controls		✓		✓
N	5177358	5177358	5177358	5177358
R^2	0.101	0.534	0.026	0.212
F	13646***	116249***	12784***	20455***

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Standard errors clustered by user in parentheses. Independent variables are indicators for each treatment group. Controls include dummy variables for week of year, hour of week, user location (geohash5), business, airport, and decile of pre-experiment lifetime rides. F statistics are scaled cluster-robust Wald statistics of the null that the coefficients on the treatment indicators are identically 0. Kleibergen and Paap (2006) statistics testing underidentification of both endogenous regressors are 40175 and 48791 for the models without and with controls, respectively.

Table 4
Second-stage regression results for Experiment 1.

	(1)	(2)
ln ETA	−0.0231*** (0.0023)	−0.0264*** (0.0018)
ln μ	−0.3601*** (0.0165)	−0.3665*** (0.0132)
ETA Elasticity	−0.0374*** (0.0037)	−0.0427*** (0.0029)
PT Elasticity	−0.5838*** (0.0268)	−0.5942*** (0.0214)
VOT	17.27*** (1.77)	19.38*** (1.39)
Control Avg. ETA	3.08	3.08
Control Avg. Price	13.83	13.83
Control Request Rate	0.620	0.620
Controls		✓
N	5177358	5177358
R^2	−0.008	0.072
Hansen J	0.970	3.769

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Standard errors clustered by user in parentheses. ln(ETA) and ln μ instrumented by experimental group indicators. Value of time expressed in 2015 US dollars per hour. Controls include dummy variables for session location (geohash5), local hour of week and week of year, decile of user lifetime rides, whether the session is at and airport, and whether the user is a business user. Hansen J statistics test overidentification and follow an asymptotic $\chi^2(df = 5 - 2)$ under the null.

by the inverse of each user’s estimated propensity to select into a given context, ensuring that user attributes are balanced between the contexts we compare. In Appendix Section E, we reweight Experiment 2 data to balance the distributions of other observable covariates across different contexts, ensuring that these other covariates are balanced in each of our comparisons. Most of the heterogeneity results are robust to these reweightings.

Trip features. Panel 1 of Table 5 summarizes our findings concerning trip features. As expected, we find evidence that the opportunity cost of time is generally higher during weekdays (average VOT approximately \$20) than weekends (average VOT is about \$18). We also find that the VOT varies more within the course of a day on weekdays than weekends. For example on weekdays, we find pronounced and statistically significant differences in VOT estimates across times within the day: during morning and evening weekday commutes (6 to 10 AM and 4 to 7 PM) we observe higher VOT (averaging above \$24) compared to non-commuting time blocks (averaging below \$17). On weekends, VOT varies less over the course of the day. These comparisons suggest that waiting time is more valuable when the objective for the trip is likely to be work related.

Market factors. Panel 2 of Table 5 summarizes the effects of market factors on the VOT. We investigate how passenger location impacts the VOT several ways: by metro region; by airports and prominent downtown locations; and by distance to the nearest public transport station/stop. Point estimates of the VOT vary between regions, though we do not reject the null of homogeneity across regions ($p = 0.154$). The region-level estimates of VOT are positively correlated with local wage levels; we return to this correlation in Section 6.1.

The VOT estimates are larger for downtown locations than for non-downtown areas,²³ and higher for passengers at airports. The estimates decrease with distance to nearest public transit stop.²⁴

Other trip characteristics. Panel 3 of Table 5 considers other features of a trip that may affect marginal sensitivity to price and waiting time. Precipitation raises the estimated VOT, suggesting that longer waiting times are less desirable in poor weather. With snow, the VOT estimate is nearly 40% larger than when there is no precipitation.

Estimates for the VOT associated with business and non-business users differ in pre-

²³Areas tagged as “downtown” are generally the city proper: for example, the “downtown” area of the San Francisco region is the city of San Francisco, while the entire region extends north through Marin County, south to Palo Alto, and east through Morgan Territory Regional Preserve. Maps are available from the authors upon request.

²⁴Counterintuitively, users are *more* price elastic when farther from public transit (Table B.6), even after correcting for observable selection (Table D.1). This pattern may arise because users far from public transit also have more ability to substitute into private vehicle travel.

Table 5
Summary of VOT results for Experiment 1.

1. Opportunity Cost		2. Market Factors		3. Other Trip Features	
Sub-sample	VOT	Sub-sample	VOT	Sub-sample	VOT
Full Sample	19.38*** (1.39)	Full Sample	19.38*** (1.39)	Full Sample	19.38*** (1.39)
Day of Week		Region		Precipitation	
Monday	20.08*** (2.47)	Austin	11.37** (4.76)	None	19.04*** (1.50)
Tuesday	17.95*** (2.22)	San Francisco	21.08*** (3.79)	Rain	23.10*** (2.26)
Wednesday	19.34*** (2.38)	Los Angeles	19.52*** (3.91)	Snow	26.56*** (7.83)
Thursday	19.44*** (2.41)	Miami	16.36*** (3.95)	User Type	
Friday	21.26*** (2.45)	Boston	23.37*** (3.79)	Business	24.42** (10.55)
Saturday	18.74*** (2.04)	Seattle	27.08*** (6.38)	Non-Business	19.22*** (1.39)
Sunday	17.12*** (2.01)	San Diego	14.80*** (4.34)		
Time of Day (Weekends)		Atlanta	11.24** (4.40)		
6–10 AM	17.40*** (3.80)	New York City	23.85*** (3.81)		
10 AM–4 PM	18.92*** (2.49)	Location Type			
4 PM–7 PM	20.16*** (2.92)	Airport	27.55 (18.02)		
7 PM–11 PM	15.04*** (2.87)	Non-Airport	18.84*** (1.35)		
11 PM–6 AM	14.71*** (3.12)	Downtown	21.64*** (1.99)		
Time of Day (Weekdays)		Non-Downtown	17.88*** (1.97)		
6–10 AM	26.71*** (2.50)	Distance to Public Transit			
10 AM–4 PM	18.76*** (2.64)	Under 50 Meters	23.81*** (2.56)		
4 PM–7 PM	22.12*** (2.70)	50 to 200 Meters	20.81*** (1.91)		
7 PM–11 PM	13.62*** (2.59)	200 to 800 Meters	16.90*** (2.03)		
11 PM–6 AM	19.03*** (2.66)	Over 800 Meters	10.88*** (2.82)		

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Standard errors clustered by user in parentheses. Controls include local week of year, local hour of week, user location (geohash5), business user, airport, and decile of pre-experiment lifetime rides.

dictable ways, with business users having a VOT about 25% higher.²⁵ This result is consistent with what we would expect if business travelers bear only part of the cost of the trip.

4.3 Results: Experiment 2

Results for Experiment 2 are presented in three parts. First, we summarize the overall estimates from this experiment and consider how the properties of the situation affect waiting-time elasticities. We then compare the results to those from Experiment 1. Finally, we use the design of the experiment to evaluate the shape of the estimated wait-time elasticities with respect to the size of the change in the wait-time.

Table 6 provides ETA elasticity estimates. Estimates from Experiment 2 are quite close to the estimates from Experiment 1; the overall (pooling across treatment levels) elasticity in Experiment 2 (-0.0432) is not significantly different from that in Experiment 1 at the 5% level. When we focus on data drawn from the roughly 200,000 customers who are in both field experiments, the estimated VOTs (\$22.15 in Experiment 1 and \$22.71 in Experiment 2) are not significantly different at the 5% level (Table B.18).

Column 1 of Table 6 explore heterogeneity in ETA elasticities by day and by various times of the week. We find qualitatively similar patterns in the ETA elasticities for different trip features as we found in Experiment 1: weekdays have larger time elasticities than weekends, and weekday elasticities are largest during peak commuting times.

Column 2 in Table 6 illustrates how the ETA elasticities vary with other trip characteristics such as inclement weather and business users. We again find that time elasticities are larger (in absolute magnitude) in rainy conditions: -0.047 versus -0.042 when there is no precipitation (Experiment 2 has no sessions with snow). We find that business users exhibit a larger (in absolute magnitude) time elasticity than non-business users (-0.049 vs. -0.043) at the $p < 0.01$ level, despite facing lower ETAs on average.

In column 3 of Table 6, we investigate how ETA elasticities vary spatially. Consistent with Experiment 1, we find differences across regions in our sample, with the absolute magnitude of the elasticities largest in New York and Washington, D.C. (-0.086 and -0.054) and smallest in Miami and San Francisco (-0.021 and -0.031). Non-downtown sessions show a larger (in absolute magnitude) time elasticity than downtown sessions (-0.048 vs. -0.039) at the $p < 0.01$ level, though the average ETA is higher in non-downtown sessions than downtown sessions.

²⁵The difference is primarily driven by business users having a smaller price elasticity: -0.398 vs. -0.605 (Table B.6).

Table 6
Summary of elasticity results: heterogeneity (Experiment 2).

Time of Week	(1)	Trip Characteristics	(2)	Location	(3)
Full Sample	−0.043*** (0.001)	Full Sample	−0.043*** (0.001)	Full Sample	−0.043*** (0.001)
Day of Week		Precipitation		Region	
Monday	−0.044*** (0.002)	None	−0.042*** (0.001)	San Francisco	−0.031*** (0.001)
Tuesday	−0.047*** (0.002)	Rain	−0.047*** (0.002)	New York City	−0.086*** (0.002)
Wednesday	−0.044*** (0.002)	User Type		Chicago	−0.039*** (0.002)
Thursday	−0.047*** (0.002)	Business	−0.049*** (0.002)	Washington, D.C.	−0.054*** (0.002)
Friday	−0.042*** (0.002)	Non-Business	−0.043*** (0.001)	Miami	−0.021*** (0.002)
Saturday	−0.039*** (0.002)			New Jersey	−0.053*** (0.004)
Sunday	−0.040*** (0.002)			Boston	−0.055*** (0.003)
Time of Day (Weekends)				Philadelphia	−0.040*** (0.003)
6–10 AM	−0.060*** (0.004)			Atlanta	−0.037*** (0.004)
10 AM–4 PM	−0.048*** (0.003)			Los Angeles	−0.033*** (0.002)
4 PM–7 PM	−0.044*** (0.003)			Location Type	
7 PM–11 PM	−0.033*** (0.003)			Downtown	−0.039*** (0.001)
11 PM–6 AM	−0.026*** (0.002)			Non-Downtown	−0.048*** (0.001)
Time of Day (Weekdays)				Distance to Public Transit	
6–10 AM	−0.061*** (0.002)			Under 50 Meters	−0.042*** (0.001)
10 AM–4 PM	−0.041*** (0.002)			50 to 200 Meters	−0.038*** (0.001)
4 PM–7 PM	−0.052*** (0.002)			200 to 800 Meters	−0.049*** (0.002)
7 PM–11 PM	−0.035*** (0.002)			Over 800 Meters	−0.084*** (0.004)
11 PM–6 AM	−0.037*** (0.002)				

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Standard errors clustered by location-hour in parentheses. Controls include local week of year, local hour of week, user location (geohash5), business user, and decile of pre-experiment lifetime rides.

Table 7
VOT estimates at different ETA treatment levels.

	All Experiment 2	Six Overlap Regions	San Francisco	Los Angeles	Miami	Boston	Atlanta	New York City
ETA Elast. (Plus 60+)	−0.027*** (0.001)	−0.028*** (0.002)	−0.015*** (0.003)	−0.027*** (0.003)	−0.015*** (0.004)	−0.038*** (0.005)	−0.025*** (0.007)	−0.054*** (0.004)
ETA Elast. (Plus 150+)	−0.037*** (0.001)	−0.036*** (0.001)	−0.025*** (0.002)	−0.026*** (0.002)	−0.019*** (0.003)	−0.046*** (0.004)	−0.032*** (0.006)	−0.080*** (0.003)
ETA Elast. (Plus 240+)	−0.053*** (0.001)	−0.050*** (0.001)	−0.039*** (0.002)	−0.040*** (0.002)	−0.024*** (0.003)	−0.067*** (0.004)	−0.044*** (0.006)	−0.099*** (0.003)
PT Elast. (2015)	—	−0.609*** (0.035)	−0.469*** (0.046)	−0.482*** (0.105)	−0.482*** (0.127)	−0.531*** (0.112)	−0.901*** (0.313)	−0.968*** (0.062)
VOT (Plus 60+)	—	11.19*** (0.90)	8.98*** (1.73)	11.73*** (2.85)	6.44*** (2.46)	16.16*** (3.98)	4.10** (1.88)	21.28*** (2.14)
VOT (Plus 150+)	—	14.45*** (0.96)	14.84*** (1.83)	11.48*** (2.68)	7.97*** (2.50)	19.45*** (4.41)	5.35** (2.12)	31.48*** (2.41)
VOT (Plus 240+)	—	20.04*** (1.24)	23.15*** (2.49)	17.36*** (3.89)	10.32*** (2.98)	27.99*** (6.09)	7.28*** (2.69)	39.09*** (2.78)
Base Avg. ETA (2017)	3.43	3.31	2.70	3.34	3.23	3.47	4.99	3.24
Control Avg. Price (2015)	—	12.50	11.45	10.90	9.65	12.23	11.53	19.90
Controls	✓	✓	✓	✓	✓	✓	✓	✓

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Standard errors clustered by location-hour (ETA elasticities) or user (PT elasticities) in parentheses. In ETA and $\ln \mu$ instrumented by experimental treatment indicators. Controls include local hour of week and week of year, session geohash5, business user, and decile of pre-experiment lifetime rides.

Finally, distance to public transit matters. The largest ETA elasticities are observed for passengers more than 800 meters from their nearest public transit stop. For Experiment 1, we focused on distance to public transit as a proxy measure for the effect of public substitutes on the price elasticity. In Experiment 2, we consider the role of waiting time and our findings are counter to what we might have expected. Passengers with better outside transportation options would be expected to respond more elastically to larger amounts of waiting times. However, these counterintuitive findings are not clear-cut. For example, the distance to public transit is correlated with other relevant characteristics of the trip, such as the base ETA. When we reweight observations in each subsample to balance these characteristics (Appendix Section E), the pattern reverses.

The design of Experiment 2 allows our analysis to gauge the implied nonlinearity in the relationship between waiting time and passenger demand, independently of the base wait time. The three treatments in the experiment allows four levels of wait time to be compared. Each can serve as an instrument to estimate the ETA elasticity of demand at different points. Table 7 presents the VOT estimates at different ETA treatment levels.

We consider three sub-samples: all Control and Plus 60 sessions; all Control and Plus 150 sessions; and all Control and Plus 240 sessions. In each sub-sample, we estimate the trip request equation (6), using the indicator of treatment (respectively Plus 60, Plus 150, and Plus 240) as a single instrument for $\ln \text{ETA}$. The resulting coefficient estimates are consistent for weighted averages of the expected derivative of the request rate with respect to $\ln \text{ETA}$ over the ETAs between the two treatment levels in the sub-sample (Angrist et al., 2000). We

divide these coefficients by the average request rate in the sub-sample to estimate an ETA elasticity. As ETAs increase from one sub-sample to the next, comparing these elasticities provides insight into how the ETA elasticity varies with ETA.

The first column of Table 7 summarizes the results. We find that for “small” increases in the amount of wait time, each 1% ETA increase translates to a 0.027% decrease in request rate; a 0.037% decrease for “moderate” increases in the ETA; and 0.053% for “large” increases in the ETA. This pattern suggests that the sensitivity to waiting time increases in magnitude as the time exceeds the base waiting time.²⁶

For the six regions which appear in both Experiment 1 and Experiment 2, we estimate average prices and price elasticities from Experiment 1 and average ETAs and ETA elasticities from Experiment 2 for each of the four treatment levels. We combine these estimates to obtain VOTs for each of the four experimental ETA levels. Figure A.4 displays these results against mean after-tax wages.²⁷ We note that for the full sample and each individual region, the estimated VOT is highest when using the largest ETA treatment as an instrument, suggesting that the value of a marginal minute increases as more minutes are added to the base wait time.

This finding is consistent with the existing literature. For example, values of (in-vehicle) travel time have been found to be positively correlated with trip length in many transportation studies (Daly and Carrasco, 2009). Our results take advantage of experimental randomization and thus rule out the possibility that the pattern is driven only by selection or unobserved confounding.

²⁶An alternative interpretation of the effect is that consumers’ responses may reflect a belief that a longer ETA implies greater congestion, and that their travel time will be longer than estimated by the app. We argue that this interpretation is not supported by the structure of the experiment. First, if ETA increases by one minute on the app, then the estimated time to destination (ETD) is also updated by one minute, to reflect the increase in pickup time. Second, it seems unreasonable to expect that consumers believe the quoted ETA but not the quoted ETD. Third, in Experiment 2, ETA elasticities do not vary with the estimated pickup-to-dropoff time; the estimate elasticity is -0.023 in both the first and fifth quintiles of pickup-to-dropoff time (Table B.19). This pattern suggests that the elasticities and VOTs we have estimated capture passengers’ responsiveness to pre-trip waiting time, not a combination of pre-trip and in-vehicle time.

²⁷To ensure that our price and time elasticity estimates are comparable yet independent, we subset both experiments to observations from the six overlapping regions, remove airport sessions from the first data, and remove from Experiment 1 observations from any users who appear in Experiment 2.

5 Issues in Identification and External Validity

5.1 Functional Form and Specification

Our primary model has wait time entering the choice equation in logarithmic form. Estimating the model with ETA entering in levels leads to a small and insignificant decrease in the estimated VOT, to \$18.01 (1.29) (see Table B.12). Because Experiment 1 involves only a binary ETA treatment, it provides no empirical basis for deciding between different functional forms that would modify the role of wait-time and assess the impact on choices with these data. Experiment 2 randomizes sessions between multiple levels of ETA increases, offering some insight into the local curvature of the request rates in response to ETA. Nonetheless, the design is not able to isolate one functional form as “best” fitting.

Our primary model is a simple linear probability model (LPM) with constant coefficients, rather than a more complex random-coefficient discrete-choice model. The decision to use an LPM rather than a constant-coefficient discrete-choice model is not consequential: using an IV-probit model with constant coefficients, we estimate a VOT of \$17.53 (1.78), close to our main LPM estimate (see Table B.11).

To assess the effects of allowing for random coefficients, we estimate a mixed logit model of the request decision in Appendix Section F. We specify the utility of a request as

$$u_{ij} = \beta_0 - \exp(\beta_{ij}^1)P_{ij} - \exp(\beta_{ij}^2)ETA_{ij} + \varepsilon_{ij}$$

where $(\beta_{ij}^1, \beta_{ij}^2)'$ is bivariate normal and ε_{ij} is logistic. The session-specific VOT is $\exp(\beta_{ij}^2 - \beta_{ij}^1)$ and follows a lognormal distribution. We estimate the mixed logit model by maximum likelihood, using a control function approach to address endogeneity in prices and wait times. Estimating such a model requires stronger functional form and distributional assumptions than the LPM, but has the advantage of estimating an entire distribution of VOTs across sessions. The mean of the VOT distribution (\$31.82) from the estimated mixed logit model is larger than our main estimate, but the median (\$22.43) is close to our full-sample LPM estimate.

5.2 Endogenous Opening of the App

Another concern arises from the fact that our VOT is estimated only when riders are considering using the service. If users only open the Lyft app when facing important time

constraints, our estimates may overestimate the overall average VOT.²⁸ As a robustness check, we use the frequency of Lyft usage before the experiment period as a proxy for the urgency of the trip (assuming that frequent Lyft users are less likely to be opening the app only when facing time constraints). When we compare VOTs across quartiles of pre-period rides taken, the results suggest the selective opening of the app does not greatly influence our estimates. The VOT estimates are gradually decreasing in the number of prior 28 day rides, with the estimate for users with more than 8 rides about 20% lower than our main estimate. We do not find any significant pattern of VOT across lifetime rides taken (see Table B.10).

5.3 Contamination

When a user decides whether to request a ride, their decision affects the available supply in their local market, which may in turn influence the ETAs and request decisions of other users. This channel introduces interaction effects in the experiment: the outcome of one session can be dependent on the treatment assignments of other sessions, since those treatment assignments affect overall available supply. While these interaction effects would violate SUTVA, we argue that the violation is relatively unimportant, because (i) these interaction effects on average uniformly affect all sessions in the market, in and outside of the experiment; (ii) the closest data available to shed light on the size of these interaction effects (Appendix E) suggests that their magnitude is small (on the order of a few seconds); and (iii) even in the presence of such interactions, we can still interpret our estimates as weighted average derivatives of individual demand functions (Appendix G).

5.4 External Validity

One concern with our elasticity and VOT estimates is that they are based on a particular sub-population (Lyft passengers) that is not representative of the broader population. To assess the severity of this external validity issue, we use data from the 2017 National Household Travel Survey (Federal Highway Administration, 2017). The survey reports person-level data on a number of demographic characteristics, including race, gender, income, and education; person-level weights, which can be used to extrapolate the survey results to the full U.S. population; and detailed information on travel patterns, including the answer to a question about use of rideshare services: *“In the past 30 days, how many times have you*

²⁸This issue highlights the importance of the weak complementarity assumption underlying our estimation of a price equivalent of a change in wait time.

purchased a ride with a smartphone rideshare app (e.g. Uber, Lyft, Sidecar)?”

We define “Rideshare Users” to be all persons in the survey who reported a positive number of rideshare trips. In Figure 2, we subset the NHTS data to persons who live in core-based statistical areas (CBSAs) corresponding to the nine regions in Experiment 1,²⁹ and compare Rideshare Users to the general population in the survey along a number of demographic dimensions. We use the NHTS person-level weights throughout this analysis.

Rideshare users have similar race and gender distributions to the full population. They are more likely to be in the upper tails of the income and education distributions; to be in the 20–40 years age bracket; and to fall within the lower end of the household size distribution. Rideshare users are also more likely to reside in urban areas.

To adjust for these demographic differences, we reweight our data by the inverse of each passenger’s propensity to use rideshare. Specifically, we proceed as follows:

1. We use the NHTS person-level data to estimate a probit model of an individual’s propensity to use rideshare in the regions in our first experiment, as a function of their CBSA, income level, and home Census tract population density.³⁰
2. For each user in Experiment 1, we impute a home location (either a latitude-longitude pair or a ZIP Code Tabulation Area), and use this home location to estimate their CBSA, Census tract, and household income.³¹
3. Using the model estimated in Step 1 and the demographic features imputed in Step 2, we calculate the propensity to use rideshare of each user in our first experiment.
4. We estimate the main model specification, using inverse propensity weights based on the propensities computed in Step 3 and selecting each user’s first session in the experiment (to account for the fact our weights are at the user-level rather than the session level).

Table 8 reports the results, both with and without the inverse probability weights.³² VOT point estimates are quite similar, suggesting that our estimates are not badly biased by the selective nature of rideshare users. We caveat, however, that we are only able to

²⁹The total population of these nine CBSAs was about 65 million in 2020.

³⁰The NHTS person-level weights are used in the estimation of the probit model.

³¹Passenger home locations are inferred through a combination of billing addresses, in-app passenger-set shortcuts, and frequent pickup and destination locations. Income is imputed as the median household income in the home Census block group (when precise home coordinates are available) or ZCTA (when only a home ZIP is available) using data from the 2017 American Community Survey (ACS). ZIP codes were mapped to CBSAs and Census tracts using the HUD-USPS ZIP Code Crosswalk data ([Office of Policy Development and Research, 2016](#)).

³²A small number of users are dropped because median income data for their home Census tracts was unavailable.

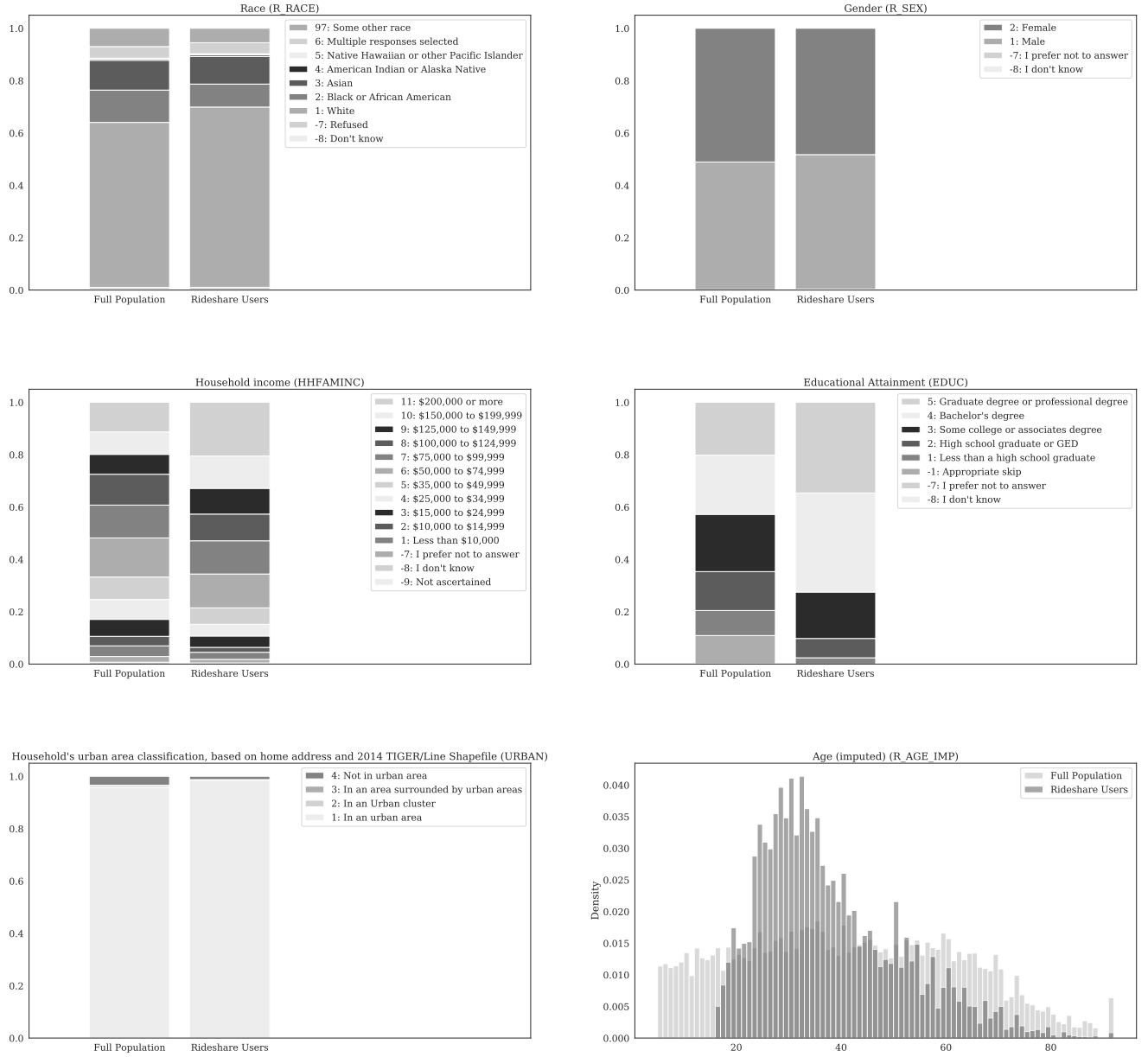


Figure 2
Comparison of demographic characteristics of rideshare users and the general population in the CBSAs sampled for Experiment 1. Based on data from the National Household Travel Survey ([Federal Highway Administration, 2017](#)).

correct for selection using the observable features of individuals who choose to use rideshare; our approach is unable to correct for any unobserved selection.

Table 8

2SLS regression results for each passenger’s first session in Experiment 1, with and without weighting by inverse propensity to use rideshare services.

	Unweighted	Weighted
ln ETA	−0.0264*** (0.0027)	−0.0230*** (0.0046)
ln μ	−0.3439*** (0.0180)	−0.3098*** (0.0353)
ETA Elasticity	−0.0447*** (0.0046)	−0.0390*** (0.0079)
PT Elasticity	−0.5833*** (0.0305)	−0.5253*** (0.0600)
VOT	23.39*** (2.50)	22.84*** (4.89)
Control Avg. ETA	3.35	3.53
Control Avg. Price	17.05	18.12
Control Req. Rate	0.593	0.594
Controls	✓	✓
N	694703	694703
R^2	0.070	0.073

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Standard errors clustered by user in parentheses. ln ETA and ln μ instrumented by experimental group indicators. Value of time expressed in 2015 US dollars per hour. Controls include local week of year, local hour of week, user location (geohash5), business user, airport, and decile of pre-experiment lifetime rides.

6 Policy Implications

6.1 The Relationship Between VOT and Wages

Our estimates allow us to test whether conventional approximations for values of travel time—frequently used in policy evaluations for cost-benefit analyses—provide reasonable approximations.

Consider the [USDOT \(2015\)](#) recommendation that the (in-vehicle) VOT be approximated as 50% of hourly earnings for personal travel (the same as [Small \(2001\)](#)). To assess the accuracy of this rule-of-thumb, we compare regional estimates from Experiment 1 to various fractions of mean and median wage rates. All tests are one-sided asymptotic z -tests based on the VOT estimates, standard errors, and wages (treated as constant) reported for each region in Table [B.8](#). We explore all approximations with specifications of the request rate equation having ETA enter in log or in level form.

Overall, we find the least evidence in favor of rejecting the 1/2 or more pre-tax mean wage and 100% or more post-tax median wage earnings rules (Table [B.9](#)). We do not reject the hypothesis that the VOT is above 1/2 wage rates for any region, but do reject the hypothesis that the VOT is above 100% the wage for some regions when using the linear specification. Importantly, the 150% rule is rejected in many regions, suggesting either that the value of in-vehicle time is *not* more than 150% of the wage rate.

Figure [3](#) plots our estimates with the post-tax mean wages (and policy-relevant fractions of these quantities) for each region. The estimates also provide insight into the cross-sectional relationship of the VOT and the mean wage rate: we estimate a correlation of 0.78 ($p = 0.014$) and an elasticity of 1.3 (Table [B.21](#)).³³

Figure [A.4](#) plots our Experiment 2 estimates alongside the post-tax mean wages. The high treatment group has a significantly higher VOT than the low treatment group in all six regions and is not significantly different than the post-tax mean wage rate of those six regions. Atlanta is the only city where the high treatment is not at or above the mean wage rate.

6.2 Applications to Public Policy

Many transportation-related cost-benefit analyses use the recommendations put forth by the [USDOT \(2015\)](#), which link the (in-vehicle) VOT to the median or mean wage in an area.

Some of our findings for the value of waiting time are consistent with this recommendation: the average values for our VOT estimates are positively correlated with the wage rates across regions. We estimate a cross-region elasticity of VOT with respect to the average

³³Price elasticities are also positively correlated with local wages (Table [B.22](#)), though the relationship is not statistically significant in our sample.

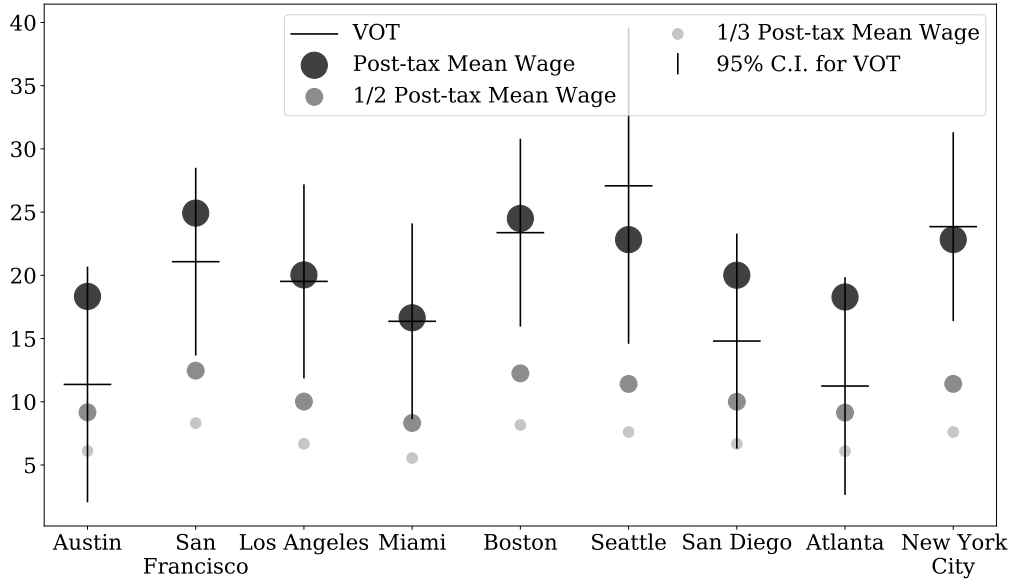


Figure 3
Estimated VOTs vs. (fractions of) mean wages for each region.

wage rate³⁴ that is approximately 1. We also find that the estimated VOT is closer to the full wage rate during the typical work commuting hours than at other times. When the primary impact of a policy is likely to affect the times in travel related to work any savings (or increases in travel time) should be valued at a rate closer to the full hourly wage rate.

Our findings also indicate that estimates of the VOT should be selected recognizing the context associated with each situation. Our analysis of how the VOT varies across trip characteristics and market factors (Sections 4.2 and 4.3) suggests that the price equivalent for a unit of waiting time varies with circumstances of choice.

7 Conclusion

Readers who have “arrived at this point” in learning about our study have surely invested a fair amount of time. We hope that your decision, ex post passes a benefit cost test. Time

³⁴Our average wage data are the all-occupations mean wages reported by the [U.S. Bureau of Labor Statistics \(2016\)](#) for the metropolitan area corresponding to each Lyft region in our experiment. To the extent Lyft users are a selected subset of the population used by the BLS to estimate these mean wages, we might expect differences from the actual average wages of Lyft passengers in each region. Our comparison is intended as an approximate gauge of the connections implied by economic theory as we explained earlier.

is the ultimate scarce resource, and its value has deep implications for a range of economic phenomena and investment decisions.

We report several insights. First, we estimate a VOT that is roughly \$19 per hour (2015 prices). This estimate is 67% of the mean pre-tax wage rate for the various regions in the experiment, which is quantitatively different from the VOT findings of previous empirical studies (Small and Verhoef, 2007). Second, consistent with standard microeconomic models (Becker, 1965; DeSerpa, 1971), the VOT is related to the available substitute set, the context for time allocation decisions, and other key features of the trip that impact either the marginal benefits or the marginal costs of the trips.

Future studies should consider the value of reliability—the extent to which waiting times vary about their mean—simultaneously with the VOT.

References

- Aguiar, M. and E. Hurst (2007). Measuring trends in leisure: The allocation of time over five decades. *Quarterly Journal of Economics* 122(3), 969–1006.
- Angrist, J. D., K. Graddy, and G. W. Imbens (2000). The interpretation of instrumental variables estimators in simultaneous equations models with an application to the demand for fish. *The Review of Economic Studies* 67(3), 499–527.
- Becker, G. S. (1965). A theory of the allocation of time. *Economic Journal*, 493–517.
- Buchholz, N., L. Doval, J. Kastl, F. Matejka, and T. Salz (2025). Personalized pricing and the value of time: Evidence from auctioned cab rides. *Econometrica* 93(3), 929–958.
- Cameron, C. A. and D. L. Miller (2015). A Practitioner’s Guide to Cluster-Robust Inference. *Journal of Human Resources* 50(2), 317–372.
- Castillo, J. C. (2024). Who Benefits from Surge Pricing? *Available at SSRN 3245533*.
- Cohen, P., R. Hahn, J. Hall, S. Levitt, and R. Metcalfe (2016). Using big data to estimate consumer surplus: The case of Uber. Technical report, National Bureau of Economic Research.
- Daly, A. and J. A. Carrasco (2009). The influence of trip length on marginal time and money values: An alternative explanation. In R. Kitamura, T. Yoshii, and T. Yamamoto (Eds.), *The Expanding Sphere of Travel Behaviour Research: Selected Papers from the Proceedings of the 11th International Conference on Travel Behaviour Research*, Bingley, United Kingdom, pp. 673–701. Emerald Group Publishing.
- De Borger, B. and M. Fosgerau (2008). The trade-off between money and travel time: A

- test of the theory of reference-dependent preferences. *Journal of Urban Economics* 64(1), 101–115.
- Deacon, R. T. and J. Sonstelie (1985). Rationing by waiting and the value of time: Results from a natural experiment. *Journal of Political Economy* 93(4), 627–647.
- DeSerpa, A. C. (1971). A theory of the economics of time. *Economic Journal* 81(324), 828–846.
- Efron, B. and R. J. Tibshirani (1994). *An introduction to the bootstrap*. CRC press.
- Federal Highway Administration (2017). National Household Travel Survey.
- Ghez, G., G. S. Becker, et al. (1975). The allocation of time and goods over the life cycle. *NBER Books*.
- Goldsmith-Pinkham, P., P. Hull, and M. Kolesár (2024, December). Contamination Bias in Linear Regressions. *American Economic Review* 114(12), 4015–51.
- Juster, F. T. and F. P. Stafford (1991). The allocation of time: Empirical findings, behavioral models, and problems of measurement. *Journal of Economic Literature* 29(2), 471–522.
- Kleibergen, F. and R. Paap (2006). Generalized reduced rank tests using the singular value decomposition. *Journal of Econometrics* 133(1), 97 – 126.
- Mäler, K.-G. (1971). A method of estimating social benefits from pollution control. *The Swedish Journal of Economics* 73(1), 121–133.
- Mohring, H. (1965). Urban highway investments. *Measuring benefits of government investments*, 231–291.
- Office of Policy Development and Research (2016). HUD USPS ZIP Code Crosswalk Files.
- Ramey, V. A. and N. Francis (2009). A century of work and leisure. *American Economic Journal: Macroeconomics* 1(2), 189–224.
- Robinson, J. P. and G. Godbey (1999). Time for life: The surprising ways americans use their time. 2nd.
- Rosaia, N. (2025). Competing Platforms and Transport Equilibrium. *Econometrica* 93(6), 2235–2271.
- Small, K. A. (2001). *Urban Transportation Economics*. Fundamentals of pure and applied economics. Routledge.
- Small, K. A. and E. T. Verhoef (2007). *The Economics of Urban Transportation*. Routledge.
- Small, K. A., C. Winston, and J. Yan (2005). Uncovering the distribution of motorists’ preferences for travel time and reliability. *Econometrica* 73(4), 1367–1382.
- Smith, V. K. and S. Banzhaf (2007). Quality adjusted price indexes and the Willig condition. *Economics Letters* 94(1), 43–48.

- Smith, V. K. and C. Mansfield (1998). Buying time: Real and hypothetical offers. *Journal of Environmental Economics and Management* 36(3), 209–224.
- Staiger, D. and J. H. Stock (1997). Instrumental variables regression with weak instruments. *Econometrica* 65(3), 557–586.
- U.S. Bureau of Labor Statistics (2016). May 2016 metropolitan and nonmetropolitan area occupation employment and wages estimates.
- USDOT (2015). Revised departmental guidance on valuation of travel time in economic analysis. *US Department of Transportation. Washington, DC.*
- Willig, R. D. (1978). Incremental consumer’s surplus and hedonic price adjustment. *Journal of Economic Theory* 17(2), 227–253.
- Wolff, H. (2014). Value of time: Speeding behavior and gasoline prices. *Journal of Environmental Economics and Management* 67(1), 71–88.

ONLINE APPENDIX

The Value of Time in the United States

Ariel Goldszmidt, John A. List, Robert D. Metcalfe,

Ian Muir, V. Kerry Smith, Jenny Wang

June 28, 2024

A Additional Figures

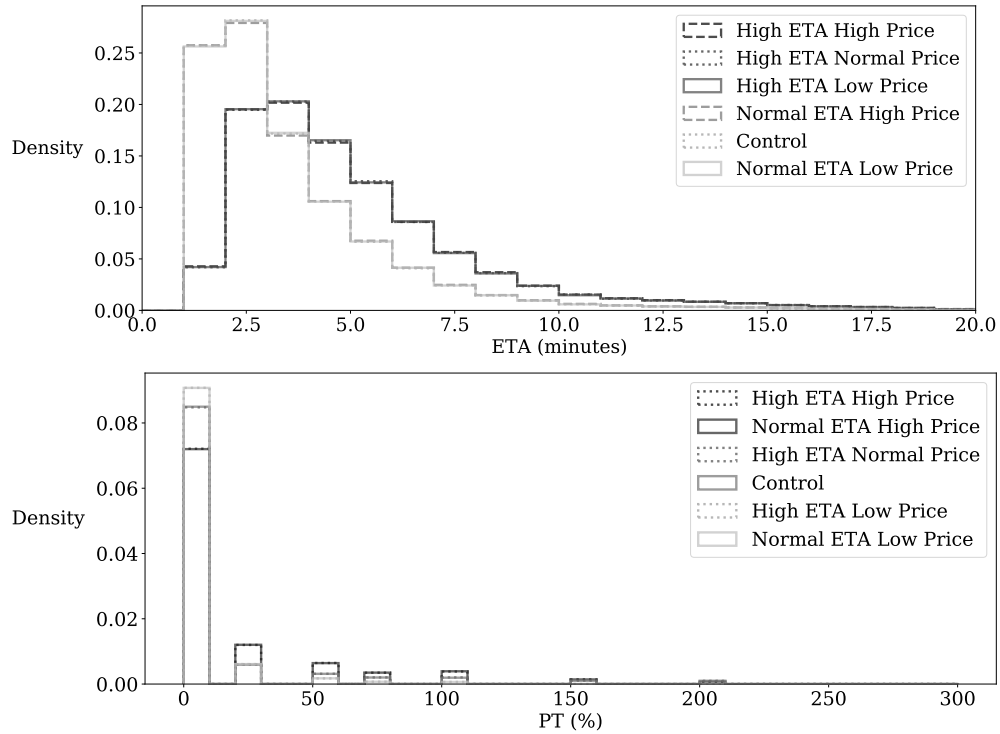


Figure A.1

Distributions of ETA and PT across sessions in each treatment group.

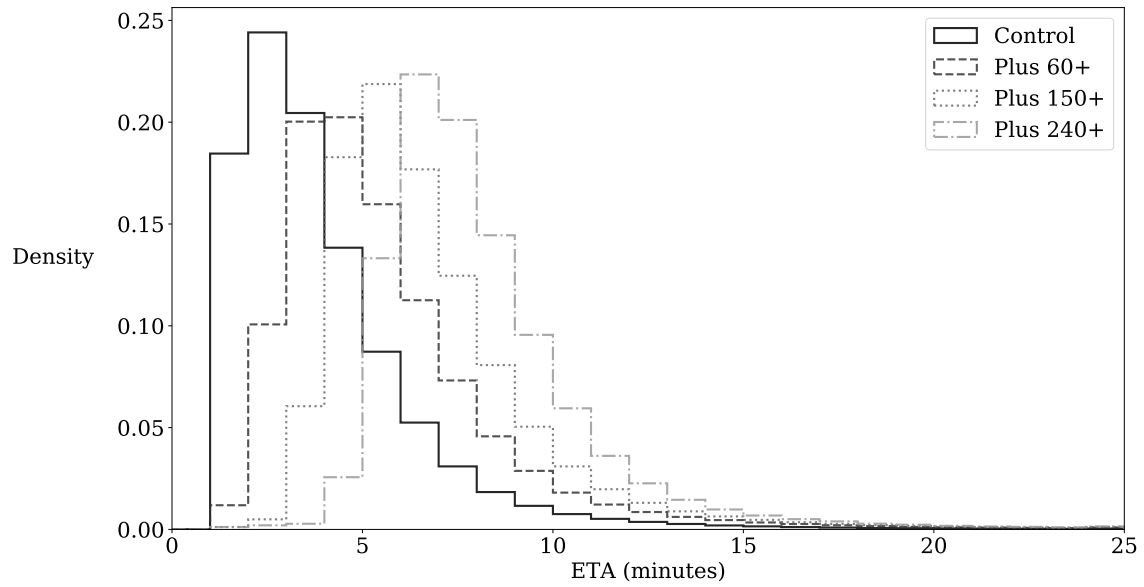


Figure A.2
Distribution of ETAs across treatments in Experiment 2.

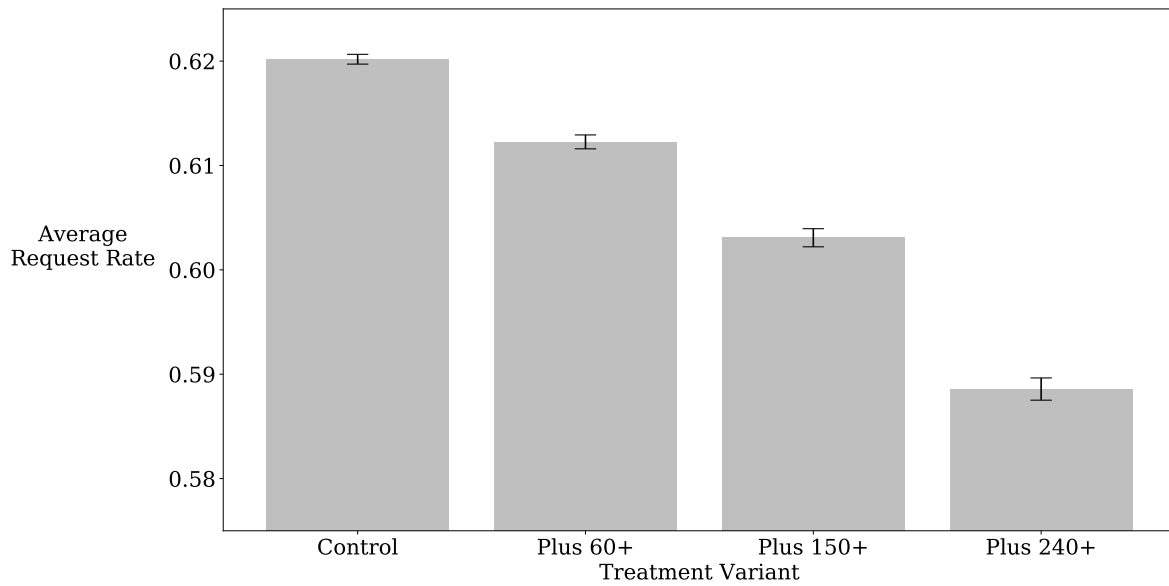


Figure A.3
Average demand across the four treatments in Experiment 2.

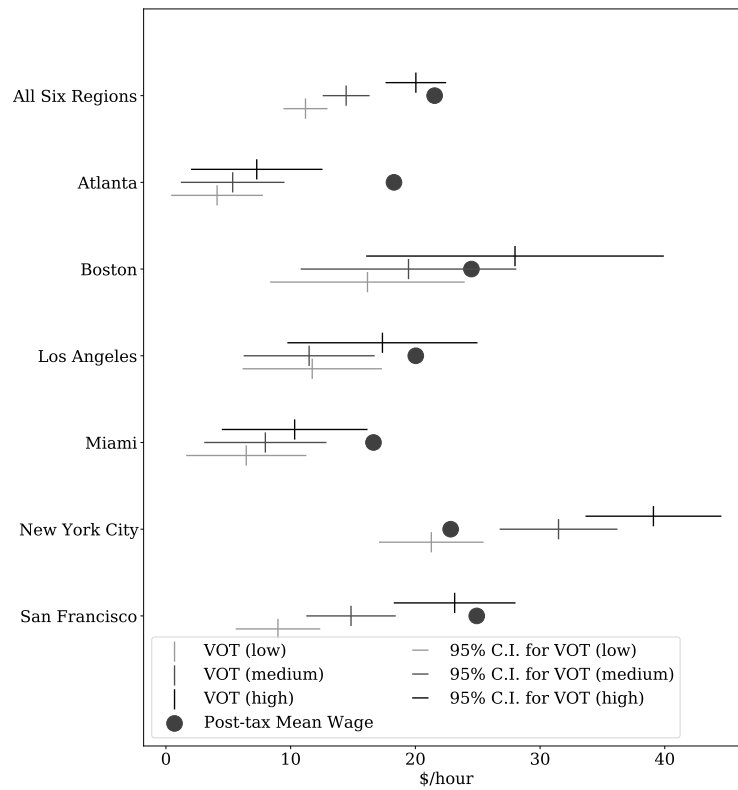


Figure A.4

Values of time estimated by three different ETA treatment level contrasts. Low, medium, and high refer to estimates derived from contrasting the following treatments: Control and Plus 60, Control and Plus 150, and Control and Plus 240, respectively.

B Additional Tables

Table B.1

Average number of sessions, sessions with requests, and sessions with completed rides per passenger in the past 28 days and in the passenger's lifetime, in each treatment group. Robust standard errors in parentheses.

	Control	High ETA High Price	High ETA Low Price	High ETA Normal Price	Normal ETA High Price	Normal ETA Low Price
# Sessions (28 days)	3.39 (0.012)	3.37 (0.033)	3.41 (0.021)	3.4 (0.021)	3.42 (0.021)	3.4 (0.021)
# Sessions (lifetime)	27.14 (0.102)	27.29 (0.28)	27.23 (0.179)	27.2 (0.179)	27.44 (0.18)	27.19 (0.179)
# Sessions w/ Request (28 days)	2.2 (0.009)	2.18 (0.025)	2.21 (0.016)	2.21 (0.016)	2.22 (0.016)	2.21 (0.016)
# Sessions w/ Request (lifetime)	17.95 (0.08)	18.01 (0.217)	17.99 (0.14)	17.95 (0.14)	18.15 (0.139)	18.0 (0.141)
# Sessions w/ Complete (28 days)	1.98 (0.008)	1.97 (0.023)	1.99 (0.015)	1.98 (0.014)	1.99 (0.015)	1.98 (0.014)
# Sessions w/ Complete (lifetime)	15.65 (0.073)	15.75 (0.198)	15.7 (0.127)	15.66 (0.127)	15.86 (0.127)	15.69 (0.129)

Table B.2

Average number of sessions, sessions with requests, and sessions with completed rides for each passenger in the experiment, by treatment group. Robust standard errors in parentheses.

	Control	High ETA High Price	High ETA Low Price	High ETA Normal Price	Normal ETA High Price	Normal ETA Low Price
# Passengers	292025	38674	97254	97051	97185	97870
# Sessions	7.27 (0.019)	6.87 (0.048)	7.13 (0.032)	7.04 (0.032)	7.09 (0.032)	7.38 (0.033)
# Sessions w/ request	4.51 (0.015)	4.05 (0.037)	4.44 (0.025)	4.28 (0.025)	4.26 (0.024)	4.68 (0.026)
# Sessions w/ complete	4.07 (0.014)	3.55 (0.034)	3.91 (0.023)	3.75 (0.023)	3.84 (0.023)	4.24 (0.025)

Table B.3
OLS regression and Durbin-Wu-Hausman test results for full data sample.

	(1)	(2)	(3)
ln ETA	−0.104*** (0.001)	−0.032*** (0.001)	−0.0264*** (0.0018)
ln μ	0.014*** (0.002)	−0.124*** (0.002)	−0.3665*** (0.0132)
ln ETA first-stage residuals			−0.0094*** (0.0018)
ln μ first-stage residuals			0.2542*** (0.0133)
Controls		✓	✓
N	5177358	5177358	5177358
R^2	0.022	0.079	0.079
F	—	—	213***

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Standard errors clustered by user in parentheses. Controls include dummy variables for local week of year, local hour of week, user location (geohash5), business user, airport, and decile of pre-experiment lifetime rides.

Column (3) conducts a Durbin-Wu-Hausman augmented test of exogeneity by including residuals from the first-stage regressions (4) and (5) as regressors. t -tests suggest that the coefficients on the residuals in this regression differ significantly from 0, and the F statistic for the joint hypothesis that both coefficients are 0 is 213 ($p < 0.0001$). This result supports the conclusion that ln(ETA) and ln μ are endogenous in equation (6), so that their coefficients cannot be consistently estimated by OLS.

Table B.4

Regressions of location-time average nearest driver ETA and average nearest-second nearest driver ETA difference on proxies for demand shocks.

	Average Nearest Driver ETA	Average ETA Spread
Number of session starts	0.675*** (0.007)	0.223*** (0.003)
Easter	11.893*** (0.732)	5.203*** (0.519)
Cinco de Mayo	-6.488*** (0.660)	-2.377*** (0.001)
Geohash6 fixed effects	✓	✓
Hour of week fixed effects	✓	✓
<i>N</i>	2467835	2467835

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Heteroskedasticity-consistent standard errors in parentheses.

Table shows results of regressing geohash6-calendar hour average nearest driver ETAs and average nearest-second nearest driver ETA difference on three proxies for exogenous demand shocks: the number of user sessions started in the geohash6-calendar hour; and indicators for the Easter and Cinco de Mayo holidays. Data is from San Francisco during the time period of Experiment 2.

We find evidence that these exogenous shocks correlate with both the level of ETAs and the variation between ETAs of different drivers available to each passenger. If these shocks (or similar unobserved shocks) also affect user conversion behavior, then estimation procedures that identify time preferences through naturally-occurring variation in ETAs may lead to biased estimates.

Table B.5

Estimated effects of $\ln$ ETA on session request rate using OLS with flexible location-time of week controls.

	(1)	(2)
$\ln$ ETA	-0.049*** (0.001)	-0.049*** (0.001)
Geohash6-hour of week fixed effects	✓	
Double ML with latitude, longitude, region, hour of week		✓
<i>N</i>	2123671	2123671

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Heteroskedasticity-consistent standard errors in parentheses.

The table shows the results of regressions of the ride request indicator on $\ln$ ETA in the control population of Experiment 1, using two methods to address possible confounding between ETAs and demand: (1) including geohash6-hour of week fixed effects and (2) a double machine learning (Chernozhukov et al., 2018) approach that controls for user latitude and longitude, region, and hour of week using gradient-boosted decision tree models. Estimates from both methods are about 85% larger than our main IV estimate in Table 4, suggesting that flexibly controlling for location and time of week is insufficient to fully address the endogeneity of ETA in the demand equation.

Table B.6
Summary of price elasticity results for Experiment 1.

1. Opportunity Cost		2. Market Factors		3. Other Trip Features	
Sub-sample	Elasticity	Sub-sample	Elasticity	Sub-sample	Elasticity
Full Sample	−0.5942*** (0.0213)	Full Sample	−0.5942*** (0.0213)	Full Sample	−0.5942*** (0.0213)
Day of Week		Region		Precipitation	
Monday	−0.627*** (0.046)	Austin	−0.709*** (0.161)	None	−0.597*** (0.024)
Tuesday	−0.676*** (0.049)	San Francisco	−0.413*** (0.026)	Rain	−0.560*** (0.026)
Wednesday	−0.653*** (0.048)	Los Angeles	−0.522*** (0.071)	Snow	−0.711*** (0.119)
Thursday	−0.585*** (0.035)	Miami	−0.595*** (0.103)	User Type	
Friday	−0.473*** (0.023)	Boston	−0.670*** (0.072)	Business	−0.398*** (0.073)
Saturday	−0.596*** (0.031)	Seattle	−0.516*** (0.077)	Non-Business	−0.605*** (0.022)
Sunday	−0.640*** (0.030)	San Diego	−0.723*** (0.152)		
Time of Day (Weekends)		Atlanta	−0.825*** (0.256)		
6–10 AM	−0.687*** (0.081)	New York City	−0.956*** (0.046)		
10 AM–4 PM	−0.727*** (0.047)	Location Type			
4 PM–7 PM	−0.616*** (0.039)	Airport	−0.553*** (0.153)		
7 PM–11 PM	−0.669*** (0.062)	Non-Airport	−0.595*** (0.021)		
11 PM–6 AM	−0.500*** (0.028)	Downtown	−0.524*** (0.019)		
Time of Day (Weekdays)		Non-Downtown	−0.762*** (0.058)		
6–10 AM	−0.594*** (0.034)	Distance to Public Transit			
10 AM–4 PM	−0.704*** (0.061)	Under 50 Meters	−0.495*** (0.026)		
4 PM–7 PM	−0.522*** (0.034)	50 to 200 Meters	−0.528*** (0.024)		
7 PM–11 PM	−0.540*** (0.041)	200 to 800 Meters	−0.802*** (0.049)		
11 PM–6 AM	−0.539*** (0.031)	Over 800 Meters	−1.162*** (0.161)		

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Standard errors clustered by user in parentheses. Controls include local week of year, local hour of week, user location (geohash5), business user, airport, and decile of pre-experiment lifetime rides.

Table B.7

Regressions of number of sessions per user on the experimental treatment indicators by week. Week 0 is first week after the end of the experiment.

	-2	-1	0	1	2	3	4	5	6	7	8	9	10	11
High ETA High Price	-0.069*** (0.010)	-0.065*** (0.010)	-0.066*** (0.010)	-0.100*** (0.010)	-0.072*** (0.010)	-0.061*** (0.010)	-0.049*** (0.010)	-0.059*** (0.010)	-0.055*** (0.010)	-0.055*** (0.010)	-0.042*** (0.010)	-0.044*** (0.009)	-0.045*** (0.009)	-0.041*** (0.009)
High ETA Normal Price	-0.040*** (0.007)	-0.031*** (0.007)	-0.039*** (0.007)	-0.044*** (0.007)	-0.043*** (0.007)	-0.037*** (0.007)	-0.034*** (0.007)	-0.032*** (0.007)	-0.030*** (0.007)	-0.033*** (0.007)	-0.030*** (0.007)	-0.018*** (0.006)	-0.014** (0.006)	-0.015** (0.006)
High ETA Low Price	-0.026*** (0.007)	-0.026*** (0.007)	-0.018*** (0.007)	-0.024*** (0.007)	-0.019*** (0.007)	-0.023*** (0.007)	-0.012* (0.007)	-0.020*** (0.007)	-0.022*** (0.007)	-0.022*** (0.007)	-0.020*** (0.007)	-0.015** (0.006)	-0.011* (0.006)	-0.016*** (0.006)
Normal ETA High Price	-0.032*** (0.007)	-0.027*** (0.007)	-0.028*** (0.007)	-0.039*** (0.007)	-0.034*** (0.007)	-0.024*** (0.007)	-0.029*** (0.007)	-0.020*** (0.007)	-0.022*** (0.007)	-0.022*** (0.007)	-0.024*** (0.007)	-0.013** (0.006)	-0.016*** (0.006)	-0.012* (0.006)
Normal ETA Low Price	0.016** (0.007)	0.016** (0.007)	0.020*** (0.007)	0.003 (0.007)	0.016** (0.007)	0.008 (0.007)	0.008 (0.007)	0.007 (0.007)	0.015** (0.007)	0.009 (0.007)	0.002 (0.007)	0.008 (0.006)	0.006 (0.006)	0.004 (0.006)
Controls	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
N	720059	720059	720059	720059	720059	720059	720059	720059	720059	720059	720059	720059	720059	720059
R^2	0.080	0.078	0.097	0.111	0.111	0.116	0.116	0.120	0.119	0.111	0.108	0.104	0.106	0.102
F	19***	16***	20***	31***	21***	16***	12***	12***	13***	12***	9***	9***	7***	6***

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Heteroskedasticity-robust standard errors in parentheses. F test is of the null that all the coefficients on the instruments equal 0. Controls include decile of lifetime rides and indicator of business user.

Table B.8
2SLS regressions on each region in the experiment.

	Austin	San Francisco	Los Angeles	Miami	Boston	Seattle	San Diego	Atlanta	New York City
ln ETA	-0.0181** (0.0078)	-0.0187*** (0.0034)	-0.0250*** (0.0039)	-0.0300*** (0.0060)	-0.0434*** (0.0062)	-0.0331*** (0.0068)	-0.0271*** (0.0067)	-0.0309*** (0.0092)	-0.0285*** (0.0045)
ln μ	-0.4593*** (0.1040)	-0.2932*** (0.0183)	-0.3241*** (0.0441)	-0.3431*** (0.0596)	-0.3922*** (0.0421)	-0.3200*** (0.0477)	-0.4419*** (0.0928)	-0.4565*** (0.1418)	-0.4896*** (0.0238)
ETA Elasticity	-0.0280** (0.0120)	-0.0263*** (0.0047)	-0.0403*** (0.0063)	-0.0519*** (0.0103)	-0.0742*** (0.0107)	-0.0534*** (0.0109)	-0.0444*** (0.0109)	-0.0558*** (0.0166)	-0.0557*** (0.0088)
PT Elasticity	-0.7088*** (0.1605)	-0.4127*** (0.0257)	-0.5217*** (0.0711)	-0.5948*** (0.1033)	-0.6704*** (0.0721)	-0.5163*** (0.0770)	-0.7232*** (0.1520)	-0.8246*** (0.2561)	-0.9564*** (0.0465)
VOT	11.37** (4.76)	21.08*** (3.79)	19.52*** (3.91)	16.36*** (3.95)	23.37*** (3.79)	27.08*** (6.38)	14.80*** (4.34)	11.24** (4.40)	23.85*** (3.81)
Median Wage	18.17	24.90	19.02	16.30	25.37	23.65	19.80	17.94	22.13
Mean Wage	24.44	33.23	26.71	22.19	32.66	30.43	26.68	24.38	30.44
Control Avg. ETA	2.79	2.12	3.02	4.39	3.65	3.41	3.43	4.92	3.09
Control Avg. Price	13.40	11.65	12.71	13.73	12.86	14.85	13.81	13.62	21.08
Control Req. Rate	0.650	0.711	0.627	0.581	0.593	0.626	0.612	0.555	0.515
Controls	✓	✓	✓	✓	✓	✓	✓	✓	✓
N	218666	1387300	1044273	501297	345974	266222	282393	222981	908252
R^2	0.060	0.077	0.068	0.041	0.054	0.073	0.073	0.050	0.029

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Standard errors clustered by user in parentheses. ln ETA and ln μ instrumented by experimental group indicators. Value of time expressed in 2015 US dollars per hour. Controls include local week of year, local hour of week, user location (geohash5), business user, airport, and decile of pre-experiment lifetime rides. Wage data are the "All Occupations" median and mean hourly wages for the metropolitan area most closely corresponding to each Lyft region, as reported in the May 2016 Occupational Employment Statistics (U.S. Bureau of Labor Statistics, 2016).

Table B.9

Number of regions for which each hypothesis test rejects the null that the VOT equals the given rule-of-thumb approximation at the 5% level, using post-tax wages and assuming a constant 25% tax rate.

1/3- Rule			1/2- Rule			1/2+ Rule		
	Median	Mean		Median	Mean		Median	Mean
Log	7	7	Log	7	6	Log	0	0
Linear	7	7	Linear	6	6	Linear	0	0
100%+ Rule			150%+ Rule					
	Median	Mean		Median	Mean			
Log	0	0	Log	4	8			
Linear	0	4	Linear	5	8			

Table B.10
Summary of elasticity and VOT results: robustness checks.

Sub-sample	ETA Elasticity	PT Elasticity	VOT
Session in experiment			
1	-0.042*** (0.004)	-0.572*** (0.030)	22.72*** (2.50)
2	-0.041*** (0.005)	-0.512*** (0.037)	22.81*** (3.21)
3	-0.045*** (0.006)	-0.545*** (0.040)	23.09*** (3.18)
4	-0.044*** (0.006)	-0.574*** (0.044)	21.16*** (3.14)
5	-0.048*** (0.007)	-0.539*** (0.047)	24.62*** (3.65)
6	-0.039*** (0.007)	-0.509*** (0.050)	20.88*** (3.90)
7	-0.043*** (0.007)	-0.449*** (0.053)	25.54*** (4.97)
8	-0.048*** (0.008)	-0.558*** (0.057)	23.03*** (4.09)
Week in experiment			
1	-0.034*** (0.005)	-0.535*** (0.033)	18.73*** (2.96)
2	-0.045*** (0.005)	-0.568*** (0.024)	22.93*** (2.46)
3	-0.047*** (0.005)	-0.566*** (0.027)	22.83*** (2.56)
4	-0.042*** (0.006)	-0.695*** (0.077)	16.92*** (3.04)
5	-0.033*** (0.006)	-0.463*** (0.037)	18.07*** (3.20)
6	-0.039*** (0.005)	-0.827*** (0.089)	13.37*** (2.20)
7	-0.043*** (0.005)	-0.622*** (0.043)	17.40*** (2.21)
8	-0.052*** (0.005)	-0.663*** (0.043)	18.07*** (1.97)
Pre-experiment lifetime rides			
0	-0.032*** (0.006)	-0.484*** (0.052)	17.62*** (3.39)
1-3	-0.073*** (0.011)	-0.942*** (0.078)	22.34*** (3.66)
4-10	-0.059*** (0.009)	-0.753*** (0.067)	22.23*** (3.68)
11-20	-0.047*** (0.009)	-0.686*** (0.064)	19.06*** (3.89)
21-50	-0.049*** (0.007)	-0.665*** (0.047)	20.41*** (2.96)
Over 50	-0.028*** (0.005)	-0.412*** (0.030)	19.38*** (3.35)
Pre-experiment 28 day rides			
0	-0.049*** (0.005)	-0.608*** (0.039)	22.11*** (2.34)
1-2	-0.053*** (0.008)	-0.716*** (0.052)	21.25*** (3.35)
3-8	-0.044*** (0.006)	-0.616*** (0.042)	19.75*** (2.83)
Over 8	-0.029*** (0.005)	-0.499*** (0.034)	15.22*** (2.71)

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Standard errors clustered by user in parentheses. Controls include local week of year, local hour of week, user location (geohash5), business user, and decile of pre-experiment lifetime rides.

Table B.11
IV-Probit estimates.

	IV-Probit (MLE)
ln ETA	−0.061*** (0.006)
ln μ	−0.934*** (0.042)
Avg. marginal effect of ln ETA	−0.023*** (0.002)
Avg. marginal effect of ln μ	−0.351*** (0.015)
VOT	17.53*** (1.78)
Controls	
N	5177358

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Standard errors clustered by user in parentheses. ln ETA and ln μ instrumented by experimental group indicators. Value of time expressed in 2015 US dollars per hour. Model is estimated using the maximum likelihood method in Stata's `ivprobit` package.

Table B.12
Checks of robustness to specification and functional form.

	Completed	Avg ETA	Linear ETA	Squared ETA
ln ETA	−0.0580*** (0.0018)	−0.0272*** (0.0018)	—	—
ln μ	−0.3690*** (0.0131)	−0.3664*** (0.0132)	−0.3664*** (0.0131)	−0.3661*** (0.0131)
ETA	—	—	−0.4770*** (0.0321)	—
ETA ²	—	—	—	−2.915*** (0.196)
VOT	42.31*** (1.83)	19.90*** (1.42)	18.01*** (1.29)	11.32*** (0.81)
Controls	✓	✓	✓	✓
N	5177358	5177358	5177358	5177358
R^2	0.092	0.072	0.072	0.070

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Standard errors clustered by user in parentheses. ln(ETA) and ln μ instrumented by experimental group indicators. Value of time expressed in 2015 US dollars per hour. Controls include dummy variables for user location (geohash5), local hour of week and week of year, decile of user lifetime rides, an indicator of whether the user is a business user, and an indicator of whether the session is at an airport.

In column (1), the dependent variable is an indicator of whether a session had a completed ride, while for the other three the dependent variable is an indicator of whether a session had a ride request. In column (2), the session-average ETA is used in place of the last-in-session ETA. In columns (3) and (4), ETA enters the equation linearly and squared, respectively, rather in logarithm. For both columns, ETA is in units of hours.

Table B.13
2SLS results by region, using linear functional form (Experiment 1).

	Austin	San Francisco	Los Angeles	Miami	Boston	Seattle	San Diego	Atlanta	New York City
ETA	−0.350** (0.150)	−0.469*** (0.084)	−0.476*** (0.075)	−0.390*** (0.078)	−0.637*** (0.092)	−0.545*** (0.112)	−0.424*** (0.104)	−0.348*** (0.103)	−0.525*** (0.083)
ln μ	−0.460*** (0.104)	−0.293*** (0.018)	−0.324*** (0.044)	−0.343*** (0.060)	−0.392*** (0.042)	−0.322*** (0.048)	−0.442*** (0.093)	−0.457*** (0.142)	−0.489*** (0.024)
VOT	10.18** (4.27)	18.65*** (3.35)	18.68*** (3.75)	15.61*** (3.77)	20.89*** (3.39)	25.18*** (5.90)	13.23*** (3.87)	10.38** (4.05)	22.60*** (3.61)
Control Avg. ETA	2.79	2.12	3.02	4.39	3.65	3.41	3.43	4.92	3.09
Control Avg. Price	13.40	11.65	12.71	13.73	12.86	14.85	13.81	13.62	21.08
Control Req. Rate	0.650	0.711	0.627	0.581	0.593	0.626	0.612	0.555	0.515
Controls	✓	✓	✓	✓	✓	✓	✓	✓	✓
N	218666	1387300	1044273	501297	345974	266222	282393	222981	908252
R^2	0.059	0.077	0.067	0.041	0.054	0.073	0.073	0.050	0.029

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Standard errors clustered by user in parentheses. ln ETA and ln μ instrumented by experimental group indicators. Value of time expressed in 2015 US dollars per hour. Controls include local week of year, local hour of week, user location (geohash5), business user, airport, and decile of pre-experiment lifetime rides.

Table B.14
2SLS results by whether session had a destination entered.

	Without Destination	With Destination
ln ETA	−0.0367*** (0.0024)	−0.0105*** (0.0022)
ln μ	−0.4554*** (0.0201)	−0.3644*** (0.0137)
ln(Price)	—	−0.1043*** (0.0014)
ETA Elasticity	−0.0721*** (0.0046)	−0.0134*** (0.0028)
PT Elasticity	−0.8938*** (0.0394)	−0.4661*** (0.0175)
VOT	21.07*** (1.53)	8.20*** (1.68)
Control Avg. ETA	3.20	2.90
Control Avg. Price	13.93	13.73
Control Req. Rate	0.514	0.783
Controls	✓	✓
N	3136534	2040824
R^2	0.112	0.064

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Standard errors clustered by user in parentheses. ln ETA and ln μ instrumented by experimental group indicators. Price is midpoint of estimated fare range. Value of time expressed in 2015 US dollars per hour. Controls include local week of year, local hour of week, user location (geohash5), business user, airport, and decile of pre-experiment lifetime rides.

Table B.15
Effect of treatments on probability of entering destination.

	Whether Destination Entered
High ETA High Price	0.010*** (0.002)
High ETA Normal Price	−0.000 (0.002)
High ETA Low Price	−0.005*** (0.002)
Normal ETA High Price	0.010*** (0.002)
Normal ETA Low Price	−0.005*** (0.002)
Controls	✓
N	5177358
R^2	0.103
F	21***

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Standard errors clustered by user in parentheses. F test is of the null that all the coefficients on the instruments equal 0. Controls include local week of year, local hour of week, user location (geohash5), business user, airport, and decile of pre-experiment lifetime rides. Approximately one-third of sessions had destinations entered.

Table B.16
Effect of treatments on ride types considered.

	Considered Shared	Considered XL
High ETA High Price	0.009*** (0.003)	0.001 (0.001)
High ETA Normal Price	0.004** (0.002)	0.000 (0.000)
High ETA Low Price	-0.006*** (0.002)	0.000 (0.000)
Normal ETA High Price	0.010*** (0.002)	0.000 (0.001)
Normal ETA Low Price	-0.006*** (0.002)	-0.000 (0.000)
Controls	✓	✓
N	5177358	5177358
R^2	0.293	0.012
F	13***	1

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Standard errors clustered by user in parentheses. F test is of the null that all the coefficients on the instruments equal 0. Controls include local week of year, local hour of week, user location (geohash5), business user, airport, and decile of pre-experiment lifetime rides. The dependent variables for the two columns are indicators of whether a session ever selected the Shared or XL ride types, respectively.

Table B.17
Regression of $\ln \text{ETA}$ and $\ln \mu$ on treatment indicators (Experiment 2).

	$\ln \text{ETA}$	$\ln \mu$
Plus 60+	0.4827*** (0.0004)	-0.0003 (0.0003)
Plus 150+	0.7522*** (0.0004)	0.0001 (0.0004)
Plus 240+	0.9699*** (0.0005)	0.0009 (0.0006)
Controls	✓	✓
N	9668820	9668820
R^2	0.608	0.144
F	1644577.2***	1.4

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Standard errors clustered by location-hour in parentheses. Controls include local week of year, local hour of week, user location (geohash5), business user, and decile of pre-experiment lifetime rides.

Table B.18
Time elasticities (Experiment 2).

	Full sample	Destination	No Destination	2017 users in 2015	2015 users in 2017
ln ETA	−0.0265*** (0.0004)	−0.0150*** (0.0005)	−0.0268*** (0.0007)	−0.0303*** (0.0029)	−0.0312*** (0.0014)
ln μ	−0.2851*** (0.0011)	−0.2002*** (0.0013)	−0.1393*** (0.0013)	−0.3785*** (0.0183)	−0.3169*** (0.0026)
ln(Price)	—	−0.1677*** (0.0003)	—	—	—
ETA Elasticity	−0.0432*** (0.0007)	−0.0202*** (0.0003)	−0.1292*** (0.0034)	−0.0460*** (0.0044)	−0.0467*** (0.0021)
VOT	—	22.15*** (0.67)	—	22.71*** (2.25)	—
Control Avg. ETA	3.43	3.33	3.76	2.72	2.82
Control Avg. Price	11.96	13.85	19.53	12.90	11.37
Control Req. Rate	0.620	0.742	0.228	0.661	0.678
N	9668820	7395984	2272839	2098259	711857
R^2	0.075	0.110	0.228	0.082	0.106

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Clustered standard errors, with clustering for the 2017 regressions at the hour-geohash7-level and clustering for the 2015 regressions at the user-level. ln ETA instrumented by experimental group indicators; ln μ is instrumented using experimental group indicators. Including controls for region, geohash5, local hour of week, local week of year, business user, and decile of user lifetime rides. Control for airports included for 2015 data. Price is the upfront price shown to passengers who enter a destination in Experiment 2. VOT for sessions with a destination entered is computed using the coefficient on ln(Price) as the price semi-elasticity and the average upfront price shown as the average price. VOT for 2015 users in 2015 is computed using the 2015 PT multiplier semi-elasticity. 2017 prices deflated to December 2015 USD using the Consumer Price Index for All Urban Consumers ([U.S. Bureau of Labor Statistics, 2020](#)).

“2017 users in 2015” refers to the sample of sessions in Experiment 1 from users who appear in both experiments. “2015 users in 2017” refers to the sample of sessions in Experiment 2 from users who appear in both experiments.

Table B.19
ETA elasticities by quintile of estimated trip duration (Experiment 2).

	Q1	Q2	Q3	Q4	Q5
ln ETA	−0.0190*** (0.0008)	−0.0146*** (0.0008)	−0.0159*** (0.0009)	−0.0163*** (0.0010)	−0.0133*** (0.0011)
ln μ	−0.2667*** (0.0020)	−0.3242*** (0.0019)	−0.3771*** (0.0019)	−0.3880*** (0.0019)	−0.3587*** (0.0021)
ETA Elasticity	−0.0233*** (0.0010)	−0.0181*** (0.0010)	−0.0210*** (0.0012)	−0.0236*** (0.0015)	−0.0233*** (0.0020)
Control Avg. ETA	3.16	3.20	3.22	3.33	3.80
Control Req. Rate	0.819	0.810	0.762	0.693	0.575
Controls	✓	✓	✓	✓	✓
N	1598111	1596340	1593377	1592525	1594672
R^2	0.073	0.077	0.085	0.081	0.070

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Standard errors clustered by location-hour in parentheses. ln ETA instrumented by experimental treatment indicators. Controls include local week of year, local hour of week, user location (geohash5), business user, airport, and decile of pre-experiment lifetime rides. Bins are quintiles of estimated in-vehicle time, in seconds. The ordered quintiles are 561, 772, 1032, 1455.

Table B.20
Convexity in marginal effects of ETA increases.

	1
Control — Plus 60	−0.290*** (0.015)
Control — Plus 150	−0.354*** (0.010)
Control — Plus 240	−0.445*** (0.008)

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Standard errors clustered by location-hour in parentheses. Estimates are the coefficients on ETA in the subsample 2SLS regressions described in Section 4.2, with ETA (measured in hours) replacing ln ETA. ETA instrumented by experimental group indicators. Controls include local week of year, local hour of week, user location (geohash5), business user, and decile of pre-experiment lifetime rides.

Table B.21
Regressions of log estimated VOT on log mean wage at the region level.

	OLS	GLS
$\ln(\text{MeanWage})$	1.73*** (0.66)	1.28*** (0.43)
N	9	9
R^2	0.586	0.534
Adj. R^2	0.527	0.467

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Heteroskedasticity-robust (HC2) standard errors in parentheses. Dependent variable is $\ln(\widehat{\text{VOT}})$. The GLS specification uses as the weighting matrix the inverse of the estimated covariance matrix of the vector of log VOT estimates across regions, which is derived from the estimated covariance matrix of the vector of price and time semi-elasticities across regions using the delta method.

Table B.22
Regressions of price elasticities on log mean wage at the region level.

	OLS	WLS
$\ln(\text{MeanWage})$	0.241 (0.434)	0.772 (0.829)
N	9	9
R^2	0.042	0.124
Adj. R^2	-0.095	0.019

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Heteroskedasticity-robust (HC2) standard errors in parentheses. Dependent variable is the estimated price elasticity of each region. The WLS specification uses as weights the inverse of the estimated variance of each price elasticity estimate.

C Random Utility Model

Our model for rideshare decisions is motivated within a utility difference framework, comparing what is realized by selecting the rideshare versus the default. We recognize that these types of decisions may be made multiple times by each individual. However, our data require we treat them as discrete choices. As we noted in the text, while it is possible to distinguish experienced rideshare users from those without experience in past uses of the Lyft platform. Nonetheless we do not have sufficient information on potential users to model these decisions as part of a full set of travel decisions.

This limitation motivates how we characterize the choice process. Rideshare trips are assumed to be undertaken to accomplish a task, designated here with T . In defining the alternative to selecting the rideshare, a potential user is assumed to have one of two default possibilities—accomplish the task with travel via another mode such as public transit or do not complete the task. Under this second possibility we are effectively assuming rideshare trips and the tasks are perfect complements.

We define the utility difference for each definition of the default alternative. Both are consistent with our empirical model. We specify the two possibilities because they provide motivation for our efforts to explore how context affects the estimates for the VOT. The equation below defines the indirect utility realized when the rideshare trip is taken.

$$v = v_1(p, ETA, P, m, T) + \epsilon \quad (8)$$

where p defines the rideshare fare, ETA is the wait time for rideshare, P is a vector of prices for other goods and services, m is the income relevant to the choice, T designates the contribution of the task, and ϵ is an unobserved error, with expected value of zero, reflecting the heterogeneity in the choice context.

Under the first default alternative, the indirect utility would be given by equation:

$$v_a = v_a(c, S, P, m, T) + \epsilon_a \quad (9)$$

where c is the fare for the travel mode alternative (e.g., bus fare), S is the schedule for the travel alternative, and ϵ_a is the unobserved heterogeneity in the first alternative. Rideshare is selected when $v - v_a \geq \epsilon_a - \epsilon$.

If the default alternative implies the task is not undertaken, then the arguments for c , S , and T in the equation given above for the alternative would be replaced by zeros. In practice, we do not know which default alternative is relevant for each potential user. We distinguish

the two alternatives here because they partially motivate our attention to the heterogeneity in our VOT estimates with context for the observed rideshare choices. Nonetheless, since our model is based on utility differences the assumption of weak complementarity allows a focus on the utility associated with the rideshare option because p and ETA enter this indirect utility and not either of the other alternatives: $E(dV) = v - v_i = dv$. Weak complementarity implies there are price changes for rideshare that are equivalent to changes in wait times so we can define $-dp/dETA$ as the marginal value of a change in wait time. In terms of model this is given in the equation:

$$-\frac{dp}{dETA} = -\frac{\left(\frac{E(dV)}{dETA}\right)}{\left(\frac{E(dV)}{dp}\right)} = -\frac{\left(\frac{dv}{dETA}\right)}{\left(\frac{dv}{dp}\right)} \quad (10)$$

Notice that while the default alternative does not influence our strategy for estimating the VOT, the task (T) does. As a result we would expect that:

$$\frac{\partial}{\partial T} \left(-\frac{dp}{dETA} \right) \neq 0 \quad (11)$$

The diversity in our VOT estimates with context confirm this possibility.

D Correcting for User Selection into Ride Contexts

The patterns of heterogeneity we document across contexts do not take into account user self-selection into these contexts. As a consequence, though our heterogeneity analysis suggests, for example, that the VOT is higher in passenger sessions occurring during commuting times than other times of the week, it does not let us say whether this context “causes” higher VOTs—it may instead be the case that passengers with higher VOTs are more likely to have sessions during commuting times.

The purpose of this analysis is to separate contextual and selection effects in the observed heterogeneity, by weighting each user i ’s sessions in context x by the inverse of an estimate of user i ’s propensity to have a session in context x , $\hat{p}_{i,x}$.

We can estimate these propensities using pre-experiment data as the observed fraction of user i ’s sessions that occur in context x . The propensity is estimated using $n_{i,x}$, the number of user i ’s pre-experiment sessions that occurred in context x :

$$\hat{p}_{x,i} = \frac{n_{x,i}}{\sum_{x' \in X} n_{x',i}}. \quad (12)$$

These estimates have high variances. The estimated propensities often equal 0 (in which case their inverses are undefined), and do not exist for users with no pre-experiment sessions.

To address these issues, we instead adopt an empirical Bayes estimation strategy. For each range of contexts X (for example, the different days of the week, or {downtown, non-downtown}), we assume that the set of vectors $\{(p_{i,x})_{x \in X}\}$ is composed of independent draws from a Dirichlet distribution with parameter vector α_X . α_X can be estimated from the data. Conditional on α_X , $\{(n_{x,i})_{x \in X}\}$ are independent draws from the Dirichlet-multinomial distribution, with joint density proportional to

$$\prod_i \left(\frac{\Gamma(\sum_x \alpha_x)}{\Gamma(\sum_x n_{x,i} + \alpha_x)} \prod_x \frac{\Gamma(n_{x,i} + \alpha_x)}{\Gamma(\alpha_x)} \right) \quad (13)$$

(Mosimann, 1962). From (13), α_X can be estimated by maximum likelihood. We then take the maximum a posteriori estimate of each vector $(p_{i,x})_{x \in X}$ based on the data $(n_{i,x})_{x \in X}$ and the prior Dirichlet($\hat{\alpha}_X$). The resulting estimates are

$$\hat{p}_{x,i} = \frac{n_{x,i} + \hat{\alpha}_x}{\sum_{x'} n_{x',i} + \hat{\alpha}_{x'}}, \quad (14)$$

so that the empirical Bayes approach amounts to shrinking the basic estimates $n_{x,i} / \sum_{x'} n_{x',i}$ toward $\hat{\alpha}_x / \sum_{x'} \hat{\alpha}_{x'}$, where estimates for users with fewer pre-experiment observations are shrunk more.²

Table D.1 reproduces heterogeneity analyses for whether downtown, whether at an airport, day and time of week, and distance to public transit using the inverses of the empirical Bayes propensity estimates as observation weights. The non-downtown vs. downtown results do not show a higher VOT in downtown sessions, suggesting that the observed higher VOT in the unweighted data may be attributable to user selection. Results for non-airport vs. airport sessions are similar, though the airport estimates remain very imprecise due to the small sample size. In contrast, the results for day and time of week are similar to those found in the unweighted data. We find that the VOT is highest in the morning commute, suggesting that this result is not primarily due to user selection. Results using inverse propensity weights based on distance to transit bins are also similar to those observed in the unweighted data.

We are unable to assess the robustness of precipitation and business/non-business results to user selection into contexts because our weather data does not extend far past the beginning of Experiment 1 and business status does not vary within user. We are unable

²For more details on empirical Bayes Dirichlet-multinomial estimation, see Maritz and Lwin (1989), section 4.5.

to effectively control for selection into regions due to limited within-user variation in region (only 19% of users have a pre-experiment session in more than one region).

We compare these inverse-propensity-weighted estimates to those obtained from the [Heckman \(1979\)](#) two-step correction procedure. For each context x in the range of contexts X , we estimate a probit regression

$$\mathbb{P}\{\text{session } ij \text{ is in context } x | W_{ij}\} = \Phi(W'_{ij}\beta_x) \quad (15)$$

where the vector W_{ij} consists of a binary indicator of whether user i has any pre-experiment sessions; the number of user i 's pre-experiment sessions; and the fraction of user i 's pre-experiment sessions which occurred in context j (set to 0 for users with no pre-experiment sessions). In the 2SLS regression on the subsample of sessions in context x , we then include the inverse Mills ratio (IMR) based on the probit (15) as a regressor in the first and second stages.

The results of this procedure are shown in Table [D.2](#). In almost all regressions, the estimated coefficient on the IMR is statistically significant, but the estimated elasticities and VOTs are little changed versus the uncorrected results.

Table D.1
VOT results using empirical Bayes inverse propensity weights.

Sub-sample	ETA Elasticity	PT Elasticity	VOT
Location Type			
Airport	-0.0301 (0.0264)	-1.0021* (0.5449)	18.07 (21.06)
Non-Airport	-0.0431*** (0.029)	-0.5956*** (0.0214)	18.85*** (1.35)
Downtown	-0.0360*** (0.0049)	-0.5238*** (0.0300)	24.17*** (3.47)
Non-Downtown	-0.0740*** (0.0125)	-0.7219*** (0.1259)	29.63*** (6.34)
Day of Week			
Monday	-0.0458*** (0.0054)	-0.6305*** (0.0473)	19.25*** (2.54)
Tuesday	-0.0474*** (0.0052)	-0.6742*** (0.0515)	18.76*** (2.38)
Wednesday	-0.0493*** (0.0052)	-0.6720*** (0.0505)	20.46*** (2.48)
Thursday	-0.0428*** (0.0050)	-0.5872*** (0.0364)	20.13*** (2.50)
Friday	-0.0409*** (0.0046)	-0.4898*** (0.0241)	21.38*** (2.44)
Saturday	-0.0393*** (0.0044)	-0.6095*** (0.0324)	18.12*** (2.09)
Sunday	-0.0381*** (0.0048)	-0.6475*** (0.0316)	15.90*** (2.07)
Time of Day (Weekends)			
6-10 AM	-0.0514*** (0.0135)	-0.546*** (0.072)	17.27*** (4.82)
10 AM-4 PM	-0.0546*** (0.0072)	-0.7258*** (0.0509)	19.52*** (2.75)
4 PM-7 PM	-0.0542*** (0.0076)	-0.6145*** (0.0441)	22.83*** (3.33)
7 PM-11 PM	-0.0365*** (0.0067)	-0.6855*** (0.0674)	16.01*** (3.11)
11 PM-6 AM	-0.0216*** (0.0059)	-0.5076*** (0.0339)	13.28*** (3.63)
Time of Day (Weekdays)			
6-10 AM	-0.0767*** (0.0067)	-0.3907*** (0.0265)	29.23*** (3.02)
10 AM-4 PM	-0.0470*** (0.0059)	-0.6969*** (0.0603)	19.62*** (2.86)
4 PM-7 PM	-0.0450*** (0.0055)	-0.5519*** (0.0370)	20.78*** (2.72)
7 PM-11 PM	-0.0244*** (0.0048)	-0.5558*** (0.0439)	13.09*** (2.63)
11 PM-6 AM	-0.0405*** (0.0062)	-0.5202*** (0.0327)	19.93*** (3.07)
Distance to Nearest Public Transit			
Under 50 Meters	-0.0412*** (0.0047)	-0.4940*** (0.0306)	25.03*** (3.02)
50 to 200 Meters	-0.0391*** (0.0035)	-0.5476*** (0.0255)	20.07*** (1.92)
200 to 800 Meters	-0.0499*** (0.0067)	-0.7261*** (0.0569)	19.87*** (2.80)
Over 800 Meters	-0.0577** (0.0291)	-1.0129*** (0.3290)	14.42* (7.94)

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Standard errors clustered by user in parentheses. Controls include local week of year, local hour of week, user location (geohash5), business user, and decile of pre-experiment lifetime rides.

Table D.2
VOT results using Heckman correction.

Sub-sample	IMR	ETA Elasticity	PT Elasticity	VOT
Location Type				
Airport	-0.800*** (0.0274)	-0.0223 (0.0155)	-0.5514*** (0.1525)	26.15 (18.24)
Non-Airport	0.077 (0.0103)	-0.0431*** (0.029)	-0.5951*** (0.0214)	18.84*** (1.35)
Downtown	-0.0079*** (0.0014)	-0.0344*** (0.0031)	-0.5241*** (0.0191)	21.65*** (1.99)
Non-Downtown	-0.0285*** (0.0014)	-0.0599*** (0.0054)	-0.7612*** (0.0574)	17.94*** (1.97)
Day of Week				
Monday	-0.2625*** (0.0172)	-0.0478*** (0.0052)	-0.6315*** (0.0461)	19.87*** (2.45)
Tuesday	-0.2370*** (0.0151)	-0.0457*** (0.0050)	-0.6806*** (0.0491)	17.82*** (2.21)
Wednesday	-0.3020*** (0.0169)	-0.0451*** (0.0050)	-0.6473*** (0.0480)	19.31*** (2.40)
Thursday	-0.0990*** (0.0281)	-0.0413*** (0.0049)	-0.5850*** (0.0352)	19.43*** (2.41)
Friday	0.3593*** (0.0262)	-0.0393*** (0.0045)	-0.4728*** (0.0233)	21.23*** (2.45)
Saturday	-0.0511*** (0.0066)	-0.0395*** (0.0041)	-0.5957*** (0.0306)	18.73*** (2.04)
Sunday	-0.1208*** (0.0092)	-0.0407*** (0.0046)	-0.6395*** (0.0300)	17.25*** (2.01)
Time of Day (Weekends)				
6–10 AM	-0.2326*** (0.0130)	-0.0551*** (0.0109)	-0.6614*** (0.0805)	17.91*** (3.95)
10 AM–4 PM	-0.2172*** (0.0131)	-0.0534*** (0.0066)	-0.7286*** (0.0465)	18.95*** (2.49)
4 PM–7 PM	-0.2204*** (0.0174)	-0.0478*** (0.0067)	-0.6150*** (0.0393)	20.18*** (2.92)
7 PM–11 PM	-0.1647*** (0.0114)	-0.0335*** (0.0061)	-0.6697*** (0.0617)	15.07*** (2.87)
11 PM–6 AM	-0.0913*** (0.0039)	-0.0239*** (0.0050)	-0.5025*** (0.0284)	14.99*** (3.10)
Time of Day (Weekdays)				
6–10 AM	-0.1783*** (0.0034)	-0.0680*** (0.0056)	-0.6033*** (0.0336)	26.23*** (2.39)
10 AM–4 PM	-0.2269*** (0.0045)	-0.0457*** (0.0055)	-0.7094*** (0.0598)	18.27*** (2.56)
4 PM–7 PM	-0.2383*** (0.0072)	-0.0461*** (0.0052)	-0.5198*** (0.0338)	22.42*** (2.70)
7 PM–11 PM	-0.1498*** (0.0087)	-0.0241*** (0.0045)	-0.5401*** (0.0405)	13.42*** (2.58)
11 PM–6 AM	-0.1351*** (0.0056)	-0.0398*** (0.0056)	-0.5434*** (0.0311)	18.55*** (2.61)
Distance to Nearest Public Transit				
Under 50 Meters	0.0148*** (0.0029)	-0.0395*** (0.0041)	-0.4953*** (0.0256)	23.91*** (2.56)
50 to 200 Meters	0.0242*** (0.0023)	-0.0393*** (0.0034)	-0.5286*** (0.0242)	20.70*** (1.91)
200 to 800 Meters	0.0162*** (0.0021)	-0.0506*** (0.0057)	-0.8015*** (0.0493)	16.95*** (2.03)
Over 800 Meters	-0.0161*** (0.0024)	-0.0575*** (0.0136)	-1.1733*** (0.1603)	10.76*** (2.78)

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Standard errors clustered by user in parentheses. Controls include local week of year, local hour of week, user location (geohash5), business user, and decile of pre-experiment lifetime rides.

E Reweighting Elasticity Estimates

In inspecting our results on heterogeneity of time elasticities, one may note that the control average ETA often varies considerably between subsamples. In light of the evidence from Experiment 2 suggesting that ETA elasticities vary over the demand curve, this observation calls into question whether our heterogeneity results are due to differences in the demand curve between subsamples, or whether they are due to us estimating elasticities at different points in different samples.

To see the issue formally, consider a simplified version of our main model, in which demand D_i in session i is a function of the (logged) waiting time $T_i \geq 0$, session and user characteristics X_i , instrument z_i , and an unobserved shock ε_i :

$$D_i = D_i(T_i, X_i, z_i, \varepsilon_i). \quad (16)$$

Suppose further that z_i is a binary instrument for T_i , satisfying the assumptions of independence (z_i is independent of $(T_i(X_i, 1), T_i(X_i, 0))$, conditional on X_i), exclusion ($D_i(T, X, z) = D_i(T, X, z')$ for all z, z'), and strict monotonicity ($T_i(X_i, 1) > T_i(X_i, 0)$). Now, consider restricting the sample to observations i for which $X_i = x$ and running an IV regression of D_i on T_i using z_i as an instrument. The resulting coefficient on T_i will be an estimate of

$$\beta_x = \frac{\mathbb{E}[D_i(T_i(1), X_i, 1) - D_i(T_i(0), X_i, 0) | X_i = x]}{\mathbb{E}[T_i(1) - T_i(0) | X_i = x]} \quad (17)$$

$$= \int_0^\infty \mathbb{E} \left[\frac{\partial D_i}{\partial T}(T) | T_i(0) < T < T_i(1), X_i = x \right] \omega(T|x) dT \quad (18)$$

where

$$\omega(T|x) \propto \mathbb{P}[T_i(0) < T < T_i(1) | X_i = x],$$

and $\omega(T|x)$ is normalized to integrate to 1. The proof is the same as that of Theorem 1 in [Angrist et al. \(2000\)](#).

The above result shows that for two values of X , say x and x' , separate IV regressions for the subsamples of observations with $X_i = x$ and $X_i = x'$ will estimate quantities that may differ for two reasons:

1. $\mathbb{E} \left[\frac{\partial D_i}{\partial T}(T) | T_i(0) < T < T_i(1), X_i = x \right]$ may differ from $\mathbb{E} \left[\frac{\partial D_i}{\partial T}(T) | T_i(0) < T < T_i(1), X_i = x' \right]$ for some fixed T ; or

2. $\omega(\cdot, x)$ may differ from $\omega(\cdot, x')$.

Differences of the first kind are economically interesting, as they indicate that the average slope of the demand curve at some fixed point differs between the two samples. Differences of the second kind are nuisances, as they are caused by our semi-elasticity estimates for the two subsamples being measured at different points.

Note that $\omega(\cdot|x)$ is determined by the distribution $(T_i(0), T_i(1))|X_i = x$. Hence, if $(T_i(0), T_i(1))|X_i = x \sim (T_i(0), T_i(1))|X_i = x'$, we would have $\omega(\cdot|x) = \omega(\cdot|x')$, and any differences between β_x and $\beta_{x'}$ would be of the first kind.

We may attempt to set $(T_i(0), T_i(1))|X_i = x \sim (T_i(0), T_i(1))|X_i = x'$ by reweighting each subsample. If such weights can be found, and can be chosen independently of z_i , we may then run weighted IV regressions on each subsample to estimate semi-elasticities which represent weighted average derivatives over the same sections of the demand curve, with the same weights, and can therefore be more readily compared.³

We attempt such reweighting for Experiment 2, in which we can observe what a session's ETA quote would have been under each treatment level. To find a reweighting scheme that (approximately) equates the joint distributions of potential ETAs, we employ the maximum entropy rebalancing idea of [Hainmueller \(2012\)](#). This entails solving a smooth, convex, constrained optimization problem to find the maximum entropy weight vector w which sets certain moments in the data equal to pre-specified target moments. For our application, we use moments on the following variables:

- the logged counterfactual ETA for each treatment level, and the squares and cross-products of these values;
- indicators of each time category of the week, precipitation type, whether the session is downtown, whether the user is a business user, distance to nearest public transit, and region; and
- log of the price multiplier and its square.

For each variable, the target moment is the moment in the full data sample. Matching moments of logged counterfactual ETA for each treatment, as well as squares and cross-products of these, is intended to approximately balance the conditional joint distributions of counterfactual ETAs in each subsample. We also add various moment conditions on

³A similar argument shows that the semi-elasticity estimates from 2SLS regressions with multiple binary instruments can also be compared across subsamples so long as the conditional joint distributions of potential ETAs for each value of the instrument vector are the same between the two subsamples: $(T_i(z_0), \dots, T_i(z_k))|X_i = x \sim (T_i(z_0), \dots, T_i(z_k))|X_i = x'$.

control variables to protect against the possibility that reweighting to balance counterfactual ETA distributions causes the samples to become unbalanced on other observables.⁴ When focusing on a particular heterogeneity dimension, we drop the associated moment condition (e.g., when comparing time elasticities across regions, we do not reweight by region.)

Table E.1 summarizes the results of the entropy-balance-reweighting exercise. The results across regions are quite similar to those found without reweighting: we still find New York City and Washington, D.C. to have the largest ETA elasticities and Miami to have the smallest ETA elasticity, suggesting that these regional variations in ETA elasticities reflect true differences in the underlying demand curves, and not just differences in prevailing ETA levels. Our observed results comparing ETA elasticities at different times of the week are also robust to reweighting. Looking at precipitation type, the difference between elasticities is no longer statistically significant.

Downtown results reverse relative to the unweighted version: we find a significantly larger ETA elasticity for downtown sessions. This suggests that the counterintuitive unweighted result of non-downtown elasticities being larger was driven by ETAs being larger in non-downtown sessions, and not by non-downtown sessions being more responsive to changes in waiting time at the same base ETA level. Similarly, with reweighting the ETA elasticity is greater when passengers are closer to public transit; this result runs counter to the unweighted estimates but is consistent with economic intuition.

The validity of this method depends on the quality of the counterfactual data, and the assumption that treatment does not affect counterfactual outcomes. For some observations, the no or only incomplete counterfactual data is available; for other sessions, the counterfactual data has apparent “errors”: the reported “counterfactual” ETA for the treatment a session actually received differs from the actual ETA. Missing and incorrect counterfactuals affect about 2% of sessions, and are attributable to small technical errors in logging. In the first three columns of Table E.2, we show that whether counterfactual data is missing and the degree of the counterfactual are not significantly predicted by treatment status.

In the last four columns of Table E.2, we regress the reported counterfactuals for each treatment group on indicator variables of treatment received. Here we find statistically significant (though economically small) evidence that sessions which had their ETAs increased more show lower counterfactual ETAs than control sessions. Since we know that the experimental treatments were assigned randomly, this effect may be caused by the following mechanism: in location-times subject to higher ETA treatments, fewer requests are made, resulting in more available drivers nearby and a lower reported counterfactual ETA. If this

⁴It is, of course, still possible that the reweighting causes the samples to become unbalanced on unobservables.

Table E.1
Reweighted heterogeneity results (Experiment 2).

Time of Week	(1)	Trip Characteristics	(2)	Location	(3)
Day of Week		Precipitation		Region	
Monday	−0.0407*** (0.0021)	None	−0.0435*** (0.0008)	San Francisco	−0.0361*** (0.0030)
Tuesday	−0.0467*** (0.0020)	Rain	−0.0396*** (0.0026)	New York City	−0.0882*** (0.0027)
Wednesday	−0.0452*** (0.0020)	User Type		Chicago	−0.0406*** (0.0022)
Thursday	−0.0474*** (0.0019)		Business	Washington, D.C.	−0.0491*** (0.0025)
Friday	−0.0419*** (0.0017)		Non-Business	Miami	−0.0257*** (0.0040)
Saturday	−0.0408*** (0.0017)			New Jersey	−0.0319*** (0.0042)
Sunday	−0.0393*** (0.0020)			Boston	−0.0440*** (0.0035)
				Philadelphia	−0.0331*** (0.0039)
Time of Day (Weekends)				Atlanta	−0.0162*** (0.0055)
6–10 AM	−0.0493*** (0.0044)			Los Angeles	−0.0345*** (0.0023)
10 AM–4 PM	−0.0448*** (0.0025)				
4 PM–7 PM	−0.0426*** (0.0032)			Location Type	
7 PM–11 PM	−0.0358*** (0.0028)			Downtown	−0.0587*** (0.0017)
11 PM–6 AM	−0.0305*** (0.0028)			Non-Downtown	−0.0288*** (0.001)
Time of Day (Weekdays)				Distance to Public Transit	
6–10 AM	−0.0558*** (0.0019)			Under 50 Meters	−0.0532*** (0.0017)
10 AM–4 PM	−0.0400*** (0.0017)			50 to 200 Meters	−0.0420*** (0.0010)
4 PM–7 PM	−0.0467*** (0.0020)			200 to 800 Meters	−0.0370*** (0.0019)
7 PM–11 PM	−0.0422*** (0.0019)			Over 800 Meters	−0.0025 (0.0215)
11 PM–6 AM	−0.0338*** (0.0025)				

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Standard errors clustered by location-hour in parentheses. Controls include local week of year, local hour of week, user location (geohash5), business user, and decile of pre-experiment lifetime rides.

Table E.2

Regressions of errors in counterfactual (CF) data and counterfactual ETAs (in minutes) on treatment indicators for Experiment 2.

	CF Missing	CF Error	(CF Error) ²	CF Control	CF Plus 60	CF Plus 150	CF Plus 240
Plus 60+	0.0000 (0.0001)	-0.0000 (0.0002)	-0.0013 (0.0012)	-0.0303*** (0.0016)	-0.0295*** (0.0018)	-0.0332*** (0.0018)	-0.0357*** (0.0018)
Plus 150+	-0.0001 (0.0001)	-0.0001 (0.0002)	-0.0006 (0.0014)	-0.0379*** (0.0019)	-0.0378*** (0.0022)	-0.0430*** (0.0021)	-0.0472*** (0.0021)
Plus 240+	0.0002 (0.0001)	-0.0003 (0.0003)	0.0039** (0.0017)	-0.0499*** (0.0024)	-0.0523*** (0.0030)	-0.0539*** (0.0028)	-0.0649*** (0.0028)
Controls	✓	✓	✓	✓	✓	✓	✓
<i>N</i>	9668820	9494639	9494639	9494639	9494639	9494639	9494639
<i>R</i> ²	0.168	0.002	0.007	0.466	0.567	0.557	0.543
<i>F</i>	0.9	0.5	2.8**	272.9***	198.8***	255.1***	329.3***

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Standard errors clustered by location-hour in parentheses. In ETA instrumented by experimental treatment indicators. Controls include local hour of week and week of year, session geohash5, business user, and decile of pre-experiment lifetime rides.

mechanism is occurring, the “counterfactuals” we observe are not truly representative of what ETAs would have been experienced under different treatment regimes.

This mechanism constitutes an interaction effect, but the estimates in Table E.2 suggest that such effects are small in magnitude, shifting ETAs by about four seconds at most.

F Mixed Logit

As a robustness check against our main linear probability model—which assumes constant effects of ETA and price on request rate, leading to potentially biased estimates of the VOT in the presence of heterogeneity—we estimate in this section a flexible mixed-logit model of rideshare demand, yielding a distribution of the value of time across user sessions.

We specify the indirect utility from requesting a ride as

$$u_{ij} = \beta_0 - \exp(\beta_{ij}^1)P_{ij} - \exp(\beta_{ij}^2)\text{ETA}_{ij} + \varepsilon_{ij}. \quad (19)$$

Here ETA_{ij} is the ETA quote for the session and P_{ij} is a *prediction* of the price of the ride request, constructed by multiplying the session’s Prime Time multiplier μ_{ij} by a prediction

$\hat{B}_{ij}$ of the session’s base price.⁵ $(\beta_{ij}^1, \beta_{ij}^2)'$ is assumed to follow a bivariate normal distribution:

$$\begin{pmatrix} \beta_{ij}^1 \\ \beta_{ij}^2 \end{pmatrix} \sim N \left(\begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}, \Sigma \Sigma' \right)$$

where Σ is the Cholesky factor of the variance-covariance matrix of $(\beta_{ij}^1, \beta_{ij}^2)'$. These parameters are exponentiated in the indirect utility specification equation to ensure that the elasticities of demand with respect to waiting time and price are always negative. The session-specific value of time in this model, $\exp(\beta_{ij}^2 - \beta_{ij}^1)$, is lognormally distributed.

To deal with endogeneity in (19), we employ a control function approach (Petrin and Train, 2010; Buchholz et al., 2025). We regress ETA and P on the experimental treatment indicators and include the residuals $\widetilde{\text{ETA}}$ and $\tilde{P}$ in the specification of the indirect utility:

$$u_{ij} = \beta_0 - \exp(\beta_{ij}^1)P_{ij} - \exp(\beta_{ij}^2)\text{ETA}_{ij} + \delta_1\tilde{P}_{ij} + \delta_2\widetilde{\text{ETA}}_{ij} + \tilde{\varepsilon}_{ij}. \quad (20)$$

The model was estimated using maximum likelihood⁶ applied to (20) on a random subsample of 200,000 observations from Experiment 1. Parameter estimates are shown in Table F.1. The significant nonzero estimates of Σ_{11} , Σ_{12} , and Σ_{22} imply that price and especially wait time sensitivity are heterogeneous across sessions, and that the sensitivities are positively correlated.

The resulting estimates imply a LogNormal(3.110, 0.699) distribution for the session-specific VOT. Though the mean of this distribution (31.82) is around 50% higher than our main estimate, its median (22.42) is quite close to the main estimate we report. The distribution also implies an interquartile VOT range of (12.76, 39.43).⁷

The main issue with this mixed-logit is the extent to which model parameters are well-identified by our experimental data. Non-parametric identification results for such random coefficient models (Ichimura and Thompson, 1998; Fox et al., 2012) generally require variation of regressors in an open set, while our experimental instruments only vary (ETA, P) between a small set of discrete levels; it is therefore likely that some functional form assumptions are necessary for identification and estimation. Moreover, unlike Buchholz et al. (2025), we are unable to rely on the panel structure of our data to identify (distributions

⁵The base price is the prediction of a LightGBM (Ke et al., 2017) machine learning model trained on completed control rides; such predictions must be used because actual base prices are not observed for sessions in which the user does not enter a destination. The predictors used include session location, session time, indicators for the various ride types considered during the session, and historical data on the passenger’s past rides and spending.

⁶We used the `mlogit` package in R (Croissant, 2020).

⁷For comparison, a linear probability model estimated by 2SLS regression of the request indicator on P and ETA in this subsample produced a VOT estimate of \$21.22.

Table F.1
Estimated parameters of mixed logit model.

	Parameter Estimate
β_0	38.70*** (11.07)
β_1	0.71*** (0.29)
β_2	3.82*** (0.30)
δ_1	1.64*** (0.48)
δ_2	-21.62*** (8.32)
Σ_{11}	-0.10*** (0.01)
Σ_{22}	1.42*** (0.18)
Σ_{12}	0.59*** (0.04)

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Standard errors in parentheses.

of) individual-specific coefficients, since our experimental treatments do not vary within individuals.

The estimator of [Dong and Lewbel \(2015\)](#) was also considered, but deemed inappropriate due to the discontinuous distribution of our price and ETA variation.

G Causality with Interaction Effects

In this section, we argue that our IV estimates maintain an [Angrist et al. \(2000\)](#)-type interpretation as causal weighted average derivatives, despite the possibility of cross-session interaction due to marketplace effects.

Let $\mathbf{Z}$ be the N -dimensional random vector giving the treatment assignment of every session in the experiment, with distribution function $F_{\mathbf{Z}}$. Let Z_i denote the i th entry of $\mathbf{Z}$ and $\mathbf{Z}_{-i}$ the vector $\mathbf{Z}$ with the i th entry deleted. Let $T_i(\mathbf{Z}) = T_i(Z_i, \mathbf{Z}_{-i})$ be the waiting time of the i th session; in general, this quantity depends on all of $\mathbf{Z}$ due to interaction effects. Also let $D_i(T, \mathbf{Z})$ be the i th session's demand (conversion rate) when the waiting time is T and the treatment vector is $\mathbf{Z}$.

The following properties hold:

1. *Independence*: $\mathbf{Z}$ is independent of $T_i(\mathbf{z})$ and $D_i(\mathbf{z}) \equiv D_i(T_i(\mathbf{z}), \mathbf{z})$ for all i and $\mathbf{z}$. This property follows because treatment assignments are fully randomized in the experiment.

A consequence is that Z_i is independent of $T_i(\mathbf{z}) = T_i(z_i, \mathbf{z}_{-i})$ and $D_i(\mathbf{z}) = D_i(z_i, \mathbf{z}_{-i})$ for all i and $\mathbf{z}$. It follows that $T(z) \equiv E[T_i(z, \mathbf{Z}_{-i})] = E[T_i(z, \mathbf{Z}_{-i})|Z_i = z] = E[T_i|Z_i = z]$ and similarly $D(z) \equiv E[D_i(z, \mathbf{Z}_{-i})] = E[D_i|Z_i = 1]$.

This result says that we can compute the average waiting time and demand for a session assigned treatment z by taking averages over all sessions assigned z . The only difference in our case vs. the no-interaction case is that the average waiting time and demand for a session assigned treatment z is an average not only over the population of sessions, but also over the possible treatment realizations for all other sessions. Hence, this average may in general depend on the distribution of $\mathbf{Z}$; for example, a distribution that assigns more sessions to a high ETA treatment $z = 1$ would result in a larger decrease in overall market demand, leaving more supply available and pushing down the average control session ETA $T(0)$. To emphasize this dependence on the distribution of $\mathbf{Z}$, we might more accurately write the average waiting time and demand for a session assigned treatment z by $T(z, F_{\mathbf{Z}})$ and $D(z, F_{\mathbf{Z}})$, but suppress this detail in our notation for simplicity.

2. *Exclusion*: $D_i(T, \mathbf{z}) = D_i(T, \mathbf{z}')$ for all i , T , $\mathbf{z}$, and $\mathbf{z}'$. This property follows from the design of the experiment: the ETA is the only aspect of the rider's decision to request that is affected by the experiment, and so holding the ETA fixed, the vector of treatment assignments for all sessions does not affect the rider's decision.
3. *Nonzero effect of treatment on ETA*: $T(z)$ is a nontrivial function of z . This property is validated empirically by noting that sessions with different ETA treatments have different average ETAs.
4. *Monotonicity*: For all z and z' , either $\Pr\{T(z) \leq T(z')\} = 1$ or $\Pr\{T(z') \leq T(z)\} = 1$. If 1 denotes the high ETA treatment, we have $\Pr\{T(0) \leq T(1)\} = 1$ by design of the experiment: since the high ETA treatment removes nearby drivers from being eligible to pick up a session, it cannot decrease a session's ETA.

These properties are exactly Assumptions 1–4 of Angrist et al. (2000), from which we conclude that the IV estimand from a regression of D_i on T_i using Z_i as an instrument,

$$\frac{E[D_i|Z_i = 1] - E[D_i|Z_i = 0]}{E[T_i|Z_i = 1] - E[T_i|Z_i = 0]},$$

can be written as

$$\int_0^\infty E \left[\frac{\partial D_i}{\partial T}(T) \middle| T_i(0) \leq T \leq T_i(1) \right] \omega(T) dT \quad (21)$$

with

$$\omega(T) = \frac{\Pr\{T_i(0) \leq T \leq T_i(1)\}}{\int_0^\infty \Pr\{T_i(0) \leq r \leq T_i(1)\} dr}. \quad (22)$$

This result illustrates that, even without strict SUTVA, our IV estimates can be interpreted as a weighted average of LATE derivatives of demand with respect to ETA. Since $T(0)$ and $T(1)$ depend on $F_{\mathbf{Z}}$, the presence of interaction effects modify the notion of “local” and the weighting of the derivatives at different points. More concretely, we conclude that under any distribution $F_{\mathbf{Z}}$ of the treatment assignments satisfying the properties 1–4 above, our IV estimates would have causal interpretations as weighted average derivatives, but the particular choice of $F_{\mathbf{Z}}$ would affect where on the demand curve these derivatives are evaluated, and how different derivatives at different points on the demand curve are weighted in the final estimate.

References

- Angrist, J. D., K. Graddy, and G. W. Imbens (2000). The interpretation of instrumental variables estimators in simultaneous equations models with an application to the demand for fish. *The Review of Economic Studies* 67(3), 499–527.
- Buchholz, N., L. Doval, J. Kastl, F. Matejka, and T. Salz (2025). Personalized pricing and the value of time: Evidence from auctioned cab rides. *Econometrica* 93(3), 929–958.
- Chernozhukov, V., D. Chetverikov, M. Demirer, E. Duflo, C. Hansen, W. Newey, and J. Robins (2018, 01). Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal* 21(1), C1–C68.
- Croissant, Y. (2020). Estimation of random utility models in r: The mlogit package. *Journal of Statistical Software* 95(11), 1–41.
- Dong, Y. and A. Lewbel (2015). A simple estimator for binary choice models with endogenous regressors. *Econometric Reviews* 34(1-2), 82–105.
- Fox, J. T., K. il Kim, S. P. Ryan, and P. Bajari (2012). The random coefficients logit model is identified. *Journal of Econometrics* 166(2), 204–212.
- Hainmueller, J. (2012). Entropy balancing for causal effects: A multivariate reweighting method to produce balanced samples in observational studies. *Political Analysis* 20(1), 25–46.

- Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica* 47(1), 153–161.
- Ichimura, H. and T. Thompson (1998, June). Maximum likelihood estimation of a binary choice model with random coefficients of unknown distribution. *Journal of Econometrics* 86(2), 269–295.
- Ke, G., Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T.-Y. Liu (2017). LightGBM: a highly efficient gradient boosting decision tree. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, NIPS’17, Red Hook, NY, USA, pp. 3149–3157. Curran Associates Inc.
- Maritz, J. S. and T. Lwin (1989). *Empirical bayes methods*. Routledge.
- Mosimann, J. E. (1962). On the compound multinomial distribution, the multivariate β -distribution, and correlations among proportions. *Biometrika* 49(1/2), 65–82.
- Petrin, A. and K. Train (2010). A control function approach to endogeneity in consumer choice models. *Journal of Marketing Research* 47(1), 3–13.
- U.S. Bureau of Labor Statistics (2016). May 2016 metropolitan and nonmetropolitan area occupation employment and wages estimates.
- U.S. Bureau of Labor Statistics (2020). Consumer Price Index for All Urban Consumers: All Items in U.S. City Average [CPIAUCSL].